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

Assessment of goodness-of-fit in quantile regression models has proven to be a persistent concern of applied researchers. In this paper we try to answer the question, "Is there an R^2 for quantile regression?" And, in the process of formulating a response, we consider several related variants of the goodness of fit process which provide useful, new hypothesis testing strategies for quantile regression

We begin, however, by trying to motivate our goodness-of-fit criterion by analogy with the familiar R^2 of classical least squares regression. Consider the linear model,

$$y_i = x_i'\beta + u_i \tag{1.1}$$

with $\{u_i\}$ iid from the distribution function F, and let $\hat{\beta}$ and $\tilde{\beta}$ denote the least squares estimators of this model, and the model under the rank q linear restriction that,

$$H_0: \quad R\beta = r \tag{1.2}$$

respectively. 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 (1.1) relative to the restricted model constrained by the hypothesis (1.2) as

$$R^2 = 1 - \hat{S}/\tilde{S}. \tag{1.3}$$

Conventionally, restricting (1.1) to include only an "intercept" parameter yields the R^2 one usually sees reported.

Version: June 25, 1998. 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 helpful comments and Carlos Arango for research assistance on the empirical growth example.

In quantile regression, we may motivate our proposed goodness of fit criterion in a rather similar manner. Let $\rho_\tau(u) = u(\tau - I(u < 0))$ denote the usual "check" function as in Koenker and Bassett (1978), and define $\hat\beta(\tau)$ as the minimizer,

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

and $\tilde\beta(\tau)$ as the minimizer for the corresponding constrained problem, with

$$\tilde V(\tau) = \min_{\{b \in \Re^p | Rb = r\}} \sum \rho_\tau(y_i - x_i' b).$$

We may now define the goodness-of-fit criterion

$$R^1(\tau) = 1 - \hat V(\tau)/\tilde V(\tau) \tag{1.4}$$

which is the natural analogue of R^2 defined above. Several points should be immediately stressed:

- Like R^2, since $\tilde\beta(\tau)$ is obtained by constraining $\hat\beta(\tau)$, it is immediately apparent that $\hat V(\tau) \le \tilde V(\tau)$ and therefore $R^1(\tau)$ lies between 0 and 1.
- Unlike R^2, which measures the relative adequacy of two models for the conditional mean function using estimates of conditional variance, $R^1(\tau)$ measures the relative success of the corresponding quantile regression models at a specific quantile. Thus, $R^1(\tau)$ constitutes a local measure of goodness-of-fit at the τ^{th} quantile rather than a global measure of goodness-of-fit like R^2. As a consequence, we obtain a goodness-of-fit *process* defined from $[0,1]$ to $[0,1]$. As we shall see, it is possible for the relaxation of the restriction to achieve large benefits in certain regions of τ in $[0,1]$ while in other regions the benefit might be negligible; $R^1(\tau)$ will reflect these differences.
- For R^2, we have a familiar connection to the classical F statistic designed to test the hypothesis $R\beta = r$. Indeed, it is easy to show that nR^2 is asymptotically χ^2_q, and thus provides a convenient approach to constructing a wide range of useful diagnostics for least squares regression. We shall see that there is an analogous connection of $R^1(\tau)$ to the large sample theory of hypothesis testing for quantile regression. In particular, we will suggest several related inference processes for quantile regression which bear a close relationship to the $R^1(\tau)$ process.
- In the median case, $\tau = 1/2$, we may interpret the goodness-of-fit criterion, R^1, as the proportion of variability – measured by the mean absolute deviation from the median – explained by available covariates. While this measure of variability is obviously different from the more familiar variance measure used by the conventional R^2, it is the natural measure associated with quantile regression. It has the added virtue of being somewhat less sensitive to outlying observations than the variance, although it is not immune from criticism on robustness grounds.

- Since $R^1(\tau)$ is tied to a particular quantile rather than playing the role of a global measure of performance for a particular parametric model, the full R^1 *process* serves as a useful diagnostic tool for a variety of regression maladies. Some interpretative guidance along these lines is provided in Section 3 below.

The use of the optimized value of the objective function defining a regression estimator as in (1.4) as a measure of goodness-of-fit is sufficiently natural that it is 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 ℓ_1 forecasting respectively. An earlier reference is McKean and Sievers (1987) who suggest $R^1(1/2)$ for the ℓ_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 (1.1)is better at the τ-quantile than the restricted model constrained by (1.2), $\hat{V}_n(\tau)$ should be significantly smaller than $\tilde{V}_n(\tau)$. 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 (1.4) 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 three such variants below. The choice of these statistics is motivated by analogies to likelihood ratio tests in the Laplace 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.

2. Inference Processes for Quantile Regression

2.1. **Motivation in Median Regression.** In Koenker and Bassett (1982) it was shown that for median regression, under the null hypothesis (1.2)

$$8(\tilde{V}(1/2) - \hat{V}(1/2))/s(1/2) \rightsquigarrow \chi^2_q \tag{2.1}$$

where $s(\tau) = (f(F^{-1}(\tau)))^{-1}$ denotes the sparsity function of F, the reciprocal of the density function. That this statistic is related to a likelihood ratio statistic is easy to see. Consider the standard Laplace (double exponential) density,

$$f(u) = \tfrac{1}{4}\exp(-\tfrac{1}{2}|u|).$$

Under the assumption that $\{u_i\}$ in (1.1) comes from this density, we have the log-likelihood,

