Goodness of Fit and Related Inference Processes for Quantile Regression

Roger KOENKER and José A. F. MACHADO

We introduce a goodness-of-fit process for quantile regression analogous to the conventional R^2 statistic of least squares regression. Several related inference processes designed to test composite hypotheses about the combined effect of several covariates over an entire range of conditional quantile functions are also formulated. The asymptotic behavior of the inference processes is shown to be closely related to earlier p-sample goodness-of-fit theory involving Bessel processes. The approach is illustrated with some hypothetical examples, an application to recent empirical models of international economic growth, and some Monte Carlo evidence.

KEY WORDS: Bessel process; Goodness of fit; Quantile regression; Rank test; Regression rankscores.

1. INTRODUCTION

Quantile regression is gradually emerging as a comprehensive approach to the statistical analysis of linear and nonlinear response models. By supplementing the exclusive focus of least squares based methods on the estimation of conditional mean functions with a general technique for estimating families of conditional quantile functions, quantile regression is capable of greatly expanding the flexibility of both parametric and nonparametric regression methods. To this end, effective assessment of goodness of fit for quantile regression models and the development of associated methods of formal inference is critical. In this article we introduce a goodness-of-fit process for quantile regression analogous to the conventional R^2 of least squares regression. Several closely related processes are then suggested that provide a foundation for a broad range of new tests and diagnostics, considerably expanding the scope of statistical inference based on quantile regression.

1.1 Quantile Treatment Effects

The simplest formulation of quantile regression is the two-sample treatment-control model, so we begin by reconsidering a model of two-sample treatment response introduced by Lehmann and Doksum that provides a natural introduction to quantile regression. In special circumstances, we may be willing to entertain the hypothesis that a treatment effect is a pure location shift; subjects with control response x would have response $x + \Delta_0$ under the treatment. This is the presumption that underlies most of conventional regression analysis. Lehmann (1974) proposed the following general model of treatment response:

> Suppose the treatment adds the amount $\Delta(x)$ when the response of the untreated subject would be x. Then the distribution G of the treatment responses is that of the random variable $X + \Delta(X)$ where X is distributed according to F (p. 68).

Special cases obviously include the location shift model $\Delta(\mathbf{X}) = \Delta_0$ and the scale shift model $\Delta(X) = \sigma X$ but the general case is well adapted to the quantile regression paradigm. Doksum (1974) provided a thorough analysis of $\Delta(x)$ showing that if we define $\Delta(x)$ as the "horizontal distance" between F and G at x, so that

$$F(x) = G(x + \Delta(x)),$$

then $\Delta(x)$ is uniquely defined and can be expressed as

$$\Delta(x) = G^{-1}(F(x)) - x. \tag{1}$$

Thus, on changing variables so that $\tau = F(x)$, we have the *quantile treatment effect*,

$$\delta(\tau) = \Delta(F^{-1}(\tau)) = G^{-1}(\tau) - F^{-1}(\tau).$$

In the two-sample setting this quantity is naturally estimable by

$$\hat{\delta}(\tau) = \hat{G}_n^{-1}(\tau) - \hat{F}_m^{-1}(\tau),$$

where G_n and F_m denote the empirical distribution functions of the treatment and control observations, based on n and m observations. If we formulate the quantile regression model for the binary treatment problem as

$$Q_{Y_i}(\tau|D_i) = \alpha(\tau) + \delta(\tau)D_i, \tag{2}$$

where D_i denotes the treatment indicator, then we may estimate the quantile treatment effect directly by solving the quantile regression problem

$$\min_{(\alpha,\delta)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \delta D_i),$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ following Koenker and Bassett (1978).

Doksum suggested that we may interpret the treatment effect in terms of a latent characteristic. For example, in survival analysis a control subject may be called weak if he is prone to die at an early age, and strong if he is prone to die at an advanced age. Strength is thus implicitly indexed by

Roger Koenker is W. B. McKinley Professor of Economics and Professor of Statistics, University of Illinois, Champaign, IL 61820. José A. F. Machado is Associate Professor of Economics, Faculty of Economics, Universidade Nova de Lisboa, Travessa Estevaõ Pinto, P-1070, Lisboa, Portugal. This research was partially supported by National Science Foundation grant SBR-9617206 to Koenker, and FEDER (Praxis XXI program grant 2/21/ECO/94) to Machado. The authors thank Steve Portnoy, a referee, and an associate editor for very helpful comments and Carlos Arango for research assistance on the empirical growth example.

Journal of the American Statistical Association
December 1999, Vol. 94, No. 448, Theory and Methods

τ, the quantile of the survival distribution at which the subject would appear if untreated; that is, $(Y_i|D_i = 0) = \alpha(\tau)$. The treatment, under the Lehmann–Doksum model, is assumed to alter the subject's control response, $\alpha(\tau)$, making it $\alpha(\tau) + \delta(\tau)$ under the treatment. If the treatment is beneficial in the sense that

$$\Delta(x) \geq 0 \quad \forall x,$$

then the distribution of treatment responses, G, is stochastically larger than the distribution of control responses, F.

1.2 Quantile Regression Goodness of Fit

Our goodness-of-fit proposal for quantile regression is motivated by the familiar R^2 of classical least squares regression. Consider the linear model for the conditional mean function of y_i given $\mathbf{x}_i$,

$$E(y_i|\mathbf{x}_i) = \mathbf{x}_i'\beta, \tag{3}$$

which we partition as

$$E(y_i|\mathbf{x}_i) = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \mathbf{x}_{i2}'\boldsymbol{\beta}_2. \tag{4}$$

Let $\hat{\boldsymbol{\beta}}$ denote the least squares estimators of the full model and let $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1', \mathbf{0}')'$ denote the estimator under the q-dimensional linear restriction that

$$H_0\colon\ \boldsymbol{\beta}_2 = \mathbf{0}. \tag{5}$$

Denoting the error sum of squares under the restricted and unrestricted forms of the model by $\tilde{S}$ and $\hat{S}$, we may define an R^2 of model (4) relative to the restricted model constrained by the hypothesis (5) as

$$R^2 = 1 - \hat{S}/\tilde{S}. \tag{6}$$

Conventionally, restricting (4) to include only an "intercept" parameter yields the R^2 usually reported.

For quantile regression, we may proceed in the same manner. Consider a linear model for the conditional quantile function,

$$Q_{y_i}(\tau|\mathbf{x}) = \mathbf{x}_{i1}'\boldsymbol{\beta}_1(\tau) + \mathbf{x}_{i2}'\boldsymbol{\beta}_2(\tau),$$

and let $\hat{\boldsymbol{\beta}}(\tau)$ denote the minimizer of

$$\hat{V}(\tau) = \min_{\{b \in \mathbb{R}^p\}} \sum \rho_\tau(y_i - \mathbf{x}_i'\mathbf{b})$$

and $\tilde{\boldsymbol{\beta}}(\tau) = (\tilde{\boldsymbol{\beta}}_1(\tau)', \mathbf{0}')'$ denote the minimizer for the corresponding constrained problem, with

$$\tilde{V}(\tau) = \min_{\{\mathbf{b}_1 \in \mathbb{R}^{p-q}\}} \sum \rho_\tau(y_i - \mathbf{x}_{1i}'\mathbf{b}_1).$$

That is, $\hat{\boldsymbol{\beta}}(\tau)$ and $\tilde{\boldsymbol{\beta}}(\tau)$ denote the unrestricted and restricted quantile regression estimates. We may now define the goodness-of-fit criterion

$$R^1(\tau) = 1 - \hat{V}(\tau)/\tilde{V}(\tau), \tag{7}$$

which is the natural analog of R^2 defined earlier.

Like R^2, because $\tilde{\boldsymbol{\beta}}(\tau)$ is obtained by constraining $\hat{\boldsymbol{\beta}}(\tau)$, it is immediately apparent that $\hat{V}(\tau) \leq \tilde{V}(\tau)$, and thus $R^1(\tau)$ lies between 0 and 1. Unlike R^2, which measures the relative success of two models for the conditional mean function in terms of residual variance, $R^1(\tau)$ measures the relative success of the corresponding quantile regression models at a specific quantile in terms of an appropriately weighted sum of absolute residuals. Thus $R^1(\tau)$ constitutes a local measure of goodness of fit for a particular quantile rather than a global measure of goodness of fit over the entire conditional distribution, like R^2.

One can easily imagine circumstances under which the covariates $\mathbf{x}_2$ might exert a significant effect on one tail of the conditional distribution of the response but might have no effect in the other tail. One example of this is the quantile regression results of Chamberlain (1994) on the U.S. labor union wage premium. Chamberlain found that the 16% mean wage premium for union membership estimated by least squares, when deconstructed by quantile regression, was seen to constitute a much larger—nearly 30%—premium in the lower tail of the conditional wage distribution, but gradually attenuated to the point that it was essentially negligible above the median.

Use of the optimized value of the objective function defining a regression estimator as in (7) as a measure of goodness of fit is sufficiently natural that it is quite commonplace in the robustness literature. Abberger (1996), Horowitz (1992), and Rousseeuw and Hubert (1997) have suggested this idea for least median of squares and l_1 forecasting. McKean and Sievers (1987) suggested $R^1(1/2)$ for the l_1 regression estimator and provided a very complete analysis of its properties, including the connection to a quasi-likelihood ratio test.

If the full model (4) is better at the τ-quantile than the restricted model constrained by (5), then $\hat{V}_n(\tau)$ should be significantly smaller than $\tilde{V}_n(\tau)$. Here "better" should be understood to mean that the τth conditional quantile function is significantly altered by the influence of the covariates $\mathbf{x}_2$. To assess the goodness of fit of the two nested models for the entire distribution, it is natural to compare the *processes* $\hat{V}_n(\tau)$ and $\tilde{V}_n(\tau)$. The $R^1(\tau)$ process (7) is one possible gauge of the differences between the $\hat{V}(\tau)$ and $\tilde{V}(\tau)$ processes for the entire distribution. But such a comparison also provides the basis for formal tests of the hypothesis H_0. We consider several such variants later. The choice of these statistics is motivated by analogies to likelihood ratio (LR) tests in the Laplacean regression model, as well as by the theory of rank tests for linear quantile regression models. These connections are explored in detail in Section 2, where the asymptotic theory of the tests is also developed. In an effort to develop some intuition about the processes as diagnostic tools in regression, we explore in Section 3 the behavior of the $R^1(\tau)$ process, and the related inference processes, using a range of artificial data. In Section 4 we illustrate the use of the methods with an analysis of a well-known model of international economic growth. In Section 5 we report on a small Monte Carlo study of the finite-sample properties of the proposed tests and in Section

6 offer some conclusions and describe some prospects for new research. We provide proofs in the Appendix.

2. INFERENCE PROCESSES FOR QUANTILE REGRESSION

The $R^1(\tau)$ process provides a natural link to two closely related versions of a LR process for quantile regression.

