INFERENCE ON THE QUANTILE REGRESSION PROCESS

ROGER KOENKER AND ZHIJIE XIAO
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Abstract. Tests based on the quantile regression process can be formulated like the classical Kolmogorov-Smirnov and Crámer-von-Mises tests of goodness-of-fit employing the theory of Bessel processes as in Kiefer (1959). However, it is frequently desirable to formulate hypotheses involving unknown nuisance parameters, thereby jeopardizing the distribution free character of these tests. We characterize this situation as "the Durbin problem" since it was posed in Durbin (1973), for parametric empirical processes.

In this paper we consider an approach to the Durbin problem involving a martingale transformation of the parametric empirical process suggested by Khmaladze (1981) and show that it can be adapted to a wide variety of inference problems involving the quantile regression process. In particular, we suggest new tests of the location shift and location-scale shift models that underlie much of classical econometric inference.

The methods are illustrated in some limited Monte-Carlo experiments and with a reanalysis of data on unemployment durations from the Pennsylvania Reemployment Bonus Experiments. The Pennsylvania experiments, conducted in 1988-89, were designed to test the efficacy of cash bonuses paid for early reemployment in shortening the duration of insured unemployment spells.

1. Introduction

Quantile regression, as introduced by Koenker and Bassett (1978), is gradually evolving into a comprehensive approach to the statistical analysis of linear and nonlinear response models for conditional quantile functions. Just as classical linear regression methods based on minimizing sums of squared residuals enable one to estimate models for conditional mean functions, quantile regression methods based on minimizing asymmetrically weighted *absolute* residuals offer a mechanism for estimating models for the conditional median function, and the full range of other conditional quantile functions. By supplementing least squares estimation of conditional mean functions with techniques for estimating a full family of conditional quantile functions, quantile regression is capable of providing a much more complete statistical analysis of the stochastic relationships among random variables.

Version: September 4, 2000. The authors wish to express their appreciation to Steve Portnoy, Hira Koul, Jim Heckman, Hannu Oja and seminar participants at the U. of Illinois, U. of Chicago, and CREST-INSEE for helpful comments. The research was partially supported by NSF grant SBR-9617206.

There is already a well-developed theory of asymptotic inference for many important aspects of quantile regression. Rank-based inference based on the approach of Gutenbrunner, Jurečková, Koenker, and Portnoy (1993) appears particularly attractive for a wide variety of quantile regression inference problems including the construction of confidence intervals for individual quantile regression parameter estimates. There has also been considerable attention devoted to various resampling strategies. See e.g. Hahn (1995), Horowitz (1998), Bilias, Chen, and Ying (1999), He and Hu (1999). In Koenker and Machado (1999) some initial steps have been taken toward a theory of inference based on the entire quantile regression process. These steps have clarified the close tie to classical Kolmogorov-Smirnov goodness of fit results, and related literature on Bessel processes initiated by Kiefer (1959).

This paper describes some further steps in this direction. These new steps depend crucially on an ingenious suggestion by Khmaladze (1981) for dealing with tests of composite null hypotheses based on the empirical distribution function. Khmaladze's results have been slow to percolate into statistics generally, but the approach has recently played an important role in work on regression diagnostics by Stute, Thies, and Zhu (1998) and Koul and Stute (1999). In econometrics, Bai (1998) seems to have been the first to recognize the potential importance of these methods.

Khmaladze's martingale transformation approach provides a general strategy for purging the effect of estimated nuisance parameters from the first order asymptotic representation of the empirical process and thereby restoring the feasibility of "asymptotically distribution free" tests. The approach seems especially attractive in quantile regression settings and is capable of greatly expanding the scope of inferential methods described in earlier work.

1.1. **Quantile Treatment Effects.** To motivate our approach it is helpful to begin by reconsidering the classical two-sample treatment-control problem. In the simplest possible setting we can imagine a random sample of size, n, drawn from a homogeneous population and randomized into n_1 treatment observations, and n_0 control observations. We have a response variable, Y_i, and are interested in evaluating the effect of the treatment on this response.

In a typical clinical trial application, for example, the treatment would be some form of medical procedure, and Y_i, might be log survival time. In our application discussed in Section 6, the treatment is an offer of a cash bonus for early exit from a spell of unemployment, and Y_i is the logarithm of individual i's unemployment duration. In the first instance we might be satisfied to know simply the mean treatment effect, that is, the difference in means for the two groups. This we could evaluate by "running the regression" of the observed y_i's on an indicator variable: $x_i = 1$, if subject i was treated, $x_i = 0$, if subject i was a control. Of course this regression would presume, implicitly, that the variability of the two subsamples was the same; this observation opens the door to the possibility that the treatment alters other features of the response distribution as well. Although we are accustomed to thinking about regression models in which the covariates affect only the location of the conditional distribution of the response – this is the force of the iid error assumption – there is no compelling reason to believe that covariates must operate in this restrictive fashion.

Lehmann (1974) introduced the following general formulation of the two sample treatment effect,

> "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."

Doksum (1974) provides a detailed axiomatic analysis of this formulation, 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

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

Changing variables, so $\tau = F(x)$ we obtain what we will call 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

$$\hat{\delta}(\tau) = G_{n_1}^{-1}(\tau) - F_{n_0}^{-1}(\tau)$$

where G_{n_1}, F_{n_0} denote the empirical distribution functions of the treatment and control observations respectively, and $F_n^{-1} = \inf\{x | F_n(x) \geq \tau\}$, as usual. Since we cannot observe subjects in *both* the treated and control states – and this platitude may be regarded as the fundamental "uncertainty principle" underlying the "causal effects" literature – it seems reasonable to regard $\delta(\tau)$ as a complete description of the treatment effect. Of course, there is no way of really *knowing* whether the treatment operates in the way prescribed by Lehmann. In fact, the treatment may make otherwise weak subjects especially robust, and turn the strong to jello. All we can observe from the experimental evidence is the difference between the two *marginal* survival distributions, so it is natural to associate the treatment effect with this difference. The quantile treatment effect provides the unexpurgated version. Of course if there are systematic differences in treatment response associated with observable covariates, then these effects can be estimated via interactions with the treatment indicator.

Heckman, Smith, and Clements (1997) consider a model which allows a specified degree of "slippage" in the quantile ranks of the two distributions in the context of a bounds analysis. Oja (1981) considers orderings of distributions based on location, scale, skewness, kurtosis, etc., based on the function $\Delta(x)$.

Of course, it is possible that the two distributions differ only by a location shift, so $\delta(\tau) = \delta_0$, or that they differ by a scale shift so $\delta(\tau) = \delta_1 F^{-1}(\tau)$ or that they differ by a location and scale shift so $\delta(\tau) = \delta_0 + \delta_1 F^{-1}(\tau)$. But these hypotheses are all nicely nested within Lehmann's general framework. And yet, as we shall see, testing these hypotheses against the general alternatives represented by the Lehmann-Doksum quantile treatment effect poses some challenging technical problems.

1.2. **Inference on the Quantile Regression Process.** In the two-sample treatment-control model there are a multitude of tests designed to answer the question: "Is the treatment effect significant." The most familiar of these, like the two-sample Student-t and Mann-Whitney-Wilcoxon tests are designed to reveal location shift alternatives. Others are dsigned for scale shift alternatives. Still others, like the two sample Kolmogorov-Smirnov test, are intended to encompass omnibus non-parametric alternatives. When the non-parametric null is posed in a form free of nuisance parameters we have an elegant distribution-free theory for a variety of tests, including the Kolmogorov-Smirnov test that are based on the empirical distribution function.

Non-parametric testing in the presence of nuisance parameters under the null, however, poses some new problems. Suppose, for example, that we wish to test the hypothesis that the response distribution under the treatment, G, differs from the control distribution, F, by a pure location shift, that is for all $\tau \in [0, 1]$,

$$G^{-1}(\tau) = F^{-1}(\tau) + \delta_0$$

for some real δ_0, or that they differ by a location-scale shift, so,

$$G^{-1}(\tau) = \delta_1 F^{-1}(\tau) + \delta_0.$$

In such cases we can easily estimate the nuisance parameters, δ_0, δ_1, but the introduction of the estimated parameters into the asymptotic theory of the empirical process destroys the distribution free character of the resulting Kolmogorov-Smirnov test. Analogous problems arise in the theory of the one sample Kolmogorov-Smirnov test when there are estimated parameters under the null, and have been considered by Durbin (1973) and others.

The two-sample treatment-control model can be considered the simplest manifestation of the quantile regression model in which the regression design matrix consists only of an intercept and a single indicator (dummy) variable. In such models it is natural to consider models in which the conditional distributions of the response at various settings of the covariates differ only by a location shift, or by only a location-scale shift. Indeed, virtually all of the econometrics literature on regression deals exclusively with just such models.

In the general linear quantile regression model specified as,

$$Q_{y|x}(\tau|x) = x^\top \beta(\tau)$$

such models may be represented by the linear hypothesis:

$$\beta(\tau) = \alpha + \gamma F_0^{-1}(\tau)$$

for α and γ in $\mathbb{R}^p$ and F_0^{-1} a univariate quantile function. Thus, all p coordinates of the quantile regression coefficient vector are required to be affine functions of the same univariate quantile function, F_0^{-1}. Such models may be viewed as arising from linearly heteroscedastic model

$$y_i = x_i^\top \alpha + (x_i^\top \gamma) u_i$$

with the $\{u_i\}$ iid from the df F_0. They can be estimated by solving the linear programming problem,

$$\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i^\top b)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$. The problem of testing such hypotheses will be considered in Sections 3 and 4, along with some special consideration for cases in which $\gamma = 0$ that lead to new tests for heteroscedasticity.

In the next section we briefly describe the nature of these problems in their canonical form, the classical one-sample goodness of fit problem. Khmaladze's martingale decomposition strategy for dealing with these problems is then introduced. Section 3 extends the Khmaladze approach to general problems of inference based on the quantile regression process. Section 4 treats some practical problems of implementing the tests. Section 5 reports the results of a limited Monte-Carlo experiment designed to evaluate the finite sample performance of the tests. Section 6 describes an empirical application to the analysis of unemployment durations. Section 7 contains some concluding remarks.

2. A Heuristic Introduction to Khmaladzation

Arguably the most fundamental problem of statistical inference is the classical goodness-of-fit problem: given a random sample, $\{y_1, \ldots, y_n\}$, on a real-valued random variable, Y, test the hypothesis that Y comes from distribution function, F_0. Tests based on the empirical distribution function, $F_n(y) = n^{-1} \sum I(Y_i \leq y)$, like the Kolmogorov-Smirnov statistic

$$K_n = \sup_{x \in \mathbb{R}} \sqrt{n} |F_n(y) - F_0(y)|,$$

are especially attractive because they are asymptotically distribution-free. The limiting distribution of K_n is the same for every continuous distribution function F_0. This remarkable fact follows by (trivially) noting that the process, $\sqrt{n}(F_n(y) - F_0(y))$, can be transformed to a test of uniformity, via the change of variable, $y \to F_0^{-1}(t)$, based on

$$v_n(t) = \sqrt{n}(F_n(F_0^{-1}(t)) - t).$$

It is well known that $v_n(t)$ converges weakly to a Brownian bridge process, $v_0(t)$, that is a mean-zero Gaussian process with covariance function

$$Ev_0(t)v_0(s) = t \wedge s - st,$$

and thus the distribution of K_n and related functionals follows from the observation of Doob (1949) and its subsequent refinements.