$$\ell(\beta) = -n \log(4) - \tfrac{1}{2} \sum_{i=1}^{n} |y_i - x_i \beta|$$

and the likelihood ratio statistic for testing H_0 would be

$$2(\ell(\hat{\beta}) - \ell(\tilde{\beta})) = 2(\tilde{V}(1/2) - \hat{V}(1/2)),$$

which we would expect to have a χ^2_q asymptotic distribution under H_0. How does this relate to the result (2.1)? Where does the factor $s(1/2)$ come from? Note that in the standard Laplace case $s(1/2) = (f(0))^{-1} = 4$, so the usual theory of the likelihood ratio is vindicated – for this very special case, twice the log likelihood ratio converges in distribution to χ^2_q under H_0. However, when the standard Laplace assumption fails to hold, $s(1/2)$ may be quite different from 4 and the likelihood ratio statistic would need to be modified accordingly.

The simplest example of this, an example which yields another variant of the likelihood ratio, is the Laplace density with free scale parameter, σ. Now,

$$f(u) = \tfrac{1}{4\sigma} \exp\{-\tfrac{1}{2\sigma}|u|\}.$$

so the log-likelihood is

$$\ell(\beta, \sigma) = -n \log(4\sigma) - \tfrac{1}{2\sigma} \sum |y_i - x_i \beta|.$$

It is easily seen that the maximum likelihood estimator of σ is

$$\hat{\sigma} = n^{-1} \hat{V}(\tfrac{1}{2}) = \tfrac{1}{2n} \sum |y_i - x_i' \hat{\beta}|$$

for the unrestricted model (1.1) and similarly, $\tilde{\sigma} = n^{-1} \tilde{V}(\frac{1}{2})$ under H_0. Concentrating the likelihood with respect to β we have

$$\ell(\hat{\beta}, \hat{\sigma}) = -n \log \hat{\sigma} - \tfrac{n}{2} - n \log 4$$

and with an analogous definition of $\ell(\tilde{\beta}, \tilde{\sigma})$ the likelihood ratio statistic becomes,

$$2(\ell(\hat{\beta}, \hat{\sigma}) - \ell(\tilde{\beta}, \tilde{\sigma})) = 2n \log(\tilde{\sigma}/\hat{\sigma})$$

Again, we are entitled to expect that the likelihood ratio statistic will have a limiting χ^2_q behavior. This can be easily checked using the earlier result by writing

$$\log(\tilde{\sigma}/\hat{\sigma}) = \log(1 + (\tilde{\sigma} - \hat{\sigma})/\hat{\sigma}) \approx (\tilde{\sigma} - \hat{\sigma})/\hat{\sigma} \tag{2.2}$$

an approximation whose validity is easy to establish under the null. Dividing numerator and denominator by $\sigma \equiv E|u|$, and arguing that $\hat{\sigma}/\sigma \to 1$, we have

$$2n \log(\tilde{\sigma}/\hat{\sigma}) = 2n(\tilde{\sigma} - \hat{\sigma})/\sigma + o_p(1) = 2(\tilde{V}(1/2) - \hat{V}(1/2))/\sigma + o_p(1) \tag{2.3}$$

Noting that $s = 4\sigma$ in this case and substituting this into (2.1) completes the argument.

While the unknown-scale form of the Laplace model is certainly preferable to the original, fixed-scale form, it is still unsatisfactory in that any departure from the condition $s = 4\sigma$ wreaks havoc with the asymptotic behavior of test statistic under the null. But this is easily rectified by defining the modified likelihood ratio statistics,

$$L_n(1/2) = 8(\tilde{V}(1/2) - \hat{V}(1/2))/s(1/2), \tag{2.4}$$

or again using (2.2),

$$\Lambda_n(1/2) = 8n\sigma \log(\tilde{\sigma}/\hat{\sigma})/s(1/2), \tag{2.5}$$

which may be shown to be asymptotically equivalent with limiting χ^2_q behavior. By directly evaluating the effect of the restriction on the mean absolute deviation from the median regression estimate this result provides a useful complement to alternative Wald and score formulations of tests of H_0. The modified likelihood ratio tests obviously require estimation of the nuisance parameters σ and s, but this is quite straightforward. For σ we may simply use $n^{-1}\hat{V}(1/2)$, while for s we may use the sparsity estimation methods described in Welsh (1988), or Koenker (1994).

2.2. **Inference for Quantile Regression.** The same approach we have elaborated for the median may be extended immediately to other quantiles. Define $\hat{V}(\tau)$ and $\tilde{V}(\tau)$ as in Section 1, and let $\hat{\sigma}(\tau) = n^{-1}\hat{V}(\tau)$, $\tilde{\sigma}(\tau) = n^{-1}\tilde{V}(\tau)$, and for the τth quantile consider the test statistics:

$$L_n(\tau) = \frac{2}{\lambda^2(\tau)s(\tau)}[\tilde{V}_n(\tau) - \hat{V}_n(\tau)]$$

and

$$\Lambda_n(\tau) = \frac{2n\hat{\sigma}(\tau)}{\lambda^2(\tau)s(\tau)}\log(\tilde{\sigma}(\tau)/\hat{\sigma}(\tau))$$

where $\lambda^2(\tau) = \tau(1-\tau)$. Tests of this sort based on the drop in the optimized value of the objective function of an M-estimator when relaxing the restriction imposed by the hypothesis, are termed ρ-tests by Ronchetti (1982). Following this terminology, we will refer below to tests based on $L_n(\tau)$ and $\Lambda_n(\tau)$ as quantile ρ-tests, although the phrase quasi-likelihood ratio processes would also be appropriate.