2.1 Two Likelihood Ratios

Consider the asymmetric Laplacean density

$$f_1(u) = \tau(1-\tau)\exp\{-\rho_\tau(u)\}.$$

The log-likelihood under the assumption that the $\{u_i\}$ come from this (rather implausible) density is

$$l(\beta) = n\log(\tau(1-\tau)) - \sum \rho_\tau(y_i - \mathbf{x}_i'\beta),$$

and thus the LR statistic for testing H_0 would be

$$-2\log\lambda_n(\tau) \equiv 2(l(\hat\beta(\tau)) - l(\tilde\beta(\tau))) = 2(\tilde V(\tau) - \hat V(\tau)).$$

Under these conditions, and H_0, conventional LR theory implies that $-2\log\lambda_n(\tau)$ will be asymptotically χ^2_q. More generally, we can ask how this statistic behaves when the $\{u_i\}$ are iid but drawn from some other distribution, say F. Adapting the arguments of Koenker and Bassett (1982) slightly, it can be shown that under H_0,

$$L_n(\tau) \equiv \frac{2(\tilde V(\tau) - \hat V(\tau))}{\tau(1-\tau)s(\tau)} \tag{8}$$

is asymptotically χ^2_q, where $s(\tau) = 1/f(F^{-1}(\tau))$. The latter quantity, variously termed the sparsity function or the quantile-density function, plays the role of a nuisance parameter, whose estimation we discuss in Section 2.5.

Similarly, we may consider a location and scale form of the asymmetric Laplacean density,

$$f_\sigma(u) = \tau(1-\tau)\exp\{-\rho_\tau(u)/\sigma\}/\sigma,$$

that yields the LR statistic

$$-2\log\lambda_n^*(\tau) \equiv 2n\log(\tilde V(\tau)/\hat V(\tau)).$$

The asymptotic behavior of this version of the LR statistic follows easily from the preceding result by writing

$$\begin{aligned} -2\log\lambda_n^*(\tau) &= 2n\log(1 + (\tilde V(\tau) - \hat V(\tau))/\hat V(\tau)) \\ &= 2n(\tilde V(\tau) - \hat V(\tau))/\hat V(\tau) + o_p(1) \\ &= 2(\tilde V(\tau) - \hat V(\tau))/\sigma(\tau) + o_p(1), \end{aligned}$$

where we assume that $\sigma(\tau) = E\rho_\tau(u) < \infty$, so $\hat\sigma(\tau) = n^{-1}\hat V(\tau) \to \sigma(\tau)$. Thus, under H_0, by (8),

$$\Lambda_n(\tau) = \frac{2n\sigma(\tau)}{\tau(1-\tau)s(\tau)}\ \log(\tilde V(\tau)/\hat V(\tau))$$

is asymptotically χ^2_q.

Tests of this sort based on the drop in the optimized value of the objective function of an M estimator after relaxation of the restriction imposed by the null hypothesis have been termed ρ tests by Ronchetti (1982). Following this terminology, we refer to these LR-type tests based on $L_n(\tau)$ and $\Lambda_n(\tau)$ as quantile ρ tests.

2.2 Likelihood Ratio Processes for Quantile Regression

In many applications it may be important to formulate joint hypotheses about the relevance of certain groups of covariates with respect to *several* quantiles. For this we require the joint asymptotic distribution of vectors of quantile ρ test statistics of the form, for example, $(L_n(\tau_1), L_n(\tau_2), \ldots, L_n(\tau_m))$. Such results are subsumed in the following theory for the ρ test *processes*, $\{L(\tau)\colon \tau\in[\varepsilon, 1-\varepsilon]\}$, and $\{\Lambda(\tau)\colon \tau\in[\varepsilon,1-\varepsilon]\}$.

In this section we restrict attention to the linear model

$$y_i = \mathbf{x}_{i1}\beta_1 + \mathbf{x}_{2i}\beta_2 + u_i \tag{9}$$

where the $\{u_i\}$'s are assumed to be iid with common distribution function, F, satisfying the following assumption:

A.1. The error distribution F has continuous Lebesgue density, f, with $f(u) > 0$ on $\{u\colon 0 < F(u) < 1\}$.

The sequence of design matrices $\{\mathbf{X}_n\} = \{(\mathbf{x}_i)_{i=1}^n\}$ is assumed to satisfy the following assumption:

A.2.

(a) $x_{i11} \equiv 1, i = 1, 2, \ldots$
(b) $\mathbf{D}_n = n^{-1}\mathbf{X}_n'\mathbf{X}_n \to \mathbf{D}$, a positive definite matrix.
(c) $n^{-1}\sum\|\mathbf{x}_i\|^4 = O(1)$
(d) $\max_{i=1,\ldots,n}\|\mathbf{x}_i\| = O(n^{1/4}/\log n)$

We consider tests of the hypothesis

$$H_0\colon\ \beta_2(\tau) = 0 \qquad \tau\in\mathcal{T} \tag{10}$$

for some index set $\mathcal{T}\subset(0,1)$ against a sequence of local alternatives formulated as the following:

A.3. There exists a fixed, continuous function, $\zeta(\tau)$: $[0,1]\to\Re^q$ such that $\beta_2(\tau) = \zeta(\tau)/\sqrt{n}$ for samples of size n.

Remarks. Conditions A.1 and A.2 are standard in the quantile regression literature. Somewhat weaker conditions on both F and $\mathbf{X}_n$ were used by Gutenbrunner, Jurečková, Koenker, and Portnoy (1993; denoted by GJKP hereafter) in an effort to extend the theory into the tails, but this doesn't seem critical here so we have reverted to conditions close to those used by Gutenbrunner and Jurečková (1992; denoted by GJ hereafter). Condition A.3 is a direct analog of condition A.3 of Koenker and Bassett (1982) and permits us to explore the local asymptotic power of the proposed tests. Note that (A.3) explicitly expands the scope of the iid error model because it permits the $\mathbf{x}_2$ effect to depend on τ.

To investigate the asymptotic behavior of the processes $L_n(\tau)$ and $\Lambda_n(\tau)$, we require some rather basic theory and notation regarding Bessel processes. Let $\mathbf{W}_q(t)$ denote a q-vector of independent Brownian motions; thus, for $t\in[0,1]$,

$$\mathbf{B}_q(t) = \mathbf{W}_q(t) - t\mathbf{W}_q(1)$$

represents a q-vector of independent Brownian bridges. Note that for any fixed $t \in (0,1)$,

$$\mathbf{B}_q(t) \sim \mathcal{N}(0, t(1-t)\mathbf{I}_q). \tag{11}$$

The normalized Euclidean norm of $\mathbf{B}_q(t)$,

$$Q_q(t) = \|\mathbf{B}_q(t)\|/\sqrt{t(1-t)},$$

is generally referred to as a Bessel process of order q. Critical values for $\sup Q_q^2(t)$ have been tabled by DeLong (1981) and, more extensively, by Andrews (1993) using simulation methods. The seminal work on Bessel processes and their connection to K-sample goodness-of-fit tests seems to be that of Kiefer (1959). Again, for any fixed $t \in (0,1)$ we have, from (11), $Q_q^2(t) \sim \chi_q^2$. Thus we may interpret $Q_q^2(t)$ as a natural extension of the familiar univariate chi-squared random variable with q degrees of freedom. Note that in the special case $q = 1, \sup Q_1^2(\cdot)$ behaves asymptotically like a squared Kolmogorov–Smirnov statistic.

To characterize the behavior of the test statistic under local alternatives, it is helpful to define a noncentral version of the squared Bessel process as an extension of the noncentral chi-squared distribution. Let $\mu(t)$ be a fixed, bounded function from $[0,1]$ to $\Re^q$. We refer to the standardized squared norm

$$\|\boldsymbol{\mu}(t) + \mathbf{B}_q(t)\|^2/(t(1-t))$$

as a squared noncentral Bessel process of order q with noncentrality function $\eta(\tau) = \boldsymbol{\mu}(t)'\boldsymbol{\mu}(t)/(t(1-t))$ and denote it by $Q_{q,\eta(t)}^2$. Of course, for any fixed $t \in (0,1), Q_{q,\eta(t)}^2 \sim \chi_{q,\eta(t)}^2$, a noncentral χ_q^2 random variable with q degrees of freedom and noncentrality parameter $\eta(t)$. We adopt the following standard notation for partitioning matrices like $\mathbf{D}$ defined in condition A.2 (b): $\mathbf{D}_{ij}$ $i,j = 1,2$ denotes the ijth block of $\mathbf{D}$, and $\mathbf{D}^{ij}$ denotes the ij block of $\mathbf{D}^{-1}$. To illustrate, recall that $\mathbf{D}^{22} = (\mathbf{D}_{22} - \mathbf{D}_{21}\mathbf{D}_{11}^{-1}\mathbf{D}_{12})^{-1}$. Finally, the symbol $\Rightarrow$ denotes weak convergence; $\rightsquigarrow$, convergence in distribution; and $\to$, convergence in probability. We can now state our first result, which is proven in the Appendix.

Theorem 1. Let $\mathcal{T} = [\varepsilon, 1-\varepsilon]$, for some $\varepsilon \in (0, \frac{1}{2})$. Under conditions A.1–A.3,

$$L_n(\tau) \Rightarrow Q_{q,\eta(\tau)}^2(\tau) \quad \text{for} \quad \tau \in \mathcal{T},$$

where $\eta(\tau) = \zeta(\tau)'(\mathbf{D}^{22})^{-1}\zeta(\tau)/\omega^2(\tau)$, and $w(t) = \sqrt{\tau(1-\tau)}s(\tau)$. Also, under the null hypothesis (10),

$$\sup_{\tau\in\mathcal{T}} L_n(\tau) \rightsquigarrow \sup_{\tau\in\mathcal{T}} Q_q^2(\tau).$$

The alternative form of the quantile ρ process based on the location-scale form of the ρ test has the same asymptotic behavior.

Corollary 1. Under conditions A.1–A.3, $\Lambda_n(\tau) = L_n(\tau) + o_p(1)$, uniformly on $\mathcal{T}$.

The foregoing results enable the investigator to test a broad range of hypotheses regarding the joint effects of covariates while also restricting attention to specified ranges of the family of conditional quantile functions. Thus, for example, we may focus attention on only one tail of the conditional density, or on just a neighborhood of the median, without any prejudgment that effects should be constant over the entire conditional density as in the conventional location-shift version of the classical linear regression model.

2.3 Wald Processes

One disadvantage of the quantile ρ tests proposed here is the required estimation of the nuisance sparsity function, $s(\tau)$. A potentially more serious disadvantage is that they require the null model to take the pure location-shift form, so the conditional density of the response at any $\tau \in \mathcal{T}$ is independent of the covariates x. This difficulty with the LR approach is familiar in many other contexts (see, e.g., Foutz and Srivastava 1977). One way to relax the latter restriction is to turn to Wald versions of the inference process.

In this section and the one that follows, we relax the iid error assumption used earlier and consider the location-scale shift model used by GJ

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \sigma_i u_i, \tag{12}$$

where $\sigma_i = \mathbf{x}_i'\gamma$ and the $\{u_i\}$ are assumed to be iid from distribution function F. We denote $\boldsymbol{\Gamma}_n = \mathrm{diag}(\sigma_i)$ and introduce the following additional assumption:

A.4. Suppose that model (12) holds with conditions A.1–A.3, that the elements σ_i are bounded away from 0 and infinity, and that $\mathbf{G}_n = n^{-1}\mathbf{X}_n'\boldsymbol{\Gamma}_n^{-1}\mathbf{X}_n$ tends to a positive definite matrix.