2.1. **The Durbin Problem.** It is rare in practice, however, that we are willing to specify F_0 completely. More commonly, our hypothesis places F in some parametric family $\mathcal{F}_\theta$ with $\theta \in \Theta \subseteq \mathbb{R}^p$. For example, we may wish to test "normality", claiming that Y has distribution $F_{\theta_0}(y) = \Phi((y - \mu_0)/\sigma_0)$, but $\theta_0 = (\mu_0, \sigma_0)$ is unknown. We are thus led to consider, following Durbin (1973), the parametric empirical process,

$$U_n(y) = \sqrt{n}(F_n(y) - F_{\hat{\theta}_n}(y)).$$

Again changing variables, so $y \to F_{\theta_0}^{-1}(t)$, we may equivalently consider

$$u_n(t) = \sqrt{n}(G_n(t) - G_{\hat{\theta}_n}(t))$$

where $G_n(t) = F_n(F_{\theta_0}^{-1}(t))$ and $G_{\hat{\theta}_n}(t) = F_{\hat{\theta}_n}(F_{\theta_0}^{-1}(t))$ so $G_{\theta_0}(t) = t$. Under mild conditions on the sequence $\{\hat{\theta}_n\}$ we have the linear (Bahadur) representation,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \int_0^1 h_0(s)dv_n(s) + o_p(1).$$

So provided the mapping $\theta \to G_\theta$ has a Fréchet derivative, $g = g_{\theta_0}$, that is, $\sup_t |G_{\theta+h}(t) - G_\theta(t) - h^\top g(t)| = o(||h||)$ as $h \to 0$, see van der Vaart (1998, p.278), we may write,

$$G_{\hat{\theta}_n}(t) = t + (\hat{\theta}_n - \theta_0)^\top g(t) + o_p(1),$$

and thus obtain, with $r_n(t) = o_p(1)$,

$$\begin{aligned} \hat{v}_n(t) &= \sqrt{n}(G_n(t) - t - (G_{\hat{\theta}_n}(t) - t)) \\ &= v_n(t) - g(t)^\top \int_0^1 h_0(s)dv_n(s) + r_n(t), \end{aligned} \tag{2.1}$$

which converges weakly to the Gaussian process,

$$u_0(t) = v_0(t) - g(t)^\top \int_0^1 h_0(s)dv_0(s).$$

The necessity of estimating θ_0 introduces the drift component $g(t)^\top \int_0^1 h_0(s)dv_0(s)$. Instead of the simple Brownian bridge process, $v_0(t)$, we obtain a more complicated Gaussian process with covariance function

$$Eu_0(t)u_0(s) = s \wedge t - ts - g(t)^\top \mathcal{H}_0(s) - g(s)^\top \mathcal{H}_0(t) + g(s)^\top \mathcal{J}_0 g(t)$$

where $\mathcal{H}_0(t) = \int_0^t h_0(s)ds$ and $\mathcal{J}_0 = \int_0^1 \int_0^1 h_0(t)h_0(s)^\top dtds$. When $\hat{\theta}_n$ is the mle, so $h_0(s) = -(E\nabla_\theta\psi)^{-1}\psi(F^{-1}(s))$ with $\psi = \nabla_\theta \log f$, the covariance function simplifies nicely to

$$Eu_0(t)u_0(s) = s \wedge t - ts - g(s)^\top \mathcal{I}_0 g(t)$$

where $\mathcal{I}_0$ denotes Fisher's information matrix. See Durbin (1973) and Shorack and Wellner (1986) for further details on this case.

The practical consequence of the drift term involving the function $g(t)$ is to invalidate the distribution-free character of the original test. Tests based on the parametric empirical process $u_n(t)$ require special consideration of the process $u_0(t)$ and its dependence on F in each particular case. Shorack and Wellner (1986) discuss several leading examples. Durbin (1973) describes a general numerical approach based on Fourier inversion, but also expresses doubts about feasibility of the method when the parametric dimension, p, of θ exceeds one. Although the problem of finding a viable, general approach to inference based on the parametric empirical process had been addressed by several previous authors, notably Darling (1955), we will, in the spirit of Stigler's law of eponymy, Stigler (1980), refer to this as "the Durbin problem."

2.2. **Martingales and the Doob-Meyer Decomposition.** Khmaladze's general approach to the Durbin problem can be motivated as a natural elaboration of the Doob-Meyer decomposition for the parametric empirical process. Recall that a stochastic process $x = \{x(t) : t \geq 0\}$ that is (i) right continuous with left limits; (ii) integrable $\sup_{0 \leq t < \infty} E|x(t)| < \infty$; and (iii) adapted to the filtration $\{\mathcal{F}_t : t \geq 0\}$, is called a submartingale if

$$E(x(t+s)|\mathcal{F}_t) \geq x(t) \qquad a.s.$$

and is called a martingale if

$$E(x(t+s)|\mathcal{F}_t) = x(t) \qquad a.s.$$

The Doob-Meyer decomposition asserts that for any nonnegative submartingale, x, there exists an increasing right continuous predictable process, $a(t)$, such that $Ea(t) < \infty$, and a right continuous martingale m, such that

$$x(t) = a(t) + m(t) \qquad a.s.$$

A process $a(t)$ is called predictable with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$ if, viewed as a mapping from $[0, \infty) \times \Omega$ to $\mathbb{R}$ it is measurable with respect to the σ-algebra generated by the filtration $\mathcal{F}_t$, that is the σ-algebra generated by all sets of the form $(r, s] \times A$ for $0 \leq r < s < 1$ and $A \in \mathcal{F}_r$. See e.g. Fleming and Harrington (1991). For an extensive account of recent developments see Prakasa Rao (1999).

Let $X_1, \ldots, X_n$ be iid from F_0, so $Y_i = F_0(X_i), \quad i = 1, \ldots, n$ are iid uniform, $U[0, 1]$. The empirical distribution function

$$G_n(t) = F_n(F_0^{-1}(t)) = n^{-1} \sum_{i=1}^n I(Y_i \leq t).$$

viewed as a process, is a submartingale. We have an associated filtration $\mathcal{F}^{G_n} = \{\mathcal{F}_t^{G_n} : 0 \leq t \leq 1\}$ and the order statistics $Y_{(1)}, \ldots, Y_{(n)}$ are Markov times with respect to $\mathcal{F}^{G_n}$, that is $\{X_{(i)} \leq t\} = \{F_n(t) \geq i/n\} \in \mathcal{F}_t^{G_n}$.

The process $G_n(t)$ is Markov; Khmaladze notes that for $\Delta t \geq 0$,

$$\begin{aligned} n\Delta G_n(t) &= n[G_n(t+\Delta t) - G_n(t)] \\ &\sim \text{Binomial}(n(1-G_n(t)), \Delta t/(1-t)) \end{aligned}$$

with $G_n(0) = 0$, thus

$$E(\Delta G_n(t)|\mathcal{F}_t^{G_n}) = \frac{1-G_n(t)}{1-t}\Delta t. \tag{2.2}$$

This suggests the decomposition

$$G_n(t) = \int_0^t \frac{1-G_n(s)}{1-s}ds + m_n(t).$$

That $m_n(t)$ is a martingale then follows from the fact that, from (2.2),

$$E(m_n(t)|\mathcal{F}_s^{G_n}) = m_n(s)$$

and integrability of $m_n(t)$ follows from the inequality

$$\int_0^t \frac{1-G_n(s)}{1-s}ds \leq -\log(1-Y_{(n)}),$$

which implies a finite mean for the compensator, or predictable component. Substituting for $G_n(t)$ in (2.2) we have the classical Doob-Meyer decomposition of the empirical process v_n

$$v_n(t) = w_n(t) - \int_0^t \frac{v_n(s)}{1-s}ds$$

where $v_n(t) = \sqrt{n}(G_n(t) - t)$ and the normalized process $w_n(t) = \sqrt{n}m_n(t)$ converges weakly to a standard Brownian motion process, $w_0(t)$, by the argument of Khmaladze(1981, §2.6).

2.3. **The Parametric Empirical Process.** To extend this approach to the general parametric empirical process, we now let $g(t) = (t, \bar{g}(t)^\top)^\top = (t, g_1(t), \ldots g_m(t))^\top$ be a $(m+1)$-vector of real-valued functions on $[0,1]$. Suppose that the functions $\dot{g}(t) = dg(t)/dt$ are linearly independent in a neighborhood of 1 so

$$C(t) \equiv \int_t^1 \dot{g}(s)\dot{g}(s)^\top ds$$

is non-singular, and consider the transformation

$$w_n(t) = v_n(t) - \int_0^t \dot{g}(s)^\top C^{-1}(s) \int_s^1 \dot{g}(r)dv_n(r)ds. \tag{2.3}$$

Here, $w_n(t)$ clearly depends upon the choice of g, and therefore differs from $w_n(t)$ defined above. But the abuse of notation maybe justified by noting that in the special case $g(t) = t$, we have $C(s) = 1-s$, and $\int_s^1 \dot{g} dv_n(r) = -v_n(s)$ yielding the Doob-Meyer decomposition (2.2) as a special case. In the general case, the transformation

$$Q_g \varphi(t) = \varphi(t) - \int_0^t \dot{g}(s)^\top C^{-1}(s) \int_s^1 \dot{g}(r) d\varphi(r) ds$$

may be recognized as the residual from the prediction of $\varphi(t)$ based on the recursive least squares estimate using information from $(t, 1]$. For functions in the span of g, the prediction is exact, that is, $Q_g g = 0$.

Now returning to the representation of the parametric empirical process, $\hat{v}_n(t)$, given in (2.1), using Khamaladze (1981, §4.2), we have,

$$\begin{aligned}
\tilde{v}_n(t) &= Q_g \hat{v}_n(t) \\
&= Q_g (v_n(t) - \bar{g}(t)^\top \int_0^1 h_0(s) dv_n(s) + r_n(t)) \\
&= Q_g (v_n(t) + r_n(t)) \\
&= w_0(t) + o_p(1).
\end{aligned}$$

The transformation of the parametric empirical process annihilates the g component of the representation and in so doing restores the feasibility of asymptotically distribution free tests based on the transformed process $\tilde{v}_n(t)$.

2.4. **The Parametric Empirical Quantile Process.** What can be done for tests based on the parametric empirical process can also be adapted for tests based on the parametric empirical *quantile* process. In some ways the quantile domain is actually more convenient. Suppose $\{y_1, \ldots, y_n\}$ constitute a random sample on Y with distribution function F_Y. Consider testing the hypothesis, $F_Y(y) = F_0((y - \mu_0)/\sigma_0)$, so,

$$\alpha(\tau) \equiv F_Y^{-1}(\tau) = \mu_0 + \sigma_0 F_0^{-1}(\tau).$$

Given the empirical quantile process

$$\hat{\alpha}(\tau) = \inf\{a \in \mathbb{R} | \sum_{i=1}^n \rho_\tau(y_i - a) = \min!\}$$

and known parameters $\theta_0 = (\mu_0, \sigma_0)$ tests may be based on

$$v_n(\tau) = \sqrt{n} \varphi_0(\tau)(\hat{\alpha}(\tau) - \alpha(\tau))/\sigma_0 \Rightarrow v_0(\tau)$$

where $\varphi_0(\tau) = f_0(F_0^{-1}(\tau))$ and $v_0(\tau)$ is the Brownian bridge process.

To test our hypothesis when θ is unknown, set $\xi(t) = (1, F_0^{-1}(t))^\top$ and for an estimator $\hat{\theta}_n$ satisfying,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \int_0^1 h_0(s) dv_n(s) + o_p(1)$$

set $\tilde{\alpha}(t) = \hat{\mu} + \hat{\sigma}_0 F_0^{-1}(t) = \hat{\theta}_n^\top \xi(t)$. Then

$$\hat{v}_n(t) = \sqrt{n}\varphi_0(t)(\hat{\alpha}(t) - \tilde{\alpha}(t)/\sigma_0 \tag{2.4}$$

$$= \sqrt{n}\varphi_0(t)(\hat{\alpha}(t) - \alpha(t) - (\tilde{\alpha}(t) - \alpha(t)))/\sigma_0$$

$$= v_n(t) - \sqrt{n}\varphi_0(t)(\hat{\theta} - \theta_0)^\top \xi(t)/\sigma_0 \tag{2.5}$$

$$= v_n(t) - \varphi_0(t)\xi(t)^\top \int_0^1 h_0(s)dv_n(s) + o_p(1)$$

Thus, if we take $g(t) = (t, \xi(t)^\top \varphi_0(t))^\top$, we obtain,

$$\dot{g}(t) = (1, \dot{f}/f, 1 + F_0^{-1}(t)\dot{f}/f)^\top$$

where $\dot{f}/f$ is evaluated at $F_0^{-1}(t)$, so for example in the Gaussian case,

$$\dot{g}(t) = (1, -\Phi^{-1}(t), 1 - \Phi^{-1}(t)^2)^\top.$$

Given the representation (2.4) and the fact that $\xi(t)$ lies in the linear span of g, we may again apply Khmaladze's martingale transformation to obtain,

$$\tilde{v}_n(t) = Q_g \hat{v}_n(t),$$

which can then be shown to converge to the standard Brownian motion process. In the next section we explore extending this approach to multidimensional quantile regression.

3. QUANTILE REGRESSION INFERENCE

The classical linear regression model asserts that the conditional mean of the response, y_i, given covariates, x_i, may be expressed as a linear function of the covariates. That is, there exists a $\beta \in \mathbb{R}^p$ such that,

$$E(y_i|x_i) = x_i^\top \beta.$$

The linear quantile regression model asserts, analogously, that the conditional quantile functions of y_i given x_i are linear in covariates,

$$F_{y_i|x_i}^{-1}(\tau|x_i) = x_i^\top \beta(\tau) \tag{3.1}$$

for τ in some index set $\mathcal{T} \subset [0,1]$. The model (3.1) will be taken to be our basic maintained hypothesis. For convenience we will restrict attention to the case that $\mathcal{T} = [\epsilon, 1-\epsilon]$ for some $\epsilon \in (0, 1/2)$, and to facilitate asymptotic local power analysis we will consider sequences of models for which $\beta(\tau) = \beta_n(\tau)$ depends explicitly on the sample size, n.

