Quantile Autoregression

Roger KOENKER and Zhijie XIAO

We consider quantile autoregression (QAR) models in which the autoregressive coefficients can be expressed as monotone functions of a single, scalar random variable. The models can capture systematic influences of conditioning variables on the location, scale, and shape of the conditional distribution of the response, and thus constitute a significant extension of classical constant coefficient linear time series models in which the effect of conditioning is confined to a location shift. The models may be interpreted as a special case of the general random-coefficient autoregression model with strongly dependent coefficients. Statistical properties of the proposed model and associated estimators are studied. The limiting distributions of the autoregression quantile process are derived. QAR inference methods are also investigated. Empirical applications of the model to the U.S. unemployment rate, short-term interest rate, and gasoline prices highlight the model's potential.

KEY WORDS: Asymmetric persistence; Autoregression; Comonotonicity; Quantile; Random coefficients.

1. INTRODUCTION

Constant-coefficient linear time series models have played an enormously successful role in statistics, and gradually various forms of random-coefficient time series models have also emerged as viable competitors in particular fields of application. One variant of the latter class of models, although perhaps not immediately recognizable as such, is the linear quantile regression model. This model has received considerable attention in the theoretical literature and can be easily estimated with the quantile regression methods proposed by Koenker and Bassett (1978). Curiously, however, all of the theoretical work dealing with this model that we are aware of focuses exclusively on the iid innovation case that restricts the autoregressive coefficients to be independent of the specified quantiles. In this article we seek to relax this restriction and consider linear quantile autoregression models with autoregressive (slope) parameters that may vary with quantiles $\tau \in [0, 1]$. We hope that these models might expand the modeling options for time series that display asymmetric dynamics or local persistence.

Considerable recent research effort has been devoted to modifications of traditional constant-coefficient dynamic models to incorporate various heterogeneous innovation effects. An important motivation for such modifications is the introduction of asymmetries into model dynamics. It is widely acknowledged that many important economic variables may display asymmetric adjustment paths (e.g., Neftci 1984; Enders and Granger 1998). The observation that firms are more apt to increase than to reduce prices is a key feature of many macroeconomic models. Beaudry and Koop (1993) have argued that positive shocks to the U.S. GDP are more persistent than negative shocks, indicating asymmetric business cycle dynamics over different quantiles of the innovation process. In addition, although it is generally recognized that output fluctuations are persistent, less persistent results are also found at longer horizons (Beaudry and Koop 1993), suggesting some form of "local persistence" (see, inter alia, Delong and Summers 1986; Hamilton 1989; Evans and Wachtel 1993; Bradley and Jansen 1997; Hess and Iwata 1997; Kuan and Huang 2001). A related development is the growing literature on threshold autoregression (TAR) (see, e.g., Balke and Fomby 1997; Tsay 1997; Gonzalez and Gonzalo 1998; Hansen 2000; Caner and Hansen 2001).

We believe that quantile regression methods can provide an alternative way to study asymmetric dynamics and local persistence in time series. We propose a new quantile autoregression (QAR) model in which autoregressive coefficients may take distinct values over different quantiles of the innovation process. We show that some forms of the model can exhibit unit root–like tendencies or even temporarily explosive behavior, but that occasional episodes of mean reversion are sufficient to ensure stationarity. The models lead to interesting new hypotheses and inference apparatus for time series.

The article is organized as follows. Section 2 introduces the model and gives some basic statistical properties of the QAR process. Section 3 develops the limiting distribution of the QAR estimator. Section 4 considers some restrictions imposed on the model by the monotonicity requirement on the conditional quantile functions. Section 5 explores statistical inference, including testing for asymmetric dynamics. Section 6 reports a Monte Carlo experiment on the sampling performance of the proposed inference procedure, and Section 7 gives an empirical application to U.S. unemployment rate time series. The Appendix provides proofs.

2. THE MODEL

There is a substantial theoretical literature, including works by Weiss (1987), Knight (1989), Koul and Saleh (1995), Koul and Mukherjee (1994), Hercé (1996), Hasan and Koenker (1997), and Hallin and Jurečková (1999), dealing with the linear quantile autoregression model. In this model the τth conditional quantile function of the response y_t is expressed as a linear function of lagged values of the response. Here we wish to study estimation and inference in a more general class of QAR models in which all of the autoregressive coefficients are allowed to be τ-dependent and thus are capable of altering the location, scale, and shape of the conditional densities.

2.1 The Model

Let $\{U_t\}$ be a sequence of iid standard uniform random variables, and consider the pth-order autoregressive process,

$$y_t = \theta_0(U_t) + \theta_1(U_t)y_{t-1} + \cdots + \theta_p(U_t)y_{t-p}, \tag{1}$$

Roger Koenker is McKinley Professor of Economics and Professor of Statistics, Department of Economics, University of Illinois, Champaign, IL 61820 (E-mail: *rkoenker@uiuc.edu*). Zhijie Xiao is Professor of Economics, Department of Economics, Boston College, Chestnut Hill, MA 02467 (E-mail: *xiaoz@bc.edu*). This research was supported in part by National Science Foundation grant SES-02-40781. The authors thank the co-editor, associate editor, two referees, and Steve Portnoy and Peter Phillips for valuable comments and discussions regarding this work.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000672

where the θ_j's are unknown functions $[0, 1] \to \mathbb{R}$ that we want to estimate. Provided that the right side of (1) is monotone increasing in U_t, it follows that the τth conditional quantile function of y_t can be written as

$$Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1} + \cdots + \theta_p(\tau)y_{t-p} \quad (2)$$

or, somewhat more compactly, as

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \mathbf{x}_t^\top \boldsymbol{\theta}(\tau), \quad (3)$$

where $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})^\top$ and $\mathcal{F}_t$ is the σ-field generated by $\{y_s, s \le t\}$. The transition from (1) to (2) is an immediate consequence of the fact that for any monotone increasing function g and standard uniform random variable, U, we have

$$Q_{g(U)}(\tau) = g(Q_U(\tau)) = g(\tau),$$

where $Q_U(\tau) = \tau$ is the quantile function of U. In the foregoing model, the autoregressive coefficients may be τ-dependent and thus can vary over the quantiles. The conditioning variables not only shift the location of the distribution of y_t, but also may alter the scale and shape of the conditional distribution. We call this model the QAR(p) model.

We argue that QAR models can play a useful role in expanding the modeling territory between classical stationary linear time series models and their unit root alternatives. To illustrate this in the QAR(1) case, consider the model

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1}, \quad (4)$$

with $\theta_0(\tau) = \sigma\Phi^{-1}(\tau)$ and $\theta_1(\tau) = \min\{\gamma_0 + \gamma_1\tau, 1\}$ for $\gamma_0 \in (0, 1)$ and $\gamma_1 > 0$. In this model, if $U_t > (1 - \gamma_0)/\gamma_1$, then the model generates the y_t according to the standard Gaussian unit root model, but for smaller realizations of U_t, we have a mean reversion tendency. Thus the model exhibits a form of asymmetric persistence in the sense that sequences of strongly positive innovations tend to reinforce its unit root-like behavior, whereas occasional negative realizations induce mean reversion and thus undermine the persistence of the process. The classical Gaussian AR(1) model is obtained by setting $\theta_1(\tau)$ to a constant.

The formulation in (4) reveals that the model may be interpreted as somewhat special form of random-coefficient autoregressive (RCAR) model. Such models arise naturally in many time series applications. Discussions of the role of RCAR models have been given by inter alia, Nicholls and Quinn (1982), Tjøstheim (1986), Pourahmadi (1986), Brandt (1986), Karlsen (1990), and Tong (1990). In contrast with most of the literature on RCAR models, in which the coefficients are typically assumed to be stochastically independent of one another, the QAR model has coefficients that are functionally dependent.

Monotonicity of the conditional quantile functions imposes some discipline on the forms taken by the $\boldsymbol{\theta}$ functions. This discipline essentially requires that the function $Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p})$ be monotone in τ in some relevant region Υ of $(y_{t-1}, \dots, y_{t-p})$ space. The correspondence between the random-coefficient formulation of the QAR model (1) and the conditional quantile function formulation (2) presupposes the monotonicity of the latter in τ. In the region Υ where this monotonicity holds, (1) can be considered a valid mechanism for simulating from the QAR model (2). Of course, model (1) can, even in the absence of this monotonicity, be taken as a valid data-generating mechanism; however, the link to the strictly linear conditional quantile model is no longer valid. At points where the monotonicity is violated, the conditional quantile functions corresponding to the model described by (1) have linear "kinks." Attempting to fit such piecewise linear models with linear specifications can be hazardous. We return to this issue in the discussion of Section 4. In the next section we briefly describe some essential features of the QAR model.

2.2 Properties of the Quantile Autoregression Process

The QAR(p) model (1) can be reformulated in more conventional random-coefficient notation as

$$y_t = \mu_0 + \alpha_{1,t}y_{t-1} + \cdots + \alpha_{p,t}y_{t-p} + u_t, \quad (5)$$

where $\mu_0 = E\theta_0(U_t)$, $u_t = \theta_0(U_t) - \mu_0$, and $\alpha_{j,t} = \theta_j(U_t)$, for $j = 1, \dots, p$. Thus $\{u_t\}$ is an iid sequence of random variables with distribution function $F(\cdot) = \theta_0^{-1}(\cdot + \mu_0)$, and the $\alpha_{j,t}$ coefficients are functions of this u_t innovation random variable. The QAR(p) process (5) can be expressed as an p-dimensional vector autoregression process of order 1,

$$\mathbf{Y}_t = \boldsymbol{\Gamma} + \mathbf{A}_t\mathbf{Y}_{t-1} + \mathbf{V}_t,$$

with

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mu_0 \\ \mathbf{0}_{p-1} \end{bmatrix}, \qquad \mathbf{A}_t = \begin{bmatrix} \mathbf{A}_{p-1,t} & \alpha_{p,t} \\ \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{bmatrix},$$

and

$$\mathbf{V}_t = \begin{bmatrix} u_t \\ \mathbf{0}_{p-1} \end{bmatrix},$$

where $\mathbf{A}_{p-1,t} = [\alpha_{1,t}, \dots, \alpha_{p-1,t}]$, $\mathbf{Y}_t = [y_t, \dots, y_{t-p+1}]^\top$, and $\mathbf{0}_{p-1}$ is the $(p-1)$-dimensional vector of 0's. In the Appendix we show that under regularity conditions given in Theorem 1, an $\mathcal{F}_t$-measurable solution for (5) can be found.

To formalize the foregoing discussion and facilitate later asymptotic analysis, we introduce the following conditions:

A.1 $\{u_t\}$ are iid random variables with mean 0 and variance $\sigma^2 < \infty$. The distribution function of u_t, F, has a continuous density f with $f(u) > 0$ on $\mathcal{U} = \{u : 0 < F(u) < 1\}$.

A.2 Let $E(\mathbf{A}_t \otimes \mathbf{A}_t) = \boldsymbol{\Omega}_{\mathbf{A}}$; the eigenvalues of $\boldsymbol{\Omega}_{\mathbf{A}}$ have moduli less than unity.

A.3 Denote the conditional distribution function $\Pr[y_t < \cdot|\mathcal{F}_{t-1}]$ as $F_{t-1}(\cdot)$ and its derivative as $f_{t-1}(\cdot)$; f_{t-1} is uniformly integrable on $\mathcal{U}$.

Theorem 1. Under assumptions A.1 and A.2, the time series y_t given by (5) is covariance stationary and satisfies a central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t - \mu_y) \Rightarrow \mathrm{N}(0, \omega_y^2),$$

where $\mu_y = \mu_0/(1 - \sum_{j=1}^{p}\mu_j)$, $\omega_y^2 = \lim n^{-1}E[\sum_{t=1}^{n}(y_t - \mu_y)]^2$, and $\mu_j = E(\alpha_{j,t})$, $j = 1, \dots, p$.

To illustrate some important features of the QAR process, we consider the simplest case of a QAR(1) process,

$$y_t = \alpha_t y_{t-1} + u_t, \quad (6)$$

where $\alpha_t = \theta_1(U_t)$ and $u_t = \theta_0(U_t)$, corresponding to (4), the properties of which are summarized in the following corollary.

Corollary 1. If y_t is determined by (6) and $\omega_\alpha^2 = E(\alpha_t)^2 < 1$, then, under assumption A.1, y_t is covariance stationary and satisfies a central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t \Rightarrow \mathrm{N}(0, \omega_y^2),$$

where $\omega_y^2 = \sigma^2(1+\mu_\alpha)/((1-\mu_\alpha)(1-\omega_\alpha^2))$ with $\mu_\alpha = E(\alpha_t) < 1$.

In the example given in Section 2.1, $\alpha_t = \theta_1(U_t) = \min\{\gamma_0 + \gamma_1 U_t, 1\} \le 1$, and $\Pr(|\alpha_t| < 1) > 0$, the condition of Corollary 1 holds, and the process y_t is globally stationary but can still display local (and asymmetric) persistency in the presence of certain type of shocks (positive shocks in the example). Corollary 1 also indicates that even with $\alpha_t > 1$ over some range of quantiles, as long as $\omega_\alpha^2 = E(\alpha_t)^2 < 1$, y_t can still be covariance stationary in the long run. Thus, a QAR process may allow for some (transient) forms of explosive behavior while maintaining stationarity in the long run.

Under the assumptions in Corollary 1, by recursively substituting in (6), we can see that

$$y_t = \sum_{j=0}^{\infty} \beta_{t,j} u_{t-j}, \quad \text{where } \beta_{t,0} = 1 \text{ and } \beta_{t,j} = \prod_{i=0}^{j-1} \alpha_{t-i}, \quad \text{for } j \ge 1, \qquad (7)$$

is a stationary $\mathcal{F}_t$-measurable solution to (6). In addition, if $\sum_{j=0}^{\infty} \beta_{t,j} v_{t-j}$ converges in L^p, then y_t has a finite pth-order moment. The $\mathcal{F}_t$-measurable solution of (6) gives a doubly stochastic $MA(\infty)$ representation of y_t. In particular, the impulse response of y_t to a shock u_{t-j} is stochastic and is given by $\beta_{t,j}$. On the other hand, although the impulse response of the QAR process is stochastic, it does converge (to 0) in mean square (and thus in probability) as $j \to \infty$, corroborating the stationarity of y_t. If we denote the autocovariance function of y_t by $\gamma_y(h)$, then it is easy to verify that $\gamma_y(h) = \mu_\alpha^{|h|}\sigma_y^2$, where $\sigma_y^2 = \sigma^2/(1-\omega_\alpha^2)$.

Remark 1. Compared with the $QAR(1)$ process y_t, if we consider a conventional AR(1) process with autoregressive coefficient μ_α and denote the corresponding process by $\underline{y}_t$, then the long-run variance of y_t (given by ω_y^2) is (as expected) larger than that of $\underline{y}_t$. The additional variance of the QAR process y_t comes from the variation of α_t. In fact, ω_y^2 can be decomposed into the summation of the long-run variance of $\underline{y}_t$ and an additional term that is determined by the variance of α_t,

$$\omega_y^2 = \underline{\omega}_y^2 + \frac{\sigma^2}{(1-\mu_\alpha)^2(1-\omega_\alpha^2)}\,\mathrm{var}(\alpha_t),$$

where $\underline{\omega}_y^2 = \sigma^2/(1-\mu_\alpha)^2$ is the long-run variance of $\underline{y}_t$.

We consider estimation and related inference on the QAR model in the next two sections.

3. ESTIMATION

Estimation of the QAR model (3) involves solving the problem

$$\min_{\boldsymbol\theta \in \mathbb{R}^{p+1}} \sum_{t=1}^{n} \rho_\tau(y_t - \mathbf{x}_t^\top \boldsymbol\theta), \qquad (8)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ as in work of Koenker and Bassett (1978). Solutions, $\widehat{\boldsymbol\theta}(\tau)$, are called autoregression quantiles. Given $\widehat{\boldsymbol\theta}(\tau)$, the τth conditional quantile function of y_t, conditional on $\mathbf{x}_t$, can be estimated by

$$\hat{Q}_{y_t}(\tau|\mathbf{x}_t) = \mathbf{x}_t^\top \widehat{\boldsymbol\theta}(\tau),$$

and the conditional density of y_t can be estimated by the difference quotients,

$$\hat{f}_{y_t}(\tau|\mathbf{x}_{t-1}) = \frac{(\tau_i - \tau_{i-1})}{\hat{Q}_{y_t}(\tau_i|\mathbf{x}_{t-1}) - \hat{Q}_{y_t}(\tau_{i-1}|\mathbf{x}_{t-1})},$$

for some appropriately chosen sequence of τ's.

If we denote $E(y_t)$ as μ_y and $E(y_t y_{t-j})$ as γ_j, and let $\boldsymbol\Omega_0 = E(\mathbf{x}_t\mathbf{x}_t^\top) = \lim n^{-1}\sum_{t=1}^{n} \mathbf{x}_t\mathbf{x}_t^\top$, then

$$\boldsymbol\Omega_0 = \begin{bmatrix} 1 & \boldsymbol\mu_y^\top \\ \boldsymbol\mu_y & \boldsymbol\Omega_y \end{bmatrix},$$

where $\boldsymbol\mu_y = \mu_y \cdot \mathbf{1}_{p\times 1}$ and

$$\boldsymbol\Omega_y = \begin{bmatrix} \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_0 \end{bmatrix}.$$

In the special case of QAR(1) model (6), $\boldsymbol\Omega_0 = E(\mathbf{x}_t\mathbf{x}_t^\top) = \mathrm{diag}[1, \gamma_0]$, $\gamma_0 = E[y_t^2]$. Let $\boldsymbol\Omega_1 = \lim n^{-1}\sum_{t=1}^{n} f_{t-1}[F_{t-1}^{-1}(\tau)] \times \mathbf{x}_t\mathbf{x}_t^\top$, and define $\boldsymbol\Sigma = \boldsymbol\Omega_1^{-1}\boldsymbol\Omega_0\boldsymbol\Omega_1^{-1}$. The asymptotic distribution of $\widehat{\boldsymbol\theta}(\tau)$ is summarized in the following theorem.

Theorem 2. Under assumptions A.1–A.3,

$$\boldsymbol\Sigma^{-1/2}\sqrt{n}(\widehat{\boldsymbol\theta}(\tau) - \boldsymbol\theta(\tau)) \Rightarrow \mathbf{B}_k(\tau),$$

where $\mathbf{B}_k(\tau)$ represents a k-dimensional standard Brownian bridge, $k = p + 1$.

By definition, for any fixed τ, $\mathbf{B}_k(\tau)$ is $\mathcal{N}(\mathbf{0}, \tau(1-\tau)\mathbf{I}_k)$. In the important special case with constant coefficients, $\boldsymbol\Omega_1 = f[F^{-1}(\tau)]\boldsymbol\Omega_0$, where $f(\cdot)$ and $F(\cdot)$ are the density and distribution functions of u_t. We state this result in the following corollary.

Corollary 2. Under assumptions A.1–A.3, if the coefficients α_{jt} are constants, then

$$f[F^{-1}(\tau)]\boldsymbol\Omega_0^{1/2}\sqrt{n}(\widehat{\boldsymbol\theta}(\tau) - \boldsymbol\theta(\tau)) \Rightarrow \mathbf{B}_k(\tau).$$

An alternative form of the model that is widely used in economic applications is the augmented Dickey–Fuller (ADF) regression,

$$y_t = \mu_0 + \delta_{0,t} y_{t-1} + \sum_{j=1}^{p-1} \delta_{j,t}\Delta y_{t-j} + u_t, \qquad (9)$$

where, corresponding to (5),

$$\delta_{0,t} = \sum_{s=1}^{p} \alpha_{s,t}, \qquad \delta_{j,t} = -\sum_{s=j+1}^{p} \alpha_{s,t}, \qquad j = 1, \dots, p-1.$$

In the foregoing transformed model, $\delta_{0,t}$ is the critical parameter corresponding to the largest autoregressive root. Letting $\mathbf{z}_t = (1, y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p+1})^\top$, we may write the quantile regression counterpart of (9) as

$$Q_{y_t}(\tau | \mathcal{F}_{t-1}) = \mathbf{z}_t^\top \boldsymbol{\delta}(\tau), \tag{10}$$

where

$$\boldsymbol{\delta}(\tau) = \big(\alpha_0(\tau), \delta_0(\tau), \delta_1(\tau), \dots, \delta_{p-1}(\tau)\big)^\top.$$

The limiting distributions of the quantile regression estimators $\widehat{\boldsymbol{\delta}}(\tau)$ can be obtained from our previous analysis. If we define

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & -1 & & -1 \\ & & & \ddots & \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Delta} = \mathbf{J\Sigma J},$$

then we have, under assumptions A.1–A.3,

$$\boldsymbol{\Delta}^{-1/2}\sqrt{n}(\widehat{\boldsymbol{\delta}}(\tau) - \boldsymbol{\delta}(\tau)) \Rightarrow \mathbf{B}_k(\tau).$$

If we focus our attention on the largest autoregressive root $\delta_{0,t}$ in the ADF-type regression (9) and consider the special case where $\delta_{j,t}$ is constant for $j = 1, \dots, p-1$, then a result similar to Corollary 1 can be obtained.

Corollary 3. Under assumptions A.1–A.3, if $\delta_{j,t}$ is constant for $j = 1, \dots, p-1$, and $\delta_{0,t} \le 1$ and $|\delta_{0,t}| < 1$ with positive probability, then the time series y_t given by (9) is covariance stationary and satisfies a central limit theorem.

4. QUANTILE MONOTONICITY

As in other linear quantile regression applications, linear QAR models should be cautiously interpreted as useful local approximations to more complex nonlinear global models. If we take the linear form of the model too literally, then obviously at some point (or points) there will be "crossings" of the conditional quantile functions, unless these functions are precisely parallel, in which case we are back to the pure location-shift form of the model. This crossing problem appears to be more acute in the autoregressive case than in ordinary regression applications, because the support of the design space (i.e., the set of $\mathbf{x}_t$'s that occur with positive probability) is determined within the model. Nevertheless, we may still regard the linear models specified earlier as valid local approximations over a region of interest.