In some applications it may be useful 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, $(L_n(\tau_1), L_n(\tau_2), ..., L_n(\tau_m))$. Such results are subsumed in the following theory for the ρ-test *processes*, $\{L(\tau) : \tau \in [\epsilon, 1-\epsilon]\}$, and $\{\Lambda(\tau) : \tau \in [\epsilon, 1-\epsilon]\}$.

We will adopt the following regularity conditions regarding the model (1.1) and the hypothesis (1.2).

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

A.2: The sequence of design matrices $\{X_n\} = \{(x_i)_{i=1}^n\}$ satisfy:
- **(i):** $x_{i1} = 1, \quad i = 1, 2, \ldots$
- **(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):** $\max_{i=1,\ldots,n} \|x_i\| = O(n^{\frac{1}{4}} / \log n)$

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

Remarks. Conditions A.1 and A.2 are by now rather standard in the quantile regression literature. Somewhat stronger 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). Condition A.3 is a direct analogue of Condition A.3 in Koenker and Bassett (1982) allowing us to explore the question of local asymptotic power of tests.

To investigate the asymptotic behavior of $L_n(\tau)$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 $t \in (0,1)$,

$$B_q(t) \sim \mathcal{N}(0, t(1-t)I_q). \tag{2.6}$$

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 (2.6), $Q_q^2(t) \sim \chi_q^2$. Thus, we may interpret $Q_q^2(t)$ as a natural extension of the familiar univariate χ^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 χ^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 non-central Bessel process of order q with noncentrality function $\eta(\tau) = \mu(t)'\mu(t)/(t(1-t))$ and will be denoted by $Q_{q,\eta(t)}^2$. Of course, for any fixed $t \in (0,1)$, $Q^2_{q,\eta(t)} \sim \chi^2_{q,\eta(t)}$, a non-central χ^2_q random variable with q degrees of freedom and non-centrality parameter $\eta(t)$. Finally, the symbol $\Rightarrow$ will denote weak convergence, $\rightsquigarrow$ convergence in distribution, and $\rightarrow$, convergence in probability. We can now state our main result which is proved 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 $\eta(\tau) = \gamma(\tau)'(RD^{-1}R')^{-1}\gamma(\tau)/\omega^2(\tau)$, and $w(t) = \lambda(\tau)s(\tau)$. And, under the null hypothesis (1.2)

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

The alternative form of the quantile ρ-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 regression model.

A disadvantage of the quantile ρ-tests proposed above is the required estimation of the nuisance sparsity function, $s(\tau)$. Clearly, analogous Wald type tests would share this disadvantage and consequently we will not pursue them here. However, an interesting connection can be drawn at this point with the theory of rank based tests for linear models proposed in Gutenbrunner, Jurečková, Koenker, and Portnoy (1993), which we will refer to hereafter as GJKP.

These tests, which may be regarded as refined forms of the simple Lagrange multiplier tests for median regression considered in Koenker and Bassett (1982), are based on the regression rank score process, introduced by Gutenbrunner and Jurečková (1992). To simplify the exposition we will, following GJKP, consider the special "exclusion form" of the hypothesis (1.2), i.e.

$$\begin{bmatrix} 0 \vdots I_q \end{bmatrix} \begin{bmatrix} \beta_1(\tau) \\ \beta_{2n}(\tau) \end{bmatrix} = 0. \tag{2.7}$$

with the understanding that, as noted in the proof of Theorem 2.1, the model may always be reparameterized so that the hypothesis can be expressed in this form. The regression rank score process for the restricted version of the model under H_0 is,

$$\hat{a}_n(\tau) = \text{argmax}\{y'a | X_1'a = (1-\tau)X_1'e, \ a \in [0,1]^n\} \tag{2.8}$$

where e denotes an n-vector of ones, and X has been partitioned as $[X_1 \vdots X_2]$ to conform with the partitioning of the hypothesis. We adopt the following standard notation for partitioning the matrix D defined in Condition A.2.ii above: D_{ij} $i, j = 1, 2$ will denote the ij^{th} block of D, and D^{ij} will denote the ij block of D^{-1}. To illustrate, recall $D^{22} = (D_{22} - D_{21}D_{11}^{-1}D_{12})^{-1}$. The problem posed in (2.8) is the formal dual problem corresponding to the (primal) quantile regression linear program. Theorem 5.1 of GJKP considers tests of the exclusion form of the null hypothesis (2.7) based on the test statistic,

$$T_n = S_n' D_n^{22} S_n / A^2(\varphi)$$

where

$$\begin{aligned} S_n &= n^{-1/2}(X_2 - \hat{X}_2)'\hat{b}_n, \\ \hat{X}_2 &= X_1(X_1'X_1)^{-1}X_1'X_2, \\ D_n^{22} &= ((X_2 - \hat{X}_2)'(X_2 - \hat{X}_2)/n)^{-1} \\ \hat{b}_n &= (-\int \varphi(t)\hat{a}_{in}(t)dt)_{i=1}^n, \end{aligned}$$

φ is a score generating function of bounded variation and $A^2(\varphi) = \int_0^1 (\varphi(t) - \bar{\varphi})^2 dt$ with $\bar{\varphi} = \int_0^1 \varphi(t)dt$. Here we restrict attention to the special quantile score function,

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

so $\bar{\varphi} = 0$, and $A^2(\varphi) = \lambda^2(\tau) = \tau(1-\tau)$. The familiar special case is sign scores with $\tau = 1/2$. The quantile score function which yields rankscores,