The conditional quantile functions of the response, y_i, under (12) may be written as

$$Q_{y_i}(\tau|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}(\tau),$$

where $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + \gamma F^{-1}(\tau)$ and which we may again partition as $\boldsymbol{\beta}(\tau) = (\boldsymbol{\beta}_1(\tau)', \boldsymbol{\beta}_2(\tau)')'$. The representation (5.25) of GJ yields, uniformly for $\tau \in \mathcal{T}$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) = s(\tau)\mathbf{G}_n^{-1}\mathbf{g}_n(\tau) + o_p(1),$$

where $\mathbf{g}_n(\tau) = n^{-1/2}\sum \mathbf{x}_i\psi_\tau(u_i - F^{-1}(\tau))$ and $\psi_\tau(u) = \tau - I(u < 0)$. Thus, setting $\boldsymbol{\Psi} = [0 \vdots \mathbf{I}_q]$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_2(\tau) - \boldsymbol{\beta}_2(\tau)) = s(\tau)\boldsymbol{\Psi}\mathbf{G}_n^{-1}\mathbf{g}_n(\tau) + o_p(1),$$

and by theorem 1 of GJ it follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_2(\tau) - \boldsymbol{\beta}_2(\tau)) \Rightarrow s(\tau)\boldsymbol{\Omega}^{1/2}\mathbf{B}_q,$$

where $\boldsymbol{\Omega} = \boldsymbol{\Psi}\mathbf{G}_n^{-1}\mathbf{D}_n\mathbf{G}_n^{-1}\boldsymbol{\Psi}'$ and $\mathbf{B}_q$ denotes a q-vector of independent Brownian bridges.

Theorem 2. Under conditions A.1–A.4,

$$\mathbf{W}_n(\tau) = n\hat{\boldsymbol{\beta}}_2(\tau)'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\beta}}_2(\tau)/\omega^2(\tau) \Rightarrow Q_{q,\eta(\tau)}^2(\tau) \quad \text{for} \quad \tau \in \mathcal{T},$$

where $\eta(\tau) = \zeta(\tau)'\boldsymbol{\Omega}^{-1}\zeta(\tau)/\omega^2(\tau)$. Also, under the null hypothesis,

$$\sup_{\tau\in\mathcal{T}} \mathbf{W}_n(\tau) \rightsquigarrow \sup_{\tau\in\mathcal{T}} Q_q^2(\tau).$$

Note that in the "homoscedastic" case $\boldsymbol{\Gamma} = \mathbf{I}$, the sandwich form of the matrix $\boldsymbol{\Omega}$ collapses to $\mathbf{D}_n^{22}$, and the Wald process has the same asymptotic behavior as the two ρ processes. More generally, the Wald process enables us to relax the stringent location-shift, iid error assumption underlying the ρ processes. In exchange, it requires us to estimate the matrix, $\mathbf{G}_n$, a task somewhat more onerous than the scalar sparsity estimate needed for the ρ processes. We defer the question of estimating $\mathbf{G}_n$ until the end of this section, and turn now to a third variant of the inference process.

2.4 Rankscore Processes

The regression rankscore process introduced by GJ arises from the formal dual problem of the linear programming formulation of the primal quantile regression problem. It may be viewed as providing a natural generalization to the linear model of the familiar statistical duality of the order statistics and ranks in the one-sample model. As such, it provides a fundamental link between quantile regression and the classical theory of rank statistics as presented by Hájek and Šidák (1967). The rankscore process may also be interpreted as an implementation of the Rao score, Lagrange multiplier principle for quantile regression inference.

The regression rankscore process for the restricted form of the linear location-scale model (12) is given by

$$\hat{\mathbf{a}}_n(\tau) = \operatorname{argmax}\{\mathbf{y}'\mathbf{a} | \mathbf{X}_1'\mathbf{a} = (1-\tau)\mathbf{X}_1'\mathbf{e}, \mathbf{a} \in [0,1]^n\}, \quad (13)$$

where $\mathbf{e}$ denotes an n-vector of 1's and the n by p matrix, $\mathbf{X}$, has been partitioned as $[\mathbf{X}_1 \vdots \mathbf{X}_2]$ to confirm with the partitioning of the hypothesis. The problem posed in (13) is the formal dual problem corresponding to the (primal) quantile regression linear program under the restriction imposed by H_0. Using theorem 1 of GJ, theorem 5.1 of GJKP may be easily extended to conform to the conditions of the location-scale shift model (12).

We can consider tests of the hypothesis H_0 based on the statistic

$$T_n = \mathbf{S}_n'\mathbf{M}_n^{-1}\mathbf{S}_n/A^2(\varphi),$$

where

$$S_n = n^{-1/2}(\mathbf{X}_2 - \hat{\mathbf{X}}_2)'\hat{\mathbf{b}}_n,$$

$$\hat{\mathbf{X}}_2 = \mathbf{X}_1(\mathbf{X}_1'\boldsymbol{\Gamma}_n^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\Gamma}_n^{-1}\mathbf{X}_2,$$

$$M_n = (\mathbf{X}_2 - \hat{\mathbf{X}}_2)'(\mathbf{X}_2 - \hat{\mathbf{X}}_2)/n,$$

$$\hat{\mathbf{b}}_n = \left(-\int \varphi(t)\, d\hat{a}_{in}(t)\right)_{i=1}^n,$$

and φ is a score-generating function of bounded variation. It may be emphasized that the only modification of the test statistic relative to the formulation of GJKP is the appearance of the matrix $\boldsymbol{\Gamma}_n$ in the expression for $\hat{\mathbf{X}}_2$. In effect we have simply replaced an ordinary least squares fit by a weighted least squares fit at this point. The following result extends theorem 5.1 of GJKP to the context of the linear heteroscedastic model (10).

Theorem 3. Let φ: $[0,1] \to \Re$ denote a function of bounded variation that is constant outside a compact subinterval, $\mathcal{T}$, of (0, 1). Set $A^2(\varphi) = \int_0^1 (\varphi(t) - \bar{\varphi}(t))^2\, dt$ with $\bar{\varphi} = \int_0^1 \varphi(t)\, dt$. Under conditions A.1–A.4,

$$T_n \rightsquigarrow \chi_q^2(\eta(\varphi,\zeta)),$$

where $\eta(\varphi,\zeta) = \zeta'\boldsymbol{\Omega}^{-1}\zeta/A^2(\varphi)$, with $\xi(\varphi,\zeta,F) = \int_0^1 \zeta(t) f(F^{-1}(t))\, d\varphi(t)$.

Here we are interested primarily in the special quantile score function,

$$\varphi_\tau(t) = \tau - I(t < \tau),$$

which focuses attention exclusively on the τth quantile. For this score function, $\bar{\varphi} = 0$ and $A^2(\varphi_\tau) = \tau(1-\tau)$. The quantile score function yields rankscores

$$\hat{\mathbf{b}}_n = \hat{\mathbf{a}}_n(\tau) - (1-\tau),$$

so $\xi(\varphi_\tau,\zeta,F) = \zeta(\tau)f(F^{-1}(\tau))$. This allows us to define the quantile rankscore process,

$$T_n(\tau) = \mathbf{S}_n(\tau)'\mathbf{M}_n^{-1}\mathbf{S}_n(\tau)/(\tau(1-\tau)),$$

which can be used in lieu of $\mathbf{W}_n(\tau)$.

Theorem 4. Under conditions A.1–A.4, $T_n(\tau) = \mathbf{W}_n(\tau) + o_p(1)$, uniformly on $\mathcal{T}$.

When $\boldsymbol{\Gamma}_n = \mathbf{I}$, a crucial feature of this form of the test that distinguishes it from the corresponding ρ test processes is that, because the rank score process $\mathbf{a}_n(\tau)$ is scale invariant, T_n does not require any estimate of the nuisance parameters $s(\tau)$ or $\sigma(\tau)$. In this case, Jurečková (1992) treated the case $q = 1$ and related the test to the Hájek and Šidák (1967) extension of the Kolmogorov–Smirnov to regression alternatives. This scale invariance is a very substantial advantage over both the ρ test and Wald approaches to testing in quantile regression, as was already stressed by Koenker and Bassett (1982). However, it should also be noted that like the Wald approach, when the null is formulated to admit heterogeneity in the matrix $\boldsymbol{\Gamma}_n$ as in the location-scale model, the rankscore process also requires an estimate of this nuisance parameter, as does the Wald process.

2.5 Estimation of Nuisance Parameters

It is an somewhat unhappy fact of life that the asymptotic precision of quantile estimates in general, and quantile regression estimates in particular, depend on the reciprocal of a density function evaluated at the quantile of interest—a quantity Tukey (1965) termed the "sparsity function" and Parzen (1979) called the quantile-density function. It is perfectly natural that the precision of quantile estimates should depend on this quantity, because it reflects the density of observations near the quantile of interest. If the data are very sparse at the quantile of interest, then it will be difficult to estimate. On the other hand, when the sparsity is low, so that observations are very dense, the quantile will be more precisely estimated. Thus, to estimate the precision of the τth quantile regression estimate directly, the nuisance quantity

$$s(\tau) = [f(F^{-1}(\tau))]^{-1}$$

must be estimated, and this leads us into the realm of density estimation and smoothing.

Luckily, there is a large literature on estimating $s(\tau)$ in the one-sample model, including works by Bofinger (1975), Sheather and Maritz (1983), Siddiqui (1960), and Welsh (1987). Siddiqui's idea is simplest and has received the most attention in the literature, so we focus on it. Differentiating the identity $F(F^{-1}(t)) = t$, we find that the sparsity function is simply the derivative of the quantile function; that is,

$$\frac{d}{dt} F^{-1}(t) = s(t).$$

So, just as differentiating the distribution function F yields the density function f, differentiating the quantile function F^{-1} yields the sparsity function s. It is thus natural, following Siddiqui, to estimate $s(t)$ using a simple difference quotient of the empirical quantile function; that is,

$$\hat{s}_n(t) = [\hat{F}_n^{-1}(t+h_n) - \hat{F}_n^{-1}(t-h_n)]/2h_n,$$

where $\hat{F}^{-1}$ is an estimate of F^{-1} and h_n is a bandwidth that may tend to 0 as $n \to \infty$. A bandwidth rule suggested by Hall and Sheather (1988) based on Edgeworth expansions for studentized quantiles is

$$h_n = n^{-1/3} z_\alpha^{2/3} [1.5 s(t)/s''(t)]^{1/3},$$

where z_α satisfies $\Phi(z_\alpha) = 1 - \alpha/2$. In the absence of other information about the form of $s(\cdot)$, we may use the Gaussian model to select the bandwidth h_n, which yields

$$h_n = n^{-1/3} z_\alpha^{2/3} [1.5 \phi^2(\Phi^{-1}(t))/(2(\Phi^{-1}(t))^2 + 1)]^{1/3}.$$