A leading special case is the *location-scale shift model*,

$$F_{y_i|x_i}^{-1}(\tau|x_i) = x_i^\top \alpha + x_i^\top \gamma F_0^{-1}(\tau). \tag{3.2}$$

where $F_0^{-1}(\tau)$ denotes a univariate quantile function. Covariates affect both the location and scale of the conditional distribution of y_i given x_i in this model, but the covariates have no effect on the *shape* of the conditional distribution. Typically, the vectors $\{x_i\}$ "contain an intercept" so e.g., $x_i = (1, z_i^\top)^\top$ and (3.2) may be seen as arising from the linear model

$$y_i = x_i^\top \alpha + (x_i^\top \gamma) u_i$$

where the "errors" $\{u_i\}$ are iid with distribution function F_0. Further specializing the model, may write,

$$x_i^\top \gamma = \gamma_0 + z_i^\top \gamma_1,$$

and the restriction, $\gamma_1 = 0$, then implies that the covariates affect only the *location* of the y_i's. We will call this model

$$F^{-1}_{y_i|x_i}(\tau|x_i) = x_i^\top \alpha + \gamma_0 F_0^{-1}(\tau) \tag{3.3}$$

the *location shift model.* Although this model underlies much of classical econometric inference, it posits a very narrowly circumscribed role for the x_i. In the remainder of this section we explore ways to test the hypotheses that the general linear quantile regression model takes either the location shift or location-scale shift form.

We will consider a linear hypothesis of the general form,

$$R\beta(\tau) - r = \Psi(\tau), \qquad \tau \in \mathcal{T}, \tag{3.4}$$

where R denotes a $q \times p$ matrix, $q \leq p, r \in \mathbb{R}^q$, and $\Psi(\tau)$ denotes a known function $\Psi : \mathcal{T} \to \mathbb{R}^q$. For example in the one sample setting of the previous section, we might take $R = \sigma^{-1}, r = \mu/\sigma$ and $\Psi(\tau) = \Phi^{-1}(\tau)$, in order to test that the y_i's were $\mathcal{N}(\mu, \sigma^2)$.

In the two sample model described in Section 1.1.

$$F^{-1}_{y_i|D_i}(\tau|D_i) = \beta_0(\tau) + \beta_1(\tau) D_i$$

we might like to test that, the treatment and control distributions differ by an affine transformation or, even more simply, that they differ by a location shift. In these cases we may take $\Psi(\tau) \equiv 0, r = \theta_0, R = (1, -\theta_1)$ in the former case, and $R = (1, -1)$ in the latter case. Of course, we could also expand these two-sample hypotheses to consider fully specified parametric models with an explicit choice of $\Psi(\tau)$, however, the semi-parametric form of the hypotheses expressed above seems more plausible for most econometric applications.

We will consider tests based on the quantile regression process,

$$\hat{\beta}(\tau) = \text{argmin}_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i^\top b)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$. Under the location-scale shift form of the quantile regression model (3.2) we will have under mild regularity conditions,

$$\sqrt{n}\varphi_0(\tau)\Omega^{-1/2}(\hat{\beta}(\tau) - \beta(\tau)) \Rightarrow v_0(\tau) \tag{3.5}$$

where $v_0(\tau)$ now denotes a p-dimensional independent Brownian bridge process,

$$\beta(\tau) = \alpha + \gamma F^{-1}(\tau),$$

and $\Omega = H_0^{-1} J_0 H_0^{-1}$ with $J_0 = \lim n^{-1} \sum x_i x_i^\top$, and $H_0 = \lim n^{-1} \sum x_i x_i^\top / \gamma^\top x_i$.

It then follows quite easily that under the null hypothesis (3.4),

$$v_n(\tau) = \sqrt{n}\varphi_0(\tau)(R\Omega R^\top)^{-1/2}(R\hat{\beta}(\tau) - r - \Psi(\tau)) \Rightarrow v_0(\tau),$$

so tests that are asymptotically distribution free can be readily constructed. Indeed, Koenker and Machado (1999) consider tests of this type when R constitutes an exclusion restriction so e.g., $R = [0\vdots I_q], r = 0$, and $\Psi(\tau) = 0$. In such cases it is also shown that the nuisance parameters $\varphi_0(\tau)$ and Ω can be replaced by consistent estimates without jeopardizing the distribution free character of the tests.

To formalize the foregoing discussion we introduce the following conditions, which closely resemble the conditions employed in Koenker and Machado. We will assume that the $\{y_i\}$'s are conditional on x_i, independent with linear conditional quantile functions given by (3.1) and local, in a sense specified in A.3, to the location-scale shift model (3.2).

A. 1. *The distribution function F_0, in (3.2) has a continuous Lesbesgue density, f_0, with $f_0(u) > 0$ on $\{u : 0 < F_0(u) < 1\}$.*

A. 2. *The sequence of design matrices $\{X_n\} = \{(x_i)_{i=1}^n\}$ satisfy:*

- **(i):** $x_{i1} \equiv 1 \qquad i = 1, 2, \ldots$
- **(ii):** $J_n = n^{-1} X_n^\top X_n \to J_0$, *a positive definite matrix.*
- **(iii):** $H_n = n^{-1} X_n^\top \Gamma_n^{-1} X_n \to H_0$, *a positive definite matrix where $\Gamma_n = diag(\gamma^\top x_i)$.*
- **(iv):** $\max_{i=1,\ldots,n} \| x_i \| = \mathcal{O}(n^{1/4} \log n)$

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

$$R\beta_n(\tau) - r - \Psi(\tau) = \zeta(\tau)/\sqrt{n}.$$

As noted in Koenker and Machado (1999), conditions A.1 and A.2 are quite standard in the quantile regression literature. Somewhat weaker conditions are employed by Gutenbrunner, Jurečková, Koenker, and Portnoy (1993) in an effort to extend the theory further into the tails. But this isn't required for our present purposes, so we have reverted to conditions closer to those of Gutenbrunner and Jurečková (1992). Condition A.3 enables us to explore local asymptotic power of the proposed tests employing a rather general form for the local alternatives.

We can now state our first result. Proofs of all results appear in the appendix.

Theorem 1. *Let $\mathcal{T}$ denote the closed interval $[\varepsilon, 1-\varepsilon]$, for some $\varepsilon \in (0, 1/2)$. Under conditions A.1-3*

$$v_n(\tau) \Rightarrow v_0(\tau) + \eta(\tau) \text{ for } \tau \in \mathcal{T}$$

where $v_0(\tau)$ denotes a q-variate standard Brownian bridge process and

$$\eta(\tau) = \varphi_0(\tau)(R\Omega R^\top)^{-1/2}\zeta(\tau).$$

Under the null hypothesis, $\zeta(\tau) = 0$, the test statistic

$$\sup_{\tau \in \mathcal{T}} \| v_n(\tau) \| \Rightarrow \sup_{\tau \in \mathcal{T}} \| v_0(\tau) \| .$$

Typically, even if the hypothesis is fully specified, it is necessary to estimate the matrix Ω and the function $\varphi_0(t) \equiv f_0(F_0^{-1}(t))$. Fortunately, these quantities can be replaced by estimates satisfying the following condition.

A. 4. *There exist estimators $\varphi_n(\tau)$ and Ω_n satisfying*

i.: $\sup_{\tau \in \mathcal{T}} |\varphi_n(\tau) - \varphi_0(\tau)| = o_p(1)$,
ii.: $||\Omega_n - \Omega|| = o_p(1)$.

Corollary 1. *The conclusions of Theorem 1 remain valid if $\varphi_0(\tau)$ and Ω are replaced by estimates satisfying condition A.4.*

Theorem 1 extends slightly the results of Koenker and Machado (1999), but it fails to answer our main question: how to deal with unknown nuisance parameters in R and r? To begin to address this question, we introduce an additional condition.

A. 5. *There exist estimators R_n and r_n satisfying $\sqrt{n}(R_n - R) = \mathcal{O}_p(1)$ and $\sqrt{n}(r_n - r) = \mathcal{O}_p(1)$.*

And we now consider the parametric quantile regression process,

$$\hat{v}_n(\tau) = \sqrt{n}\varphi_0(\tau)[R_n \Omega R_n^\top]^{-1/2}(R_n \hat{\beta}(\tau) - r_n - \Psi(\tau)).$$

The next result establishes a representation for $\hat{v}_n(\tau)$ analogous that provided in (2.2) for the univariate empirical quantile process.

Theorem 2. *Under conditions A.1-5, we have*

$$\hat{v}_n(\tau) - Z_n^\top \xi(\tau) \Rightarrow v_0(\tau) + \eta(\tau)$$

where $\xi(\tau) = \varphi_0(\tau)(1, F_0^{-1}(\tau))^\top$, and $Z_n = \mathcal{O}_p(1)$, with $v_0(\tau)$ and $\eta(\tau)$ as specified in Theorem 1.

Corollary 2. *The conclusions of Theorem 2 remain valid if $\varphi_0(\tau)$ and Ω are replaced by estimates satisfying condition A.4.*

As in the univariate case we are faced with two options. We can accept the presence of the Z_n term, and abandon the asymptotically distribution free nature of tests based upon $\hat{v}_n(\tau)$. Or we can, following Khmaladze, try to find a transformation of $\hat{v}_n(\tau)$ that annihilates the Z_n contribution, and thus restores the asymptotically distribution free nature of inference. We adopt the latter approach.

Let $g(t) = (t, \xi(t)^\top)^\top$ so $\dot{g}(t) = (1, \psi(t), \psi(t)F^{-1}(t))^\top$ with $\psi(t) = (\dot{f}/f)(F^{-1}(t))$. We will assume that $g(t)$ satisfies the following condition.

A. 6. *The function $g(t)$ satisfies:*

i: $\int \| \dot{g}(t) \|^2 \, dt < \infty$,

ii: $\{\dot{g}_i(t) : i = 1, \dots, m\}$ *are linearly independent in a neighborhood of 1.*

We may note that Khmaladze (1981, §3.3) shows that A.6.ii implies $C^{-1}(\tau)$ exists for all $\tau < 1$.

We consider the transformed process $\tilde{v}_n(\tau)$ defined as,

$$\tilde{v}_n(\tau)^\top \equiv Q_g \hat{v}_n(\tau)^\top = \hat{v}_n(\tau)^\top - \int_0^\tau \dot{g}(s)^\top C^{-1}(s) \int_s^1 \dot{g}(r) d\hat{v}_n(r)^\top ds, \tag{3.6}$$

where the recursive least squares transformation should now be interpreted as operating coordinate by coordinate on the $\hat{v}_n$ process.

Theorem 3. *Under conditions A.1 - 6, we have*

$$\tilde{v}_n(\tau) \Rightarrow w_0(\tau) + \tilde{\eta}(\tau)$$

where $w_0(\tau)$ denotes a q-variate standard Brownian motion, and $\tilde{\eta}(\tau)^\top = Q_g \eta(\tau)^\top$. Under the null hypothesis, $\zeta(\tau) = 0$,

$$\sup_{\tau \in \mathcal{T}} \| \tilde{v}_n(\tau) \| \Rightarrow \sup_{\tau \in \mathcal{T}} \| w_0(\tau) \| .$$

Typically, in applications, the function $g(t)$ will *not* be specified under the null hypothesis, but will also need to be estimated. Fortunately, only one rather mild further condition is needed to enable us to replace g by an estimate.

A. 7. *There exists an estimator, $g_n(\tau)$, satisfying* $\sup_{\tau \in \mathcal{T}} \| \dot{g}_n(\tau) - \dot{g}(\tau) \| = o_p(1)$.

Corollary 3. *The conclusions of Theorem 3 remain valid if $\varphi_0(\tau), \Omega$, and g are replaced by estimates satisfying conditions A.4 and A.7.*

In some applications, R is known and only r contains nuisance parameters that need to be estimated. In this case, we have reduced dimensionality in $\xi(t)$, Z_n, $g(t)$, and $\dot{g}(t)$. In particular, the asymptotic result of Theorem 2 reduces to

$$\hat{v}_n(\tau) - \varphi_0(\tau)[R \Omega R^\top]^{-1/2} \sqrt{n}(r_n - r) \Rightarrow v_0(\tau) + \eta(\tau),$$

and the corresponding $\dot{g}(\cdot)$ (or $g(\cdot)$) functions in transformation (3.6) are

$$\dot{g}(t) = (1, \psi(t))^\top \text{ and } g(t) = (t, \varphi_0(t))^\top.$$

An important example of this case is the heteroskedasticity test described in Section 4.2.