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

allows us to define the rank test *process*

$$T_n(\tau) = S_n(\tau)' D_n^{22} S_n(\tau)/\lambda^2(\tau).$$

For fixed $\tau \in \mathcal{T}$, Theorem 5.1 of GJKP implies that under the conditions A1-3 above, $T_n(\tau) \rightsquigarrow \chi^2_{q,\eta(\tau)}$, where $\eta(\tau) = \gamma(\tau)'(D^{22})^{-1}\gamma(\tau)/\omega^2(\tau)$. Note that $(s(\tau))^{-1} = -\int_0^1 \varphi_\tau(t) df(F^{-1}(t))$ in (5.5) of GJKP in this case. This result may be extended to the whole process for $\tau \in \mathcal{T}$ providing a rank test link to the ρ-test processes already introduced. We will refer to the process $T_n(\tau)$ form as the quantile score process in the sequel. We summarize the foregoing discussion in the following result.

Theorem 2.2. *Under the conditions A1-3, $T_n(\tau) = L_n(\tau) + o_p(1)$, uniformly on $\mathcal{T}$.*

A crucial feature of this form of the test which distinguishes it from the corresponding ρ-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 $\sigma(\tau)$. This is a very substantial advantage over both the ρ-test and Wald approaches to testing in quantile regression, as was already stressed in Koenker and Bassett (1982). Indeed, this observation leads to an important extension of the theory to cover a much broader class of nulls than those possible under the iid error assumption of model (1.1). In effect, the iid error assumption requires that in the null model the effect of the covariates is a pure location shift. However, the the theory for T_n can be extended to models in which only that under the null the conditional quantile functions are linear. This includes, but is by no means limited to, models with linear conditional scale effects.

3. Some Illustrative Examples

In this Section we continue to explore the behavior of the R^1, Λ_n, and T_n processes introduced in the previous section employing a range of artificial data. In preliminary experiments the performance of tests based on L_n and Λ_n were essentially indistinguishable, so we chose to drop L_n in the subsequent comparisons. 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 τ-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 Regressions 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).

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 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 which is also consistent with the appearance of the last two panels of the first column which represent the Λ_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 Λ_n and T_n processes both indicate a highly significant departure from the null theory over the whole 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 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 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 $q = 2$.

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 Λ_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.

It is obviously difficult to evaluate the behavior of the various processes in Figure 3.1 based on a single realization. In Figures 3.2 and 3.3 we have included two more independent realizations from the same models to provide some indication of the range of behavior which can be anticipated in each case. 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. Note, for example, that in Figure 3.3 the estimates of the quadratic effect in the heteroscedastic model are quite poor. 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