It should be stressed that the *estimated* conditional quantile functions,

$$\hat{Q}_y(\tau | \mathbf{x}) = \mathbf{x}^\top \widehat{\boldsymbol{\theta}}(\tau),$$

are guaranteed to be monotone at the mean design point, $\mathbf{x} = \bar{\mathbf{x}}$, as shown by Bassett and Koenker (1982) for linear quantile regression models. In our random-coefficient view of the QAR model,

$$y_t = \mathbf{x}_t^\top \boldsymbol{\theta}(U_t),$$

we express the observable random variable y_t as a linear function of conditioning covariates. But rather than assuming that the coordinates of the vector $\boldsymbol{\theta}$ are independent random variables, we adopt a diametrically opposite viewpoint—that they are perfectly functionally dependent, all driven by a single random uniform variable. If the functions $(\theta_0, \dots, \theta_p)$ are all monotonically increasing, then the coordinates of the random vector $\boldsymbol{\alpha}_t$ are said to be comonotonic in the sense of Schmeidler (1986). [Random variables X and Y on a probability space $(\Omega, \mathcal{A}, P)$ are said to be comonotonic if there are monotone functions g and h and a random variable Z on $(\Omega, \mathcal{A}, P)$ such that $X = g(Z)$ and $Y = h(Z)$.] This is often the case, but there are important cases in which this monotonicity fails. What then?

What really matters is that we can find a linear reparameterization of the model that does exhibit comonotonicity over some relevant region of covariate space. Because for any nonsingular matrix $\mathbf{A}$ we can write

$$Q_y(\tau | \mathbf{x}) = \mathbf{x}^\top \mathbf{A}^{-1} \mathbf{A} \boldsymbol{\theta}(\tau),$$

we can choose $p+1$ linearly independent design points $\{\mathbf{x}_s : s = 1, \dots, p+1\}$, where $Q_y(\tau | \mathbf{x}_s)$ is monotone in τ. Then, choosing the matrix $\mathbf{A}$ so that $\mathbf{A}\mathbf{x}_s$ is the sth unit basis vector for $\mathbb{R}^{p+1}$, we have

$$Q_y(\tau | \mathbf{x}_s) = \gamma_s(\tau),$$

where $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$. Now inside the convex hull of our selected points, we have a comonotonic random-coefficient representation of the model. In effect, we have simply reparameterized the design so that the $p+1$ coefficients are the conditional quantile functions of y_t at the selected points. The fact that quantile functions of sums of nonnegative comonotonic random variables are sums of their marginal quantile functions (see, e.g., Denneberg 1994; Bassett, Koenker, and Kordas 2004) allows us to interpolate inside the convex hull. Of course, linear extrapolation is also possible, but we must be cautious about possible violations of the monotonicity requirement in this region.

The interpretation of linear conditional quantile functions as approximations to the local behavior in central range of the covariate space should always be considered provisional; richer data sources can be expected to yield more elaborate nonlinear specifications that would have validity over larger regions. Figure 1 illustrates a realization of the simple QAR(1) model described in Section 2. The black sample path shows 1,000 observations generated from the model (4) with AR(1) coefficients $\theta_1(u) = .85 + .25u$ and $\theta_0(u) = \Phi^{-1}(u)$. The gray sample path

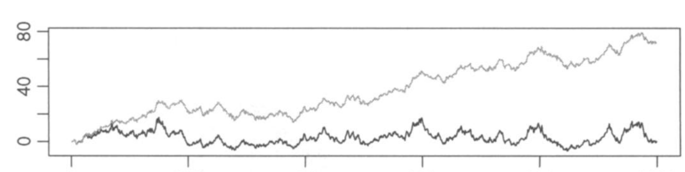

Figure 1. QAR and Unit Root Time Series. This figure contrasts two time series generated by the same sequence of innovations. The gray sample path is a random walk with standard Gaussian innovations; the black sample path illustrates a QAR series generated by the same innovations with random AR(1) coefficient .85 + .25 $\Phi(u_t)$. The latter series, although exhibiting explosive behavior in the upper tail, is stationary, as described in the text.

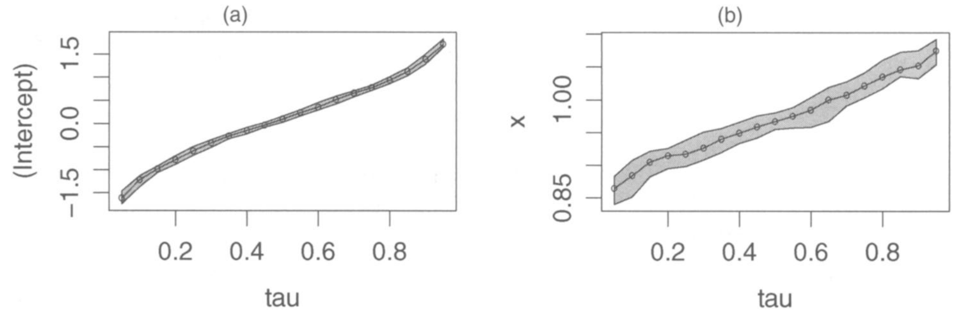

Figure 2. Estimating the QAR Model. This figure illustrates estimates of the QAR(1) model based on the black time series of the previous figure: (a) the intercept estimate at 19 equally spaced quantiles and (b) the AR(1) slope estimate at the same quantiles. The shaded region is a .90 confidence band. Note that the slope estimate quite accurately reproduces the linear form of the QAR(1) coefficient used to generate the data.

depicts the a random walk generated from the same innovation sequence, that is, the same $\theta_0(U_t)$'s but with constant θ_1 equal to 1. It is easy to verify that the QAR(1) form of the model satisfies the stationarity conditions of Section 2.2, and despite the explosive character of its upper-tail behavior, we observe that the series appears quite stationary, at least compared with the random-walk series. Estimating the QAR(1) model at 19 equally spaced quantiles yields the intercept and slope estimates depicted in Figure 2.

Figure 3 depicts estimated linear conditional quantile functions for short-term (3-month) U.S. interest rates using the QAR(1) model superimposed on the AR(1) scatterplot. In this example, the scatterplot clearly shows greater dispersion at higher interest rates, with nearly degenerate behavior at very low rates. The fitted linear quantile regression lines in Figure 3(a) show little evidence of crossing, but at rates below .04 there are some violations of the monotonicity requirement in the fitted quantile functions. Fitting the data using a somewhat more complex nonlinear (in variables) model by introducing a another additive component $\theta_2(\tau)(y_{t-1} - \delta)^2 I(y_{t-1} < \delta)$ with $\delta = 8$ in our example we can eliminate the problem of the crossing of fitted quantile functions. Figure 4 depicts the fitted coefficients of the QAR(1) model and their confidence region, showing that the estimated slope coefficient of the QAR(1) model has a somewhat similar appearance to the simulated example. Even more flexible models may be needed in other settings. A B-spline expansion QAR(1) model for Melbourne daily temperature has been described by Koenker (2000) to illustrate this approach.

The statistical properties of nonlinear QAR models and associated estimators are much more complicated than the linear QAR model that we study in the present article. Despite the possible crossing of quantile curves, we believe that the linear QAR model provides a convenient and useful local approximation to nonlinear QAR models. Such simple QAR models can still deliver important insight about dynamics (e.g., adjustment asymmetries) in economic time series and thus provides a useful tool in empirical diagnostic time series analysis.

5. INFERENCE ON THE QUANTILE AUTOREGRESSION PROCESS

In this section we turn our attention to inference in QAR models. Although other inference problems can be analyzed, here we consider the following inference problems that are of paramount interest in many applications. The first hypothesis is the quantile regression analog of the classical representation of linear restrictions on $\boldsymbol{\theta}$: (1) $H_{01}: \mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$, with known $\mathbf{R}$ and $\mathbf{r}$, where $\mathbf{R}$ denotes an $(q \times p)$-dimensional matrix and $\mathbf{r}$ is an q-dimensional vector. In addition to the classical inference problem, we are also interested in testing for asymmetric dynamics under the QAR framework. Thus we consider the hypothesis of parameter constancy, which can be formulated in

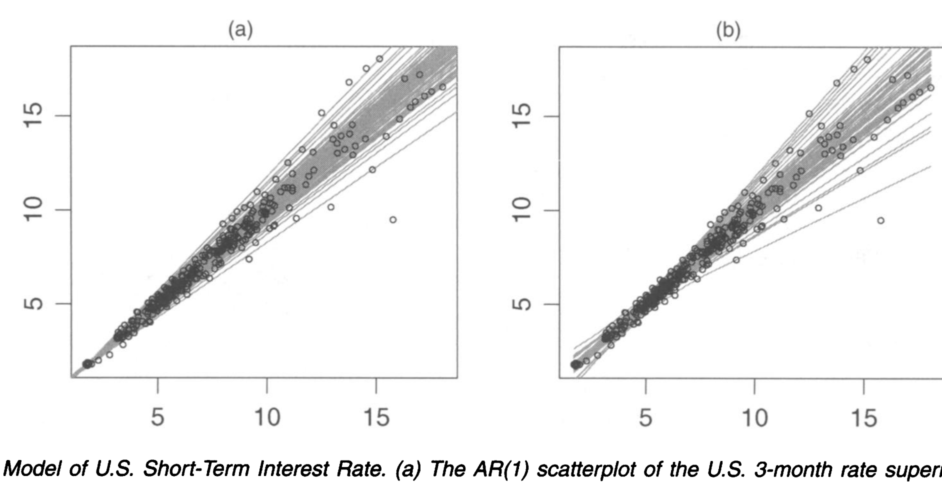

Figure 3. QAR(1) Model of U.S. Short-Term Interest Rate. (a) The AR(1) scatterplot of the U.S. 3-month rate superimposed with 49 equally spaced estimates of linear conditional quantile functions. (b) The model is augmented with a nonlinear (quadratic) component. The introduction of the quadratic component alleviates some nonmonotonicity in the estimated quantiles at low interest rates.

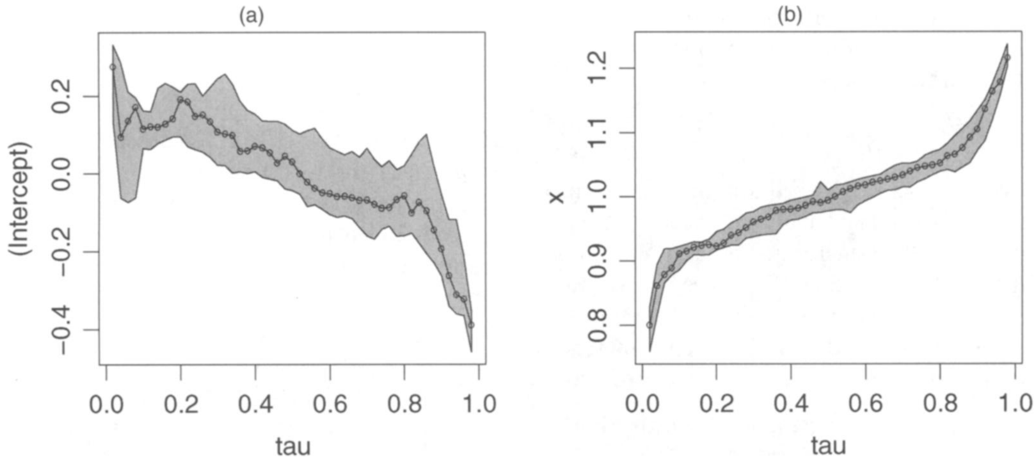

Figure 4. QAR(1) Model of U.S. Short-Term Interest Rate. The QAR(1) estimates of the intercept (a) and slope (b) parameters for 19 equally spaced quantile functions. Note that the slope parameter is, like the prior simulated example, explosive in the upper tail but mean-reverting in the lower tail.

the form of (2) $H_{02}: \mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$, with unknown but estimable $\mathbf{r}$. We consider both the cases at specific quantiles, τ (e.g., median, lower quartile, upper quartile), and the case over a range of quantiles, $\tau \in \mathcal{T}$.

5.1 The Regression Wald Process and Related Tests

Under the linear hypothesis $H_{01}: \mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$ and assumptions A.1–A.3, we have

$$\mathbf{V}_n(\tau) = \sqrt{n}[\mathbf{R}\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Omega}_1^{-1}\mathbf{R}^\top]^{-1/2}(\mathbf{R}\widehat{\boldsymbol{\theta}}(\tau) - \mathbf{r}) \Rightarrow \mathbf{B}_q(\tau), \tag{11}$$

where $\mathbf{B}_q(\tau)$ represents a q-dimensional standard Brownian bridge. For any fixed τ, $\mathbf{B}_q(\tau)$ is $\mathcal{N}(\mathbf{0}, \tau(1-\tau)\mathbf{I}_q)$. Thus the regression Wald process can be constructed as

$$W_n(\tau) = n(\mathbf{R}\widehat{\boldsymbol{\theta}}(\tau) - \mathbf{r})^\top[\tau(1-\tau)\mathbf{R}\widehat{\boldsymbol{\Omega}}_1^{-1}\widehat{\boldsymbol{\Omega}}_0\widehat{\boldsymbol{\Omega}}_1^{-1}\mathbf{R}^\top]^{-1} \times (\mathbf{R}\widehat{\boldsymbol{\theta}}(\tau) - \mathbf{r}),$$

where $\widehat{\boldsymbol{\Omega}}_1$ and $\widehat{\boldsymbol{\Omega}}_0$ are consistent estimators of $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_0$. If we are interested in testing $\mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$ over $\tau \in \mathcal{T}$, then we may consider, say, the following Kolmogorov–Smirnov (KS)-type sup-Wald test:

$$KSW_n = \sup_{\tau \in \mathcal{T}} W_n(\tau).$$

If we are interested in testing $\mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$ at a particular quantile $\tau = \tau_0$, then we can conduct a chi-squared test based on the statistic $W_n(\tau_0)$. The limiting distributions are summarized in the following theorem.

Theorem 3. Under assumptions A.1–A.3 and the linear restriction H_{01},

$$W_n(\tau_0) \Rightarrow \chi_q^2 \qquad \text{and} \qquad KSW_n = \sup_{\tau \in \mathcal{T}} W_n(\tau) \Rightarrow \sup_{\tau \in \mathcal{T}} Q_q^2(\tau),$$

where $Q_q(\tau) = \|\mathbf{B}_q(\tau)\|/\sqrt{\tau(1-\tau)}$ is a Bessel process of order q, where $\|\cdot\|$ represents the Euclidean norm. For any fixed τ, $Q_q^2(\tau) \sim \chi_q^2$ is a centered chi-squared random variable with q degrees of freedom.

5.2 Testing for Asymmetric Dynamics

The hypothesis that $\theta_j(\tau)$, $j = 1, \dots, p$, are constants over τ [i.e., $\theta_j(\tau) = \mu_j$] can be represented as $H_{02}: \mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$ by taking $\mathbf{R} = [\mathbf{0}_{p\times 1} \vdots \mathbf{I}_p]$ and $\mathbf{r} = [\mu_1, \dots, \mu_p]^\top$ with unknown parameters $\mu_1, \dots, \mu_p$. The Wald process and associated limiting theory provide a natural test for the hypothesis $\mathbf{R}\boldsymbol{\theta}(\tau) = \mathbf{r}$ when $\mathbf{r}$ is known. To test the hypothesis with unknown $\mathbf{r}$, an appropriate estimator of $\mathbf{r}$ is needed. In many econometrics applications, a $\sqrt{n}$-consistent estimator of $\mathbf{r}$ is available. If we look at the process

$$\widehat{\mathbf{V}}_n(\tau) = \sqrt{n}[\mathbf{R}\widehat{\boldsymbol{\Omega}}_1^{-1}\widehat{\boldsymbol{\Omega}}_0\widehat{\boldsymbol{\Omega}}_1^{-1}\mathbf{R}^\top]^{-1/2}(\mathbf{R}\widehat{\boldsymbol{\theta}}(\tau) - \widehat{\mathbf{r}}),$$

then, under H_{02}, we have

$$\begin{aligned}\widehat{\mathbf{V}}_n(\tau) &= \sqrt{n}[\mathbf{R}\widehat{\boldsymbol{\Omega}}_1^{-1}\widehat{\boldsymbol{\Omega}}_0\widehat{\boldsymbol{\Omega}}_1^{-1}\mathbf{R}^\top]^{-1/2}(\mathbf{R}\widehat{\boldsymbol{\theta}}(\tau) - \mathbf{r}) \\ &\quad - \sqrt{n}[\mathbf{R}\widehat{\boldsymbol{\Omega}}_1^{-1}\widehat{\boldsymbol{\Omega}}_0\widehat{\boldsymbol{\Omega}}_1^{-1}\mathbf{R}^\top]^{-1/2}(\widehat{\mathbf{r}} - \mathbf{r}) \\ &\Rightarrow \mathbf{B}_q(\tau) - f(F^{-1}(\tau))[\mathbf{R}\boldsymbol{\Omega}_0^{-1}\mathbf{R}^\top]^{-1/2}\mathbf{Z},\end{aligned}$$

where $\mathbf{Z} = \lim \sqrt{n}(\widehat{\mathbf{r}} - \mathbf{r})$. The need to estimate $\mathbf{r}$ introduces a drift component in addition to the simple Brownian bridge process, invalidating the distribution-free character of the original KS test.

To restore the asymptotically distribution-free nature of inference, we use a martingale transformation proposed by Khmaladze (1981) over the process $\widehat{\mathbf{V}}_n(\tau)$. Denote $df(x)/dx$ as $\dot{f}$ and define

$$\dot{\mathbf{g}}(r) = \left(1, (\dot{f}/f)(F^{-1}(r))\right)^\top \qquad \text{and}$$

$$\mathbf{C}(s) = \int_s^1 \dot{\mathbf{g}}(r)\dot{\mathbf{g}}(r)^\top\, dr,$$

we construct a martingale transformation $\mathcal{K}$ on $\widehat{\mathbf{V}}_n(\tau)$, defined as

$$\begin{aligned}\widetilde{\mathbf{V}}_n(\tau) &= \mathcal{K}\widehat{\mathbf{V}}_n(\tau) \\ &= \widehat{\mathbf{V}}_n(\tau) - \int_0^\tau \left[\dot{\mathbf{g}}_n(s)^\top \mathbf{C}_n^{-1}(s) \int_s^1 \dot{\mathbf{g}}_n(r)\, d\widehat{\mathbf{V}}_n(r)\right] ds,\end{aligned} \tag{12}$$

where $\dot{\mathbf{g}}_n(s)$ and $\mathbf{C}_n(s)$ are uniformly consistent estimators of $\dot{\mathbf{g}}(r)$ and $\mathbf{C}(s)$ over $\tau \in \mathcal{T}$, and propose the following KS-type test based on the transformed process:

$$KH_n = \sup_{\tau \in \mathcal{T}} \|\widetilde{\mathbf{V}}_n(\tau)\|. \tag{13}$$

(A Cramer–von Mises-type test based on the transformed process can also be constructed and analyzed in a similar way.) Under the null hypothesis, the transformed process $\widetilde{\mathbf{V}}_n(\tau)$ converges to a standard Brownian motion. (For more discussion of quantile regression inference based on the martingale transformation approach, see Koenker and Xiao 2002 and references therein.) We make the following assumption on the estimators:

A.4. There exist estimators $\dot{\mathbf{g}}_n(\tau)$, $\widehat{\mathbf{\Omega}}_0$ and $\widehat{\mathbf{\Omega}}_1$ satisfying (a) $\sup_{\tau\in} |\dot{\mathbf{g}}_n(\tau) - \dot{\mathbf{g}}(\tau)| = o_p(1)$, and (b) $\|\widehat{\mathbf{\Omega}}_0 - \mathbf{\Omega}_0\| = o_p(1)$, $\|\widehat{\mathbf{\Omega}}_1 - \mathbf{\Omega}_1\| = o_p(1)$, $\sqrt{n}(\widehat{\mathbf{r}} - \mathbf{r}) = O_p(1)$.

Theorem 4. Under the assumptions A.1–A.4 and the hypothesis H_{02},

$$\widetilde{\mathbf{V}}_n(\tau) \Rightarrow \mathbf{W}_q(\tau), \qquad KH_n = \sup_{\tau \in \mathcal{T}} \|\widetilde{\mathbf{V}}_n(\tau)\| \Rightarrow \sup_{\tau \in \mathcal{T}} \|\mathbf{W}_q(\tau)\|,$$

where $\mathbf{W}_q(r)$ is a q-dimensional standard Brownian motion.

The martingale transformation is based on the function $\dot{\mathbf{g}}(s)$, which must be estimated. There are several approaches to estimating the score $\frac{f'}{f}(F^{-1}(s))$. Portnoy and Koenker (1989) studied adaptive estimation and used kernel-smoothing method in estimating the density and score functions; they also discussed uniform consistency of the estimators. Cox (1985) proposed an elegant smoothing spline approach to the estimation of f'/f, and Ng (1994) provided an efficient algorithm for computing this score estimator. Estimation of $\mathbf{\Omega}_0$ is straightforward: $\widehat{\mathbf{\Omega}}_0 = n^{-1}\sum_t \mathbf{x}_t\mathbf{x}_t^\top$. (For estimation of $\widehat{\mathbf{\Omega}}_1$, see, inter alia, Koenker 1994; Powell 1989; Koenker and Machado 1999 for related discussions.)

6. MONTE CARLO

Here we report a Monte Carlo experiment conducted to examine the QAR-based inference procedures. We were particularly interested in time series displaying asymmetric dynamics; thus we considered the QAR model with $p = 1$ and tested the hypothesis that $\alpha_1(\tau)$ is constant over τ.

The data in our experiments were generated from model (6), where u_t are iid random variables. We consider the KS test KH_n given by (13) for different sample sizes ($n = 100$ and 300) and innovation distributions, and choose $\mathcal{T} = [.1, .9]$. Both normal innovations and Student-t innovations are considered. The number of repetitions is 1,000.