The proposed tests have non-trivial power against any local deviation $\eta(\tau)$ such that $Q_g \eta(\tau) \neq 0$. Consider the following linear operator $\mathcal{K}$:

$$\mathcal{K}x(t) = \dot{g}(t)^\top C(t)^{-1} \int_t^1 \dot{g}(r) x(r) dr$$

and denote its eigen-space with eigenvalue 1, $\{x(t) : x(t) = \mathcal{K}x(t)\}$, as $\mathcal{S}$. Our tests will have power as long as $\eta \notin \mathcal{S}$. $\mathcal{K}$ is a Fredholm integral operator defined on the space $L_2[0,1]$, and we can write,

$$\mathcal{K}x(t) = \int_0^1 K(t,s)x(s)ds$$

with Volterra kernel

$$K(t,s) = \dot{g}(t)^\top C(t)^{-1}\dot{g}(s)\mathbf{1}_{[t,1]}(s).$$

The set $\mathcal{S}$ is the set of solutions to the Volterra equation,

$$x(t) = \int_0^1 K(t,s)x(s)ds,$$

and by Khmaladze (1981, p. 252), $\mathcal{S} = \{x : x(t) = \dot{g}(t)^\top \xi\}$. Thus, the proposed tests have non-trivial power against any local deviation process $\eta(\tau)$ such that $\dot{\eta}$ is not in the linear space spanned by the elements of $\dot{g}(\tau)$.

In the above analysis, for simplicity and without loss of generality we focus on statistics of the classical Kolmogorov-Smirnov form. The results of the paper apply more generally to statistics of the form $h(\tilde{v}_n(\tau))$ for continuous function h. Besides the "sup" function, one may use other measures of the discrepancy between $\tilde{v}_n(\tau)$ and 0, depending on the alternatives of interest. For instance, we can construct a Cramer-von Mises type test statistic $\int_{\mathcal{T}} \tilde{v}_n(\tau)^\top \tilde{v}_n(\tau)d\tau$, based on $\tilde{v}_n(\tau)$. Or, more generally, $\int_{\mathcal{T}} \tilde{v}_n(\tau)^\top W(\tau)\tilde{v}_n(\tau)d\tau$ for a suitably chosen weight matrix function $W(\tau)$.

The foregoing results provide some basic machinery for a broad class of tests based on the quantile regression process. In the next section we consider several special cases including tests of the location shift hypothesis, and tests for the location-scale shift hypothesis.

4. Implementation of the Tests

Given a framework for inference based on the quantile regression process, we can now –in a somewhat more pragmatic spirit– elaborate some missing details. We will begin by considering tests of the location scale shift hypothesis against a general quantile regression alternative. Tests of the location shift hypothesis and several variants of tests for heteroscedasticity will then be considered. Problems associated with estimation of nuisance parameters are treated in the final subsection.

4.1. **The location-scale shift hypothesis.** We would like to test

$$F^{-1}_{y_i|x_i}(\tau|x_i) = x_i^\top \alpha + x_i^\top \gamma F_0^{-1}(\tau)$$

against the sequence of linear quantile regression alternatives

$$F^{-1}_{y_i|x_i}(\tau|x_i) = x_i^\top \beta_n(\tau).$$

In the simplest case the univariate quantile function is known and we can formulate the hypothesis in the (3.4) notation,

$$R\beta(\tau) - r = \Psi(\tau)$$

by setting $r_i = \alpha_i/\gamma_i$, $R = \mathrm{diag}(\gamma_i^{-1})$, and $\Psi(\tau) = \iota_{p-1}F_0^{-1}(\tau)$. Obviously, there is some difficulty if there are γ_i equal to zero. In such cases, we can take $\gamma_i = 1$, and set the corresponding elements $r_i = \alpha_i$ and $\Psi_i(\tau) \equiv 0$. How should we go about estimating the parameters α and γ? Under the null hypothesis,

$$\beta_i(\tau) = \alpha_i + \gamma_i F_0^{-1}(\tau) \qquad i = 1, \dots, p$$

so it is natural to consider linear regression. Since $\hat{\beta}_i(\tau)$ is piecewise constant with jumps at points $\{\tau_1, \dots, \tau_J)$, it suffices to consider p bivariate linear regressions of $\hat{\beta}_i(\tau_j)$ on $\{(1, F_0^{-1}(\tau_j)) : \quad j = 1, \dots, J\}$. Each of these regressions has a known (asymptotic) Gaussian covariance structure that could be used to construct a weighted least squares estimator, but pragmatism might lead us to opt for the simpler unweighted estimator. In either case we have our required $\mathcal{O}(n^{-1/2})$ estimators $\hat{\alpha}_n$ and $\hat{\gamma}_n$.

When $F_0^{-1}(\tau)$ is (hypothetically) known the Khmaladzation process is relatively painless computationally. The function $\dot{g}(t) = (1, \psi_0(t), \psi_0(t)F_0^{-1}(t))^\top$ is known and the transformation (2.3) can be carried out by recursive least squares. Again, the discretization is based on the jumps $\{\tau_1, \dots, \tau_J\}$ of the piecewise constant $\hat{\beta}(\tau)$ process. Tests statistics based on the transformed process, $\tilde{v}_n(\tau)$, can then be easily computed. The simplest of these is probably the Kolmogorov-Smirnov sup-type statistic

$$K_n = \sup_{\tau \in \mathcal{T}} \parallel \tilde{v}_n(\tau) \parallel$$

where $\mathcal{T}$ is typically of the form $[\varepsilon, 1-\varepsilon]$ with $\varepsilon \in (0, 1/2)$. The choice of the norm $||\cdot||$ is also an issue. Euclidean norm is obviously natural, but has the possibly undesirable effect of accentuating extreme behavior in a few coordinates. Instead, we will employ the ℓ_1 norm in the simulations and the empirical application below.

When $F_0^{-1}(t)$ isn't assumed to be known under the null it is convenient to choose one coordinate, typically the intercept coefficient, to play the role of numeraire. From (3.4) we can write

$$\beta_i(\tau) = \mu_i + \sigma_i \beta_1(\tau) \qquad i = 2, \dots, p \tag{4.1}$$

where $\mu_i = \alpha_i - \alpha_1\gamma_i/\gamma_1$ and $\sigma_i = \gamma_i/\gamma_1$, or in matrix notation as

$$R\beta(\tau) = r$$

where $\Psi(\tau) \equiv 0$, $R = [\sigma \vdots - I_{p-1}]$ and $r = -\mu$. Estimates of the vectors μ and σ are again obtainable by regression of $\hat{\beta}_i(\tau) \quad i = 2, \dots, p$ on the intercept coordinate $\hat{\beta}_1(\tau)$.

Finally, we must face the problem of estimating the function $\dot{g}$. Fortunately, there is already a large literature on estimation of score functions. For our purposes it is convenient to employ the adaptive kernel method described in Portnoy and Koenker (1989). An attractive alternative to this approach has been developed by Cox (1985) and Ng (1994) based on smoothing spline methods. Given a uniformly consistent estimator, $\dot{g}_n$, satisfying condition A.7, see Portnoy and Koenker (1989, Lemma 3.2), Corollary 3 implies that under the null hypothesis

$$\tilde{v}_n(t) \equiv Q_{g_n} \hat{v}_n(t) \Rightarrow w_0(t)$$

and therefore tests can be based on K_n as before. Note that in this case estimation of $\dot{g}$ provides as a byproduct an estimator of the function $\varphi_0(t)$ which is needed to compute the process $\hat{v}_n(t)$.

In applications it will usually be desirable to restrict attention to a closed interval $[\tau_0, \tau_1] \subset (0,1)$. This is easily accommodated, following Koul and Stute (1999), Remark 2.3, by considering the modified test statistic,

$$K_n = \sup_{\tau \in [\tau_0, \tau_1]} ||\tilde{v}(\tau) - \tilde{v}(\tau_0)|| / \sqrt{\tau_1 - \tau_0},$$

which converges weakly, just as in the unrestricted case, to $\sup_{[0,1]} ||w_0(\tau)||$. Indeed, it may be fruitful to consider other forms of standardization as well, and this is the subject of continuing research.

4.2. **The location shift hypothesis.** An important special case of the location-scale shift model is, of course, the pure location shift model,

$$F^{-1}_{y_i|x_i}(\tau|x_i) = x_i^\top \alpha + \gamma F_0^{-1}(\tau)$$

This is just the classical homoscedastic linear regression model,

$$y_i = x_i^\top \alpha + \gamma u_i$$

where the $\{u_i\}$ are iid with distribution function F_0. This model underlies much of classical econometric theory and practice. If it is found to be appropriate then it is obviously sensible to consider estimation by alternative methods. For F_0 Gaussian, least squares would of course be optimal. For F_0 unknown one might consider the Huber M-estimator, or its L-estimator counterpart,

$$\hat{\beta}_\alpha = (1 - 2\alpha)^{-1} \int_\alpha^{1-\alpha} \hat{\beta}(\tau) d\tau,$$

see Koenker and Portnoy (1987). In the location shift model it is also well-known from Bickel (1982), that the slope parameters, $(\beta_2, ..., \beta_p)$, are adaptively estimable provided F_0 has finite Fisher information for the location parameter. Thus, it would be reasonable to consider M-estimators like those described in Hsieh and Manski (1987) or the adaptive L-estimators described in Portnoy and Koenker (1989).

The location-shift hypothesis can be expressed in standard form,

$$R\beta(\tau) = r,$$

by setting $R = [0 \vdots I_{p-1}]$, $r = (\alpha_2, \dots, \alpha_p)^\top$. It asserts simply that the quantile regression slopes are constant, independent of τ. Again, the unknown parameters in $\{R, r\}$ are easily estimated so the process $\hat{v}_n(\tau)$ is easily constructed. The transformation is obviously somewhat simpler in this case since $g(t) = (t, \varphi_0(t))$ has one fewer coordinate than in the previous case.

We can continue to view tests of the location-shift hypothesis as tests against a general quantile regression alternative represented in (A.3), or we can also consider the behavior of the tests against a more specialized class of location-scale shift alternatives for which

$$\zeta(\tau) = \zeta_0 F_0^{-1}(\tau)$$

for some fixed vector $\zeta_0 \in \mathbb{R}^{p-1}$. In the latter setting we have a test for parametric heteroscedasticity and we can compare the performance of our very general class of tests against alternative tests designed to be more narrowly focused on heteroscedastic alternatives. We will explore this in Section 4.3 below.

An optimal (invariant) test in the parametric setting may be based on optimal L-estimator of scale with weight function,

$$\omega(\tau) = \frac{d}{dx}(x\dot{f}/f)|_{x=F_0^{-1}(\tau)}, \tag{4.2}$$

see e.g. Serfling (1980). Thus, for example, in the normal (Gaussian) model, $F_0 = \Phi$, we would have, $\omega(\tau) = \Phi^{-1}(\tau)$, so our estimator of ζ_0 would be,

$$\hat{\zeta}_n = \int_0^1 \Phi^{-1}(\tau)\hat{\beta}(\tau)d\tau,$$

and a test for heteroscedasticity could be based on the last $p-1$ coordinates of $\hat{\zeta}_n$. One way to interpret such tests is to view them as smoothly weighted linear combinations of the interquantile range tests for heteroscedasticity introduced in Koenker and Bassett (1982). Clearly, in the case of the Gaussian weight function, extreme interquantile ranges get considerable weight, so it may be prudent to consider Huberized versions of these tests that trim the influence of the tails. Alternatively, one could consider weight functions explicitly designed for more heavy tailed distributions like the Cauchy,

$$\omega(\tau) = 2\sin(2\pi\tau)(\cos(2\pi\tau) - 1).$$

4.3. **Local Asymptotic Power Comparison.** In this section we compare the heteroscedasticity tests proposed above in an effort to evaluate the cost of considering a much more general class of semiparametric alternatives instead of the strictly parametric alternatives represented by the location scale shift model.

We consider the local alternative with $\zeta(\tau) = \zeta_0 F_0^{-1}(\tau)$ in the location-shift model, and denote this hypothesis as H_n as in assumption A.3. This corresponds to the linear model with asymptotically vanishing heteroskedasticity studied by Koenker and Bassett (1982). In the location-shift model, $R = [0 \vdots I_{p-1}]$, and thus $R\Omega R^\top = \Omega_x$, where Ω_x is the lower $(p-1) \times (p-1)$ corner of Ω. Under H_n,

$$\begin{aligned}\hat{v}(\tau) &= \varphi_0(\tau)\Omega_x^{-1/2}\sqrt{n}\left(R\hat{\beta}(\tau) - r_n\right) \\ &= v_0(\tau) - \varphi_0(\tau)\Omega_x^{-1/2}\sqrt{n}\,(r_n - r) + \zeta_0\Omega_x^{-1/2}\varphi_0(\tau)F_0^{-1}(\tau) + o_p(1)\end{aligned}$$

and the transformed process is

$$\tilde{v}(\tau) = w_0(\tau) + \tilde{\zeta}(\tau)\zeta_0\Omega_x^{-1/2} + o_p(1)$$

where the noncentrality process $\tilde{\zeta}(\tau)$ is simply the martingale transformation of $\zeta(\tau) = \varphi_0(\tau)F_0^{-1}(\tau)$. Provided that $\dot{\zeta}(\tau)$ is not in the space spanned by the functions $\{1, \dot{\varphi}_0(\tau)\}$, it is clear that $\tilde{\zeta}(\tau) \neq 0$, so the proposed tests have non-trivial power. Figure 1 depicts the noncentrality function $\zeta(\tau)$ its "predicted" version, and its (residual) transformed version $\tilde{\zeta}(\tau)$ for the normal case. For any given ζ_0, the asymptotic local power of the proposed Kolmogorov-Smirnov test is given as follows

$$P(c) = \Pr\left\{\sup_{\tau\in\mathcal{T}}\left\|w_0(\tau) + \tilde{\zeta}(\tau)\zeta_0\Omega_x^{-1/2}\right\| > c_\alpha\right\},$$

where c_α is the asymptotic critical value at significance level α.