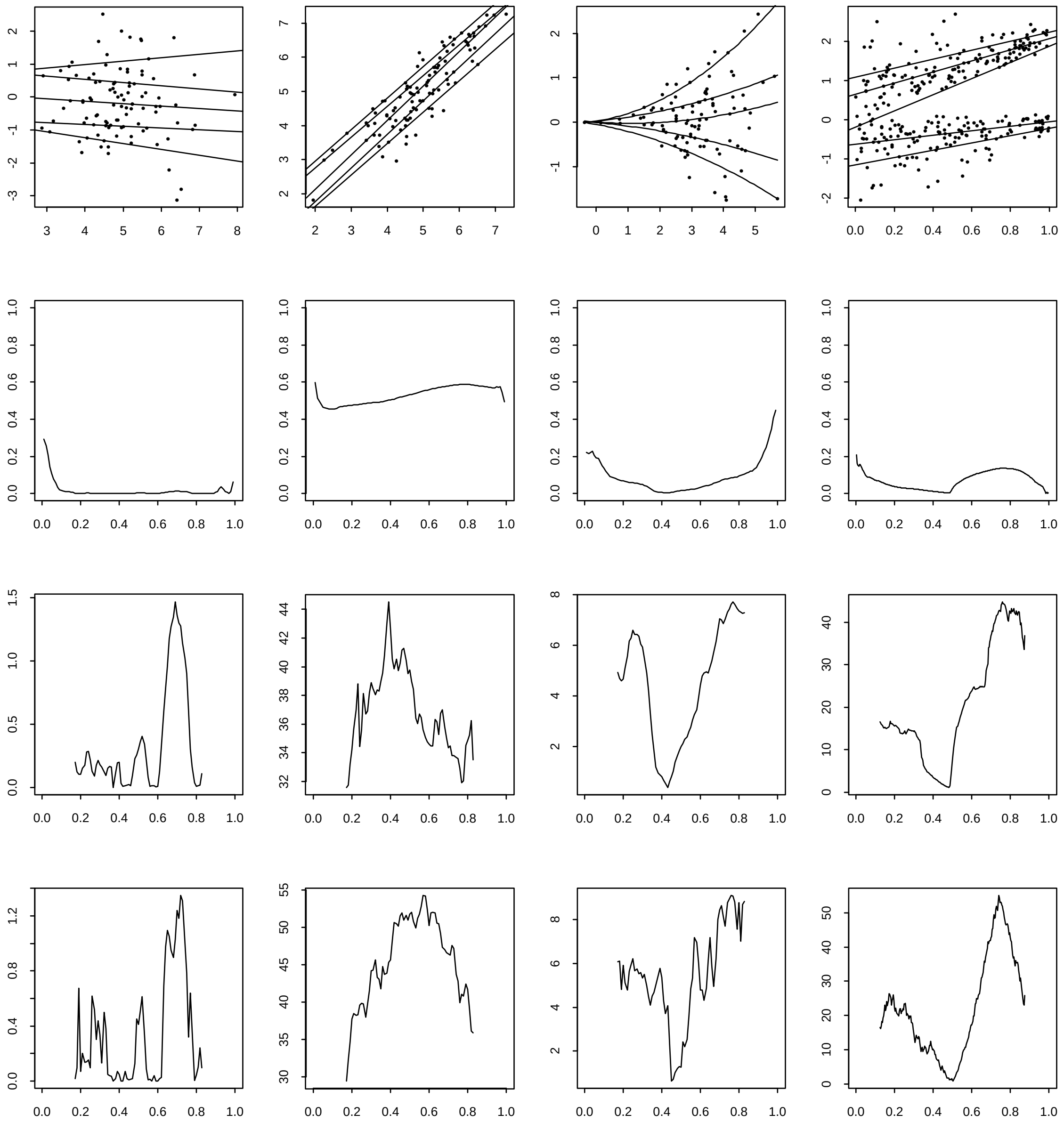

FIGURE 3.1. Some Illustrative Examples of Goodness of Fit Processes: The figure illustrates four different data configurations and their associated R^1, Λ_n and T_n functions. In the first row of the panel of figures the data is depicted, the solid lines are the fitted conditional quantile functions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

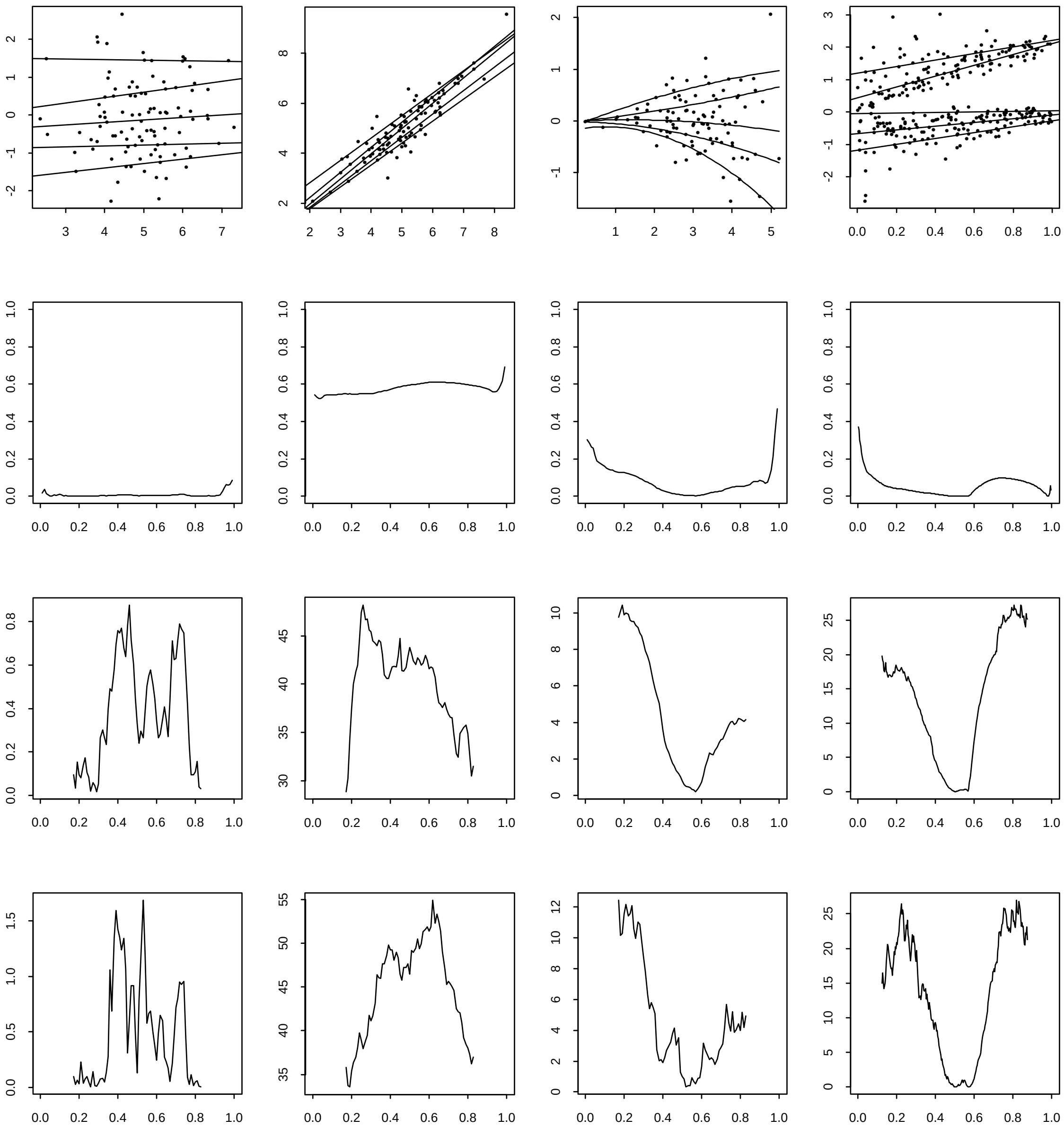

Figure 3.2. Some Illustrative Examples of Goodness of Fit Processes: The figure illustrates four different data configurations and their associated R^1, Λ_n and T_n functions. In the first row of the panel of figures the data is depicted, the solid lines are the fitted conditional quantile functions for $\tau \in \{.1, .25, .5, .75, .9\}$. The next row represents the R^1 process, and the third and fourth rows depict the Λ_n and T_n processes respectively.