Representative results of the empirical size and power of the proposed tests are reported in Tables 1–3. We report the empirical size of this test for three choices of α_t: $\alpha_t = .95$, .9, and .6. The first two choices (.95 and .9) are large and close to unity, so that the corresponding time series display a certain degree of (symmetric) persistence. For models under the alternative, we considered the following four choices of α_t:

$$\alpha_t = \varphi_1(u_t) = \begin{cases} 1, & u_t \ge 0 \\ .8, & u_t < 0, \end{cases}$$

$$\alpha_t = \varphi_2(u_t) = \begin{cases} .95, & u_t \ge 0 \\ .8, & u_t < 0, \end{cases}$$

$$\alpha_t = \varphi_3(u_t) = \min\{.5 + F_u(u_t), 1\}, \tag{14}$$

$$\alpha_t = \varphi_4(u_t) = \min\{.75 + F_u(u_t), 1\}.$$

These alternatives deliver processes with different types of asymmetric (or local) persistence. In particular, when $\alpha_t = \varphi_1(u_t)$, $\varphi_3(u_t)$, $\varphi_4(u_t)$, and y_t display unit-root behavior in the presence of positive or large values of innovations but have a mean reversion tendency with negative shocks. The alternative $\alpha_t = \varphi_2(u_t)$ has local to (or weak) unit-root behavior in the presence of positive innovations and behave more stationarily when there are negative shocks.

The construction of tests uses estimators of density and score. We estimate the density (or sparsity) function using the approach of Siddiqui (1960). The density estimation entails choosing a bandwidth. We consider the bandwidth choices suggested by Hall and Sheather (1988) and Bofinger (1975) and rescaled versions of them. A bandwidth rule that Hall and Sheather (1988) suggested based on Edgeworth expansion for studentized quantiles (and using Gaussian plug-in) is

$$h_{HS} = n^{-1/3} z_\alpha^{2/3} \left[\frac{1.5\phi^2(\Phi^{-1}(t))}{2(\Phi^{-1}(t))^2 + 1}\right]^{1/3},$$

where z_α satisfies $\Phi(z_\alpha) = 1 - \alpha/2$ for the construction of $1 - \alpha$ confidence intervals. Another bandwidth selection has been proposed by Bofinger (1975) based on minimizing the mean squared error of the density estimator and is of order $n^{-1/5}$. Plugging in the Gaussian density, we obtain the following bandwidth, which is widely used in practice:

$$h_B = n^{-1/5} \left[\frac{4.5\phi^4(\Phi^{-1}(t))}{(2(\Phi^{-1}(t))^2 + 1)^2}\right]^{1/5}.$$

Monte Carlo results indicate that the Hall–Sheather bandwidth provides a good lower bound, and the Bofinger bandwidth provides a reasonable upper bound for bandwidth in testing parameter constancy. For this reason, we considered bandwidth choices between h_{HS} and h_B. In particular, we considered rescaled versions of h_B and h_{HS} (θh_B and δh_{HS}, where $0 < \theta < 1$ and $\delta > 1$ are scalars) in our Monte Carlo experiment and report representative results. Bandwidth values that are constant over the whole range of quantiles are not recommended. The sampling performance of tests using a constant bandwidth turned out to be poor and inferior to such bandwidth choices as the Hall–Sheather and Bofinger bandwidths which vary over the quantiles. For these reason, we focus on bandwidths h_B, h_{HS}, θh_B, and δh_{HS}. The Monte Carlo results indicate that the test using a rescaled version of the Bofinger bandwidth ($h = .6h_B$) yields good performance in the cases that we study.

We estimated the score function by the method of Portnoy and Koenker (1989) and chose the Silverman (1986) bandwidth in our Monte Carlo. Our simulation results show that the test was more affected by the estimator of the density than that of the score. Intuitively, the estimator of the density plays the role of a scalar and thus has the greatest influence. The Monte Carlo results also indicate that the method of Portnoy and Koenker (1989) coupled with the Silverman bandwidth has reasonably good performance. Table 1 reports the empirical size and power for the case with Gaussian innovations and sample size $n = 100$. Table 2 reports results in the case where the u_t's are Student-t

Table 1. Empirical Size and Power of Tests of Constancy of the Coefficient α With Gaussian Innovations

	Model	$h = 3h_{HS}$	$h = h_{HS}$	$h = h_B$	$h = .6h_B$
Size	$\alpha_t = .95$	.073	.287	.018	.056
	$\alpha_t = .9$	.073	.275	.01	.046
	$\alpha_t = .6$	.07	.287	.012	.052
Power	$\alpha_t = \varphi_1(u_t)$	.474	.795	.271	.391
	$\alpha_t = \varphi_2(u_t)$	.262	.620	.121	.234
	$\alpha_t = \varphi_3(u_t)$	.652	.939	.322	.533
	$\alpha_t = \varphi_4(u_t)$	.159	.548	.046	.114

NOTE: Models for size use the indicated constant coefficient; models for power comparisons are those indicated in (14). The sample size is 100, and the number of replications is 1,000.

Table 3. Empirical Size and Power of Tests of Constancy of the Coefficient α With Gaussian Innovations

	Model	$h = 3h_{HS}$	$h = h_{HS}$	$h = h_B$	$h = .6h_B$
Size	$\alpha_t = .95$	.081	.191	.028	.049
	$\alpha_t = .90$	.098	.189	.030	.056
	$\alpha_t = .60$	.097	.160	.020	.045
Power	$\alpha_t = \varphi_1(u_t)$	.974	.992	.921	.937
	$\alpha_t = \varphi_2(u_t)$	.831	.923	.685	.763
	$\alpha_t = \varphi_3(u_t)$	.998	1.000	.971	.989
	$\alpha_t = \varphi_4(u_t)$	.557	.897	.235	.392

NOTE: Configurations are as in Table 1, except that the sample size is 300.

innovations (with 3 degrees of freedom) and $n = 100$. Results in Table 2 confirm that using the quantile regression-based approach, power gain can be obtained in the presence of heavy-tailed disturbances. (Such gains obviously depend on choosing quantiles at which there is sufficient conditional density.) Experiments based on larger sample sizes are also conducted. Table 3 reports the size and power for the case with Gaussian innovations and sample size $n = 300$. These results are qualitatively similar to those of Table 1 but also show that as the sample sizes increase, the tests do have improved size and power properties, corroborating the asymptotic theory.

7. EMPIRICAL APPLICATIONS

There have been many claims and observations that some economic time series display asymmetric dynamics. For example, it has been observed that increases in the unemployment rate are sharper than declines. If an economic time series displays asymmetric dynamics systematically, then appropriate models are needed to incorporate such behavior. In this section we apply the QAR model to two economic time series: unemployment rates and retail gasoline prices in the United States. Our empirical analysis indicate that both series display asymmetric dynamics.

7.1 Unemployment Rate

Many studies on unemployment suggest that the response of unemployment to expansionary or contractionary shocks may be asymmetric. An asymmetric response to different types of shocks has important implications for economic policy. In this section we examine unemployment dynamics using the proposed procedures.

The data that we consider are quarterly and annual rates of unemployment in the United States. In particular, we look at (seasonally adjusted) quarterly rates, starting from the first quarter of 1948 and ending at the last quarter of 2003, for a total of 224 observations, as well as the annual rates from 1890 to 1996. Many empirical studies in the unit-root literature have investigated unemployment rate data. Nelson and Plosser (1982) studied the unit-root property of annual U.S. unemployment rates in their seminal work on 14 macroeconomic time series. Evidence based on the unit-root tests suggests that the series is stationary. This series and other types of unemployment rates often have been reexamined in later analysis.

Table 2. Empirical Size and Power of Tests of Constancy of the Coefficient α With $t(3)$ Innovations

	Model	$h = 3h_{HS}$	$h = h_{HS}$	$h = h_B$	$h = .6h_B$
Size	$\alpha_t = .95$	.086	.339	.011	.059
	$\alpha_t = .9$	.072	.301	.015	.043
	$\alpha_t = .6$	.072	.305	.013	.038
Power	$\alpha_t = \varphi_1(u_t)$	.556	.819	.319	.444
	$\alpha_t = \varphi_2(u_t)$	.348	.671	.174	.279
	$\alpha_t = \varphi_3(u_t)$	.713	.933	.346	.55
	$\alpha_t = \varphi_4(u_t)$	.284	.685	.061	.162

NOTE: Configurations are as in Table 1.

We first apply regression (10) on the unemployment rates. We use the Bayes information criterion (BIC) of Schwarz (1978) and Rissanen (1978) in selecting the appropriate lag length of the autoregressions. The selected lag length is $p = 3$ for the annual data and $p = 2$ for the quarterly data. The ordinary least squares (OLS) estimation of the largest autoregressive root is .718 for the annual series and .941 for the quarterly rates. We also perform QAR for each decile. Estimates of the largest AR root at each quantile are reported in Table 4. These estimated values vary over different quantiles, displaying asymmetric dynamics.

We then test asymmetric dynamics using the martingale transformation-based KS procedure (13) based on QAR (8). Following a suggestion from the Monte Carlo results, we choose the rescaled Hall and Sheather (1988) bandwidth $3h_{HS}$ and the rescaled Bofinger (1975) bandwidth $.6h_B$ in estimating the density function. The tests are constructed over $\tau \in T = [.05, .95]$, and results are reported in Table 5. The empirical results indicate that asymmetric behavior exist in these series.

7.2 Retail Gasoline Price Dynamics

Our second application investigates the asymmetric price dynamics in the retail gasoline market. We consider weekly data of U.S. regular gasoline retail price from August 20, 1990 to Februry 16, 2004. The sample size is 699. Evidence from OLS-based ADF tests of the null hypothesis of a unit root is mixed. The unit-root null is rejected by the coefficient-based test ADF_α, with a test statistic of -17.14 and critical value of -14.1, but it cannot be rejected by the t-ratio–based test ADF_t, given the test statistic -2.67 and critical value -2.86. Again we use the BIC to select the lag length to obtain $p = 4$ for these tests.

Table 4. Estimates of the Largest AR Root at Each Decile of Unemployment

Frequency	τ	.1	.2	.3	.4	.5	.6	.7	.8	.9
Annual	$\delta_0(\tau)$	.740	.776	.929	.871	.858	.793	.727	.680	.599
Quarterly	$\delta_0(\tau)$	.912	.908	.931	.919	.951	.959	.967	.962	.953

Table 5. Kolmogorov Test of Constant AR Coefficient for Unemployment

Bandwidth	$.6h_B$	$3h_{HS}$	*5% CV*
Annual	4.89	5.12	4.523
Quarterly	4.46	5.36	3.393

We next consider quantile regression evidence based on the ADF model (9) on persistence of retail gasoline prices. Table 6 reports the estimates of the largest autoregressive roots $\widehat{\delta}_0(\tau)$ at each decile. These results suggest that the gasoline price series has asymmetric dynamics. The estimate takes quite different values over different quantiles. Estimates, $\widehat{\delta}_0(\tau)$, monotonically increase as we move from lower quantiles to higher quantiles. The AR coefficient values at the lower quantiles are relatively small, indicating that the local behavior of the gasoline price would be stationary. However, at higher quantiles, the largest AR root is close to or even slightly above unity, consequently the time series display unit root or locally explosive behavior at upper quantiles.

A formal test of the null hypothesis that gasoline prices have a constant autoregressive coefficient is conducted using the KS procedure (13) based on QAR (2) and martingale transformation (12). Constancy of coefficients is rejected. The calculated KS statistic [using the rescaled Bofinger (1975) bandwidth, $.6h_B$] is 8.347735 (lag length, $p=4$), considerably larger than the 5% level critical value of 5.56. However, taking into account the possibility of unit-root behavior under the null, we also consider the (coefficient-based) empirical quantile process $U_n(\tau) = n(\widehat{\delta}_0(\tau) - 1)$ and the KS or Cramer–von Mises (CvM)-type tests,

$$QKS_\alpha = \sup_{\tau \in \mathcal{T}} |U_n(\tau)|, \qquad QCM_\alpha = \int_{\tau \in \mathcal{T}} U_n(\tau)^2\, d\tau. \quad (15)$$

Using the results of unit-root quantile regression asymptotics provided by Koenker and Xiao (2004), we have, under the unit-root hypothesis,

$$U_n(\tau) \Rightarrow U(\tau) = \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \underline{B}_y^2\right]^{-1} \int_0^1 \underline{B}_y\, dB_\psi^\tau, \quad (16)$$

where $\underline{B}_w(r)$ and $B_\psi^\tau(r)$ are limiting processes of $n^{-1/2} \times \sum_{t=1}^{[nr]} \Delta y_t$ and $n^{-1/2} \sum_{t=1}^{[nr]} \psi_\tau(u_{t\tau})$. We adopt the approach of Koenker and Xiao (2004) and approximate the distributions of the limiting variates by resampling method and construct bootstrap tests for the unit-root hypothesis based on (15).

We consider both the QKS_α and QCM_α tests for the null hypothesis of a constant AR coefficient equal to unity. Both tests firmly reject the null with test statistics of 35.79 and 320.41 and 5% level critical values of 13.22 and 19.72. The critical values were computed based on the resampling procedure described by Koenker and Xiao (2004). These results, together with the point estimates reported in Table 6, indicate that the gasoline price series has asymmetric adjustment dynamics and thus is not well characterized as a constant coefficient unit-root process.

Table 6. Estimated Largest AR Root at Each Decile of Retail Gasoline Price

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9
$\widehat{\delta}_0(\tau)$	.948	.958	.971	.980	.996	1.005	1.016	1.024	1.047

APPENDIX: PROOFS

A.1 Proof of Theorem 2.1

Given a pth-order autoregression process (5), we write $E(\alpha_{j,t}) = \mu_j$ and assume that $1 - \sum \mu_j \neq 0$. Letting $\mu_y = \mu_0/(1 - \sum_{j=1}^p \mu_j)$ and writing

$$\underline{y}_t = y_t - \mu_y,$$

we have

$$\underline{y}_t = \alpha_{1,t}\underline{y}_{t-1} + \cdots + \alpha_{p,t}\underline{y}_{t-p} + v_t, \quad (A.1)$$

where

$$v_t = u_t + \mu \sum_{l=1}^p (\alpha_{l,t} - \mu_l).$$

It is easy to see that $Ev_t = 0$ and $Ev_tv_s = 0$ for any $t \neq s$, because $E\alpha_{l,t} = \mu_l$ and u_t are independent. To derive stationarity conditions for the process $\underline{y}_t$, we first find an $\mathcal{F}_t$-measurable solution for (A.1). We define the $p \times 1$ random vectors

$$\underline{\mathbf{Y}}_t = [\underline{y}_t, \ldots, \underline{y}_{t-p+1}]^\top \qquad \text{and} \qquad \mathbf{V}_t = [v_t, 0, \ldots, 0]^\top$$

and the $p \times p$ random matrix

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_{p-1,t} & \alpha_{p,t} \\ \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{bmatrix},$$

where $\mathbf{A}_{p-1,t} = [\alpha_{1,t}, \ldots, \alpha_{p-1,t}]$ and $\mathbf{0}_{p-1}$ is the $(p-1)$-dimensional vector of 0's. Then

$$E(\mathbf{V}_t\mathbf{V}_t^\top) = \begin{bmatrix} \sigma_v^2 & \mathbf{0}_{1\times(p-1)} \\ \mathbf{0}_{(p-1)\times 1} & \mathbf{0}_{(p-1)\times(p-1)} \end{bmatrix} = \mathbf{\Sigma},$$

and the original process can be written as

$$\underline{\mathbf{Y}}_t = \mathbf{A}_t\underline{\mathbf{Y}}_{t-1} + \mathbf{V}_t.$$

By substitution, we have

$$\begin{aligned}\underline{\mathbf{Y}}_t &= \mathbf{V}_t + \mathbf{A}_t\mathbf{V}_{t-1} + \mathbf{A}_t\mathbf{A}_{t-1}\mathbf{V}_{t-2} \\ &\quad + [\mathbf{A}_t \cdots \mathbf{A}_{t-m+1}]\mathbf{V}_{t-m} + [\mathbf{A}_t \cdots \mathbf{A}_{t-m}]\underline{\mathbf{Y}}_{t-m-1} \\ &= \underline{\mathbf{Y}}_{t,m} + \mathbf{R}_{t,m},\end{aligned}$$

where

$$\underline{\mathbf{Y}}_{t,m} = \sum_{j=0}^m \mathbf{B}_j\mathbf{V}_{t-j}, \qquad \mathbf{R}_{t,m} = \mathbf{B}_{m+1}\underline{\mathbf{Y}}_{t-m-1}, \qquad \text{and}$$

$$\mathbf{B}_j = \begin{cases} \prod_{l=0}^{j-1} \mathbf{A}_{t-l}, & j \geq 1 \\ \mathbf{I}, & j = 0. \end{cases}$$

The stationarity of an $\mathcal{F}_t$-measurable solution for y_t involves the convergence of $\{\sum_{j=0}^m \mathbf{B}_j\mathbf{V}_{t-j}\}$ and $\{\mathbf{R}_{t,m}\}$ as m increases, for fixed t. Following a similar analysis as reported by Nicholls and Quinn (1982, chap. 2), we need to verify that $\operatorname{vec} E[\underline{\mathbf{Y}}_{t,m}\underline{\mathbf{Y}}_{t,m}^\top]$ converges as $m \to \infty$. Note that $\mathbf{B}_j$ is independent with $\mathbf{V}_{t-j}$ and $\{u_t, t = 0, \pm 1, \pm 2, \ldots\}$ are independent random variables; thus $\{\mathbf{B}_j\mathbf{V}_{t-j}\}_{j=0}^\infty$ is an orthogonal sequence in the sense that $E[\mathbf{B}_j\mathbf{V}_{t-j}\mathbf{B}_k\mathbf{V}_{t-k}] = \mathbf{0}$ for any $j \neq k$. Therefore,

$$\begin{aligned}\operatorname{vec} E[\underline{\mathbf{Y}}_{t,m}\underline{\mathbf{Y}}_{t,m}^\top] &= \operatorname{vec} E\left[\left(\sum_{j=0}^m \mathbf{B}_j\mathbf{V}_{t-j}\right)\left(\sum_{j=0}^m \mathbf{B}_j\mathbf{V}_{t-j}\right)^\top\right] \\ &= \operatorname{vec} E\left[\sum_{j=0}^m \mathbf{B}_j\mathbf{V}_{t-j}\mathbf{V}_{t-j}^\top\mathbf{B}_j^\top\right].\end{aligned}$$

Noting that $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$ and $(\prod_{l=0}^{j}\mathbf{A}_l) \otimes (\prod_{k=0}^{j}\mathbf{B}_k) = \prod_{k=0}^{j}(\mathbf{A}_k \otimes \mathbf{B}_k)$, we have

$$\begin{aligned}
&\mathrm{vec}\, E\left[\sum_{j=0}^{m}\mathbf{B}_j\mathbf{V}_{t-j}\mathbf{V}_{t-j}^\top\mathbf{B}_j^\top\right] \\
&= E\left[\sum_{j=0}^{m}(\mathbf{B}_j\otimes\mathbf{B}_j)\,\mathrm{vec}(\mathbf{V}_{t-j}\mathbf{V}_{t-j}^\top)\right] \\
&= E\left[\sum_{j=0}^{m}\left(\prod_{l=0}^{j-1}\mathbf{A}_{t-l}\right)\otimes\left(\prod_{l=0}^{j-1}\mathbf{A}_{t-l}\right)\mathrm{vec}(\mathbf{V}_{t-j}\mathbf{V}_{t-j}^\top)\right] \\
&= \sum_{j=0}^{m}\prod_{l=0}^{j-1}E(\mathbf{A}_{t-l}\otimes\mathbf{A}_{t-l})\,\mathrm{vec}\,E(\mathbf{V}_{t-j}\mathbf{V}_{t-j}^\top).
\end{aligned}$$

If we write

$$\mathbf{A} = E[\mathbf{A}_t] = \begin{bmatrix}\overline{\boldsymbol{\mu}}_{p-1} & \boldsymbol{\alpha}_p \\ \mathbf{I}_{p-1} & \mathbf{0}_{p-1}\end{bmatrix},$$

where $\overline{\boldsymbol{\mu}}_{p-1} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{p-1}]$, then $\mathbf{A}_t = \mathbf{A} + \boldsymbol{\Xi}_t$, where $E(\boldsymbol{\Xi}_t) = \mathbf{0}$, and

$$\begin{aligned}
E(\mathbf{A}_{t-l}\otimes\mathbf{A}_{t-l}) &= E[(\mathbf{A}+\boldsymbol{\Xi}_t)\otimes(\mathbf{A}+\boldsymbol{\Xi}_t)] \\
&= \mathbf{A}\otimes\mathbf{A} + \mathrm{E}(\boldsymbol{\Xi}_t\otimes\boldsymbol{\Xi}_t) = \boldsymbol{\Omega}_{\mathbf{A}},
\end{aligned}$$

then

$$\mathrm{vec}\, E\left[\left(\sum_{j=0}^{m}\mathbf{B}_j\mathbf{V}_{t-j}\right)\left(\sum_{j=0}^{m}\mathbf{B}_j\mathbf{V}_{t-j}\right)^\top\right] = \sum_{j=0}^{m}\boldsymbol{\Omega}_{\mathbf{A}}^{j}\,\mathrm{vec}(\boldsymbol{\Sigma}).$$

The critical condition for the stationarity of the process $\underline{y}_t$ is that $\sum_{j=0}^{m}\boldsymbol{\Omega}_{\mathbf{A}}^{j}$ converges as $m\to\infty$.

The matrix $\boldsymbol{\Omega}_{\mathbf{A}}$ may be represented in Jordan canonical form as $\boldsymbol{\Omega}_{\mathbf{A}} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$, where $\boldsymbol{\Lambda}$ has the eigenvalues of $\boldsymbol{\Omega}_{\mathbf{A}}$ along its main diagonal. If the eigenvalues of $\boldsymbol{\Omega}_{\mathbf{A}}$ have moduli less than unity, then $\boldsymbol{\Lambda}^j$ converges to $\mathbf{0}$ at a geometric rate. Noting that $\boldsymbol{\Omega}_{\mathbf{A}}^{j} = \mathbf{P}\boldsymbol{\Lambda}^j\mathbf{P}^{-1}$, following a similar analysis as done by Nicholls and Quinn (1982, chap. 2), $\underline{\mathbf{Y}}_t$ (and thus y_t) is stationary and can be represented as

$$\underline{\mathbf{Y}}_t = \sum_{j=0}^{\infty}\mathbf{B}_j\mathbf{V}_{t-j}.$$

The central limit theorem then follows from work of Billingsley (1961) (also see Nicholls and Quinn 1982, thm. A.1.4).

