# GOODNESS OF FIT AND RELATED INFERENCE PROCESSES FOR QUANTILE REGRESSION

ROGER KOENKER AND JOSÉ A.F. MACHADO

Abstract. 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 approach is illustrated with some hypothetical examples, an application to recent empirical models of international economic growth, and some Monte-Carlo evidence.

## 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 non-parametric 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 paper 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, a model 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

Version: February 9, 1999. This research was partially supported by NSF grant SBR-9617206 to Koenker, and FEDER (Praxis XXI program grant 2/21/ECO/94) to Machado. We would like to thank Steve Portnoy, a referee, and an associate editor for very helpful comments and Carlos Arango for research assistance on the empirical growth example.

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

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

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

$$\Delta(x) = G^{-1}(F(x)) - x. \tag{1.1}$$

Thus, on changing variables so $\tau = F(x)$ 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 respectively. If we formulate the quantile regression model for the binary treatment problem as,

$$Q_{Y_i}(\tau|D_i) = \alpha(\tau) + \delta(\tau)D_i \tag{1.2}$$

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

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

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

Doksum suggests that we may interpret the treament 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 $\tau$, the quantile of the survival distribution at which the subject would appear if untreated, i.e., $(Y_i|D_i = 0) = \alpha(\tau)$. And the treatment, under the Lehmann-Doksum model, is assumed to alter the subjects control response, $\alpha(\tau)$, making it $\alpha(\tau) + \delta(\tau)$ under the treatment. If the treatment is beneficial in the sense that,

$$\Delta(x) \geq 0 \quad \text{for all } x$$

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 $x_i$,

$$E(y_i|x_i) = x_i'\beta, \tag{1.3}$$

which we partition as,

$$E(y_i|x_i) = x_{i1}'\beta_1 + x_{i2}'\beta_2. \tag{1.4}$$

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

$$H_0: \quad \beta_2 = 0. \tag{1.5}$$

Denoting the error sum of squares under the restricted and unrestricted forms of the model by $\tilde{S}$ and $\hat{S}$, respectively, we may define an $R^2$ of model (1.4) relative to the restricted model constrained by the hypothesis (1.5) as

$$R^2 = 1 - \hat{S}/\tilde{S}. \tag{1.6}$$

Conventionally, restricting (1.4) to include only an "intercept" parameter yields the $R^2$ one usually sees reported.

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

$$Q_Y(\tau|x) = x_1'\beta_1(\tau) + x_2'\beta_2(\tau),$$

and denote $\hat{\beta}(\tau)$ as the minimizer of,

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

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

$$\tilde{V}(\tau) = \min_{\{b_1 \in \mathbf{R}^{p-q}\}} \sum \rho_\tau(y_i - x_{1i}'b_1).$$

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

$$R^1(\tau) = 1 - \hat{V}(\tau)/\tilde{V}(\tau) \tag{1.7}$$

which is the natural analogue of $R^2$ defined above.

Like $R^2$, since $\tilde{\beta}(\tau)$ is obtained by constraining $\hat{\beta}(\tau)$, it is immediately apparent that $\hat{V}(\tau) \leq \tilde{V}(\tau)$ and therefore $R^1(\tau)$ 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 whole conditional distribution, like $R^2$.

One can easily imagine circumstances under which the covariates $x_2$ 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 percent mean wage premium for union membership estimated by least squares, when deconstructed by quantile regression, was seen to constitute a much larger, nearly 30 percent, premium in the lower tail of the conditional wage distribution, but gradually attentuated to the point that it was essentially negligible above the median.

The use of the optimized value of the objective function defining a regression estimator as in (1.7) as a measure of goodness-of-fit is sufficiently natural that it is quite commonplace in the robustness literature. Three recent references are Rousseeuw and Hubert (1997), Abberger (1996), and Horowitz (1992) who suggest this idea for least median of squares and $\ell_1$ forecasting respectively. An earlier reference is McKean and Sievers (1987) who suggest $R^1(1/2)$ for the $\ell_1$ regression estimator and provide a very complete analysis of its properties, including the connection to a quasi-likelihood ratio test.

If the full model (1.4) is better at the $\tau$-quantile than the restricted model constrained by (1.5), $\hat{V}_n(\tau)$ should be significantly smaller than $\tilde{V}_n(\tau)$. Here "better" should be understood to mean that the $\tau$th conditional quantile function is significantly altered by the influence of the covariates $x_2$. To assess the goodness of fit of the two nested models for the whole distribution, it is natural to compare the *processes* $\hat{V}_n(\tau)$ and $\tilde{V}_n(\tau)$. The $R^1(\tau)$ process (1.7) is one possible gauge of the diferences between the processes $\hat{V}(\tau)$ and $\tilde{V}(\tau)$ for the whole distribution. But such a comparison also provides the basis for formal tests of of the hypothesis $H_0$. We consider several such variants below. The choice of these statistics is motivated by analogies to likelihood ratio 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, employing 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. Section 5 reports a small Monte-Carlo study of the finite sample properties of the proposed tests. Section 6 offers some conclusions and describes some prospects for new research. Proofs appear 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 likelihood ratio 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,

$$\ell(\beta) = n\log(\tau(1-\tau)) - \sum \rho_\tau(y_i - x_i'\beta),$$

and thus, the likelihood ratio statistic for testing $H_0$ would be,

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

Under these conditions, and $H_0$, conventional theory of the likelihood ratio implies that $-2\log\lambda_n(\tau)$ will be asymptotically $\chi^2_q$. 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 in Koenker and Bassett (1982) slightly, it can be shown that, under $H_0$,

$$L_n(\tau) \equiv \frac{2(\tilde V(\tau) - \hat V(\tau))}{\tau(1-\tau)s(\tau)} \rightsquigarrow \chi^2_q, \tag{2.1}$$

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 will 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 likelihood ratio statistic,

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

The asymptotic behavior of this version of the likelihood ratio statistic follows easily from the preceeding 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 $\sigma(\tau) = E\rho_\tau(u) < \infty$, so $\hat\sigma(\tau) = n^{-1}\hat V(\tau) \to \sigma(\tau)$. Thus, under $H_0$, by (2.1),

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

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 are termed $\rho$-tests by Ronchetti (1982). Following this terminology, we will refer to these LR-type tests based on $L_n(\tau)$ and $\Lambda_n(\tau)$ as quantile $\rho$-tests.

2.2. **LR 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 $\rho$-test statistics of the form, for example, $(L_n(\tau_1), L_n(\tau_2), ..., L_n(\tau_m))$. Such results are subsumed in the following theory for the $\rho$-test *processes*, $\{L(\tau) : \tau \in [\epsilon, 1-\epsilon]\}$, and $\{\Lambda(\tau) : \tau \in [\epsilon, 1-\epsilon]\}$.