Having chosen this bandwidth h_n, the next question is how to compute $\hat{F}^{-1}$. One approach to estimating F^{-1} is to use the empirical quantile function suggested by Bassett and Koenker (1982). In effect this amounts to using $\hat{F}_y^{-1}(t) = \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}(t)$, where $\hat{\boldsymbol{\beta}}(\cdot)$ is the usual regression quantile process. The functions

$$\hat{Q}_y(\tau|x) = \mathbf{x}'\hat{\boldsymbol{\beta}}(\tau)$$

constitute a family of conditional quantile functions for the response variable y. At any fixed $\mathbf{x}$, we can regard $\hat{Q}_y(\tau|\mathbf{x})$ as a viable estimate of the conditional quantile function of y given $\mathbf{x}$. Of course, the precision of this estimate depends on the $\mathbf{x}$ at which we evaluate the expression, but the precision is maximized at $\mathbf{x} = \bar{\mathbf{x}}$. This makes $\hat{F}_y^{-1}(t) = \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}(t)$ an attractive choice, but we should verify that as a function of τ, this function satisfies the fundamental monotonicity requirement of a quantile function. It is clear that the estimated conditional quantile functions fitted by quantile regression may cross—indeed, this is inevitable, as the estimated slope coefficients are not identical, and thus the functions are not parallel. One might hope and expect, however, that this crossing occurs only in remote regions of design space—crossing should not occur near the centroid of the design, $\bar{\mathbf{x}}$. This is indeed the case. Theorem 2.1 of Bassett and Koenker (1982) established that $\hat{Q}_y(\tau|\bar{\mathbf{x}})$ is monotone in τ. Thus, in the iid error model, where we need only a scalar estimate of the sparsity function at each τ, we propose using

$$\hat{s}_n(t) = [\hat{Q}_y(t+h_n|\bar{\mathbf{x}}) - \hat{Q}_y(t-h_n|\bar{\mathbf{x}})]/2h_n.$$

Having described an approach to estimating the scalar sparsity parameter, we now briefly describe two approaches to estimating the matrix $\mathbf{G}_n$. One is a natural extension of sparsity estimation methods described earlier, suggested by Hendricks and Koenker (1992). The other, which is based on kernel density estimation ideas, was proposed by Powell (1989).

Provided that the τth conditional quantile function of $Q_y(\tau|\mathbf{x})$ is linear in $\mathbf{x}$, then for $h_n \to 0$, we can consistently estimate the parameters of the $\tau \pm h_n$ conditional quantile functions by $\hat{\boldsymbol{\beta}}(\tau \pm h_n)$. The density $f_i(Q_y(\tau|\mathbf{x}_i))$ can thus be estimated by the difference quotient,

$$\hat{f}_i(Q_y(\tau|\mathbf{x}_i)) = 2h_n/\mathbf{x}_i'(\hat{\boldsymbol{\beta}}(\tau+h_n) - \hat{\boldsymbol{\beta}}(\tau-h_n)).$$

Note that in the location-scale model (12) even for a fixed bandwidth we can obtain estimates of the σ_i's that are consistent up to scale, because

$$\frac{\mathbf{x}_i'(\hat{\boldsymbol{\beta}}(\tau+h_n) - \hat{\boldsymbol{\beta}}(\tau-h_n))}{2h_n} = \mathbf{x}_i'\boldsymbol{\gamma}\Delta_h = \sigma_i\Delta_h,$$

where $\Delta_h = (F^{-1}(\tau+h) - (F^{-1}(\tau-h))/2h$ is a fixed scalar independent of i. For the rankscore statistic, $T_n(\tau)$, this factor Δ_h cancels in the expression for $\hat{\mathbf{X}}_2$, so it plays no role. For the Wald statistic, we still require an estimate of the scalar sparsity parameter at each τ, but it is possible to estimate the matrix Γ_n in this fashion and then to estimate the scalar sparsity parameter using a weighted version of the approach described earlier.

A potential difficulty with the proposed estimator $\hat{f}_i$ is that there is no guarantee of positivity for every observation in the sample. Indeed, as we have just seen, the quantity

$$d_i = \mathbf{x}_i'(\hat{\boldsymbol{\beta}}(\tau+h_n) - \hat{\boldsymbol{\beta}}(\tau-h_n))$$

is necessarily positive only at $\mathbf{x} = \bar{\mathbf{x}}$. Nevertheless, in practice we find that problems due to "crossing" of the estimated conditional quantile planes occur only infrequently and then only in the extreme regions of the design space. In our implementation of this approach, we simply replace $\hat{f}_i$ by its positive part; that is, we use

$$\hat{f}_i^+ = \max\{0, 2h_n/(d_i - \varepsilon)\},$$

where $\varepsilon > 0$ is a small tolerance parameter intended to avoid dividing by 0 in the (rare) cases in which $d_i = 0$ because the ith observation is basic at both $\tau \pm h_n$. Substituting this estimate in the foregoing expression for $\mathbf{G}_n$ yields an easily implementable estimator for the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(\tau)$ in the non-iid error model.

Alternatively, Powell (1989) suggested a kernel-based method for estimating the quantile regression covariance matrix sandwich. Noting that in estimating the matrix $\mathbf{G}_n(\tau)$, what we are really after is a matrix weighted density estimator, Powell proposed a kernel estimator of the form

$$\hat{\mathbf{G}}_n(\tau) = (nh_n)^{-1} \sum K(\hat{u}_i(\tau)/h_n)\mathbf{x}_i\mathbf{x}_i',$$

where $\hat{u}_i(\tau) = y_i - \mathbf{x}_i'\hat{\beta}(\tau)$ and h_n is a bandwidth parameter satisfying $h_n \to 0$ and $\sqrt{n}h_n \to \infty$. He showed that under certain uniform Lipschitz continuity requirements on the $f_i, \hat{\mathbf{G}}_n(\tau) \to \mathbf{G}_n(\tau)$ in probability. In practice, of course, there remain a number of nettlesome questions about the choice of the kernel function K and the bandwidth parameter h_n.

Finally, we conclude this section by remarking that the nuisance parameters, $s(\tau), \sigma(\tau)$, and $\mathbf{\Gamma}_n$, appearing in the various versions of the process, $L_n, \Lambda_n, \mathbf{W}_n$, and T_n, may be replaced by the aforementioned estimates without altering the asymptotic behavior of the respective processes. (See lemma 6 of GJ for further details.)

3. SOME ILLUSTRATIVE EXAMPLES

In this section we continue to explore the behavior of the processes introduced in the previous section, using a range of artificial data. In preliminary experiments the performance of tests based on L_n and Λ_n were essentially indistinguishable, so we choose to drop L_n in the subsequent comparisons. We should emphasize that the quantile ρ tests based on L_n and Λ_n are limited in scope of application by the fact that they presume iid errors under the null hypothesis. This is reasonable in the illustrative examples of this section and in our Monte Carlo, because in such circumstances the null model contains only an intercept. However, in many empirical contexts, including the one explored in Section 4, this assumption is highly implausible. We have also chosen to defer consideration of the Wald process, $\mathbf{W}_n(\tau)$, to future work. This statistic is considerably more cumbersome to implement than the corresponding rankscore test, $T_n(\tau)$, and thus it seems prudent to focus our initial attention on $T_n(\tau)$.

To develop some experience interpreting these processes, it seems valuable to consider a variety of simple bivariate regression settings, illustrated in the top row of panels of Figure 1 and described in detail here.

Model 1: Pure Gaussian Noise. The data in the first column of Figure 1 were generated with $\{y_i\}$ iid standard normal and independent of x. The $\{x_i\}$ were generated as iid $\mathcal{N}(5,1)$ and $n = 100$.

Model 2: Gaussian Location Shift. The data for the second column of Figure 1 corresponds to the classical regression model in which

$$y_i = x_i + u_i$$

with $\{u_i\}$ iid $\mathcal{N}(0,1), \{x_i\}$ iid $\mathcal{N}(5,1)$, and $n = 100$.

Model 3: Gaussian Scale Shift. The third column of Figure 1 illustrates a pure heteroscedastic version of the regression model in which

$$y_i = \left(x_i + \frac{1}{4}\, x_i^2\right) u_i$$

with $\{u_i\}$ iid $\mathcal{N}(0,1/100), \{x_i\}$ iid $\mathcal{N}(3,1)$, and $n = 100$.

Model 4: Nonlocation Scale. Finally, the fourth column of Figure 1 has a somewhat more peculiar configuration in which for small values of x, we observe a unimodal conditional density for y, whereas for larger values of x, we have a bimodal conditional density of y. Formally, letting $Q_y(\tau|x)$ denote the τth quantile of the distribution of y given x, we may express the model as

$$Q_y(\tau|x) = x + x\Phi^{-1}(\tau) + (1-x)H^{-1}(\tau),$$

where $\{x_i\}$ iid $U[0,1], \Phi^{-1}(\tau)$ is the standard Gaussian quantile function and H^{-1} is the quantile function of a Gaussian mixture distribution of the form

$$H(u) = \frac{1}{2}\,\Phi((u-\mu_1)/\sigma_1) + \frac{1}{2}\,\Phi((u-\mu_2)/\sigma_2),$$

and we have taken $(\mu_1,\mu_2) = (-1.08, 1.08)$ and $(\sigma_1,\sigma_2) = (1/8, 1/8)$. These choices make the conditional dispersion of y given x roughly constant in x. In this example $n = 300$.

Each of these models may be expressed in the following quadratic formulation:

$$Q_y(\tau|x) = \beta_0(\tau) + \beta_1(\tau)x + \beta_2(\tau)x^2, \tag{14}$$

where the parametric functions $(\beta_0(\tau), \beta_1(\tau), \beta_2(\tau))$ are given in Table 1. Estimates of the parameters of these conditional quantile functions for $\tau \in \{.1, .25, .5, .75, .9\}$ are depicted by the solid lines in the first row of panels in Figure 1.

For models 1, 2, and 4 we compare the unrestricted linear model,

$$Q_y(\tau|x) = \beta_0(\tau) + \beta_1(\tau)x, \tag{15}$$

with the null model,

$$Q_y(\tau|x) = \beta_0(\tau). \tag{16}$$

For model 3 we compare (14) with (16). Because the null model (16) contains only an intercept, the appropriate form of the T_n process uses $\mathbf{\Gamma}_n = \mathbf{I}$ in all four models.

The last three rows of Figure 1 illustrate the realizations of three of the processes introduced in the previous section, $R^1(\tau), \Lambda_n(\tau)$, and $T_n(\tau)$, for the data depicted in the first row of panels corresponding to the four models.

For model 1 we would expect that the R^1 process to be nearly 0 over the entire range of $\tau \in (0,1)$, and this expectation is borne out in the panel below the data. Likewise, we would expect to find that the Λ_n and T_n processes behave according to their theory under that null hypothesis; that is, like the square of a normalized Brownian bridge process in this case—an expectation that is also consistent with the appearance of the last two panels of the first column, which represent the Λ_n and T_n processes.