Example. It is of obvious interest to consider alternatives for which the function $\zeta_0(t) = \varphi_0(t)F_0^{-1}(t)$ is annihilated by the martingale transformation. This entails that $\dot{\zeta}_0(t)$ lie in the linear span of $\{1, \dot{\varphi}_0(t)\}$, and is evidently satisfied for F_0 uniform. More surprisingly, it is also satisfied for triangular densities like, $f_0 = (1-|x|)I(|x| \leq 1)$. These cases may be regarded as somewhat pathological as they have infinite Fisher information for both location and scale. Thus, it may not be regarded as too disturbing that the transformation performs poorly under such circumstances.

For the alternative hypothesis H_n, optimal invariant tests can be constructed based on an optimal L-estimator of scale

$$\hat{\zeta}_n = \int_0^1 \omega(\tau)\hat{\beta}_*(\tau)d\tau,$$

where $\hat{\beta}_*(\tau) = (\hat{\beta}_2(\tau), ..., \hat{\beta}_p(\tau))^\top$, and $\omega(\tau)$ is given by (4.2). Under H_0, we have $\sqrt{n}\hat{\zeta}_n \Rightarrow N(0, V)$, where $V = \sigma^2(\omega, F)\Omega_x$, and

$$\sigma^2(\omega, F_0) = \int_0^1\int_0^1 \frac{(s\wedge t - st)}{\varphi(s)\varphi(t)}\omega(t)\omega(s)dtds.$$

A Wald type optimal invariant test can then be based on:

$$W_n = n\hat{\zeta}_n^\top V^{-1}\hat{\zeta}_n, \tag{4.3}$$

which converges to a central χ^2 random variable with $p-1$ degrees of freedom under H_0. Under H_n, we have $\sqrt{n}\hat{\zeta}_n \Rightarrow N(\mu(\omega, F_0)\zeta_0, V)$, where $\mu(\omega, F) = \int_0^1 \omega(\tau)F_0^{-1}(\tau)d\tau$, and thus

$$W_n \Rightarrow \chi^2_{p-1}(\delta),$$

where $\chi^2_{p-1}(\delta)$ is a non-central χ^2 random variable with $p-1$ degree of freedom and non-centrality parameter,

$$\delta = \mu^2(\omega, F_0)\zeta_0^\top V^{-1}\zeta_0 = \frac{\mu^2(\omega, F_0)}{\sigma^2(\omega, F_0)}\zeta_0^\top \Omega_x^{-1}\zeta_0.$$

The asymptotic variance $\sigma^2(\omega, F_0)$ parameter has the equivalent representation

$$\sigma^2(\omega, F) = \int\int [F_0(x \wedge y) - F(x)F(y)]\,\omega(F(x))\omega(F(x))dxdy$$

and may be estimated in various ways; see, *e.g.*, Koenker and Portnoy (1987) for further details.

FIGURE 4.1. Noncentrality Functions

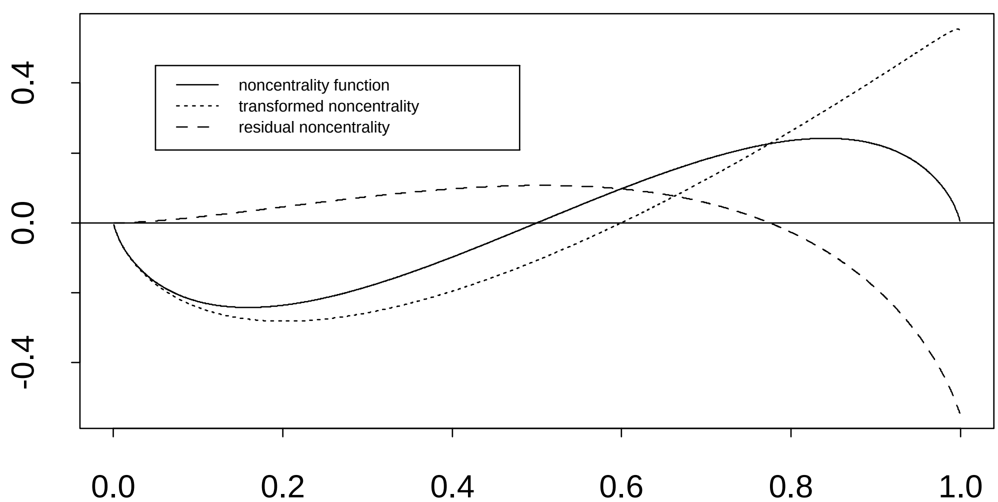
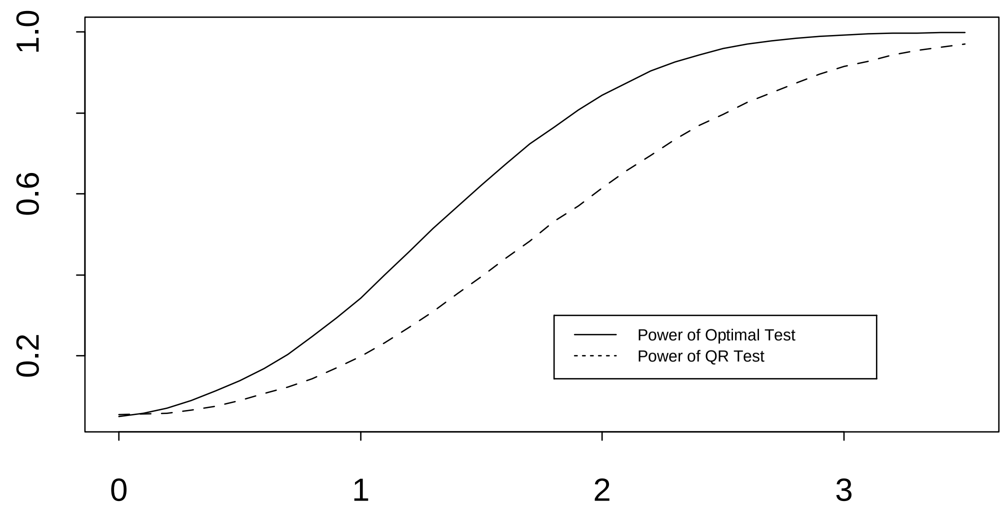


Figure 4.2. Asymptotic Power Comparison



We may now compare the proposed Kolmogorov-Smirnov type test, which was designed as a test against a general quantile regression alternative, with the above Wald test (4.3), which is an optimal invariant test against the parametric alternative H_n. For convenience of comparison, we consider the simple case where ζ_0 is a scalar (i.e. $p = 2$) so that the asymptotic power functions of these tests can be written as functions of ζ_0 and compared graphically. Denoting the α-level critical values of the limiting null distributions as c_α^1, and c_α^2 respectively, the power functions of the optimal (Wald) test and the proposed martingalized test are given by $\Pr\{\chi^2[\delta(\zeta_0)] > c_\alpha^1\}$, and $\Pr\{\sup_\tau |w_0(\tau) + \phi(\tau)\sigma_x\zeta_0| > c_\alpha^2\}$, respectively, where $\delta(\zeta_0) = \mu^2(\omega, F_0)\sigma_x^2\zeta_0^2/\sigma^2(\omega, F_0)$, and σ_x^2 is the variance of x_{i2}. Given the choice of $\mathcal{T}$, F_0, σ_x, and the significance level α, the power function can be calculated. Figure 4.2 depicts the asymptotic power functions of these tests for the Gaussian case and $\sigma_x^2 = 1$, $\mathcal{T} = [0.1, 0.9]$, at 5% level of significance.

4.4. **Generalizations.** The proposed method can be easily modified to accommodate many other inference problems, including tests for parametric conditional distributions and tests for conditional symmetry. The theory we develop also carries over with minor changes to some analytic nonlinear restrictions in place of (3.4) and associated changes in the transformation (3.6). For example, we may modify our hypothesis to the form

$$R\beta(\tau) - r = \Psi(\tau, \theta). \tag{4.4}$$

If θ were known, the previous tests would apply. More generally, if θ is unknown but may be estimated by a $\sqrt{n}$ consistent estimator θ_n, then, under regularity conditions (so that Ψ admits a Taylor expansion to the second order), the standardized empirical process $R_n\hat{\beta}(\tau) - r_n - \Psi(\tau, \theta_n)$ has the following approximation

$$\begin{aligned} \hat{v}_n(\tau) &= \sqrt{n}\varphi_0(\tau)\left[R_n \Omega R_n^\top\right]^{-1/2}\left(R_n\hat{\beta}(\tau) - r_n - \Psi(\tau, \theta_n)\right) \\ &= v_n(\tau) + \varphi_0(\tau) Z_1 + \varphi_0(\tau) F_0^{-1}(\tau) Z_2 + \varphi_0(\tau) s(\tau)^\top Z_3 + o_p(1), \end{aligned}$$

where $s(\tau) = \partial\Psi(\tau, \theta_0)/\partial\theta$, and $Z_j = O_p(1)$, $j = 1, 2, 3$. Denoting the set of linearly independent components of functions

$$(\tau, \varphi_0(\tau), \varphi_0(\tau)F_0^{-1}(\tau), \varphi_0(\tau)s(\tau)^\top)$$

as $g(\tau)$ and assuming that $\dot{g}(\tau)$ satisfies Assumption A.6, then the transformed process $\tilde{v}_n(\tau)$ defined by (3.6) can be constructed and similar results to Theorem 3 can be obtained.

The above process can be applied to testing the hypothesis that the conditional distribution of y_i, conditional on x_i, has a continuous cdf $F(\cdot, \theta)$, which depends on some parameters θ. This hypothesis can be expressed in the form of (4.4) by setting $R = (1, 0, ..., 0)$, $r = \alpha_1$, and $\Psi(\cdot, \theta_0) = F^{-1}(\cdot, \theta_0)$. In this case, $g(\tau)$ corresponds to the set of linearly independent components of functions $(\tau, \varphi_0(\tau), \varphi_0(\tau)s(\tau)^\top)$. In the case of normal hypothesis, the distribution is fully determined by the location and scale parameters which can be accomodated into R and r by re-standardization, reducing to the hypothesis to the simple formulation (3.4).

4.5. **Estimation of Nuisance Parameters.** Our proposed tests depend crucially on estimates of the quantile density and quantile score functions: $\varphi_0(\tau)$. Fortunately, there is a large related literature on estimating $\varphi_0(\tau)$, including e.g. Siddiqui (1960), Bofinger (1975), Sheather and Maritz (1983), and Welsh (1988). Following Siddiqui, since, $dF_0^{-1}(t)/dt = (\varphi_0(t))^{-1}$, it is natural to use the estimator,

$$\varphi_n(t) = \frac{2h_n}{F_n^{-1}(t + h_n) - F_n^{-1}(t - h_n)}, \tag{4.5}$$

where $F_n^{-1}(s)$ is an estimate of $F_0^{-1}(s)$ and h_n is a bandwidth which tends to zero as $n \to \infty$.

One way of estimating $F_0^{-1}(s)$ is to use a variant of the empirical quantile function for the linear model proposed in Bassett and Koenker (1982),

$$F_n^{-1}(s) = \frac{\hat{\alpha}(s) - \hat{\alpha}}{\hat{\sigma}}. \tag{4.6}$$

If we use (4.6) in the formula (4.5), the density $\varphi_0(t)$ can be estimated by

$$\varphi_n(t) = \frac{2h_n\hat{\sigma}}{\hat{\alpha}(t+h_n) - \hat{\alpha}(t-h_n)}. \tag{4.7}$$

5. Monte Carlo Results

We have conducted some limited Monte Carlo experiments to examine the finite sample performance of the proposed tests. In particular, we examine the effectiveness of the martingale transformation based on the size and power properties of the tests. The following sample sizes were considered in our experiment: $n = 100, 200, 300, 400, 500$. These sample sizes were chosen because they represent the most relevant range of sample sizes in empirical analyses.