In this subsection we restrict attention to the linear model

$$y_i = x_{i1}\beta_1 + x_{2i}\beta_2 + u_i \tag{2.2}$$

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

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

The sequence of design matrices $\{X_n\} = \{(x_i)_{i=1}^n\}$ are assumed to satisfy:

**A. 2.** **(i):** $x_{i11} \equiv 1, \quad i = 1, 2, \ldots$
**(ii):** $D_n = n^{-1}X_n'X_n \to D$, *a positive definite matrix.*
**(iii):** $n^{-1}\sum \|x_i\|^4 = O(1)$
**(iv):** $\max_{i=1,\ldots,n} \|x_i\| = O(n^{\frac{1}{4}}/\log n)$

We will consider tests of the hypothesis:

$$H_0: \qquad \beta_2(\tau) = 0 \qquad \tau \in \mathcal{T} \tag{2.3}$$

for some index set $\mathcal{T} \in [0, 1]$, against a sequence of local alternatives formulated as

**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 $X_n$ are used in Gutenbrunner, Jurečková, Koenker, and Portnoy (1993) 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 in Gutenbrunner and Jurečková (1992). We wil abbreviate these references in what follows as GJKP and GJ. Condition A.3 is a direct analogue of Condition A.3 in 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 since it permits the $x_2$ effect to depend on $\tau$.

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 $W_q(t)$ denote a $q$-vector of independent Brownian motions and thus, for $t \in [0,1]$,

$$B_q(t) = W_q(t) - tW_q(1)$$

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

$$B_q(t) \sim \mathcal{N}(0, t(1-t)I_q). \tag{2.4}$$

The normalized Euclidean norm of $B_q(t)$,

$$Q_q(t) = ||B_q(t)||/\sqrt{t(1-t)}$$

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 Kiefer (1959). Again, for any fixed $t \in (0,1)$ we have, from (2.4), $Q_q^2(t) \sim \chi_q^2$. Thus, we may interpret $Q_q^2(t)$ as a natural extension of the familiar univariate $\chi^2$ 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 non-central version of the squared Bessel process as an extension of the noncentral $\chi^2$ distribution. Let $\mu(t)$ be a fixed, bounded function from $[0,1]$ to $\Re^q$. The standardized squared norm

$$||\mu(t) + B_q(t)||^2/(t(1-t))$$

will be referred to as a squared non-central Bessel process of order $q$ with noncentrality function $\eta(\tau) = \mu(t)'\mu(t)/(t(1-t))$ and will be denoted by $Q^2_{q,\eta(t)}$. Of course, for any fixed $t \in (0,1)$, $Q^2_{q,\eta(t)} \sim \chi^2_{q,\eta(t)}$, a non-central $\chi^2_q$ random variable with $q$ degrees of freedom and non-centrality parameter $\eta(t)$. We adopt the following standard notation for partitioning matrices like $D$ defined in Condition A.2.ii above: $D_{ij}$ $i,j = 1,2$ will denote the $ij^{\text{th}}$ block of $D$, and $D^{ij}$ will denote the $ij$ block of $D^{-1}$. To illustrate, recall $D^{22} = (D_{22} - D_{21}D_{11}^{-1}D_{12})^{-1}$. Finally, the symbol $\Rightarrow$ will denote weak convergence, $\rightsquigarrow$ convergence in distribution, and $\rightarrow$, convergence in probability. We can now state our first result, which is proven in the Appendix.

**Theorem 2.1.** *Let $\mathcal{T} = [\epsilon, 1-\epsilon]$, for some $\epsilon \in (0, \frac{1}{2})$. Under conditions A.1-3,*

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

*where $\eta(\tau) = \zeta(\tau)'(D^{22})^{-1}\zeta(\tau)/\omega^2(\tau)$, and $w(t) = \sqrt{\tau(1-\tau)}s(\tau)$. And, under the null hypothesis* (2.3)

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

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

**Corollary 2.1.** *Under conditions A1-3,* $\Lambda_n(\tau) = L_n(\tau) + o_p(1)$, *uniformly on* $\mathcal{T}$.

The foregoing results enable the investigator to test a broad range of hypotheses regarding the joint effects of covariates while permitting one to restrict 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 whole conditional density as in the conventional location shift version of the classical linear regression model.

2.3. **Wald Processes.** One disadvantage of the quantile $\rho$-tests proposed above is the required estimation of the nuisance sparsity function, $s(\tau)$. A potentially more serious disadvantage is that they require that the null model 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 subsection and the one to follow, we will relax the iid error assumption employed above and consider the location-scale shift model employed by Gutenbrunner and Jurečková (1992),

$$y_i = x_i\beta + \sigma_i u_i \tag{2.5}$$

where $\sigma_i = x_i'\gamma$ and the $\{u_i\}$ are assumed to be iid from distribution function $F$. We will denote $\Gamma_n = diag(\sigma_i)$ and introduce the following additional assumption.

**A. 4.** *Suppose model* (2.5) *holds with conditions A1-3, that the elements* $\sigma_i$ *are bounded away from 0 and infinity, and that* $G_n = n^{-1}X_n'\Gamma_n^{-1}X_n$, *tends to a positive definite matrix.*

The conditional quantile functions of the response, $y_i$, under (2.5), may be written as,

$$Q_{y_i}(\tau|x_i) = x_i'\beta(\tau),$$

where $\beta(\tau) = \beta + \gamma F^{-1}(\tau)$, and which we may again partition as, $\beta(\tau) = (\beta_1(\tau)', \beta_2(\tau)')'$. The representation (5.25) of GJ yields, uniformly for $\tau \in \mathcal{T}$,

$$\sqrt{n}(\hat\beta(\tau) - \beta(\tau)) = s(\tau)G_n^{-1}g_n(\tau) + o_p(1),$$

where $g_n(\tau) = n^{-1/2}\sum x_i\psi_\tau(u_i - F^{-1}(\tau))$, and $\psi_\tau(u) = \tau - I(u < 0)$. Thus, setting $\Psi = [0\vdots I_q]$ we have,

$$\sqrt{n}(\hat\beta_2(\tau) - \beta_2(\tau)) = s(\tau)\Psi G_n^{-1}g_n(\tau) + o_p(1),$$

and by Theorem 1 of GJ it follows that,

$$\sqrt{n}(\hat{\beta}_2(\tau) - \beta_2(\tau)) \Rightarrow s(\tau)\Omega^{1/2}B_q,$$

where $\Omega = \Psi G_n^{-1} D_n G_n^{-1} \Psi'$ and $B_q$ denotes a $q$-vector of independent Brownian bridges.

**Theorem 2.2.** *Under conditions A.1-4,*

$$W_n(\tau) = n\hat{\beta}_2(\tau)'\Omega^{-1}\hat{\beta}_2(\tau)/\omega^2(\tau) \Rightarrow Q^2_{q,\eta(\tau)}(\tau) \quad \textit{for } \tau \in \mathcal{T}$$

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

$$\sup_{\tau \in \mathcal{T}} W_n(\tau) \rightsquigarrow \sup_{\tau \in \mathcal{T}} Q^2_q(\tau)$$

Note that in the "homoscedastic" case $\Gamma = I$, the sandwich form of the matrix $\Omega$ collapses to $D_n^{22}$ and the Wald process has the same asymptotic behavior as the two $\rho$-processes. More generally, the Wald process enables us to relax the stringent location shift, iid error, assumption underlying the $\rho$-processes. In exchange, it requires us to estimate the matrix, $G_n$, a task that is somewhat more onerous than the scalar sparsity estimate needed for the $\rho$-processes. We will defer the question of estimating $G_n$ 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 Gutenbrunner and Jurečková (1992), 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 (2.5) is given by,

$$\hat{a}_n(\tau) = \text{argmax}\{y'a | X_1'a = (1-\tau)X_1'e,\ a \in [0,1]^n\} \tag{2.6}$$

where $e$ denotes an $n$-vector of ones, and the $n$ by $p$ matrix, $X$, has been partitioned as $[X_1 \vdots X_2]$ to conform with the partitioning of the hypothesis. The problem posed in (2.6) is the formal dual problem corresponding to the (primal) quantile regression linear program under the restriction imposed by $H_0$. Employing Theorem 1 of GJ, Theorem 5.1 of GJKP may be easily extended to conform to the conditions of the location-scale shift model (2.5).