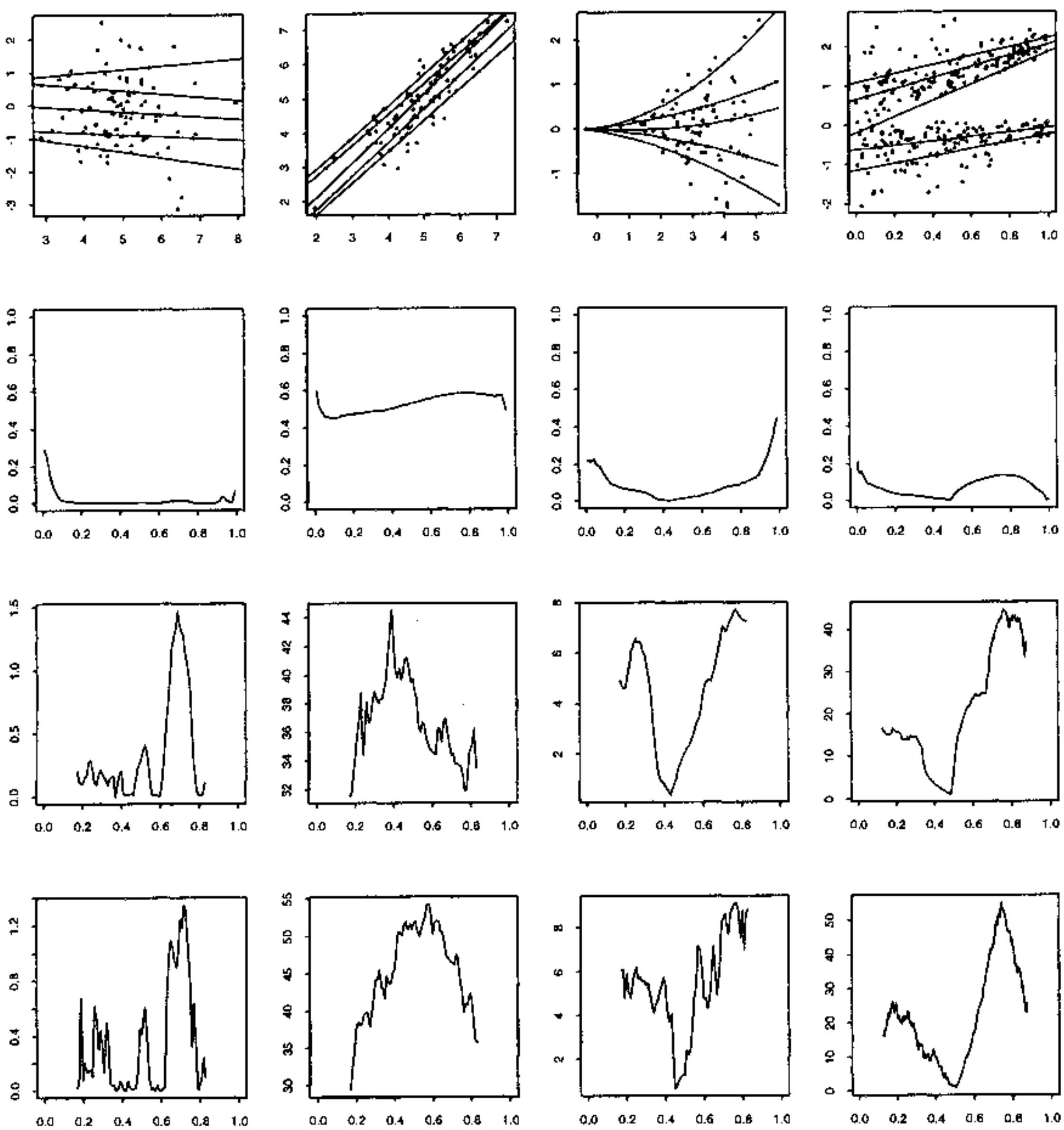

Figure 1. Some Illustrative Examples of Goodness-of-Fit Processes. The figure illustrates four different data configurations and their associated R^1; Λ_n, and T_n functions. The first row depicts the data, with the solid lines representing the fitted conditional quantile functions for $\tau \in \{.1, .25, .5, .75, .9\}$. The second row depicts the R^1 process; the third row, the Λ_n process; and the fourth row, the T_n process.

The second column of Figure 1 illustrates results for model 2. The pure location shift regression effect may be expected to yield a flat R^1 function, indicating that all conditional quantiles are equally successful in reducing variability (as measured by the weighted sum of absolute errors) relative to the unconditional counterparts. The Λ_n and T_n processes both indicate a highly significant departure from the null theory over the entire range of τ in this case, as we would expect to see.

In model 3 the conditional quantile functions fan out from the origin, and it is clear that the median fit of the full regression does not improve on the restricted fit; that is, the conditional median and the unconditional median are equal. However, for quantiles other than the median, there is a clear benefit from the quadratic form of the conditional quantile specification. This is reflected in the shallow bowl-shaped curve representing the R^1 process in this case. This pattern is also apparent in the Λ_n and T_n processes, which are marginally significant in the right tail and marginally insignificant in the left tail. Note that in this model we have estimated a quadratic quantile regression model so the L_n and T_n processes have order $q = 2$.

The last column of Figure 1 shows the results for model 4. Like model 3, this configuration also yields a much more pronounced effect away from the median, as is also apparent in the plots of the Λ_n and T_n processes. Note that in this case conventional tests of heteroscedasticity would be

Table 1. Quantile Regression Coefficients

	Model 1	Model 2	Model 3	Model 4
$\beta_0(\tau)$	0	$\Phi^{-1}(\tau)$	0	$H^{-1}(\tau)$
$\beta_1(\tau)$	0	1	$.10\Phi^{-1}(\tau)$	$1 + \Phi^{-1}(\tau) - H^{-1}(\tau)$
$\beta_2(\tau)$	0	0	$.025\Phi^{-1}(\tau)$	0

unlikely to detect any effect of the covariate x, because scale varies very little with respect to x.

As with conventional R^2, large values of R^1 and the associated test statistics correspond to situations in which *conditional* estimates differ significantly from *unconditional* estimates. Thus in the heteroscedastic cases, R^1 and the tests indicate no effect near the median, where the conditional quantile function is horizontal. However, in the tails conditional and unconditional quantiles differ substantially, as indicated by an upward curve in the tails. Similarly, in the bimodal case the conditional quantile functions at the median and in the far upper tail are essentially constant, whereas in the lower tail and upper shoulder of the distribution there is a substantial effect. In the Section 5 we report some results from a small Monte Carlo experiment that offers a somewhat more systematic evaluation of the performance of these methods.

4. AN EMPIRICAL EXAMPLE

In this section we describe a brief empirical foray into models of international economic growth designed to illustrate the use of some of the methods described earlier. The data consist of a pooled sample of 161 observations, on national growth rates for the two periods 1965–1975 and 1975–1985, taken from Barro and Lee (1994). The data are similar to those used by Barro and Sala-i-Martin (1995), but unfortunately not identical. There are initially 87 observations from 1965–1975 and 97 from 1975–1985, but only 161 after removing observations with missing data for the model estimated. This model is the same as the basic model reported by Barro and Sala-i-Martin (1995, table 12.3). The intention is to try to identify the effect of certain covariates more explicitly in terms of their influence on specific ranges of the conditional quantiles of the growth process.

Work by Barro and others has focused considerable attention on the effect of the lagged level of per-capita gross domestic product (GDP) on the growth of this variable. Figure 2 plots this effect as estimated in a quantile regression that also includes the variables indicated in Figure 3. The solid, rather jagged line in this figure is the coefficient on initial per-capita GDP from the quantile regression plotted as a function of the τ. The dotted straight line is the least squares coefficient for the same variable. The long dashed lines with intermittent dots represents a confidence band for the quantile regression process constructed from the rank test inversion approach described by Koenker (1994, 1997), but modified as in Theorem 3 to accommodate the effect of heterogeneity through the estimation of the matrix Γ_n as described in Section 2.5. The solid horizontal line at the top of the plot represents the null effect of initial GDP on growth over the entire range of τ. Clearly, Figure 2 suggests that the effect of the initial level of GDP is relatively constant over the entire distribution, with perhaps a slightly stronger effect in the upper tail. Barro and Sala-i-Martin (1995) commented that, given the 10-year increments of the data, the least squares coefficient of $-.026$ "implies that convergence occurs at the rate of 3.0 percent per year." Note that to convert to annual rates from 10-year increments (Barro and Sala-i-Martin 1995, p. 422), compute $-(1 - e^{-10\beta})/10$.

The notion of convergence implicit in this interpretation may seem a bit strained, as the model underlying the least squares fit of a pure location shift effect of initial GDP implies a stationary distribution for national incomes, a situation in which there is no convergence. Stigler (1996) considered a rather amusing, but similar empirical example from the early industrial organization literature. Note, however, that if the initial GDP effect in Figure 2 could be interpreted as suggesting a downward sloping effect, this would imply a stronger sense of convergence in which the scale of the international distribution of per-capita income would shrink over time. Obviously, the evidence for this in the figure is tenuous at best.

The effects of the remaining variables are quite varied and thus perhaps more interesting. We will try to isolate a few of these effects in our discussion. Consider first the effect of public consumption as a share of GDP. The least squares effect of $-.11$ suggests, according to Barro and Sala-i-Martin, that an increase of 6.5%, or one standard deviation, in G/Y reduces the expected growth rate by .7% per year. The quantile regression results indicate that this effect is essentially constant over the upper half of the distribution, but may be considerably larger in the lower tail. Improving the terms of trade appears to exert a monotonically increasing effect in τ, suggesting that this variable tends to help faster-growing countries proportionally more than those countries in the lower tail of the growth distribution, thus tending to accentuate the inequality among nations. In contrast, a larger black market premium appears to slow the growth of the upper tail countries more than that of the lower tail countries. None of the education effects is particularly clear. There is some indication that the effect

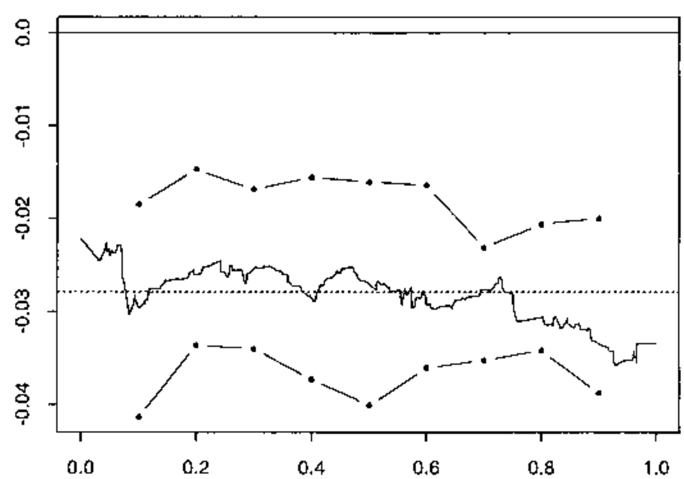

Figure 2. Estimated Quantile Regression Effect of Initial Per-Capita GDP on GDP Growth. The figure illustrates the estimated quantile regression process for initial per capita GDP in a model for GDP growth that also includes the variables indicated in Figure 3. The rather jagged, solid line is the quantile regression process for the initial GDP effect, the horizontal dotted line indicates the estimated conditional mean (location-shift) effect estimated by ordinary least squares, and the dashed lines indicate a rank test—based confidence band for the quantile regression effect. The band clearly excludes the null effect indicated by the solid horizontal line at 0 at the top of the figure.