First of all, to investigate the effectiveness of the martingale transformation on quantile regression inference, we examine the size and power properties of the infeasible version tests where the true density and score functions are used in the standardization and the martingale transformation. We start with the heteroscedasticity test. The data were generated from

$$y_i = \alpha + \beta x_i + \sigma(x_i)u_i, \tag{5.1}$$

where x_i and u_i are iid $\mathcal{N}(0,1)$ random variates and are mutually independent, $\alpha = 0$, and $\beta = 1$. $\sigma(x_i) = \gamma_0 + \gamma_1 x_i$, $\gamma_0 = 1$. We examined the empirical rejection rates of the test for different choices of sample sizes and γ_1 values, at 5% level of significance. In constructing the test, we used the OLS estimator for $\widehat{\beta}$, and the truncation parameter value $\delta = 0.05$ (i.e. $\mathcal{T} = [0.05, 0.95]$). Since x_i is a scalar, the limiting null distribution of the test statistic is $\sup_{0.05 \le \tau \le 0.95} |W(\tau)|$. The 5% level critical value is 2.14. For the choices of the heteroscedasticity parameter γ_1, we consider $\gamma_1 = 0, 0.1, 0.2, 0.3, 0.5, 1, 2, 5$. When $\gamma_1 = 0$, the model is homoscedastic and the rejection rates give the empirical sizes. When $\gamma_1 \neq 0$, the model is heteroscedastic and the rejection rates deliver the empirical powers. Table 1 reports the empirical rejection rates for different values of γ_1 and n. Other values of the truncation parameter δ were also tried and quantitatively similar results were obtained. These Monte Carlo results indicate that, given information on the density and score, the martingale transformation brings pretty good size and power to the proposed testing procedure in finite sample.

The remaining Monte-Carlo experiments are based on the even simpler two sample model,

$$\begin{cases} y_{1i} = \alpha_1 + \sigma_1 u_i, \ i = 1, \ldots\ldots, n_1, \\ y_{2i} = \alpha_2 + \sigma_2 v_i, \ i = 1, \ldots\ldots, n_2, \end{cases} \tag{5.2}$$

In particular, we considered the following two sets of parameter values

$$\text{Location Shift:} \quad \alpha_1 = 1, \alpha_2 = 0, \sigma_1 = \sigma_2 = 1, \tag{5.3}$$

$$\text{Location-Scale Shift:} \quad \alpha_1 = 1, \alpha_2 = 0, \sigma_1 = 2, \sigma_2 = 1, \tag{5.4}$$

where u_i, v_i are iid $\mathcal{N}(0,1)$ random variates. When the parameters take the first set of values, (5.2) represents a pure location shift model. The null hypothesis of a shift model can be tested by the procedure given in Section 4.2. When the data is generated from the second set parameters, (5.2) is a location-scale shift model. The location-scale hypothesis can be tested by the procedure given in Section 4.1. Table 2 reports the empirical size of these tests for different combinations of n_1 and n_2. We can see that the test has good size properties in finite samples. These Monte Carlo results, together with the results on the heteroscedasticity test in Table 1, confirm the effectiveness of the martingale transformation in quantile regression inference.

The above Monte Carlo experiments use the true density and score. It is obviously important to evaluate the effect of nonparametric nuisance parameter estimation on the performance of the proposed tests. In our next Monte Carlo experiments, we estimated $F^{-1}(s)$ using the empirical quantile function approach given by formula (4.6). For the density function, we use procedure (4.7) as an estimator of $\varphi_0(s)$. The quantile score process, and thereby the function g, is estimated by the adaptive kernel estimator of Portnoy and Koenker (1989).

The kernel estimation procedures for these nuisance functions are nonparametric and therefore obviously entail choices of bandwidth values. Unsuitable bandwidth selection can produce poor estimates. However, under broad conditions on the convergence rate of the bandwidth parameters, the nonparametric estimates are consistent and testing procedures using different bandwidth choices are (first order) asymptotically equivalent, although the finite sample performance of these tests can vary considerably with bandwidth choice. Extensive simulations have been conducted in the literature to show the importance of bandwidth choice on estimation and testing procedure that use nonparametric estimates.

It was anticipated that the estimation of $\varphi_0(s)$ would exert important influence on the finite sample performance of our tests. This is confirmed in the simulations. For this reason, we pay particular attention to the bandwidth choice in density estimation. Hall and Sheather (1988) suggested a bandwidth rule based on Edgeworth expansion for studentized quantiles. This bandwidth is of order $n^{-1/3}$ and we denote it as h_{HS}. Another bandwidth selection has been proposed by Bofinger (1975) is of order $n^{-1/5}$. We denote it by h_B. We have considered both of these bandwidth choices for our tests. In addition, notice that the Bofinger bandwidth is eventually much larger than the Hall and Sheather bandwidth, we have also considered the following bandwidth choice which takes values between h_{HS} and h_B, it is denoted as h_θ, $h_\theta = \theta h_B$, where h_B is the Bofinger bandwidth and θ is a scalar. We report the results for the case $\theta = 0.6$ here. The score function was estimated by the method of Portnoy and Koenker (1989) and we simply choose the Silverman (1986) bandwidth.

Tables 3a, 3b, 3c report the Monte Carlo results for the heteroscedasticity test with different bandwidth selections and Tables 4a, 4b, 4c give the result of the location-scale test. The Monte Carlo evidence indicates that the bandwidth choice does have an important influence on the finite sample performance of these tests. It also shows that, by choosing appropriate bandwidth, the proposed tests have reasonable size and power properties. In general, we found over-rejection when the Hall-Sheather bandwidth was used. For the other two bandwidth, h_θ and h_B, the relative performance depends on which test we consider. For the heteroscedasticity test, we found under-rejection when the Bofinger bandwidth was used. In this test, at least for the model and the nonparametric methods used here, the bandwidth choice h_θ provides pretty good finite sample performance. However, for the location-scale test, h_θ tends to over-reject and h_B seems to be a relatively better bandwidth choice. To focus our attention on the effect of $\varphi_n(s)$, we have also conducted Monte Carlo experiments where only the density function is estimated (and use the true score function), the Monte Carlo results reconfirmed our findings on the three bandwidth choices.

TABLE 1: Size and Power of the Heteroskedasticity Test (Truncated, $\delta = 0.05$)

	Size	Power						
n	$\gamma_1 = 0$	$\gamma_1 = 0.1$	$\gamma_1 = 0.2$	$\gamma_1 = 0.3$	$\gamma_1 = 0.5$	$\gamma_1 = 1$	$\gamma_1 = 2$	$\gamma_1 = 5$
100	0.006	0.134	0.377	0.729	0.974	0.981	0.990	0.999
200	0.054	0.269	0.77	0.977	0.999	1.000	1.000	1.000
300	0.052	0.383	0.931	1.000	1.000	1.000	1.000	1.000
400	0.052	0.549	0.989	1.000	1.000	1.000	1.000	1.000
500	0.052	0.616	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 2: Application to The Two-Sample Models

Case 1: Location Shift						Case 2: Location-Scale Shift					
$\alpha_1 = 1, \alpha_2 = 0, \sigma_1 = \sigma_2 = 1$						$\alpha_1 = 1, \alpha_2 = 0, \sigma_1 = 2, \sigma_2 = 1$					
n_1	n_2	size	n_1	n_2	size	n_1	n_2	size	n_1	n_2	size
100	100	0.074	100	200	0.060	100	100	0.153	100	200	0.179
150	150	0.080	100	300	0.086	150	150	0.158	100	300	0.196
200	200	0.064	150	300	0.055	200	200	0.169	150	300	0.175
250	250	0.054	200	300	0.056	250	250	0.172	200	300	0.183

TABLE 3a

(The Heteroskedasticity Test. Bandwidth in Density Estimation: h_{HS};
Kernel Estimation of Score with Silverman Bandwidth)

	Size	Power		
n	$\gamma_1 = 0$	$\gamma_1 = 0.2$	$\gamma_1 = 0.5$	$\gamma_1 = 1$
100	0.45	0.723	0.99	1.000
200	0.21	0.877	1.000	1.000
300	0.195	0.952	1.000	1.000
400	0.186	0.995	1.000	1.000
500	0.173	1.000	1.000	1.000

TABLE 3b
(The Heteroskedasticity Test. Bandwidth in Density Estimation: h_B;
Kernel Estimation of Score with Silverman Bandwidth)

	Size		Power	
n	$\gamma_1 = 0$	$\gamma_1 = 0.2$	$\gamma_1 = 0.5$	$\gamma_1 = 1$
100	0.009	0.053	0.197	0.545
200	0.013	0.109	0.772	0.949
300	0.019	0.229	0.985	0.992
400	0.023	0.412	0.997	0.998
500	0.029	0.565	1.000	1.000

TABLE 3c
(The Heteroskedasticity Test. Bandwidth in Density Estimation: h_θ;
Kernel Estimation of Score with Silverman Bandwidth)

	Size		Power	
n	$\gamma_1 = 0$	$\gamma_1 = 0.2$	$\gamma_1 = 0.5$	$\gamma_1 = 1$
100	0.035	0.211	0.755	0.820
200	0.041	0.406	0.990	0.989
300	0.043	0.665	1.000	1.000
400	0.043	0.809	1.000	1.000
500	0.045	0.911	1.000	1.000

TABLE 4a
(Location-Scale Test. Bandwidth in Density Estimation: h_{HS};
Kernel Estimation of Score with Silverman Bandwidth)

n_1	n_2	size	n_1	n_2	size
100	100	0.589	50	50	0.616
150	150	0.538	75	75	0.603
200	200	0.511	250	250	0.507
500	500	0.406	300	300	0.456

TABLE 4b
(Location-Scale Test. Bandwidth in Density Estimation: h_B;
Kernel Estimation of Score with Silverman Bandwidth)

n_1	n_2	size	n_1	n_2	size
100	100	0.037	50	50	0.028
150	150	0.079	75	75	0.033
200	200	0.079	250	250	0.065
500	500	0.105	300	300	0.078

TABLE 4c
(Location-Scale Test. Bandwidth in Density Estimation: h_θ;
Kernel Estimation of Score with Silverman Bandwidth)

n_1	n_2	size	n_1	n_2	size
100	100	0.097	50	50	0.063
150	150	0.112	75	75	0.086
200	200	0.123	250	250	0.126
500	500	0.145	300	300	0.135

6. A Reappraisal of the Pennsylvania Reemployment Bonus Experiments

A common concern about unemployment insurance (UI) systems has been the suggestion that the insurance benefit acts as a disincentive for job-seekers and thus prolongs the duration of unemployment spells. During the 1980's several controlled experiments were conducted in the U.S. to test the incentive effects alternative compensation schemes for UI. In these experiments, UI claimants were offered a cash bonus if they found a job within some specified period of time and if the job was retained for a specified duration. The question addressed by the experiments was: would the promise of such a monetary lump-sum benefit provide a significant inducement for more intensive job-seeking and thus reduce the duration of unemployment?

In the first experiments conducted in Illinois a random sample of new UI claimants were told that they would receive a bonus of $500 if they found full-time employment within 11 weeks after filing their initial claim, and if they retained their new job for at least 4 months. These "treatment claimants" were then compared with a control group of claimants who followed the usual rules of the Illinois UI system. The Illinois experiment provided very encouraging initial indication of the incentive effects of such policies. They showed that bonus offers resulted in a significant reduction in the duration of unemployment spells and consequent reduction of the regular amounts paid by the state to UI beneficiaries. This finding led to further "bonus experiments" in the states of New Jersey, Pennsylvania and Washington with a variety of new treatment options. An excellent review of the experiments, some general conclusions about their efficacy and a critique of their policy relevance can be found in Meyer (1995), and Meyer (1996). In this section we will focus more narrowly on a reanalysis of data from the Pennsylvania Reemployment Bonus Demonstration described in detail in Corson, Decker, Dunstan, and Keransky (1992).

The Pennsylvania experiments were conducted by the U.S. Department of Labor between July 1988 and October 1989. During the enrollment period, claimants who became unemployed and registered for unemployment benefits in one of the 12 selected local offices throughout the state were *randomly assigned* either to a control group or one of six experimental treatment groups. In the control group the existing rules of the unemployment insurance system applied. Individuals in the treatment groups were offered a cash bonus if they became reemployed in a full-time job, working more than 32 hours per week, within a specified qualification period. Two bonus levels and two qualification periods were tested, but we will restrict attention to the high bonus, long qualification period treatment which offered a cash of bonus of six times the weekly benefit for claimants establishing reemployment within 12 weeks. A detailed description of the characteristics of claimants under study is presented in Koenker and Bilias (1999) which has information on age, race, gender, number of dependents, location in the state, existence of recall expectations, and type of occupation.

Since a large portion of spells end in either the first week or the twenty seventh week, it should be stressed that the definition of the first spell of UI in the Pennsylvania study includes a waiting week and that the maximum number of uninterrupted full weekly benefits is 26. This implies that many subjects did not receive any weekly benefit and that many other claimants received continuously their full, entitled unemployment benefit. Again, Koenker and Bilias (1999) contains further details.