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

$$T_n = S_n' M_n^{-1} S_n / A^2(\varphi)$$

where

$$\begin{aligned} S_n &= n^{-1/2}(X_2 - \hat{X}_2)'\hat{b}_n, \\ \hat{X}_2 &= X_1(X_1'\Gamma_n^{-1}X_1)^{-1}X_1'\Gamma_n^{-1}X_2, \\ M_n &= (X_2 - \hat{X}_2)'(X_2 - \hat{X}_2)/n, \\ \hat{b}_n &= (-\int \varphi(t)d\hat{a}_{in}(t))_{i=1}^n, \end{aligned}$$

and $\varphi$ is a score generating function of bounded variation. It may be emphasized that the only modification of the test statistic relative to the formulation in GJKP is the appearance of the matrix $\Gamma_n$ in the expression for $\hat{X}_2$. 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 (2.3)

**Theorem 2.3.** *Let $\varphi : [0,1] \to \Re$ 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 A1-4,*

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

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

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

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

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

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

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

$$T_n(\tau) = S_n(\tau)' M_n^{-1} S_n(\tau)/(\tau(1-\tau)).$$

which can be used in lieu of $W_n(\tau)$.

**Theorem 2.4.** *Under the conditions A1-4, $T_n(\tau) = W_n(\tau) + o_p(1)$, uniformly on $\mathcal{T}$.*

When $\Gamma_n = I$ a crucial feature of this form of the test which distinguishes it from the corresponding $\rho$-test processes is that, since the rank score process $a_n(\tau)$ is scale invariant, $T_n$ does not require any estimate of the nuisance parameters $s(\tau)$ or $\sigma(\tau)$. In this case, Jurečková (1992) treats the case $q = 1$ and relates 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 $\rho$-test and Wald approaches to testing in quantile regression, as was already stressed in 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 $\Gamma_n$ as in the location-scale model, then 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 upon the reciprocal of a density function evaluated at the quantile of interest – a quantity Tukey (1965) has termed the "sparsity function" and Parzen (1979) calls the quantile-density function. It is perfectly natural that the precision of quantile estimates should depend on this quantity since it reflects the density of observations near the quantile of interest. If the data is very sparse at the quantile of interest it will be difficult to estimate. On the other hand, when the sparsity is low, so observations are very dense, the quantile will be more precisely estimated. Thus, to estimate the precision of the $\tau$th 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 Siddiqui (1960), Bofinger (1975), Sheather and Maritz (1983), and Welsh (1987). Sidddiqui's idea is simplest and has received the most attention in the literature so we will 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, i.e.,

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

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 therefore natural following Siddiqui to estimate $s(t)$ using a simple difference quotient of the empirical quantile function, i.e.,

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

where $\hat{F}^{-1}$ is an estimate of $F^{-1}$ and $h_n$ is a bandwidth which may tend to zero as $n \to \infty$. A bandwidth rule suggested 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 $z_\alpha$ satisfies $\Phi(z_\alpha) = 1 - \alpha/2$ for the construction of $1-\alpha$ confidence intervals. 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 should we compute $\hat{F}^{-1}$? One approach to estimating $F^{-1}$ is to employ the empirical quantile function suggested in Bassett and Koenker (1982). In effect this amounts to using $\hat{F}_y^{-1}(t) = \bar{x}'\hat{\beta}(t)$ where $\hat{\beta}(\cdot)$ is the usual regression quantile process. The functions

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

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

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

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

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

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

Note that in the location-scale model (2.5) even for a fixed bandwidth we can obtain estimates of the $\sigma_i$'s that are consistent up to scale, since

$$\frac{x_i'(\hat{\beta}(\tau+h_n)-\hat{\beta}(\tau-h_n))}{2h_n} = x_i'\gamma\Delta_h = \sigma_i\Delta_h$$

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 $\Delta_h$ cancels in the expression for $\hat{X}_2$, so it plays no role. For the Wald statistic we still require an estimate of the scalar sparsity parameter at each $\tau$, but it is possible to estimate the matrix $\Gamma_n$ in this fashion and then to estimate the scalar sparsity parameter using a weighted version of the approach described above.

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

$$d_i = x_i'(\hat{\beta}(\tau+h_n)-\hat{\beta}(\tau-h_n))$$

is *necessarily* positive only at $x=\bar{x}$. Nevertheless, in practice we find that problems due to "crossing" of the estimated conditional quantile planes occurs only infrequently and then in the extreme regions of the design space. In our implementation of this approach we simply replace $\hat{f}_i$ by its positive part, i.e., 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 zero in the (rare) cases in which $d_i=0$ because the $i$th observation is basic at both $\tau \pm h_n$. Substituting this estimate in the expression for $G_n$ above yields an easily implementable estimator for the asymptotic covariance matrix of $\hat{\beta}(\tau)$ in the non-iid error model.

Alternatively, Powell (1989) has suggested a kernel-based method to estimate the quantile regression covariance matrix sandwich. Noting that in estimating the matrix $G_n(\tau)$ we are really after is a matrix weighted density estimator, he proposes a kernel estimator of the form

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

where $\hat{u}_i(\tau) = y_i - x_i'\hat{\beta}(\tau)$ and $h_n$ is a bandwidth parameter satisfying $h_n \to 0$ and $\sqrt{n}h_n \to \infty$. He shows that under certain uniform Lipschitz continuity requirements on the $f_i$, $\hat{G}_n(\tau) \to G_n(\tau)$ 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 may conclude this Section by remarking that the nuisance parameters, $s(\tau)$, $\sigma(\tau)$, $\Gamma_n$ appearing in the various versions of the process, $L_n, \Lambda_n, W_n, T_n$, may be replaced by the estimates described above 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 employing a range of artificial data. In preliminary experiments the performance of tests based on $L_n$ and $\Lambda_n$ were essentially indistinguishable, so we chose to drop $L_n$ in the subsequent comparisons. We should emphasize that the quantile $\rho$-tests based on $L_n$ and $\Lambda_n$ 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, since in such circumstances the null model contains only an intercept. However, in many empirical contexts, including the one explored in Section 4 below, this assumption will be highly implausible. We have also chosen to defer consideration of the Wald process, $W_n(\tau)$, 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 3.1 and described in detail below.

**Model 1: Pure Gaussian Noise.** The data in the first column of Figure 3.1 was 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 3.1 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 3.1 illustrates a pure heteroskedastic version of the regression model in which,

$$y_i = (x_i + \tfrac{1}{4}x_i^2)u_i$$

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

**Model 4: Non-location-scale.** Finally, the fourth column of Figure 3.1 has a somewhat more peculiar configuration in which for small values of $x$ we observe a unimodal conditional density for $y$, while for larger values of $x$ we have a bimodal conditional density of $y$. Formally, denoting by $Q_y(\tau|x)$ the $\tau$-th 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,

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

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 over $x$. In this example $n = 300$.