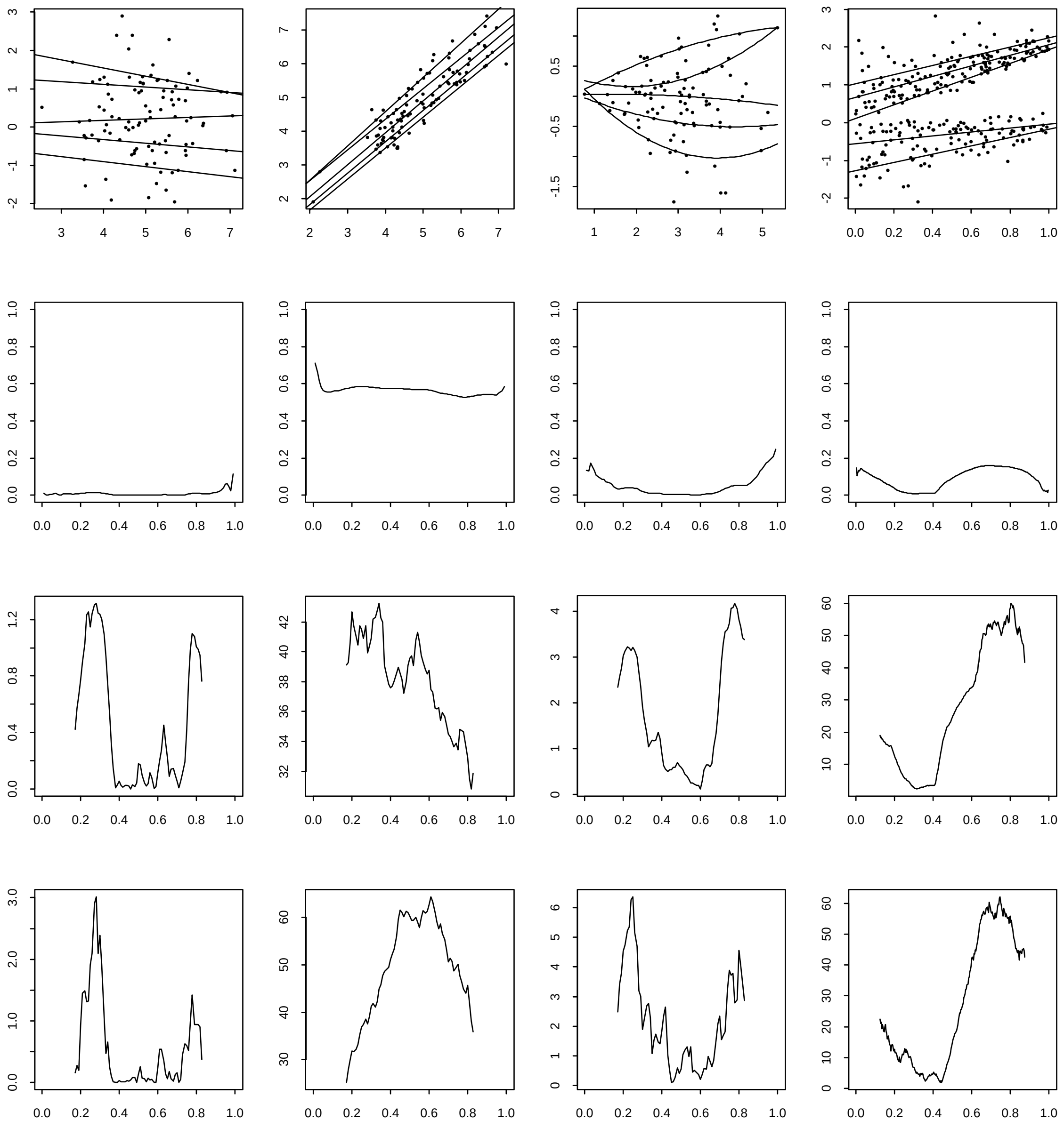

FIGURE 3.3. Some Illustrative Examples of Goodness of Fit Processes: The figure illustrates four different data configurations and their associated R^1, Λ_n and T_n functions. In the first row of the panel of figures the data is depicted, the solid lines are the fitted conditional quantile functions for $\tau \in \{.1, .25, .5, .75, .9\}$. The next row represents the R^1 process, and the third and fourth rows depict the Λ_n and T_n processes respectively.