6.1. **The Model.** Our basic model for analyzing the Pennsylvania experiment presumes that the logarithm of the duration (in weeks) of subjects' spells of UI benefits have linear conditional quantile functions of the form

$$Q_{\log(T)}(\tau|x) = x^\top \beta(\tau).$$

The choice of the log transformation is dictated primary by the desire to achieve linearity of the parametric specification and by its ease of interpretation. Multiplicative covariate effects are widely employed throughout survival analysis, and they are certainly more plausible in the present application than the assumption of additive effects. It is perhaps worth reiterating that the role of the transformation is completely transparent in the quantile regression setting, where $Q_{h(T)}(\tau|x) = x^\top \beta(\tau)$ implies $Q_T(\tau|x) = h^{-1}(x^\top \beta(\tau))$. In contrast, the role of transformations in models of the conditional mean function are rather complicated since the transformation affects not only location, but scale and shape of the conditional distribution of the response. Our (provisional) model includes the following effects:

- Indicator for the treatment group.
- Indicators for female, black and hispanic respondents.

- Number of dependents, with 2 indicating two or more dependents.
- Indicators for the 5 quarters of entry to the experiment.
- Indicator for whether the claimant "expected to be recalled" to a previous job.
- Indicators for whether the respondent was "young" – less than 35, or "old" – indicating age greater than 54.
- Indicator for whether claimant was employed in the durable goods industry.
- Indicator for whether the claimant was registered in one of the low unemployment short unemployment duration districts: Coatesville, Reading, or Lancaster.

In Figure 6.1 we present a concise visual representation of the results from the estimation of this model. Each of the panels of the Figure illustrate one coordinate of the vector-valued function, $\hat{\beta}(\tau)$, viewed as a function of $\tau \in [\tau_0, \tau_1]$. Here we choose $\tau_0 = .20$ and $\tau_1 = .80$, effectively neglecting the proportion of the sample that are immediately reemployed in week one and those whose unemployment spell exceeds that insured limit of 26 weeks. The lightly shaded region in each panel of the figure represents a 90 percent confidence band. We omit the plots for the 5 quarter of entry indicators, and for the for the low unemployment district variable to conserve space.

Before turning to interpretation of specific coefficients, we will try to offer some brief general remarks on how to interpret these figures. The simplest case is the pure location shift model for which we would have the classical accelerated failure time (AFT) model,

$$\log T_i = x_i^\top \beta + u_i$$

with $\{u_i\}$'s iid from some F. For F of the form $F(u) = 1 - \exp(-\exp(u))$, this is the well known Cox proportional hazard model with Weibull baseline hazard. In this case we would expect to see slope coefficients $\hat{\beta}_j(\tau)$ that oscillate around a constant value indicating that the shift in the response due to a change in the covariate is constant over the entire estimated range of the distribution. The conditional mean effect estimated by least squares is asymptotically equivalent in this case to integrating the estimated coefficients over the unit interval.

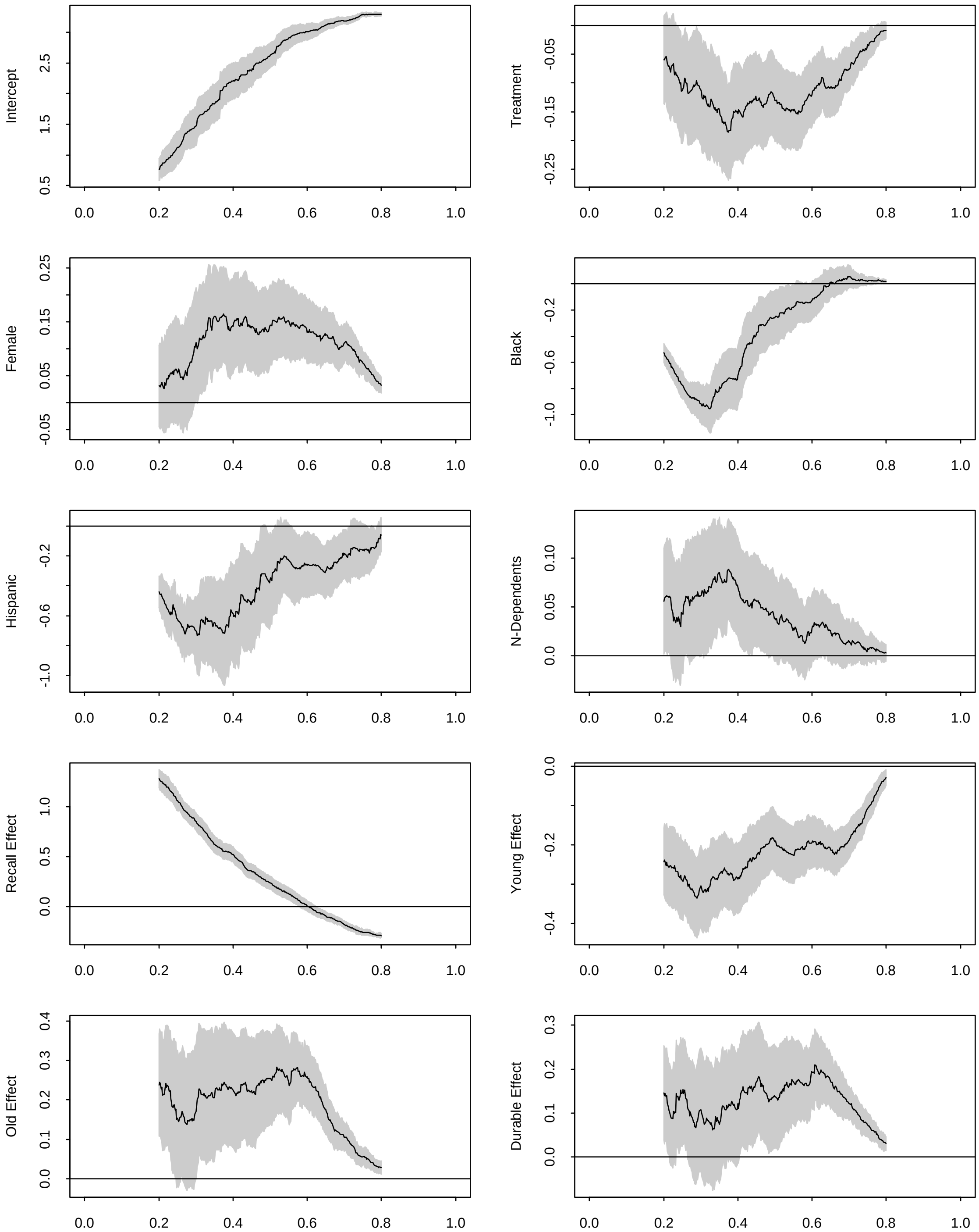
Another conventional model with linear quantile functions is the linear location-scale model,

$$\log T_i = x_i^\top \beta + (x_i^\top \gamma) u_i$$

where again, u_i is taken to be iid. Now the covariates are allowed to influence the scale as well as the location of the conditional distribution of durations. In this case the plots of the "slope" coefficients $\hat{\beta}_j(\tau)$ should look just like the "intercept" coefficient up to a location and scale shift. The intercept coefficient estimates a normalized version of the quantile function of the u_i's and all the other coefficients are simply location and scale shifts of this function.

6.2. **Interpretation of the Estimated Effects.** No treatment effect is observed in either tail implying that the treatment had no effect in changing the probability of

FIGURE 6.1. Quantile Regression Process for Log Duration Model



immediate reemployment (in week one), or in effecting the probability of durations beyond the 26 week maximum. The high bonus and long qualification period treatment, yielded roughly a 15% reduction in median duration. This effect is considerably stronger statistical significance than that seen in the other treatments.

The randomization of the experiment was quite effective in rendering the potentially confounding effects of other covariates orthogonal to the treatment indicator. Nevertheless, it is of some interest to explore the effect of other covariates in an effort to better understand determinants of the duration of unemployment.

Women are 5 to 15% slower than men to exit unemployment. Blacks and Hispanics appear much quicker than whites to become reemployed. This effect is particularly striking in the case of blacks for whom median duration is roughly half ($\approx e^{-.75}$) that of whites, and only 30% as long as controls at $\tau = .33$. The number of dependents appears to exert a rather weak positive effect on unemployment durations. The quarter-of-entry variables are inherently not very interesting, but it appears that late entry into the experiment improved one's chances for early reemployment. The recall indicator is considerably more interesting; anticipated recall to one's prior job has a very strong and very precisely estimated detrimental effect over the entire lower tail of the distribution. However, beyond quantile $\tau = .6$, which corresponds to about 20 weeks duration for white, male controls, the anticipated recall appears to be abandoned and beyond this point expected recall becomes a significant force for early reemployment in the upper tail of the distribution.

Not surprisingly the young (those under 34) tend to find reemployment earlier than their middle aged counterparts, while the old (those over 54) do significantly worse. In both cases the effects are highly significant throughout the entire range of quantiles we have estimated. Prior employment in durable manufacturing has a weakly disadvantageous effect on reemployment, but residing in a low unemployment district is, not surprisingly, helpful in facilitating more rapid reemployment.

The treatment effect of the bonus offer clearly does not conform well to the location shift paradigm of the conventional models. After the log transformation of durations, a location shift would imply that the treatment exerts a constant *percentage* change in all durations. In the present instance this implication is particularly unpalatable since the entire point of the experiment was to alter the shape of the conditional duration distribution, concentrating mass within the qualification period, and reducing it beyond this period. In the treatment panel of Figure 6.1 we have seen that the bonus effect gradually reduces durations from a null effect in the lower tail to a maximum reduction of 15% at the median, and then gradually again returning to a null effect in the upper tail. This finding accords perfectly with the timing imposed by the qualification period of the experiment. It might be thought that the bonus should not affect durations at all beyond the qualification period, but further consideration suggests that accelerated search in an effort to meet the qualification period deadline could easily yield "successes" that extended beyond the qualification period due to decision delay by potential employers, or other factors.

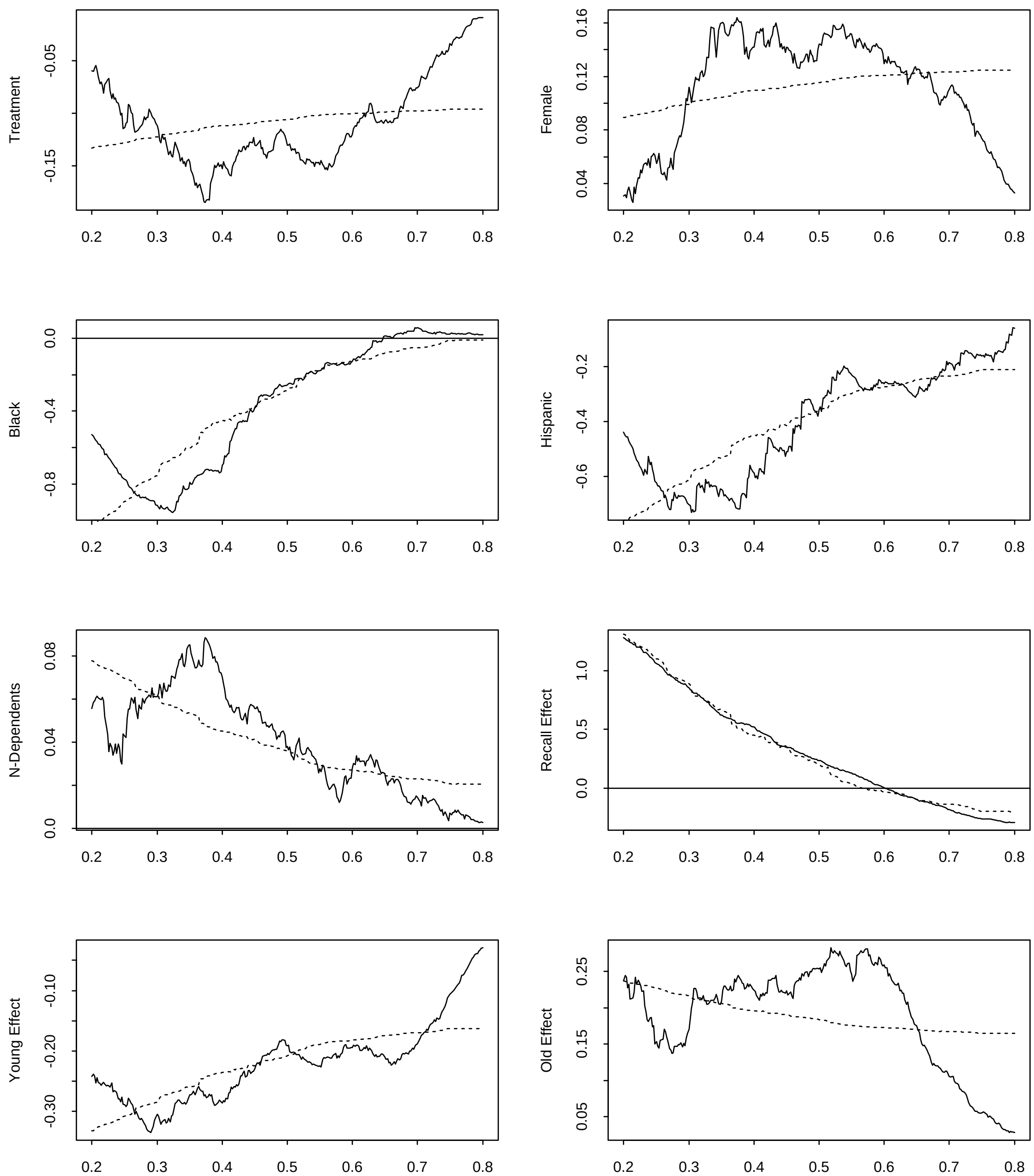
Taken together, the results presented in Figure 6.1 do not seem to lend much support to either the location shift, or to the location-scale shift, hypotheses of the conventional regression model. In the former case we would expect to see plots that appeared essentially constant in τ while in the latter, we expect to see plots that mimic the shape of the intercept plot. Neither of these expectations are fulfilled. However, as we have emphasized earlier, it is crucial to be able evaluate these impressions by more formal statistical methods, a task that is undertaken in the next subsection.