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

$$Q_y(\tau|x) = \beta_0(\tau) + \beta_1(\tau)x + \beta_2(\tau)x^2 \tag{3.1}$$

where the parametric functions $(\beta_0(\tau), \beta_1(\tau), \beta_2(\tau))$ are given in Table 3.1. 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 3.1.

| | 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 |

Table 3.1. Quantile Regression Coefficients

For Models 1,2, and 4 we would like to compare the unrestricted linear model,

$$Q_y(\tau|x) = \beta_0(\tau) + \beta_1(\tau)x \tag{3.2}$$

with the null model

$$Q_y(\tau|x) = \beta_0(\tau) \tag{3.3}$$

For model 3 we compare the unrestricted model (3.1) with model (3.3). Since the null model (3.3) contains only an intercept, the appropriate form of the $T_n$ process employes $\Gamma_n = I$ in all four models.

The last three rows of Figure 3.1 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 would be nearly zero 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 $\Lambda_n$ 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 which is also consistent with the appearance of the last two panels of the first column which represent the $\Lambda_n$ and $T_n$ processes, respectively.

In the second column of Figure 3.1 we illustrate 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 $\Lambda_n$ and $T_n$ processes both indicate a highly significant departure from the null theory over the whole range of $\tau$ 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 doesn't improve upon the restricted fit, i.e., that 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 $\Lambda_n$ 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 $q = 2$.

In the last column we show the results for Model 4. Like the scale-shift model this configuration also yields a much more pronounced effect away from the median and this is also apparent in the plots of the $\Lambda_n$ and $T_n$ processes. Note that in this case conventional tests of heteroscedasticity would be unlikely to detect any effect of the covariate $x$, since 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* ones. Thus in the heteroscedastic cases $R^1$ and the tests indicate no effect near the median where the conditional quantile function is horizontal, but in the tails where conditional and unconditional quantiles differ substantially they indicate this by curving upward in the tails. Similarly, in the the bimodal case, the conditional quantile functions at the median and in the far upper tail are essentially constant, while 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 which offers a somewhat more systematic evaluation of the 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 above. The data consists of a pooled sample of 161 observations, on national growth rates for the two periods 1965-75 and 1975-85, taken from the Barro and Lee (1994). The data is similar to that used in Barro and Sala-i-Martin (1995), but unfortunately not identical. There are initially 87 observations from 1965-75 and 97 from 1975-85, but only 161 after removing observations with missing data for the model estimated. This model is the same as the basic model reported in 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 GDP on the growth of this variable. Figure 4.1 plots this effect as estimated in a quantile regression which also includes the variables indicated in Figure 4.2. 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 $\tau$. The dotted

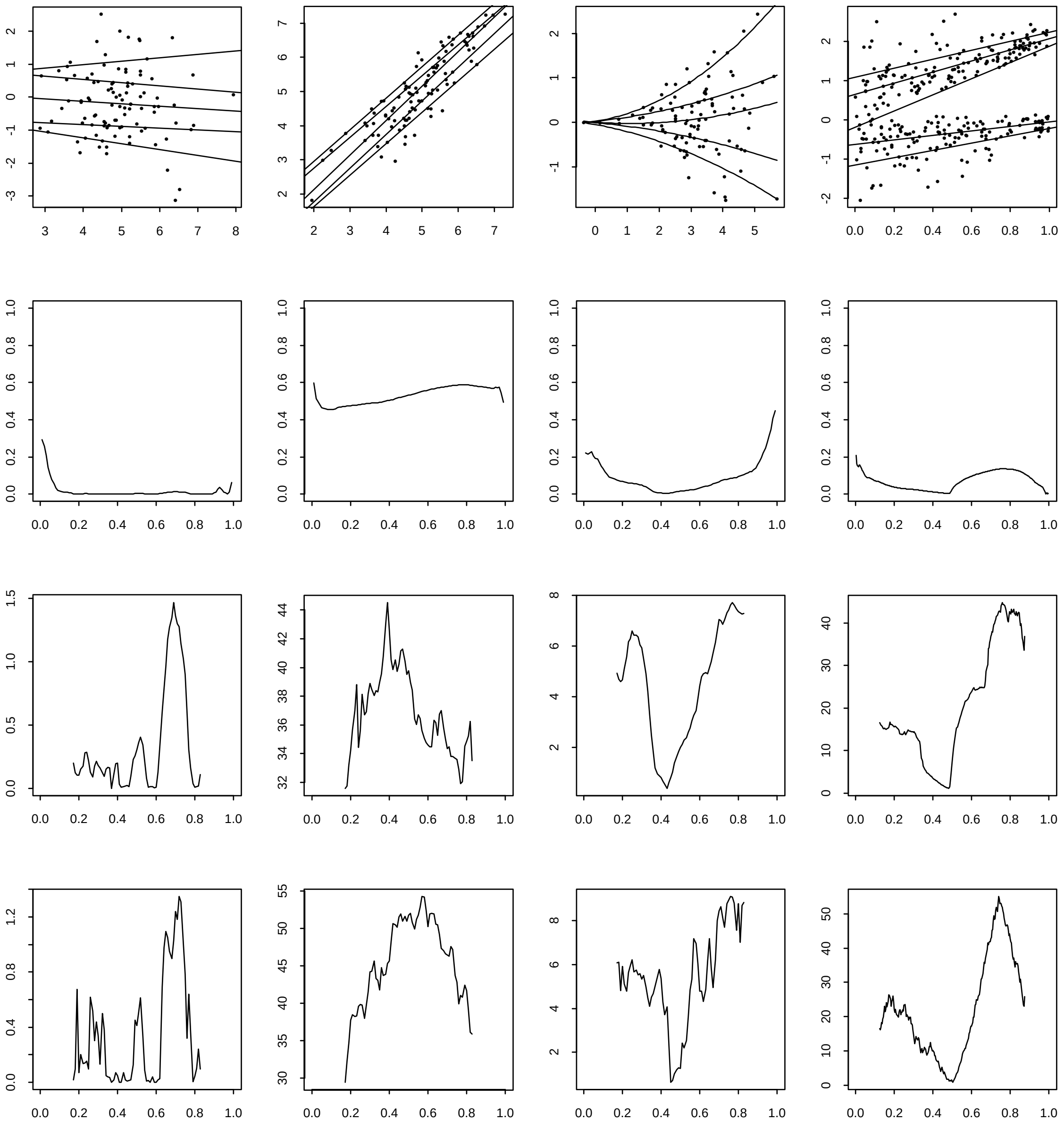

Figure 3.1. Some Illustrative Examples of Goodness of Fit Processes: The figure illustrates four different data configurations and their associated $R^1$, $\Lambda_n$ and $T_n$ functions. In the first row of the panel of figures the data is depicted, the solid lines are the fitted conditional quantile functions for $\tau \in \{.1, .25, .5, .75, .9\}$. The next row represents the $R^1$ process, and the third and fourth rows depict the $\Lambda_n$ and $T_n$ processes respectively.