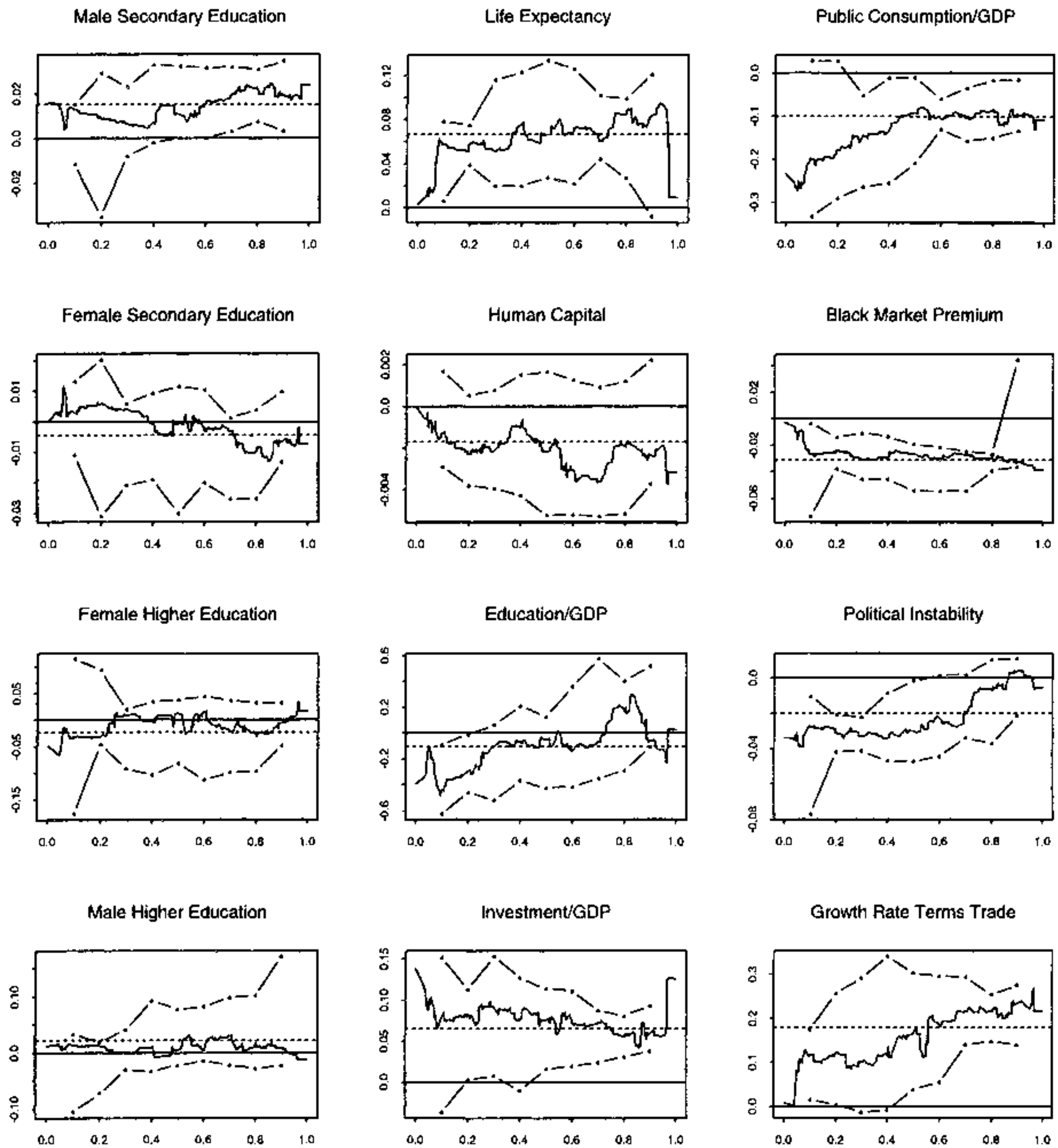

Figure 3. Estimated Quantile Regression Effects of Other Variables on GDP Growth. The figure illustrates the estimated quantile regression process for 12 other variables in a model for GDP growth. The format of the plot follows that of Figure 2.

of male secondary education is increasing in τ, again suggesting that the effect of an across-the-board increase in this variable would be to accentuate international inequality. Curiously, the human capital variable appears to exhibit the opposite effect, but, as the confidence band suggests, this effect is rather weak.

Individually, the five education variables seem to exert a rather weak effect on growth, but, not surprisingly, it is difficult to distinguish their separate effects. An obvious application of the hypothesis testing apparatus introduced in Section 2 is to consider the joint effect of the education variables in the foregoing model. This is done in Figure 4, where we plot the $R^1(\tau)$ function and the corresponding $\Lambda_n(\tau)$ and $T_n(\tau)$ processes corresponding to the null hypothesis of no joint effect of the education variables. Here, because the null hypothesis imposes no presumption of homogeneity on the elements of Γ_n, we have estimated this matrix as described in Section 2.5. In this plot any departure from the null, whether positive or negative, appears as a positive effect, so we can no longer distinguish the nature of the effects in quite the detail provided by the plots of the individual effects of the variables. From work of Andrews (1993), we find that for $q = 5$, the critical value for the $\sup Q^2(\tau)$ over the interval $[.05, .95]$ at $\alpha = .05$ is given by 19.61. Thus the Λ_n version of the test fails to reject the null of no joint significance; however, the rank-based T_n version of the test does, marginally, reject. Because iid error conditions are quite clearly implausible here, the validity of the Λ_n test is highly questionable. We try to shed

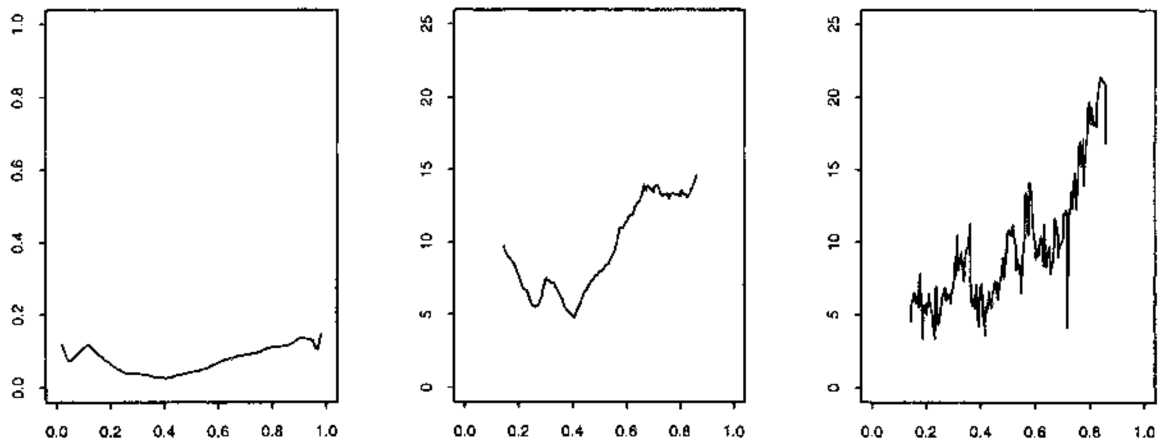

Figure 4. Joint Effect of Education Variables on GDP Growth. (a) The $R^1(\tau)$ function corresponding to the null hypothesis of no effect of any of the five education variables in Figure 3. Note that the somewhat bowl-shaped R^1 plot suggests that these variables taken together have somewhat larger positive effects on growth in both tails of the conditional growth distribution than they do in the center of the distribution. This feature is also apparent in the adjacent Λ_n process plot (b). However, the corresponding rank-test, T_n, (c) appears to indicate a monotonically increasing effect of the education variables.

more light on this conflict between the tests in the next section.

5. MONTE CARLO

In this section we report on a small Monte Carlo experiment designed to evaluate the performance of the asymptotic theory developed in Section 2 as a guide to the finite-sample behavior of the proposed tests. To this end, we reconsider the four models introduced in Section 3 and try to evaluate size and power of the Λ_n and T_n versions of the tests for several different sample sizes. At this stage we make no effort to compare power of these tests with other tests available for these hypotheses, but it is clear that because the alternative is rather general, we would expect to be able to provide more powerful tests for each of explicit alternatives.

Table 2 reproduces a small extract from Andrews's (1993) table of critical values for the $\sup_{\tau\in\mathcal{T}} Q_q^2(\tau)$ process. We consider only the case of $\mathcal{T} = [.05, .95]$ and $q \in \{1, 2, 4\}$. To help calibrate the values in this table, we might recall that the critical values for chi-squared for 1, 2, and 4 degrees of freedom are 3.84, 5.99, and 9.49 at $\alpha = .05$.

To evaluate the size of the sup tests, we consider the model

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + u_i, \tag{17}$$

Table 2. Critical Values for sup Λ_n and sup T_n Tests

q	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$
1	8.19	9.84	13.01
2	11.20	12.93	16.44
4	15.62	17.56	21.54

NOTE: The entries constitute a small extract from a much larger table of Andrews (1993). Each entry is c_α such that $\Pr(\sup_{\{\tau\in[.05,.95]\}} Q_q^2 < c_\alpha] = \alpha$. The "degrees of freedom" parameter q is the number of linear restrictions imposed by the null hypothesis. These critical values are used in the Monte Carlo evaluation of the size and power performance of the tests.

where $x_{i1} \equiv 1$ and $\{x_{ij}: j = 2, \ldots, p\}$ are iid from the standard normal, $\mathcal{N}(0,1)$, distribution. The $\{u_i\}$'s are drawn as iid from $\mathcal{N}(0,1)$, Cauchy, or χ_4^2 distributions. For the size simulations reported in Table 3, we take $\beta = (\beta_i) = 0$. The null hypothesis for the size simulations is, of course, $\beta_j = 0, j = 2, \ldots, p$, and we take $q = p - 1$ with $p = 2, 3$, or 5. The proportion of rejections in 500 replications for each of the two tests is reported in Table 3. It is clear that both tests are undersized, rejecting at less than the specified nominal rates. The T_n test performs somewhat better in this respect than the Λ_n test, but it is clear that both tests would benefit from some size correction.

Table 4 considers the power of the $\sup \Lambda_n$ and $\sup T_n$ tests in a setting that differs from the size simulations in only one aspect: The parameter β_2 is now set to 1/2, rather than 0. To provide some comparison with conventional hypothesis testing methods in this iid error linear model setting, the table includes entries reporting the rejection frequencies of the classical F test. We expect to find better performance from the F test in the Gaussian and chi-squared cases, but in the Cauchy case the F test is poor, whereas the $\sup \Lambda_n$ and $\sup T_n$ tests perform quite well.