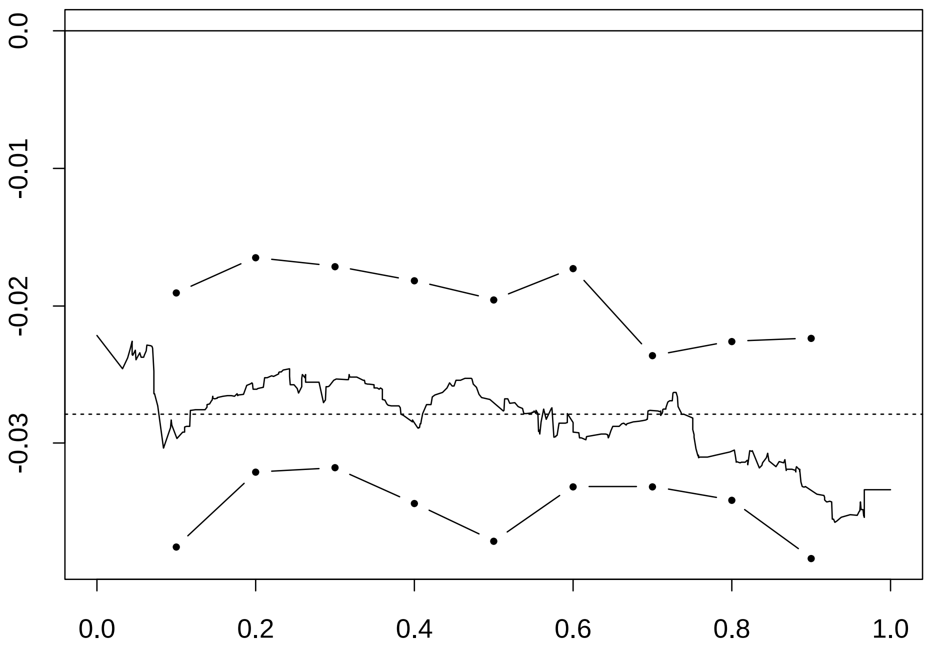

Figure 4.1. Estimated Quantile Regression Effect of Initial Per Capita GDP on GDP Growth: The figure illustrates the estimated quantile regression process for initial per capita GDP in a model for GDP growth which also includes the variables indicated in Figure 4.2. The rather jagged, solid line is the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

effect as estimated in a quantile regression which also includes the variables indicated in Figure 4.2. The solid, rather jagged line in this 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 in

Koenker (1994) and Koenker (1997). 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, 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 $-(1-e^{-\beta T})/T$.

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 similar empirical example from the early industrial orgranization literature, and 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 τ, 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 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.

An obvious application of the joint 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 $T_n(\tau)$ processes corresponding to the null hypothesis of no joint effect of

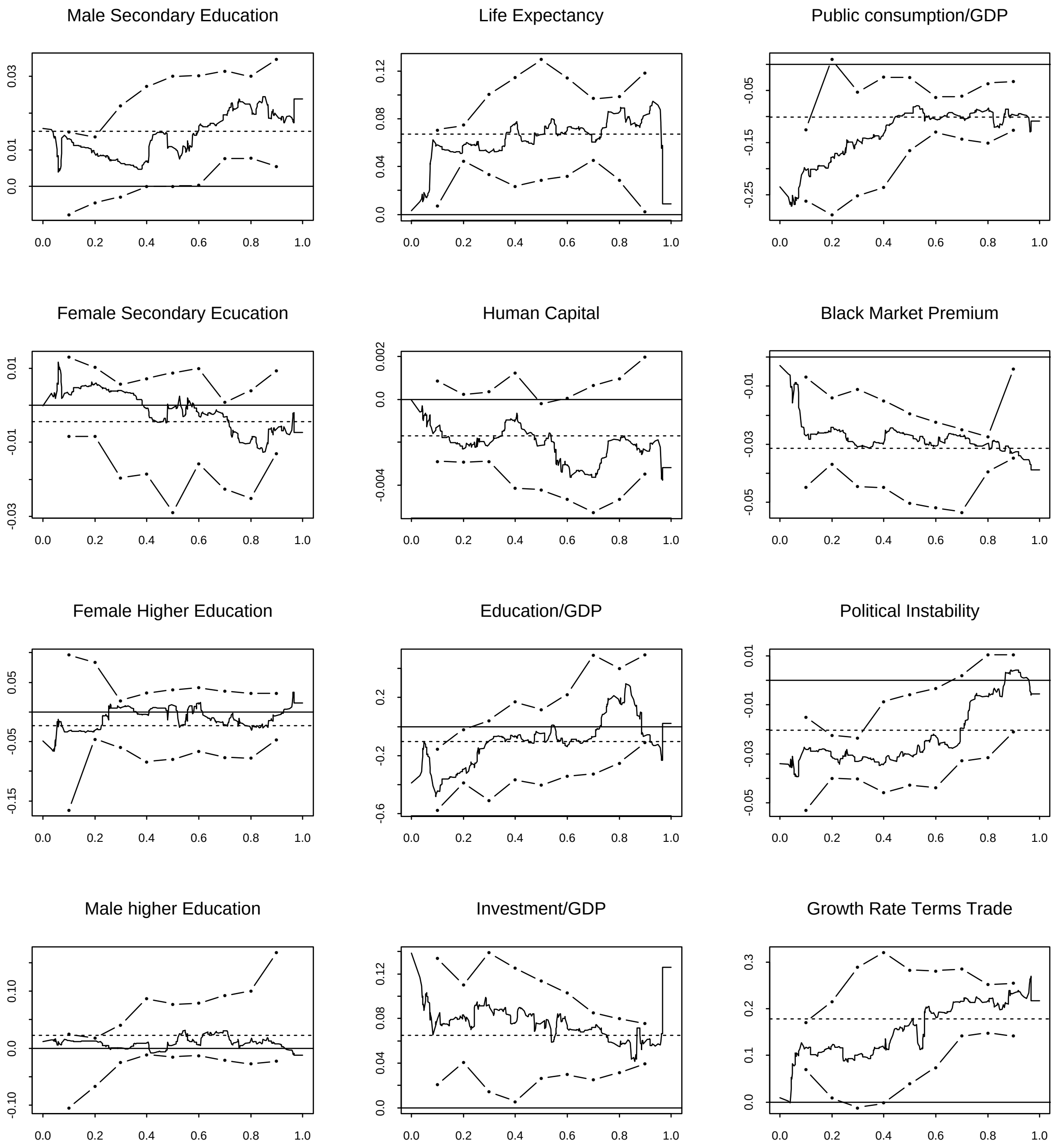

FIGURE 4.2. Estimated Quantile Regression Effects of other Variables on GDP Growth: The figure illustrates the estimated quantile regression process for 12 other variables in a model for GDP growth. The format of the plot follows that of the previous Figure.