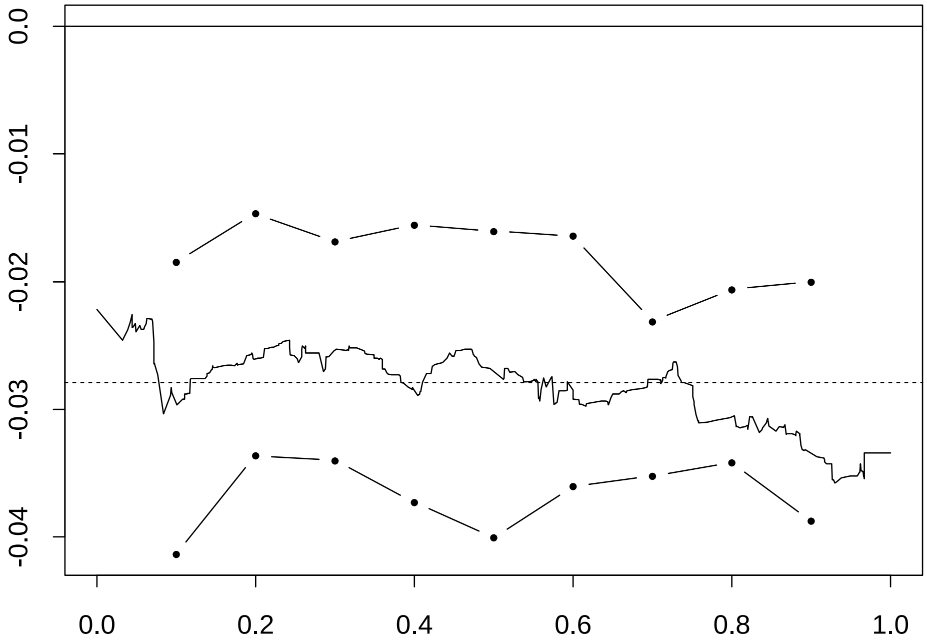


FIGURE 4.1. Estimated Quantile Regression Effect of Initial Per Capita GDP on GDP Growth: The figure illustrates the estimated quantile regression process for initial per capita GDP in a model for GDP growth which also includes the variables indicated in Figure 4.2. The rather jagged, solid line is the 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 zero at the top of the Figure.

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 in Koenker (1994) and Koenker (1997), but modified as in Theorem 2.3 above to accomodate the effect of heterogeneity through the estimation of the matrix $\Gamma_n$ 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 $\tau$. Clearly, the figure 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 comment that given the ten 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 ten 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 since 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) considers a rather amusing, but similar, empirical example from the early industrial orgranization literature. Friedman (1992) and Quah (1993) have made similar observations regarding the use of the term convergence in the empirical growth literature. Note, however, that if the initial GDP effect in Figure 4.1 could be interpreted as suggesting a downwardly 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 therefore 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 percent, or one standard deviation, in G/Y, reduces the expected growth rate by .7 percent per year. The quantile regression results indicate that this effect is essentially constant over the upper half of the distribution, but in the lower tail the effect may be considerably larger. Improvement of the terms of trade appears to exert a monotonically increasing effect in $\tau$, suggesting that this variable tends to help faster growing countries proportionally more than those 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 those of the lower tail. None of the education effects are particularly clear. There is some indication that the effect of male secondary education is increasing in $\tau$, 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 4.3, where we plot the $R^1(\tau)$ function and the corresponding $\Lambda_n(\tau)$ and

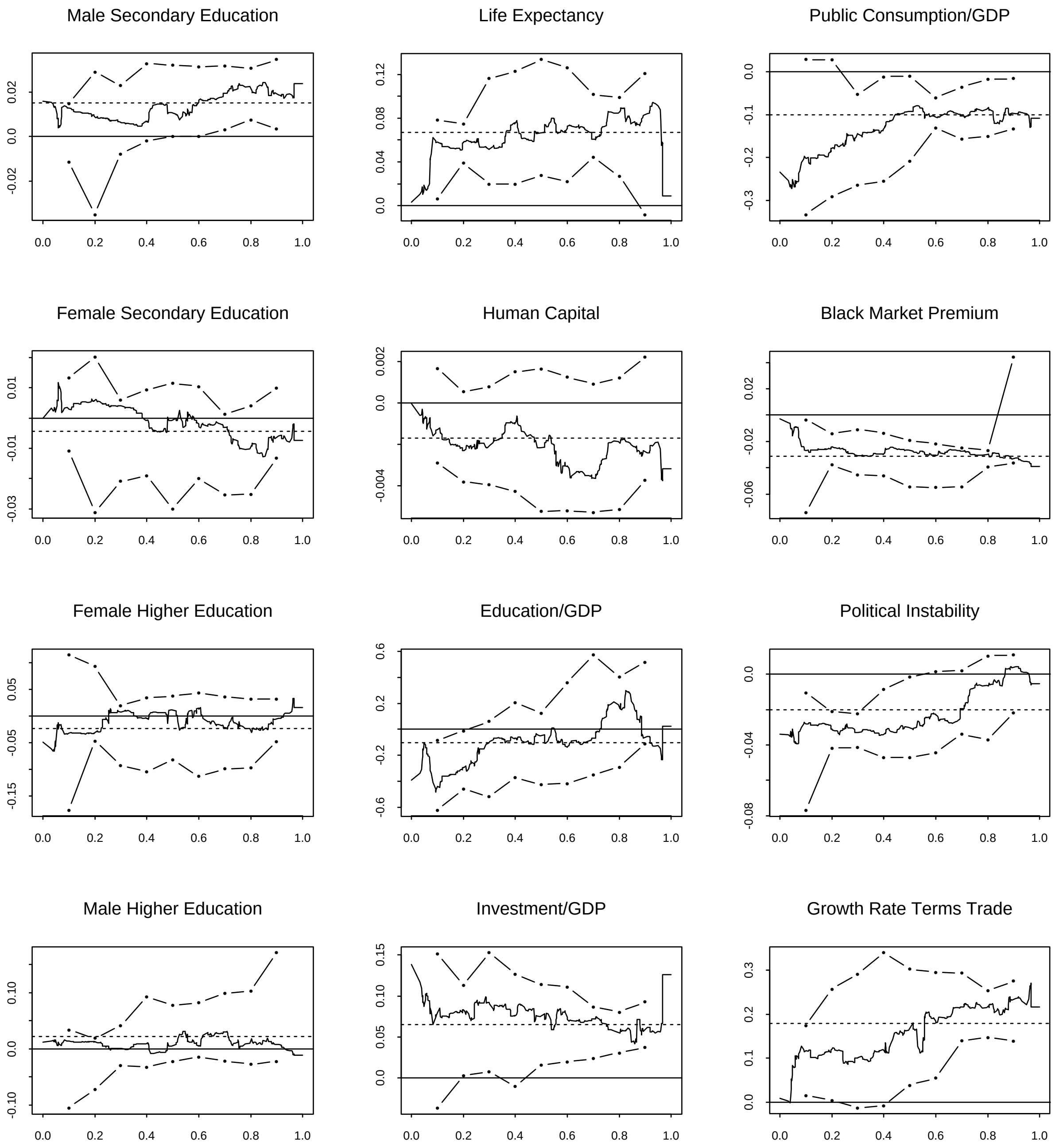


FIGURE 4.2. Estimated Quantile Regression Effects of other Variables on GDP Growth: The figure illustrates the estimated quantile regression process for 12 other variables in a model for GDP growth. The format of the plot follows that of the previous Figure.