To further illustrate the power of the new tests in cases where the F test is unsuccessful, we consider in Table 5 two variants of the heteroscedastic and bimodal examples of Section 3. For the heteroscedastic model (model 3 of Sec. 3), we adopt the quadratic specification used in Figures 1,

$$y_i = \alpha + (\beta x_{2i} + \gamma x_{2i}^2)u_i, \tag{18}$$

with $\{x_{1i} = 1\}, \{x_{ji}\}$ iid $\mathcal{N}(3,1)$, for $j = 2, \ldots, p$, and $\{u_i\}$ iid $\mathcal{N}(0, 1/100)$, with $\alpha = 0, \beta = 1$, and $\gamma = 1/4$. In this case,

$$E(y_i|x_i) = 0,$$

so the classical regression F test has asymptotic power equal size, but the quantile regression tests have power from

Table 3. Monte Carlo Sizes of the sup Λ_n and sup T_n Tests

			Gaussian			Cauchy			χ^2_4		
Test	n	q	.1	.05	.01	.1	.05	.01	.1	.05	.01
Λ_n	100	1	.000	.000	.000	.004	.000	.000	.002	.002	.000
T_n	100	1	.042	.018	.002	.044	.022	.006	.044	.008	.006
Λ_n	100	2	.004	.002	.000	.002	.002	.000	.004	.002	.000
T_n	100	2	.042	.022	.002	.040	.024	.008	.038	.018	.002
Λ_n	100	4	.000	.000	.000	.002	.000	.000	.000	.000	.000
T_n	100	4	.038	.018	.002	.038	.022	.002	.038	.018	.002
Λ_n	500	1	.022	.010	.002	.006	.004	.000	.028	.010	.000
T_n	500	1	.058	.032	.012	.056	.022	.006	.078	.036	.006
Λ_n	500	2	.024	.010	.002	.004	.000	.000	.012	.004	.000
T_n	500	2	.068	.040	.012	.058	.020	.004	.056	.028	.006
Λ_n	500	4	.012	.006	.002	.004	.000	.000	.012	.006	.000
T_n	500	4	.066	.042	.010	.066	.034	.006	.054	.026	.004

NOTE: The table is based on 500 replications per cell. Each cell reports the proportion of rejections of the designated test at the designated level of significance using the critical values appearing in Table 2 taken from Andrews (1993). The model in each case is (17) with $x_{1i} \equiv 1$ and $\{x_{ij}: j = 2, \ldots, p\}$ iid standard Gaussian and $\{y_i\}$ iid from one the three distributions indicated at the top of the table. In each case $\mathcal{T} = [.05, .95]$ and $q = p - 1$.

both tails of the distribution, because the conditional quantiles process is $Q_y(\tau|x) = (x + 1/4x^2)\Phi^{-1}(\tau)$.

In the bimodal example (model 4 of Sec. 3), we set

$$y_i = (1 - x_{2i})\Phi^{-1}(u_i) + x_{2i}H^{-1}(u_i), \tag{19}$$

where $\{x_{1i} = 1\}, x_{ji} \sim \mathcal{U}[0,1]$, for $j = 2, \ldots, p, u_i \sim \mathcal{U}[0,1], \Phi^{-1}$ is the standard Gaussian quantile function, and H^{-1} is the quantile function of the Gaussian mixture with distribution function,

$$H(x) = \frac{1}{2}\ \Phi((x - \mu_1)/\sigma_1) + \frac{1}{2}\ \Phi((x - \mu_2)/\sigma_2),$$

which needs to be computed numerically. We again chose $(\mu_1, \mu_2) = (-1.08, 1.08)$ and $(\sigma_1, \sigma_2) = (1/8, 1/8)$ so H is symmetric and bimodal. As in the heteroscedastic case, for this model,

$$E(y_i|x_i) = 0,$$

so the F test has asymptotic power equal size, but the quantile regression tests may be expected to perform well, because there are linear conditional quantile functions with nonzero slope for all $\tau \neq 1/2$ (notice that $Q_y(1/2|x) = 0$). Several features of Table 5 merit attention. The new tests are a clear success at $n = 500$, where they have power near 1 in virtually all cases. In contrast, the F test performs abysmally as expected; with larger p, there is some apparent power from the F test, but this is spurious, because $E(y|x) = 0$ in these cases. With $n = 100$, the alternative is obviously more difficult to discern. Here it is clear that the sup T test has a real advantage over the sup Λ_n. This is presumably largely a consequence of the size distortion that we have already observed in Table 3. Because both the T_n and Λ_n tests are significantly undersized, effective size correction would presumably yield improvement in power in both Tables 4 and 5. This is a topic that deserves further investigation.

Table 4. Monte Carlo Power of the sup Λ_n and sup T_n Tests in Linear Models with iid Errors

			Gaussian			Cauchy			χ^2_4		
Test	n	q	.1	.05	.01	.1	.05	.01	.1	.05	.01
Λ_n	100	1	.916	.820	.546	.472	.354	.126	.336	.202	.054
T_n	100	1	.988	.960	.850	.750	.616	.384	.536	.418	.206
F	100	1	1.000	1.000	1.000	.354	.254	.130	.964	.940	.842
Λ_n	100	2	.788	.640	.314	.312	.194	.054	.180	.084	.020
T_n	100	2	.952	.906	.708	.604	.468	.282	.368	.270	.100
F	100	2	1.000	1.000	1.000	.528	.446	.310	.982	.974	.938
Λ_n	100	4	.514	.348	.116	.120	.072	.010	.078	.036	.008
T_n	100	4	.888	.776	.514	.474	.334	.136	.288	.182	.064
F	100	4	1.000	1.000	1.000	.840	.796	.646	.996	.992	.982
Λ_n	500	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.994
T_n	500	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.996
F	500	1	1.000	1.000	1.000	.318	.226	.118	1.000	1.000	1.000
Λ_n	500	2	1.000	1.000	1.000	1.000	1.000	.996	.998	.998	.984
T_n	500	2	1.000	1.000	1.000	1.000	1.000	.998	.998	.996	.984
F	500	2	1.000	1.000	1.000	.514	.430	.278	1.000	1.000	1.000
Λ_n	500	4	1.000	1.000	1.000	1.000	.994	.988	.992	.986	.926
T_n	500	4	1.000	1.000	1.000	1.000	1.000	.996	1.000	.994	.972
F	500	4	1.000	1.000	1.000	.812	.758	.652	1.000	1.000	1.000

NOTE: The table is based on 500 replications per cell. Each cell reports the proportion of rejections of the designated test at the designated level of significance using the critical values appearing in Table 2 taken from Andrews (1993). The model in each case is (17) with $x_{1i} \equiv 1$ and $\{x_{ij}: j = 2, \ldots, p\}$ iid standard Gaussian, $\beta = (0, .5, 0, \ldots, 0) \in R^p$, and $\{u_i\}$ iid from one the three distributions indicated at the top of the table. In each case $\mathcal{T} = [.05, .95]$ and $q = p - 1$.

Table 5. Monte Carlo Power of the sup Λ_n and sup T_n Tests in a Bi-Modal Transition Model and a Purely Heteroscedastic Model

			Bimodal			Heteroscedastic		
Test	n	q	.1	.05	.01	.1	.05	.01
Λ_n	100	1	.234	.142	.030	.226	.098	.006
T_n	100	1	.314	.174	.054	.836	.722	.472
F	100	1	.116	.066	.026	.196	.122	.050
Λ_n	100	2	.104	.050	.006	.070	.014	.000
T_n	100	2	.192	.114	.036	.706	.582	.332
F	100	2	.298	.222	.106	.346	.282	.140
Λ_n	100	4	.024	.008	.000	.002	.000	.000
T_n	100	4	.130	.072	.008	.534	.388	.154
F	100	4	.686	.608	.466	.798	.722	.540
Λ_n	500	1	1.000	1.000	.994	1.000	1.000	1.000
T_n	500	1	1.000	1.000	.992	1.000	1.000	1.000
F	500	1	.096	.050	.010	.164	.100	.024
Λ_n	500	2	1.000	.994	.958	1.000	1.000	1.000
T_n	500	2	1.000	.998	.962	1.000	1.000	1.000
F	500	2	.314	.216	.102	.404	.318	.148
Λ_n	500	4	.994	.962	.812	1.000	1.000	1.000
T_n	500	4	.996	.978	.900	1.000	1.000	1.000
F	500	4	.738	.650	.480	.760	.706	.540

NOTE: The table is based on 500 replications per cell. Each cell reports the proportion of rejections of the designated test at the designated level of significance using the critical values appearing in Table 2. The model in the heteroscedastic cases is (18) and in the bimodal case is (19). In each case $\mathcal{T} = [.05, .95]$, and $q = p - 1$.

6. CONCLUSIONS

We have suggested a goodness-of-fit criterion for quantile regression that generalizes the proposal of McKean and Sievers (1987) for l_1 regression and adapts it to an arbitrary linear hypothesis. We have introduced three related goodness-of-fit processes, $\Lambda_n(\tau)$, $\mathbf{W}_n(\tau)$, and $T_n(\tau)$, which significantly expand the available toolkit for inference in quantile regression models. For a given τ, the proposed test may be regarded as variants of the LR, Wald, and Lagrange multiplier tests for l_1 regression. The tests have an asymptotic theory involving Bessel processes that is closely related to earlier work on inference for goodness of fit as well as to more recent work on inference in changepoint models. We have focused on $\sup \Lambda_n(\tau)$ and $\sup T_n(\tau)$ forms of the tests, but there is considerable scope for expanding the range of tests. Extensions to Cramér–von Mises forms are obviously possible along the lines of work done by Portnoy (1992). It would also be useful to investigate extensions to null hypotheses that involve unknown nuisance parameters along the lines of work of Durbin (1973) and others.

APPENDIX: PROOFS

Proof of Theorem 1

Let

$$V_n(\delta,\tau) = \sum \rho_\tau(u_i(\tau) - \omega(\tau)\mathbf{x}_i'\delta/\sqrt{n}),$$

where $\omega(\tau) = \lambda(\tau)s(\tau), \lambda^2(\tau) = \tau(1-\tau)$, and $u_i(\tau) = u_i - F^{-1}(\tau)$. Because in model (1), by A.2 (a), $\beta(\tau) = \beta + F^{-1}(\tau)\mathrm{e}$, we may interpret δ as $\delta = \sqrt{n}(\mathbf{b} - \beta)/\omega(\tau)$ for some choice of b. Lemma 3.1 of GJKP may be recast slightly to provide the following uniform asymptotically quadratic approximation for $V_n(\delta, t)$: Under conditions A.1–A.3, for any fixed $C > 0$ and $\alpha \in (0, 1/2)$,

$$\sup\{|\mathbf{W}_n(\delta,\tau)|\colon \|\delta\| \le C\sqrt{\log\log n}, \tau \in \mathcal{T}\} \to 0$$

and

$$\mathbf{W}_n(\delta,\tau) = (\lambda^2(\tau)s(\tau))^{-1}\{V_n(\delta,\tau) - V_n(0,\tau)\} - \left[\frac{1}{2}\,\delta'\mathbf{D}\delta - (\lambda(t)^{-1}\delta'\mathbf{g}_n)\right],$$

where $\mathbf{g}_n = n^{-1/2}\sum \mathbf{x}_i\psi_\tau(u_i(\tau))$ and $\psi_\tau(u) = \tau - I(u < 0)$.