A.2 Proof of Theorem 3.1

If we denote $\widehat{\mathbf{v}} = \sqrt{n}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau))$, then $\rho_\tau(\mathbf{y}_t - \widehat{\boldsymbol{\theta}}(\tau)^\top\mathbf{x}_t) = \rho_\tau(u_{t\tau} - (n^{-1/2}\widehat{\mathbf{v}})^\top\mathbf{x}_t)$, where $u_{t\tau} = y_t - \mathbf{x}_t^\top\boldsymbol{\theta}(\tau)$. Minimization of (8) is equivalent to minimizing

$$Z_n(\mathbf{v}) = \sum_{t=1}^{n}\left[\rho_\tau\left(u_{t\tau} - (n^{-1/2}\mathbf{v})^\top\mathbf{x}_t\right) - \rho_\tau(u_{t\tau})\right]. \tag{A.2}$$

If $\widehat{\mathbf{v}}$ is a minimizer of $Z_n(\mathbf{v})$, then we have $\widehat{\mathbf{v}} = \sqrt{n}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau))$. The objective function $Z_n(\mathbf{v})$ is a convex random function. Knight (1989) (also see Pollard 1991; Knight 1998) showed that if the finite-dimensional distributions of $Z_n(\cdot)$ converge weakly to those of $Z(\cdot)$ and $Z(\cdot)$ has a unique minimum, then the convexity of $Z_n(\cdot)$ implies that $\widehat{\mathbf{v}}$ converges in distribution to the minimizer of $Z(\cdot)$.

We use the following identity: If we denote $\psi_\tau(u) = \tau - I(u<0)$, then for $u \neq 0$,

$$\begin{aligned}
&\rho_\tau(u-v) - \rho_\tau(u) \\
&= -v\psi_\tau(u) + (u-v)\{I(0>u>v) - I(0<u<v)\} \\
&= -v\psi_\tau(u) + \int_0^v \{I(u\le s) - I(u<0)\}\,ds.
\end{aligned} \tag{A.3}$$

Thus the objective function of minimization problem can be written as

$$\begin{aligned}
&\sum_{t=1}^{n}\left[\rho_\tau\left(u_{t\tau} - (n^{-1/2}\mathbf{v})^\top\mathbf{x}_t\right) - \rho_\tau(u_{t\tau})\right] \\
&\quad = -\sum_{t=1}^{n}(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t\psi_\tau(u_{t\tau}) \\
&\qquad + \sum_{t=1}^{n}\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}\{I(u_{t\tau}\le s) - I(u_{t\tau}<0)\}\,ds.
\end{aligned}$$

We first consider the limiting behavior of

$$W_n(\mathbf{v}) = \sum_{t=1}^{n}\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}\{I(u_{t\tau}\le s) - I(u_{t\tau}<0)\}\,ds.$$

For convenience of asymptotic analysis, we denote

$$W_n(\mathbf{v}) = \sum_{t=1}^{n}\xi_t(\mathbf{v}),$$

$$\xi_t(\mathbf{v}) = \int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}\{I(u_{t\tau}\le s) - I(u_{t\tau}<0)\}\,ds.$$

We further define $\overline{\xi}_t(\mathbf{v}) = E\{\xi_t(\mathbf{v})|\mathcal{F}_{t-1}\}$ and $\overline{W}_n(\mathbf{v}) = \sum_{t=1}^{n}\overline{\xi}_t(\mathbf{v})$. Then $\{\xi_t(\mathbf{v}) - \overline{\xi}_t(\mathbf{v})\}$ is a martingale difference sequence.

Note that

$$u_{\tau t} = y_t - \mathbf{x}_t^\top\boldsymbol{\alpha}(\tau) = y_t - F_{t-1}^{-1}(\tau),$$

$$\begin{aligned}
\overline{W}_n(\mathbf{v}) &= \sum_{t=1}^{n}E\left\{\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}[I(u_{t\tau}\le s) - I(u_{t\tau}<0)]\Big|\mathcal{F}_{t-1}\right\} \\
&= \sum_{t=1}^{n}\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}\left[\int_{F_{t-1}^{-1}(\tau)}^{s+F_{t-1}^{-1}(\tau)}f_{t-1}(r)\,dr\right]ds \\
&= \sum_{t=1}^{n}\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}\frac{F_{t-1}(s+F_{t-1}^{-1}(\tau)) - F_{t-1}(F_{t-1}^{-1}(\tau))}{s}s\,ds.
\end{aligned}$$

Under assumption A.3,

$$\begin{aligned}
\overline{W}_n(\mathbf{v}) &= \sum_{t=1}^{n}\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}f_{t-1}(F_{t-1}^{-1}(\tau))s\,ds + o_p(1) \\
&= \frac{1}{2n}\sum_{t=1}^{n}f_{t-1}(F_{t-1}^{-1}(\tau))\mathbf{v}^\top\mathbf{x}_t\mathbf{x}_t^\top\mathbf{v} + o_p(1).
\end{aligned}$$

By our assumptions and stationarity of y_t, we have

$$\overline{W}_n(\mathbf{v}) \Rightarrow \frac{1}{2}\mathbf{v}^\top\boldsymbol{\Omega}_1\mathbf{v}.$$

Using the same argument as that of Herce (1996), the limiting distribution of $\sum_t\xi_t(\mathbf{v})$ is the same as that of $\sum_t\overline{\xi}_t(\mathbf{v})$.

For the behavior of the first term, $n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t\psi_\tau(u_{t\tau})$, in the objective function, note that $\mathbf{x}_t\in\mathcal{F}_{t-1}$ and $E[\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}] = 0$, $\mathbf{x}_t\psi_\tau(u_{t\tau})$ is a martingale difference sequence, and thus $n^{-1/2}\times\sum_{t=1}^{n}\mathbf{x}_t\psi_\tau(u_{t\tau})$ satisfies a central limit theorem. Following the arguments of Portnoy (1984) and Gutenbrunner and Jurečková (1992), the QAR process is tight, and thus the limiting variate, viewed as a random function of τ, is a Brownian bridge over $\tau\in\mathcal{T}$,

$$n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t\psi_\tau(u_{t\tau}) \Rightarrow \boldsymbol{\Omega}_0^{1/2}\mathbf{B}_k(\tau).$$

For each fixed τ, $n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t\psi_\tau(u_{t\tau})$ converges to a q-dimensional vector-normal variate with covariance matrix $\tau(1-\tau)\boldsymbol{\Omega}_0$. Thus

$$\begin{aligned}
Z_n(\mathbf{v}) &= \sum_{t=1}^{n}\big[\rho_\tau\big(u_{t\tau}-(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t\big)-\rho_\tau(u_{t\tau})\big] \\
&= -\sum_{t=1}^{n}(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t\psi_\tau(u_{t\tau}) \\
&\quad+\sum_{t=1}^{n}\int_0^{(n^{-1/2}\mathbf{v})^\top\mathbf{x}_t}\{I(u_{t\tau}\le s)-I(u_{t\tau}<0)\}\,ds \\
&\Rightarrow -\mathbf{v}^\top\boldsymbol{\Omega}_0^{1/2}\mathbf{B}_k(\tau)+\frac{1}{2}\mathbf{v}^\top\boldsymbol{\Omega}_1\mathbf{v}=Z(\mathbf{v}).
\end{aligned}$$

By the convexity lemma of Pollard (1991) and arguments of Knight (1989), note that $Z_n(\mathbf{v})$ and $Z(\mathbf{v})$ are minimized at $\widehat{\mathbf{v}}=\sqrt{n}(\widehat{\boldsymbol{\alpha}}(\tau)-\boldsymbol{\alpha}(\tau))$ and $\boldsymbol{\Sigma}^{1/2}\mathbf{B}_k(\tau)$. By lemma A of Knight (1989), we have

$$\boldsymbol{\Sigma}^{-1/2}\sqrt{n}(\widehat{\boldsymbol{\alpha}}(\tau)-\boldsymbol{\alpha}(\tau))\Rightarrow\mathbf{B}_k(\tau).$$

[Received October 2005. Revised October 2006.]

REFERENCES

Balke, N., and Fomby, T. (1997), "Threshold Cointegration," *International Economic Review*, 38, 627–645.

Bassett, G., and Koenker, R. (1982), "An Empirical Quantile Function for Linear Models With iid Errors," *Journal of the American Statistical Association*, 77, 407–415.

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

Beaudry, P., and Koop, G. (1993), "Do Recessions Permanently Change Output?" *Journal of Monetary Economics*, 31, 149–163.

Billingsley, P. (1961), "The Lindeberg–Levy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788–792.

Bofinger, E. (1975), "Estimation of a Density Function Using Order Statistics," *Australian Journal of Statistics*, 17, 1–7.

Bradley, M. D., and Jansen, D. W. (1997), "Nonlinear Business Cycle Dynamics: Cross-Country Evidence on the Persistence of Aggregate Shocks," *Economic Inquiry*, 35, 495–509.

Brandt, A. (1986), "The Stochastic Equation $Y_{n+1}=A_nY_n+B_n$ With Stationary Coefficients," *Advances in Applied Probability*, 18, 211–220.

Caner, M., and Hansen, B. (2001), "Threshold Autoregression With a Unit Root," *Econometrica*, 69, 1555–1596.

Cox, D. (1985), "A Penalty Method for Nonparametric Estimation of the Logarithmic Derivative of a Density Function," *Annals of the Institute of Mathematical Statistics*, 37, 271–288.

Delong, J. B., and Summers, L. H. (1986), "Are Business Cycles Symmetrical?" in *American Business Cycle*, ed. R. J. Gordon, Chicago: University of Chicago Press, pp. 166–178.

Denneberg, D. (1994), *Non-Additive Measure and Integral*, Dordrecht: Kluwer.

Enders, W., and Granger, C. (1998), "Unit Root Tests and Asymmetric Adjustment With an Example Using the Term Structure of Interest Rates," *Journal of Business & Economic Statistics*, 16, 304–311.

Evans, M., and Wachtel, P. (1993), "Inflation Regions and the Sources of Inflation Uncertainty," *Journal of Money, Credit, and Banking*, 25, 475–511.

Gonzalez, M., and Gonzalo, J. (1998), "Threshold Unit Root Models," working paper, University Carlos III de Madrid.

Gutenbrunner, C., and Jurečková, J. (1992), "Regression Rank Scores and Regression Quantiles," *The Annals of Statistics*, 20, 305–330.

Hall, P., and Sheather, S. (1988), "On the Distribution of a Studentized Quantile," *Journal of the Royal Statistical Society*, Ser. B, 50, 381–391.

Hallin, M., and Jurečková, J. (1999), "Optimal Tests for Autoregressive Models Based on Autoregression Rank Scores," *The Annals of Statistics*, 27, 1385–1414.

Hamilton, J. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hansen, B. (2000), "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.

Hasan, M. N., and Koenker, R. (1997), "Robust Rank Tests of the Unit Root Hypothesis," *Econometrica*, 65, 133–161.

Hess, G. D., and Iwata, S. (1997), "Asymmetric Persistence in GDP? A Deeper Look at Depth," *Journal of Monetary Economics*, 40, 535–554.

Hercé, M. (1996), "Asymptotic Theory of LAD Estimation in a Unit Root Process With Finite Variance Errors," *Econometric Theory*, 12, 129–153.

Karlsen, H. A. (1990), "Existence of Moments in a Stationary Stochastic Difference Equation," *Advances in Applied Probability*, 22, 129–146.

Khmaladze, E. (1981), "Martingale Approach to the Goodness of Fit Tests," *Theory of Probability and Its Applications*, 26, 246–265.

Knight, K. (1989), "Limit Theory for Autoregressive-Parameter Estimates in an Infinite-Variance Random Walk," *Canadian Journal of Statistics*, 17, 261–278.

——— (1998), "Asymptotics for L1 Regression Estimates Under General Conditions," *The Annals of Statistics*, 26, 755–770.

Koenker, R. (1994), "Confidence Intervals for Regression Quantiles," in *Proceedings of the 5th Prague Symposium on Asymptotic Statistics*, Berlin: Springer-Verlag, pp. 349–359.

——— (2000), "Galton, Edgeworth, Frisch and Prospects for Quantile Regression in Econometrics," *Journal of Econometrics*, 95, 347–374.

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

Koenker, R., and Machado, J. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 81, 1296–1310.

Koenker, R., and Xiao, Z. (2002), "Inference on the Quantile Regression Processes," *Econometrica*, 70, 1583–1612.

——— (2004), "Unit Root Quantile Regression Inference," *Journal of the American Statistical Association*, 99, 775–787.

Koul, H., and Mukherjee, K. (1994), "Regression Quantiles and Related Processes Under Long Range Dependent Errors," *Journal of Multivariate Analysis*, 51, 318–317.

Koul, H., and Saleh, A. K. (1995), "Autoregression Quantiles and Related Rank-Scores Processes," *The Annals of Statistics*, 23, 670–689.

Kuan, C. M., and Huang, Y. L. (2001), "The Semi-Nonstationary Process: Model and Empirical Evidence, preprint, Univ. Hong Kong.

Neftci, S. (1984), "Are Economic Time Series Asymmetric Over the Business Cycle?" *Journal of Political Economy*, 92, 307–328.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

Ng, P. (1994), "Smoothing Spline Score Estimation," *SIAM Journal of Scientific and Statistical Computing*, 15, 1003–1025.

Nicholls, D. F., and Quinn, B. G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Berlin: Springer-Verlag.

Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199.

Portnoy, S. (1984), "Tightness of the Sequence of Empiric cdf Processes Defined From Regression Fractiles," in *Robust and Nonlinear Time Series Analysis*, eds. J. Franke, W. Härdle, and D. Martin, New York: Springer-Verlag, pp. 231–246.

Portnoy, S., and Koenker, R. (1989), "Adaptive L-Estimation of Linear Models," *The Annals of Statistics*, 17, 362–381.

Pourahmadi, M. (1986), "On Stationarity of the Solution of a Doubly Stochastic Model," *Journal of Time Series Analysis*, 7, 123–131.

Powell, J. (1989), "Estimation of Monotonic Regression Models Under Quantile Restrictions," in *Nonparametric and Semiparametric Methods in Econometrics*, eds. J. Powell and G. Tauchen, Cambridge, U.K.: Cambridge University Press, pp. 357–386.

Rissanen, J. (1978), "Modelling by Shortest Data Description," *Automatica*, 14, 465–471.

Schmeidler, D. (1986), "Integral Representation Without Additivity," *Proceedings of the American Mathematical Society*, 97, 255–261.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Siddiqui, M. (1960), "Distribution of Quantiles From a Bivariate Population," *Journal of Research of the National Bureau of Standards*, 64, 145–150.

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

Tjøstheim, D. (1986), "Some Doubly Stochastic Time Series Models," *Journal of Time Series Analysis*, 7, 51–72.

Tong, H. (1990), *Nonlinear Time Series: A Dynamical Approach*, Oxford, U.K.: Oxford University Press.

Tsay, R. (1997), "Unit Root Tests With Threshold Innovations," preprint, University of Chicago.

Weiss, A. (1987), "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation," *Econometric Theory*, 7, 46–68.

Comment

Jianqing FAN and Yingying FAN

We congratulate Koenker and Xiao on their interesting and important contribution to quantile autoregression (QAR). The article provides a comprehensive overview of the QAR model, from probabilistic aspects to model identification, statistical inferences, and empirical applications. The attempt to integrate the quantile regression and the QAR process is intriguing. It demonstrates that, surprisingly, nonparametric coefficient functions can be estimated at a root-n rate for the QAR processes. The authors then put forward some useful tools for testing the significance of lag variables and asymmetric dynamics of time series. We appreciate the opportunity to comment several aspects of this article.

1. CONNECTIONS WITH VARYING–COEFFICIENT MODELS

QAR is closely related to the functional-coefficient autoregressive (FCAR) model. In the time series context, Cai, Fan, and Li (2000b) proposed the following FCAR model for capturing the nonlinearity of a time series:

$$Y_t = \alpha_0(U_t) + \alpha_1(U_t)Y_{t-1} + \cdots + \alpha_p(U_t)Y_{t-p} + \varepsilon_t, \qquad (1)$$

where U_t is a thresholding variable and $\{\varepsilon_t\}$ is a sequence of independent innovations. In particular, when $U_t = Y_{t-d}$ for some lag d, the model was called a functional autoregressive model (FAR) by Chen and Tsay (1993). Varying-coefficient models have been widely used in many aspects of statistical modeling; see, for example, the work of Hastie and Tibshirani (1993), Carroll, Ruppert, and Welsh (1998), and Cai et al. (2000a) for applications to generalized linear models; Brumback and Rice (1998), Fan and Zhang (2000), and Chiang, Rice, and Wu (2001) for analysis of functional data; Lin and Ying (2001) and Fan and Li (2004) for analysis of longitudinal data; Tian, Zucker, and Wei (2005) and Fan, Lin, and Zhou (2006) for applications to the Cox hazards regression model; and Fan, Jiang, Zhang, and Zhou (2003), Hong and Lee (2003), and Mercurio and Spokoiny (2004) for applications to financial modeling. These are just a few examples that testify to the flexibility, popularity, and explanatory power of the varying-coefficient models. In the same vein, they reflect the importance of the QAR model.

What makes QAR different from the FCAR model or, more generally, the varying coefficient model is that the variable U_t is unobservable and $\varepsilon_t = 0$. This makes estimating techniques completely different. For example, in the varying-coefficient model, the coefficient functions in (1) are estimated through localizing on U_t (which are observable), whereas in the QAR model, the coefficient functions are estimated through quantile regression techniques. As a result, two completely different sets of rates of convergence are obtained. The former model admits a nonparametric rate, whereas the latter reveals the parametric rate.

Despite their differences in statistical inferences, QAR is a subfamily of models of FCAR as far as probabilistic aspects are concerned. Hence the stochastic properties established in FCAR are applicable directly to QAR. Chen and Tsay (1993) have given sufficient conditions for the solution to (1) to admit a stationary and ergodic solution. With some modifications of their proof, it can be shown that if $\alpha_j(\cdot)$ is bounded by c_j for all j and if all roots of the characteristic function

$$\lambda^p - c_1\lambda^{p-1} - \cdots - c_p = 0$$

are inside the unit circle, then there exists a stationary solution that is geometrically ergodic.

2. IDENTIFIABILITY OF THE MODEL

An important observation made by Koenker and Xiao is that if, given $Y_{t-p}, \dots, Y_{t-1}$, the function

$$\beta_t(u) = \theta_0(u) + \theta_1(u)Y_{t-1} + \cdots + \theta_p(u)Y_{t-p} \qquad (2)$$

is strictly increasing in u, then $\beta_t(\tau)$ is the conditional τ-quantile of Y_t given $Y_{t-p}, \dots, Y_{t-1}$. Because the conditional τ-quantile is identifiable under some mild conditions, the identifiability condition becomes that with probability 1, the QAR model generates at least $(p+1)$ linearly independent vectors of form $\mathbf{Y}_t = (1, Y_{t-1}, \dots, Y_{t-p})^T$. In other words, letting

$$\mathcal{T} = \{t : \beta_t(u) \text{ is strictly increasing in } u\}, \qquad (3)$$

there are at least $(p+1)$ distinct time points $t_i \in \mathcal{T}$ such that $\mathbf{Y}_{t_i}$ are linearly independent for each realization. A natural and open question is what kind of population would generate, with probability 1, the samples that satisfy the foregoing condition.

The aforementioned identifiability conditions are hard to check. However, they are needed not only for connections to the quantile regression, but also for identifiability. To see this, look at the specific case where $p = 0$, in which $Y_t = \theta_0(U_t)$. Clearly, $\theta_0(\cdot)$ is the quantile function of Y_t only when $\theta_0(\cdot)$ is monotone increasing. When this condition is violated, the model is not necessarily identifiable. For example, $Y_t = |U_t - .5|$ and $Z_t = U_t - .5I(U_t > .5)$ have identically the same distribution but very different $\theta_0(\cdot)$.

We note that the QAR(p) model

$$Y_t = \theta_0(U_t) + \theta_1(U_t)Y_{t-1} + \cdots + \theta_p(U_t)Y_{t-p} \equiv \boldsymbol{\theta}(U_t)^T\mathbf{X}_t$$

is not differentiable from the model

$$Y_t = \boldsymbol{\theta}(1 - U_t)^T\mathbf{X}_t,$$

where $\mathbf{X}_t = (Y_{t-1}, \dots, Y_{t-p})^T$. Thus if $\boldsymbol{\theta}(\tau)$ is a solution, then so is $\boldsymbol{\theta}(1-\tau)$.

Jianqing Fan is Professor (E-mail: *jqfan@princeton.edu*) and Yingying Fan is a Doctoral Candidate (E-mail: *yingying@princeton.edu*), Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544. The authors thank the associate editor for helpful comments. The work was supported in part by National Science Foundation grant DMS-05-32370 and National Institutes of Health grant R01 GM072611.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000681

3. FITTING AND DIAGNOSTICS

Koenker and Xiao estimate the coefficient functions $\boldsymbol{\theta}(\tau)$ with the quantile regression

$$\min_{\boldsymbol{\theta}} \sum_t \rho_\tau(Y_t - \mathbf{X}_t^T \boldsymbol{\theta}). \tag{4}$$

This convex optimization usually exists. The resulting estimates $\hat{\boldsymbol{\theta}}(\tau)$ are consistent estimates of the parameter

$$\boldsymbol{\theta}^*(\tau) = \arg\min_{\boldsymbol{\theta}} E\rho_\tau(Y_t - \mathbf{X}_t^T \boldsymbol{\theta}) \tag{5}$$

under some mild conditions. Without some technical conditions, $\boldsymbol{\theta}^*(\tau)$ and $\boldsymbol{\theta}(\tau)$ are not necessarily the same. This is evidenced by the example given in the previous section in which $\theta^*(\tau) = \tau/2$ with no ambiguity, whereas $\theta(\tau) = |\tau - .5|$ or $\tau - .5I(\tau > .5)$.

The foregoing argument suggests that the results of Koenker and Xiao should replace $\boldsymbol{\theta}(\tau)$ by $\boldsymbol{\theta}^*(\tau)$ unless the conditions under which they are identical are clearly imposed. If the primary interest is really on $\boldsymbol{\theta}(\tau)$, then the conditional quantile regression should be replaced by the restricted conditional quantile regression (RCQR),

$$\min_{\boldsymbol{\theta}} \sum_t I(t \in \mathcal{T}) \rho_\tau(Y_t - \mathbf{X}_t^T \boldsymbol{\theta}). \tag{6}$$

This avoids some samples in which the monotonicity condition is violated that create inconsistent estimators. However, the set $\mathcal{T}$ is unknown and depends on the value at another quantile τ. This creates some difficulties in implementation.