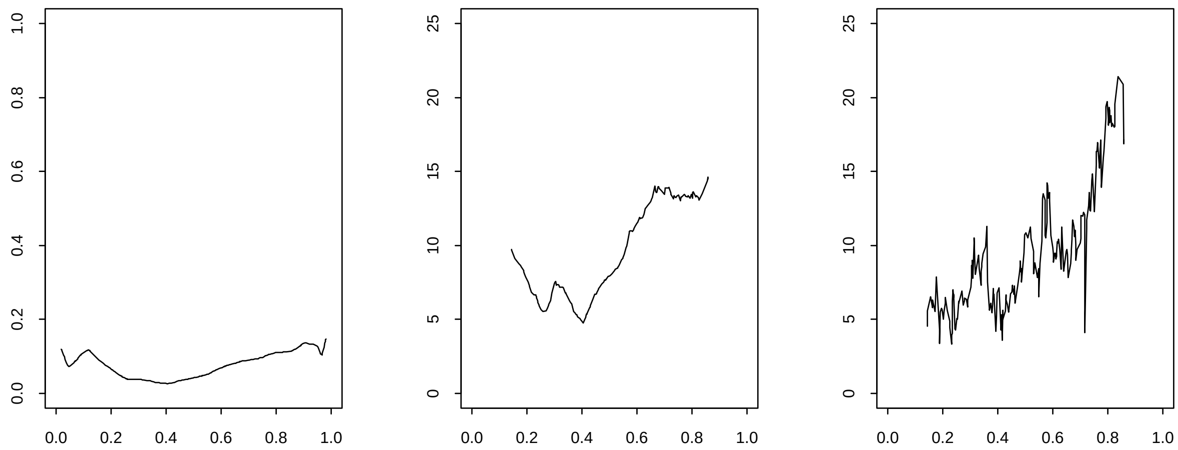


Figure 4.3. Joint Effect of Education Variables on GDP Growth: The the first panel of the figure illustrates the $R^1(\tau)$ function corresponding to the null hypothesis of no effect of any of the 5 education variables in the previous model. Note that the somewhat bowl shaped $R^1$ plot suggests that 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 $\Lambda_n$ process plot. However, the corresponding rank-test, $T_n$, in the third panel of the Figure appears to indicate a monotonically increasing effect of the education variables.

$T_n(\tau)$ processes corresponding to the null hypothesis of no joint effect of the education variables. Here, since the null hypothesis imposes no presumption of homogeneity on the elements of $\Gamma_n$ 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 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 $\Lambda_n$ 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. Since iid error conditions are quite clearly implausible here the validity of the $\Lambda_n$ test is highly questionable. We try to shed some further 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 $\Lambda_n$ 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 since the alternative is rather general, we would expect to be able to provide more powerful tests for each of explicit alternatives.

In Table 5.1 we reproduce a small extract from Andrews (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 $\chi^2$ 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,

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + u_i \tag{5.1}$$

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 $\chi_4^2$ distributions. For the size simulations reported in Table 5.2 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 5.2. It is clear that both tests are under-sized, rejecting at less than the specified nominal rates. The $T_n$ test performs somewhat better in this respect than does the $\Lambda_n$ test, but it is clear that both tests would benefit from some size correction.

| $q$ | $\alpha$ =0.10 | $\alpha$ =0.05 | $\alpha$ =0.01 |
|---|---|---|---|
| 1 | 8.19 | 9.84 | 13.01 |
| 2 | 11.20 | 12.93 | 16.44 |
| 4 | 15.62 | 17.56 | 21.54 |

TABLE 5.1. Critical Values for $\sup \Lambda_n$ and $\sup T_n$ Tests: The entries constitute a small extract from a much larger table of Andrews (1993). Each entry is $c_\alpha$ such that $\text{Prob}(\sup_{\{\tau \in [.05,.95]\}} Q_q^2 < c_\alpha) = \alpha$. The "degrees of freedom" parameter $q$ is the number of linear restrictions imposed by the null hypothesis. These critical values are employed below in the Monte-Carlo evaluation of the size and power performance of the tests.

| Test | | | Gaussian | | | Cauchy | | | $\chi^2_4$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | n | q | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $\Lambda_n$ | 100 | 1 | 0.000 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.002 | 0.002 | 0.000 |
| $T_n$ | 100 | 1 | 0.042 | 0.018 | 0.002 | 0.044 | 0.022 | 0.006 | 0.044 | 0.008 | 0.006 |
| $\Lambda_n$ | 100 | 2 | 0.004 | 0.002 | 0.000 | 0.002 | 0.002 | 0.000 | 0.004 | 0.002 | 0.000 |
| $T_n$ | 100 | 2 | 0.042 | 0.022 | 0.002 | 0.040 | 0.024 | 0.008 | 0.038 | 0.018 | 0.002 |
| $\Lambda_n$ | 100 | 4 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $T_n$ | 100 | 4 | 0.038 | 0.018 | 0.002 | 0.038 | 0.022 | 0.002 | 0.038 | 0.018 | 0.002 |
| $\Lambda_n$ | 500 | 1 | 0.022 | 0.010 | 0.002 | 0.006 | 0.004 | 0.000 | 0.028 | 0.010 | 0.000 |
| $T_n$ | 500 | 1 | 0.058 | 0.032 | 0.012 | 0.056 | 0.022 | 0.006 | 0.078 | 0.036 | 0.006 |
| $\Lambda_n$ | 500 | 2 | 0.024 | 0.010 | 0.002 | 0.004 | 0.000 | 0.000 | 0.012 | 0.004 | 0.000 |
| $T_n$ | 500 | 2 | 0.068 | 0.040 | 0.012 | 0.058 | 0.020 | 0.004 | 0.056 | 0.028 | 0.006 |
| $\Lambda_n$ | 500 | 4 | 0.012 | 0.006 | 0.002 | 0.004 | 0.000 | 0.000 | 0.012 | 0.006 | 0.000 |
| $T_n$ | 500 | 4 | 0.066 | 0.042 | 0.010 | 0.066 | 0.034 | 0.006 | 0.054 | 0.026 | 0.004 |

Table 5.2. Monte Carlo Size of the $\sup \Lambda_n$ and $\sup T_n$ Tests: 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 5.1 taken from Andrews (1993). The model in each case is (5.1) with $x_{1i} \equiv 1$ and $\{x_{ij} : \; j = 2, ..., p\}$ 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 $q = p - 1$.

In Table 5.3 we consider the power of the sup $\Lambda_n$ and sup $T_n$ tests in a setting which differs from the size simulations in only one aspect: the parameter $\beta_2$ is now set to 1/2, not 0. To provide some comparison with conventional hypothesis testing methods in this iid error linear model setting we include entries in the Table reporting the rejection frequencies of the classical F-test. We expect to find better performance from the F-test in the Gaussian and $\chi^2$ cases, but in the Cauchy case the F-test is poor, while 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 5.4 two variants of the heteroscedastic and bimodal examples of Section 3. For the heteroscedastic model (model 3 of Section 3) we adopt the quadratic specification used in Figures 3.1-3