Theorem 1 of GJ established that

$$\hat\delta_n(\tau) = \sqrt{n}(\hat\beta_n(\tau) - \beta(\tau))/\omega(\tau) = \lambda(\tau)^{-1}\mathbf{D}^{-1}\mathbf{g}_n + o_p(1)$$

where the representation holds uniformly on $\mathcal{T}$. Thus

$$\hat\delta_n \Rightarrow \lambda^{-1}\mathbf{D}^{-1/2}\mathbf{B}_p$$

Similarly, for the restricted estimator, we have

$$\tilde\delta_n(\tau) = \begin{pmatrix} \mathbf{D}_{11}^{-1}\mathbf{D}_{12}\zeta(\tau)/\omega(\tau) + \mathbf{D}_{11}^{-1}\mathbf{g}_{1n}/\lambda(\tau) \\ -\zeta(\tau)/\omega(\tau) \end{pmatrix},$$

where we partition $\mathbf{g}_n = (\mathbf{g}_{1n}', \mathbf{g}_{2n}')'$ to conform with $\beta = (\beta_1', \beta_2')'$. Now substituting back into the quadratic approximation, we may write (see, e.g., Koenker and Bassett 1982)

$$\begin{aligned} L_n &= 2(\tilde V - \hat V)/(\lambda\omega) \\ &= (\tilde\delta'\mathbf{D}\tilde\delta - 2\tilde\delta'\mathbf{g}_n/\lambda) - (\hat\delta'\mathbf{D}\hat\delta - 2\hat\delta'\mathbf{g}_n/\lambda) + o_p(1) \\ &= \mathbf{h}'\mathbf{D}^{22}\mathbf{h} + o_p(1), \end{aligned}$$

where $\mathbf{h}(\tau) = (\mathbf{D}^{22})^{-1}\zeta(\tau)/\omega(\tau) + (\mathbf{g}_{2n} - \mathbf{D}_{21}\mathbf{D}_{11}^{-1}\mathbf{g}_{1n}/\lambda(\tau)$.

Finally, note that

$$h \Rightarrow (\mathbf{D}^{22})^{-1}\zeta/\omega + (\mathbf{D}^{22})^{-1/2}\mathbf{B}_q/\lambda,$$

so we may write

$$L_n = \|\boldsymbol{\mu} + \mathbf{B}_q\|^2/\lambda^2 + o_p(1),$$

where $\mu = \lambda\omega^{-1}(\mathbf{D}^{22})^{-1/2}\zeta$, and this yields the noncentrality function

$$\eta = \|\boldsymbol{\mu}\|^2/\lambda^2 = \boldsymbol{\zeta}'(\mathbf{D}^{22})^{-1}\zeta/\omega^2.$$

Proof of Theorem 2

The result follows from the argument given in the text.

Proof of Theorem 3

Let $\tilde{\mathbf{X}}_2 = \mathbf{X}_2 - \hat{\mathbf{X}}_2$, so $\tilde{\mathbf{x}}_{2i} = \mathbf{x}_{2i} - \hat{\mathbf{x}}_{2i}$, and denote

$$\hat{\mathbf{W}}_{n2}(\tau) = n^{-1/2} \sum \tilde{\mathbf{x}}_{2i}(\hat{a}_i(\tau) - (1-\tau)),$$

$$\mathbf{W}_{n2}(\tau) = n^{-1/2} \sum \tilde{\mathbf{x}}_{2i}(\tilde{a}_i(\tau) - (1-\tau)),$$

and

$$\mathbf{W}_{n1}(\tau) = n^{-1/2} \sum \mathbf{x}_{1i}(\tilde{a}_i(\tau) - (1-\tau)),$$

where $\tilde{a}_i(\tau) = I(u_i > F^{-1}(\tau))$. By (5.26) of GJ, we have the uniform (on $\mathcal{T}$) representation

$$\begin{aligned}\hat{\mathbf{W}}_{n2}(\tau) &= \mathbf{W}_{n2}(\tau) - \tilde{\mathbf{X}}_2'\Gamma^{-1}\mathbf{X}_1(\mathbf{X}_1'\Gamma^{-1}\mathbf{X}_1)^{-1}\mathbf{W}_{n1}(\tau) + o_p(1) \\ &= \mathbf{W}_{n2}(\tau) + o_p(1).\end{aligned}$$

So, integrating as for (4.2) of GJKP, we have

$$\mathbf{S}_n = n^{-1/2} \sum \tilde{\mathbf{x}}_{2i}\varphi(F(u_i)) + o_p(1)$$

and, consequently, $T_n \rightsquigarrow \chi_q^2$ under H_0.

Under H_n, by contiguity we have the representation on $\mathcal{T}$,

$$\begin{aligned}S_n &= n^{-1/2} \sum \tilde{\mathbf{x}}_{2i}\varphi(F(u_i - n^{-1/2}\mathbf{x}_{2i}'\zeta(F(u_i))/\sigma_i)) + o_p(1) \\ &= n^{-1} \sum (\tilde{\mathbf{x}}_{2i}\mathbf{x}_{2i}/\sigma_i)\varphi'(F(u_i))f(u_i)\zeta(F(u_i)) \\ &\quad + n^{-1/2} \sum \tilde{\mathbf{x}}_{2i}\varphi(F(u_i)) + o_p(1).\end{aligned}$$

Using conventional matrix partitioning notation,

$$n^{-1} \sum \tilde{\mathbf{x}}_{2i}\mathbf{x}_{2i}/\sigma_i = \mathbf{G}_{n22} - \mathbf{G}_{n12}\mathbf{G}_{n11}^{-1}\mathbf{G}_{n21} \equiv (\mathbf{G}_n^{22})^{-1}.$$

Integrating yields, $ES_n = (\mathbf{G}_n^{22})^{-1}\lambda(\varphi, \zeta, F)$, and the result then follows by verifying the rather tedious partitioned matrix identity $(\mathbf{G}_n^{22})\mathbf{M}_n(\mathbf{G}_n^{22}) = \mathbf{\Omega}$.

Proof of Theorem 4

Note that with the quantile score function $\varphi_\tau(t) = \tau - I(t < \tau)$,

$$\mathbf{S}_n(\tau) \equiv \hat{\mathbf{W}}_{n2}(\tau) = \mathbf{W}_{n2}(\tau) + o_p(1)$$

uniformly on $\mathcal{T}$ under H_0. By lemma 2 of GJ,

$$\mathbf{W}_{n2}(\tau) \Rightarrow \mathbf{M}^{1/2}\mathbf{B}_q(t),$$

which proves the result under H_0. Under H_n, write

$$\Delta_n(\tau) \equiv \sqrt{n}(\hat{\beta}_1(\tau) - \beta_1(\tau), -\zeta(\tau)/\sqrt{n}).$$

The representation (5.15) of GJ yields

$$\begin{aligned}\mathbf{S}_n(\tau) &= \mathbf{W}_{n2}(\tau) - (1/s(\tau))[(1/n)\sum \tilde{\mathbf{x}}_{2i}\mathbf{x}_i'/\sigma_i]\Delta_n(\tau) + o_p(1) \\ &= \mathbf{W}_{n2}(\tau) + (1/s(\tau))(\mathbf{G}^{22})^{-1}\zeta(\tau) + o_p(1),\end{aligned}$$

because $(1/n)\sum \tilde{\mathbf{x}}_{2i}\mathbf{x}_{1i}'/\sigma_i = 0$. The result then follows by lemma 2 of GJ.

[Received February 1998. Revised April 1999.]

REFERENCES

Abberger, K. (1996), "Nonparametric Quantile Forecasting and Its Application to Stock Returns," Discussion Paper SFB 178, University of Konstanz.

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856.

Barro, R., and Lee, J.-W. (1994), "Data Set for a Panel of 138 Countries," discussion paper, NBER, http://www.nber.org/pub/barro.lee.

Barro, R. J., and Sala-i-Martin, X. (1995), *Economic Growth*, New York: McGraw-Hill.

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

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

Chamberlain, G. (1994), "Quantile Regression Censoring and the Structure of Wages," in *Advances in Econometrics*, ed. C. Sims, New York: Elsevier.

DeLong, D. M. (1981), "Crossing Probabilities for a Square Root Boundary By a Bessel Process," *Communications in Statistics, Part A—Theory and Methods*, 10, 2197–2213.

Doksum, K. (1974), "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, 2, 267–277.

Durbin, J. (1973), *Distribution Theory for Tests Based on the Sample Distribution Function*, Philadelphia: SIAM.

Foutz, R. V., and Srivastava, R. C. (1977), "The Performance of the Likelihood Ratio Test When the Model Is Incorrect," *The Annals of Statistics*, 5, 1183–1194.

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

Gutenbrunner, C., Jurečková, J., Koenker, R., and Portnoy, S. (1993), "Tests of Linear Hypotheses Based on Regression Rank Scores," *Journal of Nonparametric Statistics*, 2, 307–331.

Hájek, J., and Šidák, Z. (1967), *Theory of Rank Tests*. New York: Academic Press.

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

Hendricks, W., and Koenker, R. (1992), "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association*, 87, 58–68.

Horowitz, J. L. (1992), "The Role of the List Price in Housing Markets: Theory and an Econometric Model," *Journal of Applied Econometrics*, 7, 115–129.

Jurečková, J. (1992), "Tests of Kolmogorov–Smirnov Type Based on Regression Rank Scores," *Transactions of the 11th Prague Conference on Information Theory, Random Processes and Statistical Decision Functions*, B, 41–49.

Kiefer, J. (1959), "K-Sample Analogues of the Kolmogorov–Smirnov and Cramér–von Mises Tests," *Annals of Mathematical Statistics*, 30, 420–447.

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

——— (1997), "Rank Tests for Linear Models," in *Handbook of Statistics*, eds. G. Maddala and C. Rao, Amsterdam: North-Holland.

Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

——— (1982), "Tests of Linear Hypotheses and l_1 Estimation," *Econometrica*, 50, 1577–1584.

Lehmann, E. (1974), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

McKean, J. W., and Sievers, G. L. (1987), "Coefficients of Determination for Least Absolute Deviation Analysis," *Statistical and Probability Letters*, 5, 49–54.

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105–121.

Portnoy, S. (1992), "A Regression Quantile-Based Statistic for Testing Nonstationary Errors," in *Nonparametric Statistics and Related Topics*, ed. A. Saleh, North Holland: Amsterdam, pp. 191–203.

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

Ronchetti, E. (1982), "Robust Alternatives to the F Test for the Linear

Model," in *Probability and Statistical Inference*, eds. W. G. P. Grossmann and W. Wertz, Dordrecht: Reidel, pp. 329–342.

Rousseeuw, P., and Hubert, M. (1997), "Recent Developments in PROGRESS," in L_1*-Statistical Procedures and Related Topics*, ed. Y. Dodge, Hayward, CA: IMS.

Sheather, S., and Maritz, J. (1983), "An Estimate of the Asymptotic Standard Error of the Sample Median," *Australian Journal of Statistics*, 25, 109–122.

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

Stigler, S. M. (1996), "The History of Statistics in 1933," *Statistical Science*, 11, 244–252.

Tukey, J. (1965), "Which Part of the Sample Contains the Information?," *Proceedings of the National Academy*, 53, 127–134.

Welsh, A. (1987), "Kernel Estimates of the Sparsity Function," in *Statistical Data Analysis Based on the* L_1*-Norm and Related Methods*, ed. Y. Dodge, New York: Elsevier, pp. 369–377.