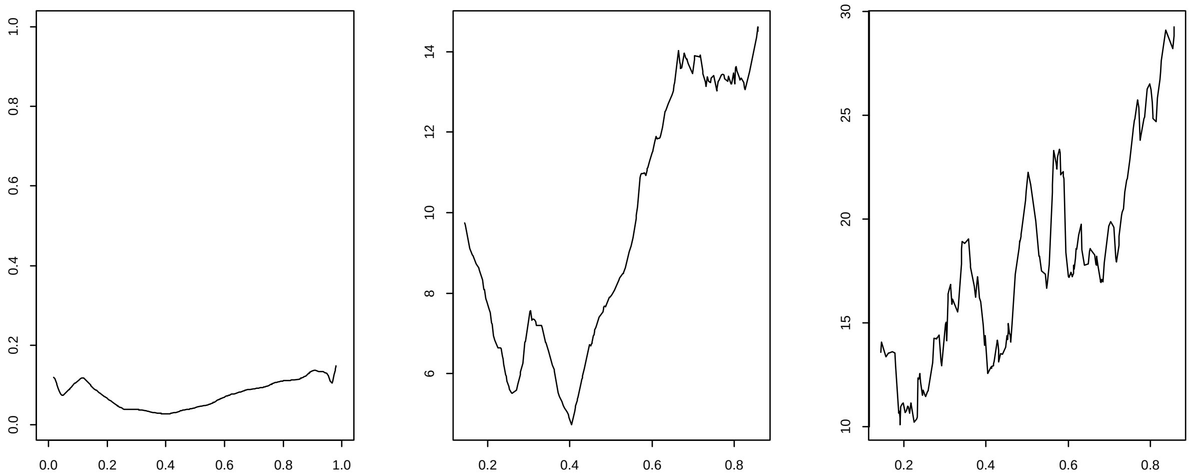

FIGURE 4.3. Joint Effect of Education Variables on GDP Growth: The the first panel of the figure illustrates the $R^1(\tau)$ function corresponding to the null hypothesis of no effect of any of the 5 education variables in the previous model. Note that the somewhat bowl shaped R^1 plot suggests that 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 Λ_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.

the education variables. Since in this plot any departure from the null whether positive or negative appears as a positive effect, 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 Λ_n version of the test fails to reject the null of no joint significance, while the rank based T_n version of the test does reject. 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 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 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 χ^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 χ^2_4 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 Λ_n test, but it is clear that both tests would benefit from some size correction.

q	α =0.10	α =0.05	α =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_α 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.

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 β_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 χ^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.

Test			Gaussian			Cauchy			χ^2_4		
	n	q	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
Λ_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
Λ_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
Λ_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
Λ_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
Λ_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
Λ_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 $q = p - 1$.

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 $\{x_{1i} = 1\}$, with $\{x_{ji} = 1\}$ 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,

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

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

Test			Gaussian			Cauchy			χ^2_4		
	n	q	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
Λ_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
Λ_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
Λ_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
Λ_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
Λ_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
Λ_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 Λ_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 $q = p - 1$.

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

Test			Bimodal			Heteroscedastic		
	n	q	0.1	0.05	0.01	0.1	0.05	0.01
Λ_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
Λ_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
Λ_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
Λ_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
Λ_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
Λ_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 $q = p - 1$.

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 Λ_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 ℓ_1 regression and adapts it to an arbitrary linear hypothesis. Two related goodness of fit processes, $\Lambda_n(\tau)$ and $T_n(\tau)$, are introduced 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 likelihood ratio and and Lagrange multiplier tests for ℓ_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 of Theorem 2.1

Proof: By a standard reduction to canonical form, as in for example (Amemiya (1985), Section 1.4.2) or (Gutenbrunner, Jurečková, Koenker, and Portnoy (1993), p. 312), the general linear hypothesis (1.2) may be reformulated as an "exclusion" hypothesis, of the form given in (2.7). Let

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

where $\omega(\tau) = \lambda(\tau) s(\tau)$ 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 δ, 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 $\alpha \in (0, \frac{1}{2})$,

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

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

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

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

Now, note that,

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

so we may write

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

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

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

Finally, note that if the null hypothesis (1.2) is of the exclusion form (2.7), i.e., with $R = [0 \vdots I_q]$, then the fact that $nR(X'X)^{-1}R' \to D^{22}$ follows immediately. And, in the general case, the reparameterization employed in the reduction to canonical form implies the same result after a brief computation. ■

REFERENCES

ABBERGER, K. (1996): "Non-parametric Quantile Forecasting and its Application to Stock Returns," Discussion Paper SFB 178, U. of Konstanz.

AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard, Cambridge, MA.

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.

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.

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

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.

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.

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.

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

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.

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 *ProbStInfr*, 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.

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

Welsh, A. H. (1988): "Asymptotically Efficient Estimation of the Sparsity Function At a Point," *Stat. and Prob. Letters*, 6, 427–432.