One possible way out is to replace $\mathcal{T}$ by one of its subsets. For example, if all $\theta_j(\cdot)$'s are monotonically increasing, then we can replace $\mathcal{T}$ by the subset in which all components of $\mathbf{X}_t$ are nonnegative. Another possibility is to use (4) to get an initial estimate and then check whether the functions $\{\hat{\beta}_t(\tau), t = 1, \ldots, T\}$ are strictly increasing at some percentiles (e.g., $\tau = .05, .1, .15, \ldots, .95$). Delete the cases in which the monotonicity is violated and use RCQR (6). The process can be iterated.

To illustrate the problem using the conditional quantile regression (4) and to address the issue of identifiability, we generate 2,000 data points from the QAR(1) model

$$Y_t = \Phi^{-1}(U_t) + (1.8U_t - 1.7)Y_{t-1}. \tag{7}$$

Hence we have $\theta_0(\tau) = \Phi^{-1}(\tau)$ and $\theta_1(\tau) = 1.8\tau - 1.7$. Fit the data using (4) and (6). The resulting estimates are depicted in Figure 1. The estimates (dot-dashed curve) obtained using RCQR (6) are very close to the true coefficient functions (thin solid curve), whereas the conditional quantile method (4) results in estimates (dashed curve) far away from the true functions. Indeed, the latter estimates are for the functions $\theta_0^*(\tau)$ and $\theta_1^*(\tau)$ defined by (5), which were computed numerically and depicted in Figure 1 by the thick solid curve. This example shows that even if monotonicity conditions are not fulfilled at all t, the coefficient functions can still be identifiable and consistently estimated, but the conditional quantile regression estimate, defined by maximizing (4), can be inconsistent.

A related question is how robust the fitting techniques are to model misspecification. For example, if the data-generating process is FCAR (1) without observing U_t, but we still use the conditional quantile regression (4) or its modification (6) to fit the data, how robust is the quantile estimate? To quantify this, we simulate the 2,000 data points from the model

$$Y_t = \Phi^{-1}(U_t) + (.85 + .25U_t)Y_{t-1} + \varepsilon_t, \tag{8}$$

where $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. Figure 2 shows the plots for small noise, $\sigma = 0$; moderate noise, $\sigma = .8$, and relatively large noise, $\sigma = 2$. The fitting techniques are very sensitive to the noise level. The estimates differ substantially from the true coefficient functions for even moderate $\sigma = .8$.

Checking monotonicity of $\hat{\beta}_t(\tau)$ is one aspect of model diagnostics. Another aspect is to check whether the distribution of $\hat{U}_t = \hat{\beta}_t^{-1}(Y_t)$ for $t \in \mathcal{T}$ is uniform. There are many approaches to this kind of testing problem, including the Kolmogorov–Smirnov test and visual inspection of the estimated density. For example, one can create the normally transformed data $\hat{Z}_t = \Phi^{-1}(\hat{U}_t)$, then use the normal-reference rule of the kernel density estimate to see whether, the transformed residuals $\{\hat{Z}_t\}$ are normally distributed. Alternatively, one can use the quantile–quantile plot to accomplish a similar task.

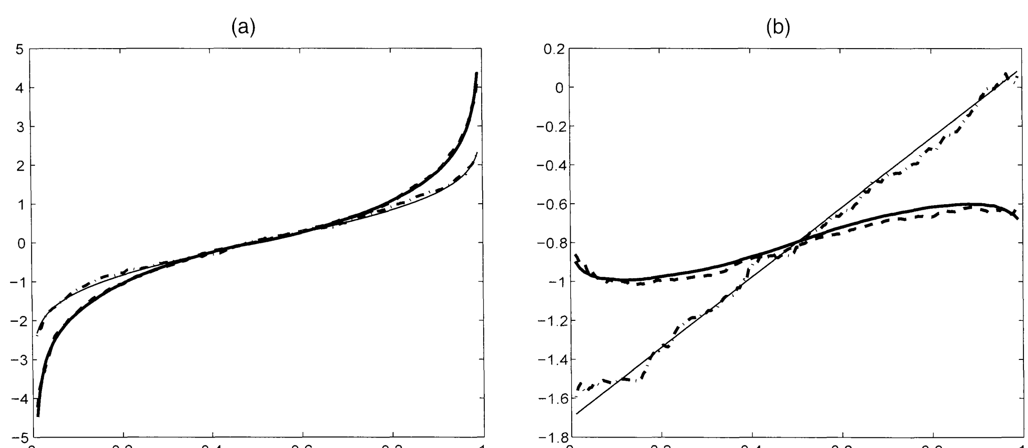

Figure 1. Estimates of $\theta_0(\tau) = \Phi^{-1}(\tau)$ (a) and $\theta_1(\tau) = 1.8\tau - 1.7$ (b) in Model (7). The thin curves are the true coefficient functions $\theta_0(\tau)$ and $\theta_1(\tau)$, the dashed curves are the estimates obtained by using conditional quantile regression (4), the dot-dashed curves are the estimates obtained by using restricted conditional quantile regression (6), and the thick solid curves are $\theta_0^*(\tau)$ and $\theta_1^*(\tau)$.

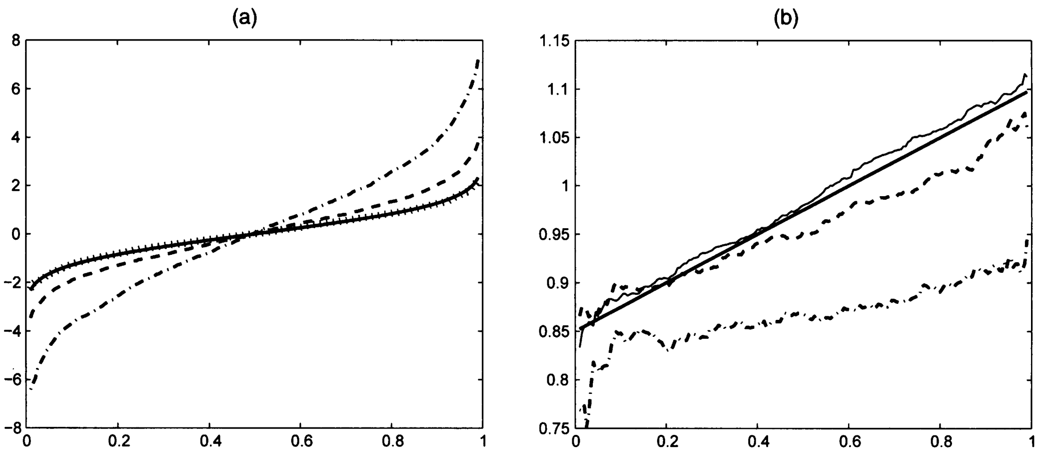

Figure 2. The Influences of the Error ε_t on the Estimation of $\theta_0(\tau) = \Phi^{-1}(\tau)$ (a) and the Estimation of $\theta_1(\tau) = .85 + .25\tau$ (b) in Model (8) for Different Noise Levels σ [(a) ——, true; ıııııı, $\sigma = 0$; - - -, $\sigma = .8$; ·-·-, $\sigma = 2$. (b) ——, true; ——, $\sigma = 0$; - - -, $\sigma = .8$; ·-·-, $\sigma = 2$].

For the data generated from (7) used in Figure 1, Figure 3 presents the histograms of $\hat{U}_t$ and the quantile–quantile plots of $\hat{Z}_t$. From these plots, we can see that the distribution of $\hat{U}_t = \hat{\beta}_t^{-1}(Y_t)$ for $t \in \mathcal{T}$ by using the conditional quantile regression method (4) is not uniform, whereas $\hat{U}_t$ obtained using the RCQR method (6) is uniformly distributed. These results are also supportive of our previous points.

Note that when the data are generated from model (1) without observing U_t, the model still can be identifiable when the density f_ε of innovation ε_t is known. The problem is more com-

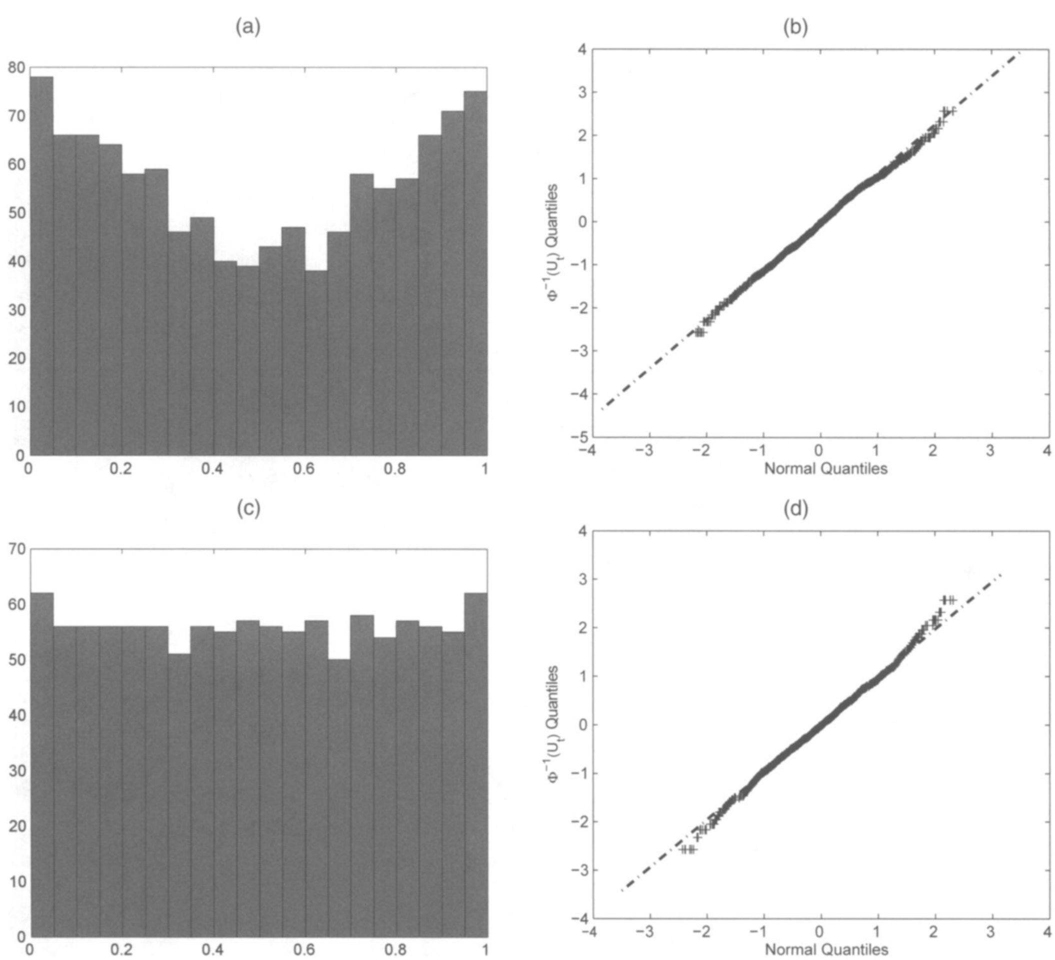

Figure 3. Histograms of $\hat{U}_t$ and Quantile–Quantile Plots of $\hat{Z}_t$. (a) Histogram plot of $\hat{U}_t = \hat{\beta}_t^{-1}(Y_t)$ for $t \in \mathcal{T}$, where $\hat{\beta}_t(\tau)$ is estimated using the conditional quantile regression (4). (b) Quantile–quantile plot of $\Phi^{-1}(\hat{U}_t)$ versus the standard normal distribution, where $\hat{U}_t$ is the same as in (a). (c) and (d) The same as (a) and (b), but with restricted conditional quantile regression (6) as the estimation method.

plicated but similar to a deconvolution problem. The estimation procedure can be quite complicated. To see this, note that

$$P(Y_t \le a | Y_{t-1}, \ldots, Y_{t-p}) = P(\beta_t(U_t) + \varepsilon_t \le a) = \int \beta_t^{-1}(a-x) f_\varepsilon(x)\, dx.$$

Thus letting $Q_t(\tau)$ be the conditional τ-quantile, we have

$$\int \beta_t^{-1}(Q_t(\tau) - x) f_\varepsilon(x)\, dx = \tau.$$

Let us denote the solution by $Q_t(\tau) = h(\beta_t(\cdot), \tau)$ for some function h. Then the coefficient functions can be estimated by minimizing a quantity similar to (6),

$$\min_{\theta} \int_0^1 \sum_t I(t \in \mathcal{T}) \rho_\tau\big(Y_t - h(\beta_t(\cdot), \tau)\big)\, d\tau,$$

where $\boldsymbol{\theta}(\cdot)$ and $\beta_t(\cdot)$ are related through (2). This is indeed a complicated optimization problem. The method of Koener and Xiao is a specific case of this method with $\varepsilon_t = 0$.

4. CONCLUDING REMARKS

Koenker and Xiao have developed a nice scheme for conditional quantile inference and made insightful connections with the QAR model. However, the issues of identifiability and possible misspecification of models suggest that extra care should be taken in making this kind of link. In particular, the conditional quantile method does not always produce a consistent estimate for the random-coefficient functions $\boldsymbol{\theta}(\cdot)$ when the monotonicity conditions are not satisfied. Further studies are needed.

ADDITIONAL REFERENCES

Brumback, B., and Rice, J. A. (1998), "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves" (with discussion), *Journal of the American Statistical Association*, 93, 961–994.

Cai, Z., Fan, J., and Li, R. (2000a), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.

——— (2000b), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–956.

Carroll, R. J., Ruppert, D., and Welsh, A. H. (1998), "Nonparametric Estimation via Local Estimating Equations," *Journal of the American Statistical Association*, 93, 214–227.

Chen, R., and Tsay, R. J. (1993), "Functional-Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001), "Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables," *Journal of the American Statistical Association*, 96, 605–619.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003), "Time-Dependent Diffusion Models for Term Structure Dynamics and the Stock Price Volatility," *Statistica Sinica*, 13, 965–992.

Fan, J., and Li, R. (2004), "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis," *Journal of the American Statistical Association*, 99, 710–723.

Fan, J., Lin, H., and Zhou, Y. (2006), "Local Partial-Likelihood Estimation for Life Time Data," *The Annals of Statistics*, 34, 290–325.

Fan, J., and Zhang, J. T. (2000), "Functional Linear Models for Longitudinal Data," *Journal of the Royal Statistical Society*, Ser. B, 62, 303–322.

Hastie, T. J., and Tibshirani, R. J. (1993), "Varying-Coefficient Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Hong, Y., and Lee, T.-H. (2003), "Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models," *Review of Economics and Statistics*, 85, 1048–1062.

Lin, D. Y., and Ying, Z. (2001), "Semiparametric and Nonparametric Regression Analysis of Longitudinal Data" (with discussion), *Journal of the American Statistical Association*, 96, 103–126.

Mercurio, D., and Spokoiny, V. (2004), "Statistical Inference for Time-Inhomogeneous Volatility Models," *The Annals of Statistics*, 32, 577–602.

Tian, L., Zucker, D., and Wei, L. J. (2005), "On the Cox Model With Time-Varying Regression Coefficients," *Journal of the American Statistical Association*, 100, 172–183.

Comment

Keith KNIGHT

First, I would like to congratulate the authors for a truly interesting and stimulating article. They have presented a very elegant methodology that should prove useful in many disciplines in which time series analysis is used. I find myself with really nothing to criticize; however, I comment on two issues: (a) the relationship between QAR(p) and AR(p) processes and (b) asymptotics for estimation in the infinite-variance case.

1. QAR VERSUS AR

QAR processes appear to be a very useful complement to AR processes, particularly in identifying local behavior in time series having different structure than the global behavior. However, viewed under the lens of classical time series analysis, the two processes are quite similar. In particular, it is interesting to note that the autocovariance function of a stationary QAR(p) process is simply that of a stationary (fixed) parameter AR(p) process; we have

$$y_t = \theta_0(U_t) + \theta_1(U_t) y_{t-1} + \cdots + \theta_p(U_t) y_{t-p} = E[\theta_1(U_t)] y_{t-1} + \cdots + E[\theta_p(U_t)] y_{t-p} + V_t,$$

where

$$V_t = \theta_0(U_t) + \big\{\theta_1(U_t) - E[\theta_1(U_t)]\big\} y_{t-1} + \cdots + \big\{\theta_p(U_t) - E[\theta_p(U_t)]\big\} y_{t-p}.$$

It is easy to verify that $\{V_t\}$ is a sequence of uncorrelated (but not independent) random variables. Thus classical model identification techniques based on autocorrelations, partial autocorrelations, and so on will tend to identify data generated from a QAR(p) process as a (fixed-parameter) AR(p) model. Although this is probably acceptable from certain viewpoints (e.g., prediction), it would also fail to identify structure in the

Keith Knight is Professor, Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, Canada (E-mail: *keith@utstat.toronto.edu*).

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000690

data that is potentially very informative and likely would be revealed by looking at autogressive quantiles.

In some data, it may be the extreme quantiles (i.e., τ close to 0 or 1) that reveal unusual structure. These cases are not discussed in this article, but there is a growing literature on extreme regression quantiles (see, e.g., Feigen and Resnick 1994; Smith 1994; Portnoy and Jurečková 1999; Knight 2001; Chernozhukov 2005).

2. INFINITE VARIANCE

The article assumes throughout that $\mathrm{var}[\theta_j(U_t)] < \infty$ for $j = 0, 1, \dots, p$. However, it is worthwhile to entertain situations in which at least $\mathrm{var}[\theta_0(U_t)]$ is infinite; this, of course, implies that the QAR process has infinite variance. (Whether it makes sense to allow $\mathrm{var}[\theta_j(U_t)]$ to be infinite for $j \geq 1$ is open to discussion.) In the case of (fixed-parameter) stationary AR(p) models, asymptotics of estimation in the infinite-variance case has been studied by a number of authors, including Kanter and Steiger (1974), Hannan and Kanter (1977), Gross and Steiger (1979), An and Chen (1982), and Davis, Knight, and Liu (1992). An interesting feature of estimation in infinite-variance AR models is that the autoregressive parameters can be estimated at a faster rate than the $O_p(n^{-1/2})$ rate obtained in the finite-variance case.

When $\mathrm{var}[\theta_0(U_t)] = \infty$, the dynamics of the QAR process are somewhat different than in the finite-variance case, and this difference will greatly affect the properties of the estimator $\widehat{\theta}(\tau)$. In particular, its asymptotic properties will be driven largely by the extreme values of $\{\theta_0(U_t)\}$.

The results of Davis et al. (1992) can be extended to autogressive quantile estimation for the QAR(p) process using the point process asymptotic techniques pioneered by Davis and Resnick (1985a,b; 1986). In particular, we assume that

$$P\big[|\theta_0(U_t)| > x\big] = x^{-\alpha}L(x)$$

for $0 < \alpha < 2$, where L is a slowly varying function, and that

$$\lim_{x\to\infty} \frac{P[\theta_0(U_t) > x]}{P[|\theta_0(U_t)| > x]} = \gamma \in [0, 1].$$

These two conditions are of course necessary and sufficient for the iid sequence $\{\theta_0(U_t)\}$ to lie in the domain of attraction of a stable law with index α. However, these conditions also imply that we can approximate distribution of the number of "large" $\theta_0(U_t)$'s by a Poisson distribution. More precisely, if $\{a_n\}$ satisfies

$$nP\big[|\theta_0(U_t)| > a_n x\big] = x^{-\alpha},$$

in which case $a_n = n^{1/\alpha}L^*(n)$, where L^* is another slowly varying function, then, for example, for any set A that does not contain 0, we have

$$\sum_{t=1}^{n} I[a_n^{-1}\theta_0(U_t) \in A] \xrightarrow{d} \mathrm{Poisson}(\mu(A)),$$

where

$$\mu(dx) = \alpha[\gamma x^{-\alpha-1}I(x > 0) + (1-\gamma)(-x)^{-\alpha-1}I(x < 0)]\,dx.$$

The dynamics of a QAR(p) process are determined largely by the extreme values of $\{\theta_0(U_t)\}$, that is, those values for which $|\theta_0(U_t)| > ka_n$. The number of such extreme values has approximately a Poisson distribution. In contrast, when $\mathrm{var}[\theta_0(U_t)]$ is finite, we have

$$\max_{1\leq t\leq n} n^{-1/2}|\theta_0(U_t)| \xrightarrow{p} 0,$$

which means that the number of $|\theta_0(U_t)|$'s exceeding $kn^{1/2}$ tends to 0.

Using point process asymptotics, we can sketch an analog to theorem 2 in the infinite-variance case. Under appropriate regularity conditions (similar to those given by Davis et al. 1992), we have that

$$\sqrt{n}\{\widehat{\theta}_0(\tau) - \theta_0(\tau)\} \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\right)$$

and

$$a_n\{\widehat{\theta}_1(\tau) - \theta_1(\tau), \dots, \widehat{\theta}_p(\tau) - \theta_p(\tau)\} \xrightarrow{d} \arg\min(Z),$$

where the two limits are independent. The objective function Z is actually an integral with respect to the (random) measure resulting from the limiting Poisson process representation as in the work of Davis et al. (1992). Here Z can be written as

$$Z(\mathbf{v}) = \sum_{s=1}^{\infty}\sum_{k=1}^{\infty}\big\{\rho_\tau\big(V_s(\mathbf{U}_k) - \mathbf{c}_s(\mathbf{U}_k)^\top \mathbf{v}\Delta_k\Gamma_k^{-1/\alpha}\big) - \rho_\tau(V_s(\mathbf{U}_k))\big\},$$

where $\{U_{k,s}: k \geq 1; -\infty < s < \infty\}$ is an array of iid uniform (on [0, 1]) random variables [with $\mathbf{U}_k = (\dots, U_{k,-1}, U_{k,0}, U_{k,1}, U_{k,2}, \dots)$], $\{\Delta_k\}$ is a sequence of iid random variables with

$$P(\Delta_k = 1) = 1 - P(\Delta_k = -1) = \gamma,$$

and $\Gamma_k = E_1 + \cdots + E_k$, where $\{E_k\}$ are independent exponential random variables with mean 1; the sequences $\{\Delta_k\}$, $\{E_k\}$, and $\{U_{k,s}\}$ are mutually independent. The vector $\mathbf{c}_s(\mathbf{U}) = (c_{s-1}(\mathbf{U}), \dots, c_{s-p+1}(\mathbf{U}))^\top$, where $\{c_s(\mathbf{U})\}$ are the coefficients in the (random) linear process representation of $\{Y_t\}$,

$$Y_t = \sum_{s=0}^{\infty} c_s(\mathbf{U})\theta_0(U_{t-s}).$$

Likewise, $V_s(\mathbf{U}_k)$ is also a linear process with random coefficients. The positive density assumption guarantees that Z will have a unique minimizer.