$$y_i = \alpha + (\beta x_{2i} + \gamma x_{2i}^2) u_i \tag{5.2}$$

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

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

| **Test** | | | **Gaussian** | | | **Cauchy** | | | $\chi^2_4$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | n | q | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $\Lambda_n$ | 100 | 1 | 0.916 | 0.820 | 0.546 | 0.472 | 0.354 | 0.126 | 0.336 | 0.202 | 0.054 |
| $T_n$ | 100 | 1 | 0.988 | 0.960 | 0.850 | 0.750 | 0.616 | 0.384 | 0.536 | 0.418 | 0.206 |
| F | 100 | 1 | 1.000 | 1.000 | 1.000 | 0.354 | 0.254 | 0.130 | 0.964 | 0.940 | 0.842 |
| $\Lambda_n$ | 100 | 2 | 0.788 | 0.640 | 0.314 | 0.312 | 0.194 | 0.054 | 0.180 | 0.084 | 0.020 |
| $T_n$ | 100 | 2 | 0.952 | 0.906 | 0.708 | 0.604 | 0.468 | 0.282 | 0.388 | 0.270 | 0.100 |
| F | 100 | 2 | 1.000 | 1.000 | 1.000 | 0.528 | 0.446 | 0.310 | 0.982 | 0.974 | 0.938 |
| $\Lambda_n$ | 100 | 4 | 0.514 | 0.348 | 0.116 | 0.120 | 0.072 | 0.010 | 0.078 | 0.036 | 0.008 |
| $T_n$ | 100 | 4 | 0.888 | 0.776 | 0.514 | 0.474 | 0.334 | 0.136 | 0.288 | 0.182 | 0.064 |
| F | 100 | 4 | 1.000 | 1.000 | 1.000 | 0.840 | 0.796 | 0.646 | 0.996 | 0.992 | 0.982 |
| $\Lambda_n$ | 500 | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.994 |
| $T_n$ | 500 | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.996 |
| F | 500 | 1 | 1.000 | 1.000 | 1.000 | 0.318 | 0.226 | 0.118 | 1.000 | 1.000 | 1.000 |
| $\Lambda_n$ | 500 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.998 | 0.998 | 0.984 |
| $T_n$ | 500 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.998 | 0.996 | 0.984 |
| F | 500 | 2 | 1.000 | 1.000 | 1.000 | 0.514 | 0.430 | 0.278 | 1.000 | 1.000 | 1.000 |
| $\Lambda_n$ | 500 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.988 | 0.992 | 0.986 | 0.926 |
| $T_n$ | 500 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 1.000 | 0.994 | 0.972 |
| F | 500 | 4 | 1.000 | 1.000 | 1.000 | 0.812 | 0.758 | 0.652 | 1.000 | 1.000 | 1.000 |

TABLE 5.3. Monte Carlo Power of the sup $\Lambda_n$ and sup $T_n$ Tests in linear models with iid errors: 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 5.1 taken from Andrews (1993). The model in each case is (5.1) with $x_{1i} \equiv 1$ and $\{x_{ij} : \ j = 2, ..., p\}$ iid standard Gaussian, $\beta = (0, .5, 0, ..., 0) \in R^p$, and $\{u_i\}$ iid from one the three distributions indicated at the top of the table. In each case $\mathcal{T} = [.05, .95]$, and $q = p - 1$.

so the classical regression F-test has asymptotic power equal size, but the quantile regression tests have power from both tails of the distribution since the conditional quantiles process is $Q_y(\tau|x) = (x + 1/4x^2)\Phi^{-1}(\tau)$.

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

$$y_i = (1 - x_{2i})\Phi^{-1}(u_i) + x_{2i}H^{-1}(u_i) \tag{5.3}$$

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

| Test | | | Bimodal | | | Heteroscedastic | | |
|---|---|---|---|---|---|---|---|---|
| | n | q | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $\Lambda_n$ | 100 | 1 | 0.234 | 0.142 | 0.030 | 0.226 | 0.098 | 0.006 |
| $T_n$ | 100 | 1 | 0.314 | 0.174 | 0.054 | 0.836 | 0.722 | 0.472 |
| F | 100 | 1 | 0.116 | 0.066 | 0.026 | 0.196 | 0.122 | 0.050 |
| $\Lambda_n$ | 100 | 2 | 0.104 | 0.050 | 0.006 | 0.070 | 0.014 | 0.000 |
| $T_n$ | 100 | 2 | 0.192 | 0.114 | 0.036 | 0.706 | 0.582 | 0.332 |
| F | 100 | 2 | 0.298 | 0.222 | 0.106 | 0.346 | 0.282 | 0.140 |
| $\Lambda_n$ | 100 | 4 | 0.024 | 0.008 | 0.000 | 0.002 | 0.000 | 0.000 |
| $T_n$ | 100 | 4 | 0.130 | 0.072 | 0.008 | 0.534 | 0.388 | 0.154 |
| F | 100 | 4 | 0.688 | 0.608 | 0.466 | 0.798 | 0.722 | 0.540 |
| $\Lambda_n$ | 500 | 1 | 1.000 | 1.000 | 0.994 | 1.000 | 1.000 | 1.000 |
| $T_n$ | 500 | 1 | 1.000 | 1.000 | 0.992 | 1.000 | 1.000 | 1.000 |
| F | 500 | 1 | 0.096 | 0.050 | 0.010 | 0.164 | 0.100 | 0.024 |
| $\Lambda_n$ | 500 | 2 | 1.000 | 0.994 | 0.958 | 1.000 | 1.000 | 1.000 |
| $T_n$ | 500 | 2 | 1.000 | 0.998 | 0.962 | 1.000 | 1.000 | 1.000 |
| F | 500 | 2 | 0.314 | 0.216 | 0.102 | 0.404 | 0.318 | 0.148 |
| $\Lambda_n$ | 500 | 4 | 0.994 | 0.962 | 0.812 | 1.000 | 1.000 | 1.000 |
| $T_n$ | 500 | 4 | 0.996 | 0.978 | 0.900 | 1.000 | 1.000 | 1.000 |
| F | 500 | 4 | 0.738 | 0.650 | 0.480 | 0.760 | 0.706 | 0.540 |

Table 5.4. Monte Carlo Power of the sup $\Lambda_n$ and sup $T_n$ Tests in a bimodal transition model and a purely heteroscedastic model: 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 5.1. The model in the heteroscedastic cases is (5.2) bimodal case is (5.3). In each case $\mathcal{T} = [.05, .95]$, and $q = p - 1$.

with df,

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

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, since there are linear conditional quantile functions with non-zero slope for all $\tau \neq 1/2$ (notice that, $Q_y(1/2|x) = 0$). Several features of Table 5.4 merit attention. The new tests are a clear success at $n = 500$ where they have power near one 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 since $E(y|x) = 0$ 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 $\sup \Lambda_n$. This is presumably largely a consequence of the size distortion which we have already observed in Table 5.2. Since both $T_n$ and $\Lambda_n$ tests are significantly undersized, effective size correction would presumably yield improvement in power in both Table 5.3 and 5.4. This is a topic which deserves further investigation.

## 6. Conclusion

We have suggested a goodness of fit criterion for quantile regression which generalizes the proposal of McKean and Sievers (1987) for $\ell_1$ regression and adapts it to an arbitrary linear hypothesis. Three related goodness of fit processes, $\Lambda_n(\tau)$, $W_n(\tau)$, and $T_n(\tau)$, are introduced which significantly expand the available toolkit for inference in quantile regression models. For a given $\tau$, the proposed test may be regarded as variants of the likelihood ratio, Wald, and and Lagrange multiplier tests for $\ell_1$ regression. The tests have an asymptotic theory involving Bessel processes which is closely related to earlier work on inference for goodness of fit as well as to more recent work on inference in change-point 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 Portnoy (1992). It would also be useful to investigate extensions to null hypotheses which involve unknown nuisance parameters along the lines of Durbin (1973) and others.