Asymptotic inference in the infinite-variance case is extremely difficult, because the limiting distribution is difficult to compute and depends on infinite-dimensional nuisance parameters. This is certainly in contrast with the finite-variance case, where asymptotic inference as outlined by the authors is (dare I say it) quite beautiful.

ADDITIONAL REFERENCES

An, H. Z., and Chen, Z. G. (1982), "On the Convergence of LAD Estimates in Autoregression With Infinite Variance," *Journal of Multivariate Analysis*, 12, 335–345.

Chernozhukov, V. (2005), "Extremal Quantile Regression," *The Annals of Statistics*, 33, 806–839.

Davis, R. A., Knight, K., and Liu, J. (1992), "M-Estimation for Autogressions With Infinite Variance," *Stochastic Processes and Their Applications*, 40, 145–180.

Davis, R. A., and Resnick, S. (1985a), "Limit Theory for Moving Averages of Random Variables With Regularly Varying Tail Probabilities," *The Annals of Probability*, 13, 179–195.
——— (1985b), "More Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages," *Stochastic Processes and Their Applications*, 20, 257–279.
——— (1986), "Limit Theory for the Sample Correlation Functions of Moving Average," *The Annals of Statistics*, 14, 533–558.
Feigen, P. D., and Resnick, S. I. (1994), "Limit Distributions for Linear Programming Time Series Estimators," *Stochastic Processes and Their Applications*, 51, 135–166.
Gross, S., and Steiger, W. L. (1979), "Least Absolute Deviation Estimates in Autoregression With Infinite Variance," *Journal of Applied Probability*, 16, 104–116.
Hannan, E. J., and Kanter, M. (1977), "Autoregressive Processes With Infinite Variance," *Journal of Applied Probability*, 14, 411–415.
Kanter, M., and Steiger, W. L. (1974), "Regression and Autoregression With Infinite Variance," *Advances in Applied Probability*, 6, 768–783.
Knight, K. (2001), "Limiting Distributions of Linear Programming Estimators," *Extremes*, 4, 87–104.
Portnoy, S., and Jurečková, J. (1999), "On Extreme Regression Quantiles," *Extremes*, 2, 227–243.
Smith, R. L. (1994), "Nonregular Regression," *Biometrika*, 81, 173–183.

Comment

Marc HALLIN and Bas J. M. WERKER

1. INTRODUCTION

Koenker and Bassett's 1978 *Econometrica* article (Koenker and Bassett 1978) has been rightly acclaimed as a major statistical breakthrough, restoring, after more than a century of Gaussian L_2 hegemony, the merits and popularity of the Laplacian L_1 views. Since then, quantile regression and its many refinements have ranked among the most fecund areas of theoretical as well as applied statistics.

Quantile regression is a major step toward the old statistical dream of describing the entire distribution of some variable of interest Y conditional on a set of covariates (possibly including Y's own past), as a substitute to the classical method of conditional location and/or variance modeling. The present article by Koenker and Xiao is a most welcome attempt to concretize that dream in a time series setting, by proposing a model [eq. (KX2); references to Koenker and Xiao's article throughout are labeled KX, to avoid confusion with our own labels] in which each conditional quantile depends on the past through a (quantile-specific) linear autoregressive relation.

The terms *quantile (auto)regression* and *(auto)regression quantile*, here and in the literature, are used for two distinct if closely related things: (a) a type of model, of which (KX2) is an example, and (b) an estimation method, which is used in section 3 of Koenker and Xiao's article. Our discussion successively addresses these two aspects: first quantile autoregression models, then autoregression quantile estimators.

Equation (KX2) is much more than a traditional model equation, as the solution $\{Y_t\}$ on the left side appears through its conditional quantiles. This fact induces a number of restrictions (related to the intrinsic monotonicity of quantile functions) that do not explicitly appear in the equation but play a major role at the modeling level. In Section 2 we discuss some of these implications.

Empirical autoregression quantiles in the context of (KX2) appear as a "natural" tool for estimation and indeed, in view of theorem 3.1, are doing a very good job. One may be concerned, however, about their behavior when (KX2) does not adequately describe the actual data-generating process. All the more so that, as the authors rightly mention at the beginning of their section 4, "linear quantile autoregression models should be cautiously interpreted as useful local approximations to more complex nonlinear global models." Thus in Section 3 we try to answer the question: What are empirical autoregression quantiles really estimating in such case?

Finally, in many applications, time series are either obtained through or subjected to cross-sectional or temporal aggregation. In Section 4 we discuss temporal aggregation issues in the context of quantile autoregression.

2. MODELING ISSUES

The article deals with two models: a random coefficient autoregressive (RCAR) model (KX1), and a quantile autoregression (QAR) model (KX2). Equation (KX1), as a stochastic difference equation, constitutes a traditional time series model, with Y_t expressed in terms of Y_{t-1} and an unpredictable "innovation." Under assumptions A1 and A2 (these conditions are presented as sufficient; under A1, however, one may wonder whether A2 is not also necessary), this model admits (thm. 2.1) a unique stationary solution $\{Y_t\}$.

Equation (KX2) is of a somewhat different nature. Here the solution $\{Y_t\}$ appears, on the left side, through its conditional quantile functions $\tau \mapsto Q_{Y_t}(\tau | Y_{t-1} = y_{t-1}, \dots, Y_{t-p} = y_{t-p})$, which by nature are monotonic increasing functions, whereas on the right side, the same solution appears through its realization at time $t-1$. On that right side, $\tau \mapsto \theta_0(\tau) + \sum_{j=0}^{p} \theta_j(\tau) Y_{t-j}$ thus also should be monotonic, $(Y_{t-1}, \dots, Y_{t-p})$ a.s.

This monotonicity constraint places restrictions on the functions $\tau \mapsto \theta_j(\tau)$. As explained by the authors, for $\{Y_t\}$ described by (KX1) to satisfy (KX2), (KX1) should be flanked with side conditions ensuring that this monotonicity constraint is satisfied. Conversely, it can be easily verified that a process $\{Y_t\}$ sat-

Marc Hallin is Professor Ordinaire, Institut de Statistique, ECARES, and Department of Mathematics, Free University of Brussels, Brussels, Belgium (E-mail: *mhallin@ulb.ac.be*). Bas J. M. Werker is Professor of Econometrics and Finance at CentER, Tilburg University, Tilburg, The Netherlands (E-mail: *Werker@TilburgUniversity.nl*). This research was supported by a PAI contract from the Belgian Federal Government and an Action de Recherche Concertée of the Communauté Française de Belgique.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000546

isfying (KX2) also satisfies (KX1) for some adequate iid uniform sequence $\{U_t\}$. Indeed, defining

$$U_t := Q^{\leftarrow}_{Y_t}(Y_t|Y_{t-1}, \ldots, Y_{t-p}) := F_{Y_t|Y_{t-1},\ldots,Y_{t-p}}(Y_t),$$

where $F_{Y_t|Y_{t-1},\ldots,Y_{t-p}}(y) := \mathrm{P}[Y_t \leq y|Y_{t-1}, \ldots, Y_{t-p}]$, the U_t's obviously (throughout, we assume that Y_t is absolutely continuous) are iid, uniform (over [0, 1]), independent of Y_{t-j}, $j \geq 1$, and such that under (KX2), the relation (KX1) between Y_t and U_t holds. It follows that the two models are equivalent if and only if the monotonicity constraints are added to (KX1). In their section 4 Koenker and Xiao discuss some consequences of these constraints; we develop some further consequences here.

For simplicity, we restrict attention to the AR(1) case,

$$Y_t = \theta_0(U_t) + \theta_1(U_t)Y_{t-1}, \qquad t \in \mathbb{Z}, \tag{1}$$

where θ_0 and θ_1 are such that a stationary solution $\{Y_t\}$ with absolutely continuous distribution exists. Monotonicity Y_{t-1} a.s. of $\tau \mapsto \theta_0(\tau) + \theta_1(\tau)Y_{t-1}$ implies that two distinct quantile functions $y \mapsto \theta_0(\tau_1) + \theta_1(\tau_1)y$ and $y \mapsto \theta_0(\tau_2) + \theta_1(\tau_2)y$ $(0 < \tau_1 \neq \tau_2 < 1)$ cannot cross over the support of the stationary distribution of Y_t. Therefore, one of the following holds:

a. $\theta_1(\tau)$ is a constant $\theta_1 \in [-1, 1]$, $\tau \mapsto \theta_0(\tau)$ is strictly monotone increasing, and model (1) reduces to the traditional fixed-coefficient AR(1) model.

b. The support of Y_t is a finite interval that, possibly after adequate linear transformation, can be assumed to be [0, 1] without loss of generality.

c. The support of Y_t is a half-line that, possibly after adequate linear transformation, can be assumed to be $[0, \infty)$ without loss of generality.

Case a is of limited interest in this context. Turning to case b, eq. (1) rewrites as

$$Y_t = \theta_0(U_t)(1 - Y_{t-1}) + (\theta_0(U_t) + \theta_1(U_t))Y_{t-1}, \qquad t \in \mathbb{Z}. \tag{2}$$

Because Y_t has support [0, 1], this implies that $\tau \mapsto \theta_0(\tau)$ and $\tau \mapsto \theta_0(\tau)+\theta_1(\tau)$ are two monotone increasing functions, with values in [0, 1], such that

$$0 = \min\bigl(\theta_0(0), \theta_0(0) + \theta_1(0)\bigr)$$

and

$$\max\bigl(\theta_0(1), \theta_0(1) + \theta_1(1)\bigr) = 1.$$

In view of (2), the stationary solution of (1) thus is obtained as a convex linear combination, with random coefficients $(1 - Y_{t-1})$ and Y_{t-1} depending on the past, of two perfectly dependent iid sequences, $\theta_0(U_t)$ and $(\theta_0(U_t) + \theta_1(U_t))$, where U_t is independent of $Y_{t-j}, j \geq 1$.

As for case c, strict monotonicity of the quantiles conditional on $Y_{t-1} = 0$ requires that $\tau \mapsto \theta_0(\tau)$ be nonnegative and strictly increasing, with $\theta_0(0) = 0$; moreover, because Y_{t-1} can be arbitrary large, $\tau \mapsto \theta_1(\tau)$ also must be nonnegative and nondecreasing. Recall, however, that $\mathrm{E}[\theta_1^2(U_t)]$ must be less than 1 if a stationary solution is to exist.

3. CONSISTENCY ISSUES

The proposed estimators are the sample autoregression quantiles. The remarks in the previous section indicate that (KX1) and (KX2)–(KX3) are not necessarily equivalent, and hence the conditional quantile interpretation of $y \mapsto \theta_0(\tau) + \theta_1(\tau)y$ might not be valid. Therefore, consistency issues under general assumptions deserve some attention. What do $\hat{\theta}_0(\tau)$ and $\hat{\theta}_1(\tau)$ estimate (in a consistent way) when computed from a general stationary first-order Markovian process $\{Y_t\}$ with transition probabilities

$$F(y|x) = \mathrm{P}\{Y_t \leq y|Y_{t-1} = x\}?$$

Here we denote by $x \mapsto F(y|x)$ the value at y of the distribution function of Y conditional on $X = x$. Let (Y, X) denote a generic random variable with the same distribution as (Y_t, Y_{t-1}). Estimation of $\theta_0(\tau)$ and $\theta_1(\tau)$ for given τ involves minimization of the sample equivalent of

$$\mathrm{E}\bigl\{(Y - \theta_0 - \theta_1 X)[\tau - I(Y - \theta_0 - \theta_1 X < 0)]\bigr\}, \tag{3}$$

which can be rewritten as

$$\mathrm{E}\left\{\int_{y=-\infty}^{\theta_0+\theta_1 X} (\tau - 1)(y - \theta_0 - \theta_1 X)\,\mathrm{d}F(y|X) + \int_{y=\theta_0+\theta_1 X}^{\infty} \tau(y - \theta_0 - \theta_1 X)\,\mathrm{d}F(y|X)\right\}. \tag{4}$$

Differentiating with respect to (θ_0, θ_1), we find that the proposed autoregression quantiles generally estimate the value of (θ_0, θ_1) that solves (still for given τ)

$$0 = \mathrm{E}\left\{[F(\theta_0 + \theta_1 X|X) - \tau]\begin{pmatrix}1\\X\end{pmatrix}\right\}. \tag{5}$$

One easily verifies that under model (KX2), condition (5) is satisfied, because in that case

$$F\bigl(\theta_0(\tau) + \theta_1(\tau)X|X\bigr) - \tau = 0 \quad \text{a.s.} \tag{6}$$

However, even when (KX2) does not hold—that is, when $Q_{Y_t}(\tau|Y_{t-1})$ is not linear in Y_{t-1}—eq. (5) still often has a unique solution for (θ_0, θ_1), that does not satisfy (6). In that case, estimation of the (affine) quantile autoregressive model clearly amounts to finding an affine approximation to the nonlinear quantile function $y \mapsto Q_{Y_t}(\tau|Y_{t-1} = y)$—an objective fully in line with the authors' comments at the beginning of their section 4. A natural question then is: which is the criterion implicitly considered in this linear approximation?

For a positive weight function $x \mapsto w_\tau(x)$, consider the minimization problem

$$\min \frac{1}{2}\mathrm{E}\bigl\{\bigl(F(\theta_0 + \theta_1 X|X) - \tau\bigr)^2 w_\tau(X)\bigr\}. \tag{7}$$

The first-order conditions for this problem are

$$0 = \mathrm{E}\left\{\bigl(F(\theta_0 + \theta_1 X|X) - \tau\bigr)F_y(\theta_0 + \theta_1 X|X) \times w_\tau(X)\begin{pmatrix}1\\X\end{pmatrix}\right\}, \tag{8}$$

where $F_y(y|x)$ denotes the value at y of the conditional density of Y given $X = x$. Note that if the affine approximation $\theta_0 + \theta_1 x$ to the true quantile function $Q_Y(\tau|X = x)$ is sufficiently accurate (uniform in x), we have that $F_y(\theta_0 + \theta_1 x|x)$ is close to the conditional density of Y given $X = x$ at the τth (conditional on $X = x$) quantile.

Comparing with (5), we see that (affine) quantile autoregression, as proposed by Koenker and Xiao in their section 3, actually uses in (7) a weighting function inversely related to the conditional density of Y given X, evaluated at the τth quantile—which, intuitively, is not the best choice.

4. TEMPORAL AGGREGATION ISSUES

Time series models are often considered in the analysis of data that are, at least theoretically, available at several frequencies. Given that it is often difficult to motivate the use of a model at a specific frequency, it is interesting to see how those models behave under temporal aggregation.

Consider once more a stationary process $\{Y_t\}$ satisfying (KX1). Clearly, we have

$$Y_t = \tilde{\theta}_0(U_t, U_{t-1}) + \tilde{\theta}_1(U_t, U_{t-1})Y_{t-2}, \tag{9}$$

where $\tilde{\theta}_0(U_t, U_{t-1}) := \theta_0(U_t) + \theta_0(U_{t-1})\theta_1(U_t)$ and $\tilde{\theta}_1(U_t, U_{t-1}) := \theta_1(U_{t-1})\theta_1(U_t)$. The dependence between $\tilde{\theta}_0(U_t, U_{t-1})$ and $\tilde{\theta}_1(U_t, U_{t-1})$ in general is not perfect. As a result, the aggregated process $\{Y_{2t}\}$ cannot be written in the form (KX1). It would be interesting to investigate whether the class of time series model described by (KX1) can be extended to a class that is closed under temporal aggregation.

Comment

Christian M. HAFNER and Oliver B. LINTON

In our discussion of the article of Koenker and Xiao (2006), henceforth KX, we emphasize some applications of this methodology to financial econometrics. A central model there is the GARCH(1, 1) process (Bollerslev, Engle, and Nelson 1994) in which $y_t = \varepsilon_t\sigma_t$, where the ε_t's are iid with mean 0 and variance 1, whereas $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \gamma y_{t-1}^2$. We investigate the relationship between the QAR class of processes and the generalized autoregressive conditional heteroscedasticity (GARCH) process.

Consider the QAR(1) process with

$$y_t = \theta_0(U_t) + \theta_1(U_t)y_{t-1}, \tag{1}$$

where θ_0 and θ_1 are functions on $[0, 1]$. If the function $u \mapsto \theta_0(u) + \theta_1(u)y_{t-1}$ is monotone increasing (for the given y_{t-1}), then the conditional quantile function is linear in y_{t-1}, $Q_\tau(y_t|y_{t-1}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1}$; if it is decreasing, then $Q_\tau(y_t|y_{t-1}) = \theta_0(1-\tau) + \theta_1(1-\tau)y_{t-1}$. Furthermore,

$$E[y_t|y_{t-1}] = E[\theta_0(U_t)] + E[\theta_1(U_t)]y_{t-1}$$

and

$$\begin{aligned}\operatorname{var}[y_t|y_{t-1}] &= \operatorname{var}[\theta_0(U_t)] + \operatorname{var}[\theta_1(U_t)]y_{t-1}^2 \\ &\quad + 2\operatorname{cov}[\theta_0(U_t), \theta_1(U_t)]y_{t-1}.\end{aligned}$$

Therefore, if $\operatorname{cov}[\theta_0(U_t), \theta_1(U_t)] = 0$ and $\operatorname{var}[\theta_j(U_t)] > 0$, $j = 0, 1$, then this process has the same conditional variance as an ARCH(1) process. If also $E[\theta_0(U_t)] = E[\theta_1(U_t)] = 0$, then it has conditional mean 0. We separate the question of mean and variance dynamics in the sequel and so assume that the process has conditional mean 0.

Note that in this example, y_t is a semistrong ARCH(1) process in the sense of Drost and Nijman (1993); that is, it is like a strong ARCH(1) in terms of its first two conditional moments. However, its higher-order properties may be quite different from those of the strong ARCH model. The conditional skewness of (4) is

$$\begin{aligned}&\operatorname{skew}[y_t|y_{t-1}] \\ &\quad = E[\theta_0^3(U_t)] + E[\theta_1^3(U_t)]y_{t-1}^3 \\ &\qquad + 3E[\theta_0(U_t)\theta_1^2(U_t)]y_{t-1}^2 + 3E[\theta_0^2(U_t)\theta_1(U_t)]y_{t-1}.\end{aligned}$$

This is quite a flexible function of the past information set; it is not restricted to be an even or odd function of the past, whereas the conditional skewness of the strong ARCH(1) process, $\operatorname{skew}[y_t|y_{t-1}] = \sigma_t^3E(\varepsilon_t^3) = E(\varepsilon_t^3)(\omega + \gamma y_{t-1}^2)^{3/2}$, is an even function of lagged values.

The QAR(p) can be seen to nest the ARCH(p) in terms of second-order properties. To nest the GARCH(1, 1) process [or ARCH(∞)], one has to extend the QAR processes to infinite order. Thus, suppose that

$$y_t = \theta_0(U_t) + \sum_{j=1}^{\infty}\theta_j(U_t)y_{t-j}, \tag{2}$$

where the $\theta_j(\cdot)$'s are such that the process is weakly stationary and mixing. A strong condition that would ensure this is $\sup_{\tau\in[0,1]}\sum_{j=1}^{\infty}|\theta_j(\tau)| < \infty$, but clearly some moment conditions would suffice. If also the function $u \mapsto \theta_0(u) + \sum_{j=1}^{\infty}\theta_j(u)y_{t-j}$ is increasing for given $\{y_{t-j}\}_{j=1}^{\infty}$, then the corresponding conditional quantile function for all $\tau \in [0, 1]$ satisfies

$$Q_\tau(y_t|y_{t-1}, \ldots) = \theta_0(\tau) + \sum_{j=1}^{\infty}\theta_j(\tau)y_{t-j}. \tag{3}$$

Christian M. Hafner is Professor, Institute of Statistics, Catholic University of Louvain, 1348 Louvain-la-Neuve, Belgium. Oliver B. Linton is Professor of Econometrics, Department of Economics, London School of Economics, London WC2A 2AE, U.K. (E-mail: *o.linton@lse.ac.uk*). This research was supported by the Economic and Social Science Research Council of the United Kingdom.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000717

Again, one might ask for functions θ_j that are mutually orthogonal, in which case

$$\mathrm{var}[y_t|y_{t-1},\ldots] = \mathrm{var}[\theta_0(U_t)] + \sum_{j=1}^{\infty} \mathrm{var}[\theta_j(U_t)]y_{t-j}^2,$$

and so, provided that $\mathrm{var}[\theta_j(U_t)] = \gamma\beta^{j-1}$ for suitable β, $\gamma \in (0,1)$, one would achieve the same conditional variance as in a GARCH(1, 1) process. Therefore, the process (2) nests the GARCH(1, 1) as far as second-order properties, but has different higher-order properties.

In the foregoing discussion we have assumed some properties of θ_j to guarantee the quantile interpretation and the variance nesting. As discussed by KX, a property that would guarantee (3) for all positive $\{y_{t-j}\}_{j=1}^{\infty}$ is comonotonicity, where all of the functions θ_j are strictly increasing. However, then it is impossible for the functions to be mutually orthogonal, as the following lemma shows.

Lemma. There do not exist strictly increasing functions f and g on $[0,1]$ such that: $\int_0^1 f(x)\,dx = 0$, $\int_0^1 g(x)\,dx = 0$, and $\int_0^1 g(x)f(x)\,dx = 0$.

Proof. Let x_f be such that $f(x_f) = 0$ and for all $x > x_f$, $f(x) > 0$, whereas for all $x < x_f$, $f(x) < 0$, and let x_g be such that $g(x_g) = 0$ and for all $x > x_g$, $g(x) > 0$, whereas for all $x < x_g$, $g(x) < 0$. If $x_f = x_g$, then $g(x), f(x) < 0$ for all $x < x_f$, and so $\int_0^1 g(x)f(x)\,dx = \int_0^{x_f} g(x)f(x)\,dx + \int_{x_f}^1 g(x)f(x)\,dx > 0$. Suppose, that without loss of generality, $x_f < x_g$ and write

$$\int_0^1 g(x)f(x)\,dx = \int_0^{x_f} g(x)f(x)\,dx + \int_{x_f}^{x_g} g(x)f(x)\,dx + \int_{x_g}^{1} g(x)f(x)\,dx.$$

Note that $\int_{x_f}^{x_g} g(x)f(x)\,dx > g(x_f)\int_{x_f}^{x_g} f(x)\,dx$, $\int_{x_g}^1 g(x)f(x)\,dx > g(x_g)\int_{x_g}^1 f(x)\,dx$, and $\int_0^{x_f} g(x)f(x)\,dx > g(x_f)\int_0^{x_f} f(x)\,dx$, and because $g(x_g) > g(x_f)$, we have $\int_0^1 g(x)f(x)\,dx > g(x_f) \times \int_0^1 f(x)\,dx = 0$, a contradiction.

In conclusion, either orthogonality or comonotonicity must be sacrificed. In the former case, it seems impossible to obtain the semistrong GARCH(1, 1) conditional variance, and conditions for weak stationarity are more complicated. Nevertheless, the model is interesting and important in its own right. In the latter case, it is harder to justify the quantile interpretation. In the special case of (4), one can find examples that satisfy the weaker condition that the right side function is monotonic in τ. For example, suppose that $\theta_0(u) = a(u - 1/2)$ and

$$\theta_1(u) = \begin{cases} b(u - 1/4) & \text{if } u \le 1/2 \\ b(3/4 - u) & \text{if } u > 1/2, \end{cases}$$

where a and b are positive constants. This satisfies $\int_0^1 \theta_0(u) \times \theta_1(u)\,du = 0$ for any a and b. This process is stationary iff $E[\theta_1(U_t)^2] < 1$, which holds if $b < \sqrt{48}$. By further restricting b, one can get monotonicity of Q_τ. Note that $\theta_0(\tau) + \theta_1(\tau)y_{t-1}$ is monotone in τ as long as $|y_{t-1}| \le a/b$. Thus y_t should be bounded. To find conditions for this, note that if $y_0 = 0$, then $|y_1| \le a/2$, $|y_2| \le a/2(1 + b/4)$, $|y_3| \le a/2(1 + b/4 + b^2/16)$, and so on, where the bounds converge iff $b < 4$, so that $|y_t| \le a/(2 - b/2)$. To satisfy the monotonicity condition on the whole support of y_t, we further must impose $b < 4/3$. This example process has quantile dependence at lag 1. Figure 1 shows the quantilogram introduced by Linton and Whang (2006) for the process with $a = b = 1$ at lag 1 as a function of τ. It is interesting that the median has positive dependence and the outer quantiles have negative dependence. There is no dependence at the quartiles, $\tau = 1/4$ and $\tau = 3/4$, because $\theta_1(\cdot)$ is 0 in that case. The quantilogram is 0 for all higher-order lags.

Is it necessary to have the function $u \mapsto \theta_0(u) + \sum_{j=1}^{\infty} \theta_j(u) \times y_{t-j}$ strictly increasing for the conditional quantile interpretation? A sufficient condition for the quantile of a nonlinear function of a random variable to be the nonlinear function of the

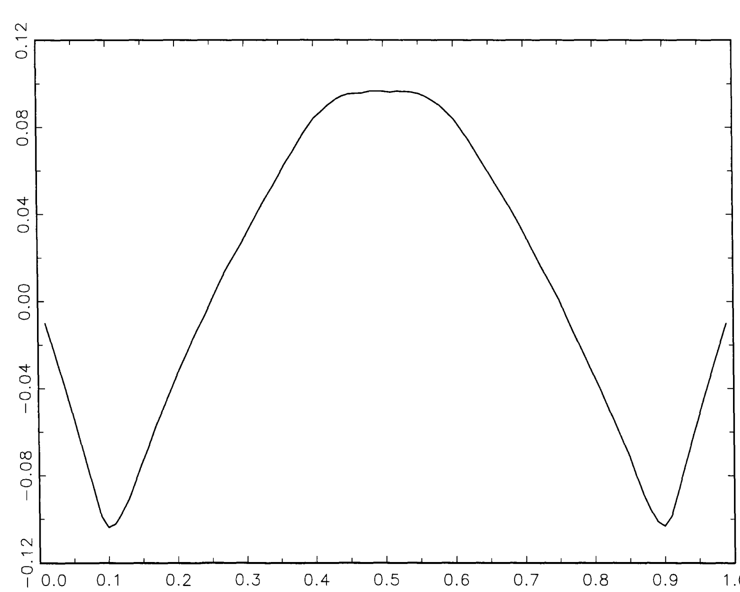

Figure 1. Quantilogram at Lag One as a Function of the Quantile τ for the QAR(1) Process With $\theta_0(u) = (u - 1/2)$ and $\theta_1(u) = (u - 1/4)\mathbb{1}_{u\le 1/2} + (3/4 - u)\mathbb{1}_{u>1/2}$. The graph was produced by generating a realization of length 500,000 of this process and computing its quantilogram for $\tau = .01, .02, \ldots, .99$.

quantile of the random variable is that the function be strictly increasing throughout its domain. But this is far from necessary. Let U be the given random variable and let g be a function. Then suppose that Q_τ is the τ quantile of U and that g is strictly increasing at Q_τ with the further property that

$$\inf_{u-Q_\tau \geq \epsilon} g(u) > g(Q_\tau) \tag{4}$$

and

$$\sup_{u-Q_\tau \leq -\epsilon} g(u) < g(Q_\tau). \tag{5}$$

Then it can be seen that the τ quantile of $g(U)$ is $g(Q_\tau)$. For example, take the function $\Phi(u)+u\phi(u)$; this is strictly increasing at $u=0$, but at $u=\pm\sqrt{2}$, the function turns around and asymptotics to 1 or 0. But this does not affect the mid-range quantiles. Suppose that we drop the requirement of monotonicity in the QAR(1) case and consider the orthogonal functions

$$\theta_0(\tau) = \sqrt{3\alpha_0}(2\tau - 1) \qquad \text{and}$$
$$\theta_1(\tau) = \sqrt{5\alpha_1}(6\tau^2 - 6\tau + 1),$$

where α_0 and α_1 are parameters of a corresponding ARCH(1) model. Obviously, θ_1 is not monotone on [0, 1]. Nevertheless, $Q_\tau(y_t|y_{t-1}=x) = \theta_0(\tau) + \theta_1(\tau)x$ satisfies (4) and (5) for x not too large. For example, if $\alpha_0 = .9$ and $\alpha_1 = .1$ [so that $\text{var}(y_t) = 1$], then this holds for $\tau \in [.225, .775]$ and $x \in [-1, 1]$ (which covers about 68% of the distribution of y_t), so that Q_τ is the conditional quantile for these ranges of τ and x. Reducing the range of x to $[-.8, .8]$, covering about 46% of the distribution, Q_τ is well defined for $\tau \in [.031, .969]$. Finally, note that even when no form of monotonicity is satisfied, the conditional quantiles are well defined but may be just a more complicated functional of θ_j. For example, suppose that $y_t = \theta_0(U_t)$, where $\theta_0(u) = \sin^2(2\pi\Phi^{-1}(u))$, where Φ is the standard normal cdf; then $\theta_0(\tau)$ is not the τ-quantile of y_t. This can be shown to be $.5(1-\cos\tau\pi)$.

The model (3) is very general and quite difficult to estimate without further restrictions on the functions θ_j; thus so we investigate some special cases that might be more amenable. Suppose that $\theta_j(u) = c_{kj}m_k(u)$ for functions m_k, $k = 1, \ldots, K$ and for all $j \geq 1$, in which case

$$y_t = \theta_0(U_t) + \sum_{k=1}^{K} m_k(U_t) \sum_{j=1}^{\infty} c_{kj} y_{t-j}. \tag{6}$$

This is a sort of factor structure. In the special case where $K = 1$ with θ_0 and m mutually orthogonal and $c_j \geq 0$ for all j, we can justify the quantile interpretation in some examples, and we have

$$E[y_t|y_{t-1},\ldots] = E[\theta_0(U_t)] + E[m(U_t)] \sum_{j=1}^{\infty} c_j y_{t-j}$$

and

$$\text{var}[y_t|y_{t-1},\ldots] = \text{var}[\theta_0(U_t)] + \text{var}[m(U_t)] \left(\sum_{j=1}^{\infty} c_j y_{t-j}\right)^2.$$

Similarly, the third cumulant depends on the first three powers of $\sum_{j=1}^{\infty} c_j y_{t-j}$. The conditional variance in this case is similar to that of the volatility model considered by Robinson (1991, 8.16).

For some special cases, we can also find stationarity conditions of the QAR(∞) model. For example, if $c_j = \gamma\beta^{j-1}$, then the model can be written as a bivariate Markov chain,

$$\begin{pmatrix} y_t \\ h_t \end{pmatrix} = \mathbf{A}_t \begin{pmatrix} y_{t-1} \\ h_{t-1} \end{pmatrix} + \mathbf{V}_t,$$

with stochastic coefficient matrix

$$\mathbf{A}_t = \begin{pmatrix} \gamma m(U_t) & m(U_t) \\ \gamma\beta & \beta \end{pmatrix}$$

and $\mathbf{V}_t = (\theta_0(U_t), 0)^\top$. Following the same line of proof as in theorem 1 of KX, a necessary and sufficient condition for covariance stationarity is that all eigenvalues of $E[\mathbf{A}_t \otimes \mathbf{A}_t]$ have modulus <1.

We now turn to estimation of the QAR(∞) model. Estimation of the general model (2) is quite challenging. One could try to fit a long QAR(p) for each quantile τ by the QAR methods of KX, but this is likely to work poorly without some restrictions on the functions $\theta_j(\cdot)$. One possibility is to assume that $\sup_{u\in[0,1]} |\theta_j(u)| \leq \gamma\beta^{j-1}$ for suitable $\beta, \gamma \in (0, 1)$, thereby reducing the size of the parameter space; one may be able to use the methods of Koenker and Ng (2005) here. The special case (6) is more amenable. By parameterizing $c_j = c_j(\boldsymbol{\eta})$ with $\boldsymbol{\eta} \in \mathbb{R}^p$ and $c_j(\boldsymbol{\eta})$ decaying geometrically, one can estimate $\mathbf{m}$ and $\boldsymbol{\eta}$ by pooling information across quantiles. Thus let $\tau_1, \ldots, \tau_K$ be some given set of quantiles and define the objective function

$$S_T(\boldsymbol{\theta}_0, \mathbf{m}, \boldsymbol{\eta}) = \sum_{k=1}^{K} \sum_{t=1}^{T} \rho_{\tau_k}\left(y_t - \theta_{0k} + m_k \sum_{j=1}^{t-1} c_j(\boldsymbol{\eta}) y_{t-j}\right) w_k,$$

where $\mathbf{m} = (m_1, \ldots, m_K)^\top = (m(\tau_1), \ldots, m(\tau_K))^\top$ and $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0K})^\top$ are parameter vectors, whereas $\rho_\tau(e) = e(\tau - \mathbb{1}(e < 0))$. We introduce an additional weighting scheme w_k to improve efficiency. We have a total of $p + 2K$ parameters with KT "observations," and so the parameters should be identified under some restrictions. Perhaps one can impose the restrictions $\theta_{01} < \cdots < \theta_{0K}$ and $m_1 < \cdots < m_K$ using the methods of Koenker and Ng (2005).

We end with some questions and comments. The first is about diagnostics for the QAR processes. For linear processes, there is a relationship between the correlogram and the parameters of the process, but there does not seem to be such a simple relation here. So how does one identify the correct lag order to use? It seems ironic that one needs a finite variance in the error terms (assumption A.3) for estimation of a quantile regression parameter, usually this is not required. In the testing part, is it possible to bypass the estimation of the error density f? Is there anything to say about forecasting? Estimation of value at risk (VAR) is a natural application of this model, and in that case one is interested in small quantiles (i.e., $\tau = .01$ and .05). It is not clear that the moderate quantile asymptotics are the right ones to use in this case; some extreme quantile asymptotics may be needed.

In conclusion, we think this is a very important article that opens up many future lines of research.

ADDITIONAL REFERENCES

Bollerslev, T., Engle, R. F., and Nelson, D. (1994), "ARCH Models," in *The Handbook of Econometrics*, Vol. IV, eds. D. F. McFadden and R. F. Engle III, Amsterdam: North-Holland.

Drost, F., and Nijman, T. (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

Koenker, R., and Ng, P. (2005), "Inequality Constrained Quantile Regression," *The Indian Journal of Statistics*, 67, 418–440.

Linton, O., and Whang, Y.-J. (2006), "The Quantilogram: With an Application to Evaluating Directional Predictability," *Journal of Econometrics, Special Issue*, forthcoming.

Robinson, P. M. (1991), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67–84.

Comment

P. M. Robinson

My remarks about this article are organized within four points. The first of these notes the close connection between the authors' random-coefficient model and some earlier models that were not formulated in terms of "random-coefficients." The second comments on the model's identification. The third point queries the transition from the authors' basic model (1) to their quantile version (2). Finally, I request a clarification of the sufficient conditions for the theorems justifying quantile-based inference.

1. Write (1)/(5) as

$$y_t = \mu_0 + \sum_{j=1}^{p} a_j y_{t-j} + u_t + \sum_{j=1}^{p} \eta_{jt} y_{t-j},$$

where $a_j = E\theta_j(u_t)$ and $\eta_{jt} = \theta_j(U_t) - a_j$. The conditional mean of y_t given $\mathcal{F}_{t-1}$ is

$$E(y_t|\mathcal{F}_{t-1}) = \mu_0 + \sum_{j=1}^{p} a_j y_{t-j}. \tag{A}$$

The conditional mean for both standard constant-coefficient AR(p) models, and the "usual" random-coefficient AR(p) models (in which coefficients are independent of errors) is also of form (A). The "usual" random-coefficient models go back at least as far as Anděl (1976), but my impression is that they have not been greatly used in practice; if I am correct here, a partial explanation may be the inability to distinguish them from constant-coefficient models at the conditional mean level (typically, using second-moment information).

On the other hand, because the "usual" random-coefficient models were first developed, conditional heteroscedasticity and asymmetry (despite possibly symmetric innovations) have emerged as an important feature of many time series, notably financial ones. Dependence of coefficients on errors allows the authors' model to describe both phenomena, thereby distinguishing it from both the constant-coefficient and "usual" random-coefficient models. However, the authors' model overlaps with others already in the literature. Assuming $E\theta_j^2(U_t) < \infty, j = 0, 1, \dots, p$, the conditional variance of y_t given $\mathcal{F}_{t-1}$ is

$$V(y_t|\mathcal{F}_{t-1}) = \gamma_{00} + 2\sum_{j=1}^{p} \gamma_{0j} y_{t-j} + \sum_{j=1}^{p}\sum_{k=1}^{p} \gamma_{jk} y_{t-j} y_{t-k}, \tag{B}$$

where $\gamma_{jk} = \mathrm{cov}(\theta_j(U_t), \theta_k(U_t))$. Suppose first that

$$\theta_j(\tau) = b_j \theta_0(\tau), \qquad 1 \le j \le p, \tag{C}$$

for some constants b_j, and let $V(\theta_0(U_t)) = 1$ with no loss of generality. Then the $(p+1) \times (p+1)$ matrix with (j,k)th element $\gamma_{j-1,k-1}$ has unit rank, and (B) reduces to

$$V(y_t|\mathcal{F}_{t-1}) = \left(b_0 + \sum_{j=1}^{p} b_j y_{t-j}\right)^2. \tag{D}$$

This is a special case of a model proposed by Robinson [1991, formula (16)]. With respect to the authors' discussion of asymmetry, Robinson (1991) pointed out that in (D), with martingale difference levels y_t, $\mathrm{cov}(y_{t+j}^2, y_t)$ can be nonzero for $j \ge 1$ even when $\theta_0(U_t)$ is distributed symmetrically. His model was formulated principally in terms of the conditional-variance property, but subsequently Giraitis, Leipus, Robinson, and Surgailis (2004) considered a corresponding dynamic model for levels that is analogous to a special case of that of Koenker and Xiao, calling

$$y_t = \left(b_0 + \sum_{j=1}^{p} b_j y_{t-j}\right)\eta_{0t} \tag{E}$$

a linear ARCH (LARCH) model; they studied in detail its leverage effect, as well as conditions for stationarity. There have been other developments of this kind of model. Giraitis, Robinson, and Surgailis (2000) established a long-memory property of powers generated by an infinite autoregressive extension of (E) that started from the long-memory version that Robinson (1991) described for his conditional-variance form. Giraitis and Surgailis (2002) then replaced the left side of (E) by an infinite AR structure, also to model (long-memory) autocorrelation in levels; despite Giraitis and Surgailis' focus on long memory, their model (1.1) formally covers short memory. I emphasize that all of these models are based on the structure (C), which is special relative to Koenker and Xiao's $\theta_j(\tau)$. However, Sentana (1995) extended (D) to a "quadratic ARCH" model with conditional variance function of identical generality to (B), which is a consequence of the authors' model (1) without the restriction (C), and also discussed asymmetry.

P. M. Robinson is Tooke Professor of Economic Science and Statistics, London School of Economics, London WC2A 2AE, U.K. (E-mail: *p.m.robinson@lse.ac.uk*). This research was supported by ESRC grant R000239936.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000708

2. The formulas (A), (B), and (D) also prompt a point about model choice. In (1) and (2), p represents a maximal order, due to the possibility of some a_j vanishing, or degeneracy of some $\theta_j(U_t)$. (If this is known, it should be imposed for the sake of parsimony.) In their first empirical example, the authors apparently apply a rule for selecting p that pertains to (A) and is based on second moments; if the order in (B) is greater than in (A), then this rule will tend to underfit.

3. Now I want to move on from (1) to ask precisely how it leads to the conditional quantile function (2), which is then used in constructing rules of large sample inference. The authors assume that the right side of (1) is monotone increasing in U_t. However, it also depends on $y_{t-1}, \ldots, y_{t-p}$. Certainly, nonnegativity of all $y_{t-1}, \ldots, y_{t-p}$ is sufficient for a monotone increase of all $\theta_j(\tau)$ to imply the same of the right side of (1). But a monotone increase of $\theta_j(\tau)$ implies that $\theta_j(\tau)y_{t-j}$ is monotone decreasing when $y_{t-j} < 0$, and in general for some configurations of $y_{t-1}, \ldots, y_{t-p}$, a monotone increase in all of the $\theta_j(\tau)$ does not imply the same for the right side of (1). Much raw data are nonnegative, but negative values can occur when transformations (such as logs) are used. Gaussian errors $\theta_0(U_t)$, as in the authors' special case (4), would imply that arbitrarily large negative y_{t-j} can occur with positive probability.

4. The authors discuss in section 4 monotonicity issues in relation to their QAR model (2), but I could not see there or elsewhere in the article an answer to the following query, which relates to the preceding point. In their statements of theorem 1 and corollaries 1 and 2, which concern the behavior of partial sums of y_t, the authors complete their assumptions A.1–A.3 by indicating the underlying model that is assumed, (1), (6), and (9). In contrast, there are no such model assumptions in the statements of theorem 2 or of the theorems in section 5. These results give useful asymptotic properties of quantile estimates and consequent test statistics and are surely the main theoretical results of interest to practitioners. Thus I would be grateful if the authors would complete their statements by confirming sufficient conditions on the $\theta_j(\tau)$. Nonnegativity of the y_t's can be seen in simple cases to follow from certain restrictions on the support of the $\theta_j(\tau)$, but it would be nice to elucidate conditions for the general model (1). But because the authors do not mention nonnegativity of y_t, perhaps they can characterize a more general setting.

ADDITIONAL REFERENCES

Anděl, J. (1976), "Autoregressive Series With Random Parameters," *Mathematiche Operationsforschung Statistik*, 7, 735–741.

Giraitis, L., Leipus, R., Robinson, P. M., and Surgailis, D. (2004), "LARCH, Leverage and Long Memory," *Journal of Financial Econometrics*, 2, 177–210.

Giraitis, L., Robinson, P. M., and Surgailis, D. (2000), "A Model for Long-Memory Conditional Heteroscedasticity," *The Annals of Applied Probability*, 10, 1002–1024.

Giraitis, L., and Surgailis, D. (2002), "Arch-Type Bilinear Models With Double Long Memory," *Stochastic Processes and Their Applications*, 100, 275–300.

Robinson, P. M. (1991), "Testing for Strong Autocorrelation and Dynamic Conditional Heteroscedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67–89.

Sentana, E. (1995), "Quadratic ARCH Models," *Review of Economic Studies*, 62, 639–661.

Rejoinder

Roger KOENKER and Zhijie XIAO

We are very grateful to the editors and all of the discussants for their valuable contributions and their encouragement. We have tried to adhere carefully to the dictum of Donoho (2002): "In developing methodology, leave room for improvement. It is absolutely crucial not to kill a field by doing too good a job on the first outing." In this, at least, we feel we have succeeded, as the discussants have amply demonstrated. There is much more to be done!

1. MODELS: CAN'T LIVE WITH THEM, CAN'T LIVE WITHOUT THEM

Statistical models almost inevitably inspire a love–hate ambivalence; we would like to believe in them absolutely, but we are schooled to be skeptical. The QAR model exemplifies this tension. It arose as an attempt to broaden the scope of classical time series models that rely too heavily on additive iid innovations, moment information, and location-shift linear structure. Relaxing the iid innovation assumption while retaining the linearity of the specification poses some inherent conflicts, as several of the discussants have noted. We begin by addressing these monotonicity concerns, and then turn to other issues.