## Appendix

**Proof:** (Theorem 2.1) Let

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

where $\omega(\tau) = \lambda(\tau)s(\tau)$, $\lambda^2(\tau) = \tau(1-\tau)$ and $u_i(\tau) = u_i - F^{-1}(\tau)$. Since, in model (1.1), by A.2 (i), $\beta(\tau) = \beta + F^{-1}(\tau)e$, we may interpret $\delta$, as $\delta = \sqrt{n}(b - \beta)/\omega(\tau)$, for some choice of $b$. Lemma 3.1 of Gutenbrunner, Jurečková, Koenker, and Portnoy (1993) may be slightly recast to provide the following uniform asymptotically quadratic approximation for $V_n(\delta, t)$: under Conditions A.1-3, for any fixed $C > 0$ and $\alpha \in (0, \frac{1}{2})$,

$$\sup\{|W_n(\delta, \tau)| : \|\delta\| \leq C\sqrt{\log\log n}, \ \ \tau \in \mathcal{T}\} \to 0,$$

and

$$W_n(\delta, \tau) = (\lambda^2(\tau)s(\tau))^{-1}[V_n(\delta, \tau) - V_n(0, \tau)] - [\tfrac{1}{2}\delta' D\delta - (\lambda(t)^{-1}\delta' g_n)$$

where $g_n = n^{-1/2}\sum x_i \psi_\tau(u_i(\tau))$, and $\psi_\tau(u) = \tau - I(u < 0)$.

Theorem 1 of Gutenbrunner and Jurečková (1992) establishes that

$$\hat{\delta}_n(\tau) = \sqrt{n}(\hat{\beta}_n(\tau) - \beta(\tau))/\omega(\tau) = \lambda(\tau)^{-1} D^{-1} g_n + o_p(1)$$

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

$$\hat{\delta}_n \Rightarrow \lambda^{-1} D^{-1/2} B_p$$

Similarly, for the restricted estimator, we have

$$\tilde{\delta}_n(\tau) = \begin{pmatrix} D_{11}^{-1} D_{12} \zeta(\tau)/\omega(\tau) + D_{11}^{-1} g_{1n}/\lambda(\tau) \\ -\zeta(\tau)/\omega(\tau) \end{pmatrix}$$

where we partition, $g_n = (g_{1n}', g_{2n}')'$ 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),

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

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

Finally, note that,

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

so we may write

$$L_n = ||\mu + B_q||^2/\lambda^2 + o_p(1)$$

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

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

■

**Proof:** (Theorem 2.2) The result follows from the argument given in the text. ■

**Proof:** (Theorem 2.3) Let $\tilde{X}_2 = X_2 - \hat{X}_2$, so $\tilde{x}_{2i} = x_{2i} - \hat{x}_{2i}$, and denote,

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

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

$$W_{n1}(\tau) = n^{-1/2} \sum x_{1i}(\tilde{a}_i(\tau) - (1-\tau)),$$

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

$$\begin{aligned} \hat{W}_{n2}(\tau) &= W_{n2}(\tau) - \tilde{X}_2'\Gamma^{-1} X_1 (X_1'\Gamma^{-1} X_1)^{-1} W_{n1}(\tau) + o_p(1) \\ &= W_{n2}(\tau) + o_p(1). \end{aligned}$$

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

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

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

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

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

Using conventional matrix partitioning notation,

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

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

**Proof:** (Theorem 2.4) Note that with the quantile score function $\varphi_\tau(t) = \tau - I(t < \tau)$

$$S_n(\tau) \equiv \hat{W}_{n2}(\tau) = W_{n2}(\tau) + o_p(1)$$

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

$$W_{n2}(\tau) \Rightarrow M^{1/2} B_q(t),$$

this 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

$$\begin{aligned} S_n(\tau) &= W_{n2}(\tau) - (1/s(\tau))[(1/n) \sum \tilde{x}_{2i} x_i'/\sigma_i] \Delta_n(\tau) + o_p(1) \\ &= W_{n2}(\tau) + (1/s(\tau))(G^{22})^{-1} \zeta(\tau) + o_p(1) \end{aligned}$$

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

## REFERENCES


ABBERGER, K. (1996): "Non-parametric Quantile Forecasting and its Application to Stock Returns," Discussion Paper SFB 178, U. 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 J.-W. LEE (1994): "Data Set for a Panel of 138 Countries," Discussion paper, NBER, `http://www.nber.org/pub/barro.lee`.

BARRO, R. J., AND X. SALA-I-MARTIN (1995): *Economic Growth*. McGraw-Hill, New York.

BASSETT, GILBERT, J., AND R. KOENKER (1982): "An Empirical Quantile Function for Linear Models With Iid Errors," *J. of Am. Stat. Assoc.*, 77, 407–415.

BOFINGER, E. (1975): "Estimation of a density function using order statistics," *Australian J. of Statistics*, 17, 1–7.

CHAMBERLAIN, G. (1994): "Quantile regression censoring and the structure of wages," in *Advances in Econometrics*, ed. by C. Sims, New York. Elsevier.

DeLong, D. M. (1981): "Crossing Probabilities for a Square Root Boundary By a Bessel Process (Corr: V12 P1699)," *Communications in Stat. (A)*, 10, 2197–2213.

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

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

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

Friedman, M. (1992): "Do Old Fallacies ever Die," *J. of Economic Lit.*, pp. 2129–2132.

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

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

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

Hall, P., and S. Sheather (1988): "On the distribution of a studentized quantile," *J. of Royal Stat Society. (B)*, 50, 381–391.

Hendricks, W., and R. Koenker (1992): "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *J. of Am. Stat. Assoc.*, 87, 58–68.

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

Jurečková , J. (1992): "Tests of Kolmogorov-Smirnov type based on regression rank scores," *Trans. 11th Prague Conf. on Inform. Theory, Random Proc.and Statistical Decision Functions*, B, 41–49.

Kiefer, J. (1959): "K-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises Tests," *Annals of Math. Stat.*, 30, 420–447.

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

——— (1997): "Rank Tests for Linear Models," in *Handbook of Statistics*, ed. by G. Maddala, and C. Rao, vol. 15. North Holland.

Koenker, R., and G. Bassett (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.* Holden-Day, San Francisco.

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

Parzen, E. (1979): "Nonparametric statistical data modeling," *J. of the American Statistical Association*, 74, 105–121.

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

Powell, J. (1989): "Estimation of monotonic regression models under quantile restrictions," in *Nonparametric and Semiparametric Methods in Econometrics*, ed. by B. W.A., J. Powell, and G. Tauchen, Cambridge. Cambridge U. Press.

Quah, D. (1993): "Galton's Fallacy and Tests of the Convergence Hypothesis," *Scand. J. of Econ.*, 95, 427–43.

Ronchetti, E. (1982): "Robust Alternatives to the $F$-test for the Linear Model," in *Probability and Statistical Inference*, ed. by W. G. P. Grossmann, and W. Wertz, pp. 329– 342. Reidel.

ROUSSEEUW, P., AND M. HUBERT (1997): "Recent developments in PROGRESS," in $L_1$-*Statistical Procedures and Related Topics*, ed. by Y. Dodge, vol. 31 of *IMS Lecture Notes - Monograph Series*. IMS.

SHEATHER, S., AND J. MARITZ (1983): "An estimate of the asymptotic standard error of the sample median," *Australian J. 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," *Stat. 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. by Y. Dodge, pp. 369–377, New York. Elsevier.