2. MONOTONICITY

From the outset we have stressed that the functions

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \mathbf{x}_t^\top \boldsymbol{\theta}(\tau)$$

must be monotone increasing in τ on $[0, 1]$. Given linearity in $\mathbf{x}_t \equiv (1, y_{t-1}, \ldots, y_{t-p})^\top$, this necessarily restricts the domain of $\mathbf{x}_t$, except for the (rather uninteresting) special case in which the $\theta_i(\tau)$'s are independent of τ for $i = 2, 3, \ldots, p$, which throws us back into the location-shift model from which we were trying to escape. Theorems 2, 3, and 4 all apply to the model (1) under the explicit proviso that $\mathbf{x}_t$ is restricted to a domain that ensures that $\mathbf{x}_t^\top \boldsymbol{\theta}(\tau)$ is monotone increasing in τ.

For the QAR(1) model considered by Fan and Fan (FF),

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1}, \tag{1}$$

with $\theta_0(\tau) = F^{-1}(\tau)$ and $\theta_1(\tau) = \alpha_0 + \alpha_1\tau$, it is relatively easy to compute a bound on the region of monotonicity. We require that for any $\tau_1 < \tau_2$,

$$F^{-1}(\tau_1) + (\alpha_0 + \alpha_1\tau_1)y_{t-1} \le F^{-1}(\tau_2) + (\alpha_0 + \alpha_1\tau_2)y_{t-1},$$

© 2006 American Statistical Association
Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000726

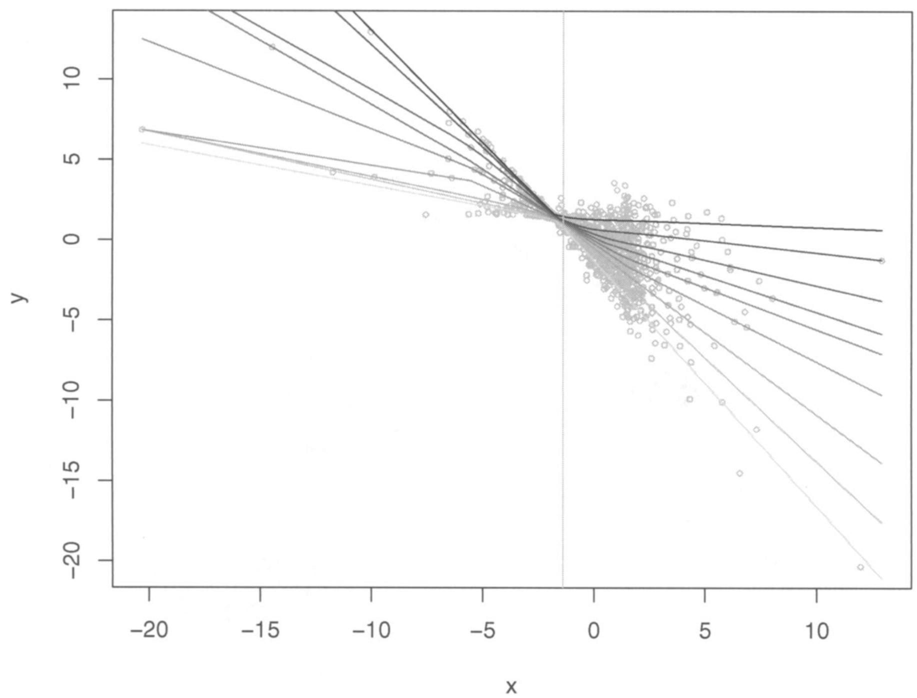

Figure 1. Fan and Fan Model. Scatterplot of 1,000 observations from the model of Fan and Fan with estimates of the conditional deciles superimposed (——, $\tau = .9$; ——, $\tau = .8$; ——, $\tau = .7$; ——, $\tau = .6$; ——, $\tau = .5$; ——, $\tau = .4$; ——, $\tau = .3$; ——, $\tau = .2$; ——, $\tau = .1$). Computation of the piecewise linear conditional quantile function estimates is based on the R function rqss (Koenker 2006) implementing the total variation penalized quantile smoothing splines proposed by Koenker, Ng, and Portnoy (1994).

so it suffices that

$$y_{t-1} \geq \frac{F^{-1}(\tau_1) - F^{-1}(\tau_2)}{\alpha_1(\tau_2 - \tau_1)}$$
$$\geq -\Big(\alpha_1 \sup_u f(u)\Big)^{-1},$$

where f is the density of the distribution function, F, provided that it exists.

Of course, data can always be generated according to the random-coefficient specification,

$$y_t = \theta_0(U_t) + \theta_1(U_t) y_{t-1}, \tag{2}$$

with U_t iid U[0, 1], but failure of the monotonicity condition implies that the conditional quantile functions are then no longer linear. Thus, for FF's example with their choice of $\boldsymbol{\alpha} = (-1.7, 1.8)^\top$ and $F = \Phi$, we obtain an AR(1) scatterplot like the one in Figure 1. The "bowtie" shape clearly indicates that the conditional quantile functions of the random-coefficient model (2) do not conform to the linear conditional quantile functions of (1); instead, they exhibit approximately *piecewise* linear behavior. As such, they are relatively easy to estimate with the nonparametric quantile regression methods. Figure 1 also illustrates an example of such fitting; the estimated conditional quantile functions are piecewise linear and have a sharp kink in the neighborhood of the point $-(\sqrt{2\pi}/1.8) \approx -1.4$, indicated by the vertical line.

Resorting to nonparametric fitting may appear to be too radical a step away from our convenient linear specification. There are certainly appealing parametric alternatives as well. We are currently exploring parametric nonlinear models suggested by stationary copula specifications. If we consider a stationary Markov process of order p, then the probabilistic features of the process are completely determined by the joint distribution of vectors $(Y_t, Y_{t-1}, \ldots, Y_{t-p})$, say $H(y_0, y_1, \ldots, y_p)$. By Sklar's theorem (see, e.g., Joe 1997), there exists a copula function $C(u_0, u_1, \ldots, u_p)$ such that

$$H(y_0, y_1, \ldots, y_k) = C\big(F(y_0), F(y_1), \ldots, F(y_p)\big),$$

where F denotes the marginal distribution of Y_t. Differentiating the copula function with respect to u and evaluating at $u_j = F(y_{t-j})$ for $j = 0, \ldots, p$, we obtain the conditional distribution of Y_t given $Y_{t-1}, \ldots, Y_{t-p}$, say $F_t(y|y_{t-1}, \ldots, y_{t-p})$. Thus, for any τ, we can solve

$$\tau = F_t(y|y_{t-1}, \ldots, y_{t-p}),$$

for y to obtain the τth conditional quantile function of Y_t. For various parametric copula functions, this approach provides a convenient strategy for developing semiparametric nonlinear QAR models. Copula models may also suggest transformations of the observed process that restrict the domain of the transformed process, thereby making it easier to impose the monotonicity requirement by, for example, considering conditional quantiles of the copula function itself on the unit cube.

In some instances, linear specification of QAR models is quite adequate for the simple reason that the process stays within the region of monotonicity. An early example motivated by our interest in "near–unit-root" models was the specification

$$Q_{y_t}(\tau|y_{t-1}) = 2 + \min\left\{\frac{3}{4} + \tau, 1\right\} y_{t-1} + 3\Phi^{-1}(\tau). \tag{3}$$

Substituting U_t for τ, we see that with probability $\frac{3}{4}$, we get realizations from a simple unit-root model, but when $U_t < \frac{1}{4}$, we

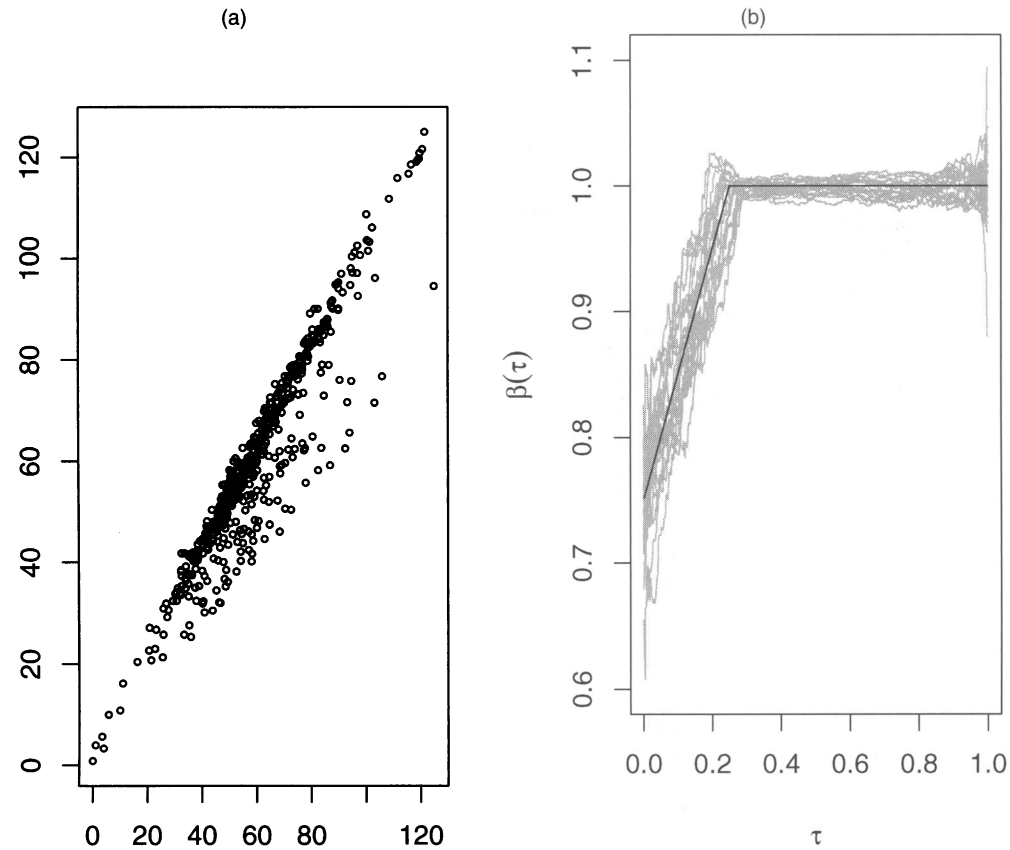

Figure 2. Near-Unit-Root Model. (a) A typical AR(1) scatterplot with 500 observations from the model (3). (b) Twenty sample paths of estimated QAR processes (in gray) superimposed (in black) is the true $\theta_1(\tau)$ function. In this model y_t tends to stay above 0 in the region in which conditional quantile functions are linear in y_{t-1} and monotone in τ.

get observations that encourage mean reversion. For this model, even when started at $y_0 = 0$, the probability that y_t strays into the nonmonotone region is tiny, and we can quite accurately estimate the coefficient functions, $\theta_0(\tau) = 2 + 3\Phi^{-1}(\tau)$ and $\theta_1(\tau) = \min\{\frac{3}{4} + \tau, 1\}$, as illustrated in Figure 2.

Whenever we attempt to estimate linear conditional quantiles and there is significant nonlinearity in the true conditional quantile functions, we are in a state of sin. The first step in the expiation of this sin is a better understanding of the consequences of such misspecification. As noted by Hallin and Werker, QAR estimation attempts to find the best linear approximation of the nonlinear function in the sense that $\hat{\boldsymbol{\theta}}_n$ tends to the solution of

$$E_{\mathbf{X}}(F(\mathbf{X}^\top \boldsymbol{\theta}) - \tau)\mathbf{X} = \mathbf{0}.$$

There are various ways to rewrite this. Angrist, Chernozhukov, and Fernandez-Val (2006) showed that it can be written as

$$E_{\mathbf{X}} w(\mathbf{X}, \boldsymbol{\theta})\Delta^2(\mathbf{X}, \boldsymbol{\theta}) \equiv E_{\mathbf{X}} w(\mathbf{X}, \boldsymbol{\theta})\big(\mathbf{X}^\top \boldsymbol{\theta} - Q_{y_t}(\tau|\mathbf{X})\big)^2 = 0,$$

where $w(x, \theta) = \int_0^1 (1 - u) f_{y_t}(Q_{y_t}(\tau|x) + u\Delta(x, \theta)|x)\, du$. So the limiting form of the estimator minimizes a weighted $\mathcal{L}_2$ discrepancy between the true quantile function and its linear approximant. The weighting is somewhat complicated, but provided that the conditional density does not vary too wildly, its role is minor, and the approximation is similar to the well-known least squares situation. Angrist et al. (2006) have provided further details and several nice examples.

Even when monotonicity holds incontrovertibly in a parametric family of QAR models, we may worry that *estimation* of such models in the unrestricted τ by τ fashion that we have described may prove to be a recipe for disaster. Will estimation error produce crossing quantile functions even in situations in which the population quantile functions are nicely monotone? The short answer is, inevitably, "yes." But in our experience, such difficulties are no more severe than a host of other potentially embarrassing statistical eventualities: negative variance estimates in autoregressive conditional heteroscedasticity (ARCH) models, negative variance component estimates in random-effects models, negative density estimates with higher-order kernel methods, and so on. The estimated conditional quantile function,

$$\hat{Q}_{y_t}(\tau|\mathbf{x}_t) = \mathbf{x}_t^\top \hat{\boldsymbol{\theta}}(\tau),$$

is ensured to be monotone in τ at $\mathbf{x}_t = \bar{\mathbf{x}}$, as we noted in Section 4 of our article. But even quite near $\bar{\mathbf{x}}$, there can be violations of monotonicity; however, we would argue that these violations are generally quite innocuous. Recently, Neocleous and Portnoy (2006) have shown that if one considers grids for $\tau \in [0, 1]$ with spacings somewhat wider than the $\mathcal{O}(1/(n \log n))$ spacings of the full quantile regression process, with spacings δ_n satisfying $\limsup \delta_n n^\eta > 0$ and $\liminf \delta_n n^{1/2}/\log n > 0$ for some $\eta > 0$, then with probability tending to 1, $\hat{Q}_{y_t}(\tau|x)$ is strictly monotone for $\epsilon \le \tau \le 1 - \epsilon$ and bounded domain for x.

In some circumstances, it may be desirable to impose monotonicity by estimating quantiles jointly. This is relatively easy to implement along the lines described by Koenker (2005, sec. 6.8). The very intriguing proposals of Hafner and Linton for estimating QAR(∞) processes lend themselves to this approach. Imposing linear inequality restrictions is easily accomplished given the linear programming structure of the underlying problem, and the sparsity of the linear algebra allows simultaneous efficient estimation of many parameters. Some details of this approach have been provided in recent work of Takeuchi, Le, Sears, and Smola (2006).

3. OTHER RAMIFICATIONS

3.1 ARCH

We welcome the connections between QAR and ARCH models developed by Hafner and Linton. Although pure location and location-scale models have played an enormously important role in the development of time series analysis, there is surely scope for broader classes of models that introduce asymmetry and more flexible tail behavior. Clearly, the orthogonality requirement on the QAR coefficients is extremely attractive from an interpretation-of-moments perspective, but it is difficult to reconcile with the monotonicity requirements already discussed. Restricting the domain of parametric models in τ as discussed by Hafner and Linton and FF, seems to be a prudent tactic. Flexible models designed for restricted domains may serve as fruitful complements to restrictive models for larger domains.

The model (E) suggested by Robinson is clearly an important special case of the QAR model. In the terminology of Koenker and Xiao (2002), Robinson's hypothesis (C) leading to (E) gives a pure scale-shift model, and like the location-shift hypothesis treated in our theorem 4, it can be tested against more general QAR alternatives within the framework of that theorem. Appropriately modifying the form of the hypothesis as described in section 5.1 of Koenker and Xiao (2002) and replacing the location score function $\dot{f}(\theta_0(\tau))/f(\theta_0(\tau))$ by the scale score function $1+\theta_0(\tau)\dot{f}(\theta_0(\tau))/f(\theta_0(\tau))$ yields a test with the same asymptotic behavior as already described for the location-shift form of the hypothesis. Indeed, tests for more general location-scale shift versions of Robinson's hypothesis (C),

$$\theta_i(\tau) = a_i + b_i\theta_0(\tau), \qquad i = 1, \dots, p,$$

can also be formulated in the same way as described by Koenker and Xiao (2002).

3.2 RCAR and FCAR Models

The parallels developed between QAR models and functional coefficient autoregressive (FCAR) models by FF are also very valuable, particularly, as FF note, due to the similarities in their theoretical analysis. Many of these similarities are already apparent in the comparison with RCAR models as can be seen in our proof of theorem 1. Robinson's "partial explanation" of the scarcity of applications of RCAR models as resting in the "inability to distinguish them from constant coefficient models at the conditional mean level" strikes us as entirely plausible. Likewise, the expanded focus on models of conditional variances in recent years has perhaps restricted our appreciation of asymmetry and other potential interesting data features. Rather than continuing to climb Mosteller and Tukey's (1977) "misty staircase of moments," it seems worth considering a detour through the hallway of quantiles.

3.3 Identification

Identification is an issue that is always on the statistical table. In the QAR context, comonotonicity of the coefficients ensures a unique association of the model

$$y_t = \mathbf{x}_t^\top \boldsymbol{\theta}(U_t) \tag{4}$$

with the conditional quantile functions

$$Q_{y_t}(\tau|\mathbf{x}_t) = \mathbf{x}_t^\top \boldsymbol{\theta}(\tau). \tag{5}$$

However, as we have seen, in nonmonotone situations, one can easily construct distinct nonlinear functions, $\boldsymbol{\theta}(\tau)$, that give rise to a single distribution for y_t and thus to a unique quantile function. As FF observe, with $U_t \sim \mathrm{U}[0,1]$, the model (4) is indistinguishable from

$$y_t = \mathbf{x}_t^\top \boldsymbol{\theta}(1-U_t), \tag{6}$$

but we do not consider this a failure of identification because U_t and $1-U_t$ have the same distribution.

3.4 Infinite Variance

We were reluctant to venture into the land of infinite variances, but we are very grateful for Knight's comments on this curious region. Given the influential role of extreme values in such cases, it seems that it may be worthwhile to explore variants of QAR estimation that are less sensitive to outlying values of $\mathbf{x}_t$ than the usual ones. Some alternatives have been briefly described by Koenker (2005, sec. 8.5), but much more could be done to tailor these methods to time series applications.

3.5 Extremes

Extreme value theory in its conditional (regression) manifestation is another topic that we have conscientiously avoided, but there is an important body of work to which Knight is an important contributor that provides much improved guidance on asymptotic behavior and inferential matters for extreme quantile regression. Because extremes are often of crucial practical concern in applications, it is particularly important to find better ways to deal with inference in these regions.

3.6 Weighting

In most QAR settings, variation of the conditional density function along the conditional quantile functions suggests that efficiency gains in estimation are available by considering weighted versions of the quantile regression objective function. Hallin and Werker noted that reweighting may also be a valuable tool in misspecified QAR settings. In some respects, weighted quantile regression is just like weighted least squares estimation, except that the weights that we would like to choose should be based on conditional densities rather than on conditional variances. But far less is known about density weighting for quantile regression models, and it would be very valuable to have further theoretical guidance and practical experience with these methods.

3.7 Model Selection

Order selection in QAR models, as emphasized by Robinson, is an important problem. Our use of BIC, adapted to QAR along the lines suggested by Machado (1993), was based not on second moments, but rather on absolute first moments. At the median, we use the criterion

$$BIC = n \log \hat{\sigma} - \frac{p}{2} \log n,$$

where $\hat{\sigma} = n^{-1} \sum |y_t - \mathbf{x}_t^\top \hat{\boldsymbol{\theta}}(1/2)|$. For other quantiles, the obvious asymmetric modification of this expression can be used. It is convenient to impose the restriction that order is not τ dependent, but it is certainly plausible that there will be applications in which this is not desirable.

3.8 Temporal Aggregation

Hallin and Werker also raised the challenging problem of temporal aggregation for QAR models. Considerable theoretical and data-analytic experience has accumulated on this topic based on classical linear and bilinear time series models, but extending these ideas to the QAR context appears to be difficult. Closure under either point-sampled or unit-averaged forms of temporal aggregation imposes a heavy burden on such models, but it is definitely worth exploring because the effects of such aggregation are commonly encountered in applications.

A related challenge involves QAR models for irregularly spaced time series data. Wei and He (2006) suggested an innovative approach in which QAR coefficients depend on spacings between successive observations. Their estimated conditional growth charts clearly indicate the desirability of quantile-specific QAR models. They also considered diagnostic evaluation of the linearity assumption and suggested rank-based tests for goodness of fit.

3.9 Forecasting

Hafner and Linton's query about forecasting also deserves an extended treatment, but we might briefly observe that one-step-ahead density forecasting is immediately available from the estimated form of the QAR model. Multistep forecasting is not quite so straightforward; instead of generating a single forecasted sample path, we can instead generate ensembles of forecast paths from simulation of the random-coefficient form of the estimated model. From these, we can compute confidence bands. We hope to explore this approach in future work.

4. CONCLUSION

Statistical models are at best crude approximations, revealing some things and concealing others. QAR models allow us to explore some features of time series that are inaccessible through classical methods, but they suffer from their own limitations, many of which have been brought out in this fruitful discussion. Much room is left for improvement; we look forward to the continuing discussion.

ADDITIONAL REFERENCES

Angrist, J. D., Chernozhukov, V., and Fernandez-Val, I. (2006), "Quantile Regression Under Misspecification, With an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Donoho, D. (2002), "How to Be a Highly Cited Author," available at *http://www.in-cites.com/scientists/DrDavidDonoho.html*.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, New York: Chapman & Hall.

Koenker, R. (2005), *Quantile Regression*, Cambridge, U.K.: Cambridge University Press.

——— (2006), *Quantreg: A Quantile Regression R Package*, version 3.89, available at *http://www.r-project.org*.

Koenker, R., Ng, P., and Portnoy, S. (1994), "Quantile Smoothing Splines," *Biometrika*, 81, 673–680.

Machado, J. A. F. (1993), "Robust Model Selection and M-Estimation," *Econometric Theory* 9, 478–493.

Mosteller, F., and Tukey, J. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.

Neocleous, T., and Portnoy, S. (2006), "On Monotonicity of Regression Quantile Functions," preprint, University of Illinois, Dept. of Statistics.

Takeuchi, I., Le, Q. V., Sears, T., and Smola, A. J. (2006), "Nonparametric Quantile Estimation," *Journal of Machine Learning Research*, 7, 1231–1264.

Wei, Y., and He, X. (2006), "Conditional Growth Charts," *The Annals of Statistics*, forthcoming.