6.3. **Inference on the Quantile Regression Process.** To illustrate our proposed inference strategy we have decomposed the test of the location scale shift hypothesis based on the full model represented in Figure 6.1, into several intermediate steps. In each of these steps we present results for only a subset of eight selected covariate effects in an effort to conserve space, but all 15 covariate effects are handled in an identical fashion. In Figure 6.2 we present, for each of our selected covariates, the prediction of the process $\hat{\beta}_i(\tau)$ based on the regression onto the estimated "intercept process", $\hat{\beta}_1(\tau)$ as indicated by (4.1). Each of the fitted curves is based on least squares estimation using the 301 estimated points of the quantile regression process for each coordinate. The solid lines in these panels are the same as those appearing in the previous figure; the dotted lines represents the fitted curve. With the possible exception of the recall effect, none of these fits look very compelling, but at this stage we are already deeply mired in the Durbin problem and so it is difficult to judge the significance of departures from the fitted relationships.

Taking the residuals from the panels of Figure 6.2, and standardizing by the Cholesky decomposition of their (inverse) covariance matrix yields the parametric quantile regression process, $\hat{v}_n(\tau)$. It is misleading, of course, to associate the coordinates of this process with the original labeling of the coordinates of $\hat{\beta}(\tau)$, since the matrix transformation of the process mixes the coordinates thoroughly. Had we specified hypothetical values for the coefficients rather than estimating them for Figure 6.2, we could of course treat the resulting process as a vector of independent Brownian bridges under the null. However, the effect of the estimation is to distort the variability of the process, as we have seen in Section 3. At this point we estimate the function $\dot{g}$ and perform the martingale transformation on each slope coordinate. The transformation is applied on the restricted subinterval, $[\tau_0, \tau_1]$, as described at the end of Section 4.1, yielding the new process, $\tilde{v}_n(\tau)$, The transformed coordinates of this process are, under the null hypothesis, asymptotically, independent Brownian motions. We consider the test statistic,

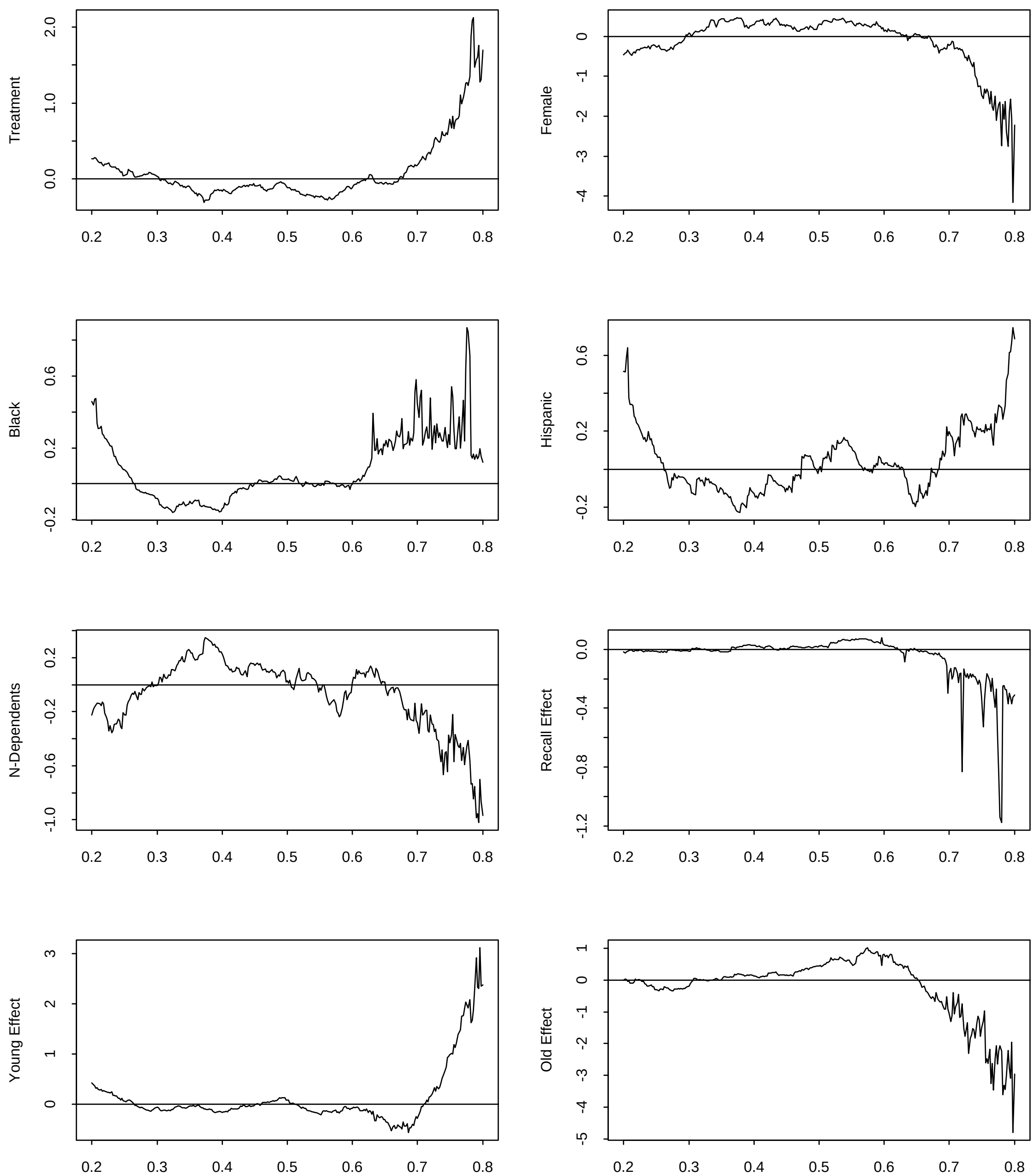
$$K_n = \sup_{\tau \in \mathcal{T}} ||\tilde{v}_n(\tau) - \tilde{v}_n(\tau_0)|| / \sqrt{\tau_1 - \tau_0}$$

Figure 6.2. Quantile Regression Process for Log Duration Model



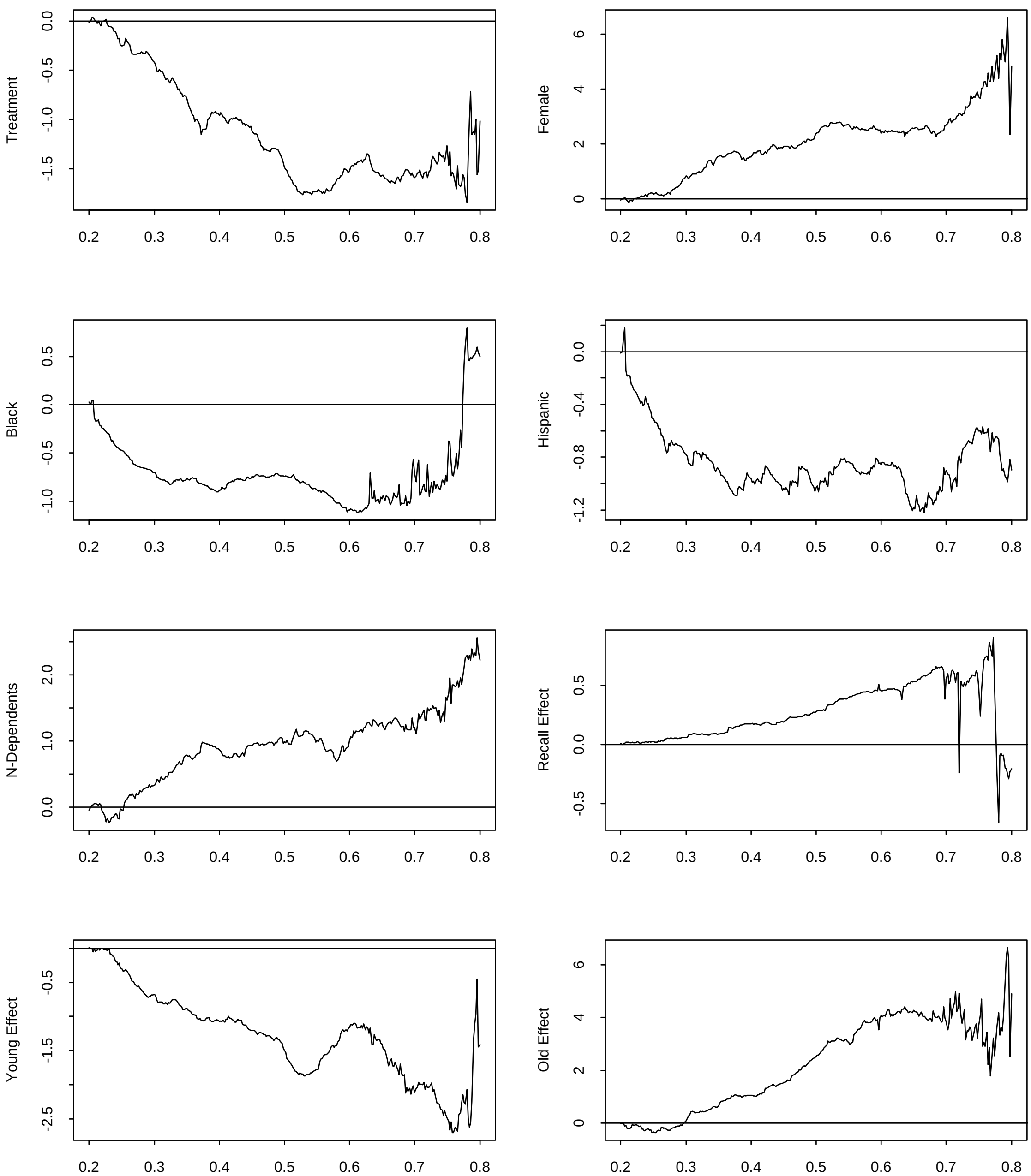
which takes the value 82.42. Here $\mathcal{T} = [.25, .75]$ so there is an additional .05 trimming to mollify the extreme behavior of the transformation in the tails. The critical value

FIGURE 6.3. Parametric Quantile Regression Process for Log Duration Model



for this test is 16.55, employing the ℓ_1 norm, so the location-scale-shift hypothesis is decisively rejected.

Figure 6.4. Transformed Parametric Quantile Regression Process



It is of some independent interest to investigate which of the coordinates contribute most to the joint significance of our K_n statistic. This inquiry is fraught with all the usual objections since the coordinates are not independent, but we plunge ahead, nevertheless. In place of the joint hypothesis we can consider univariate sub-hypotheses of the form,

$$\beta_i(\tau) = \mu_i + \sigma_i \beta_1(\tau)$$

for each "slope" coefficient. In effect, this approach replaces the matrix standardization used for the joint test by a scalar standardization. The martingale transformation is then applied just as in the previous case. Now, because there is no matrix standardization the original labeling of the coordinates *is* meaningful. In Figure 6.3 we replot the standardized residuals for our eight selected covariate effects using this coordinatewise approach. And in Figure 6.4 we plot these processes after the martingale transformation. In Table 6.1 we report the test statistics,

$$K_{ni} = \sup_{\tau \in \mathcal{T}} |\tilde{v}_{ni}(\tau) - \tilde{v}_{ni}(\tau_0)| / \sqrt{\tau_1 - \tau_0}$$

for each of the covariates. Effects for the 5 quarters of entry are not reported. The critical values for these coordinatewise tests are 1.923 at .05, and 2.420 at .01, as given in Appendix B, so the gender and age effects are highly significant and the treatment and dependents effects are weakly significant.

Variable	Location Scale Shift	Location Shift
Treatment	1.76	1.82
Female	3.91	3.61
Black	1.12	12.98
Hispanic	1.22	1.59
N-Dependents	1.66	1.88
Recall Effect	0.66	9.01
Young Effect	2.48	3.15
Old Effect	5.00	5.82
Durable Effect	1.85	2.22
Lusd Effect	2.05	2.61
Joint Effect	82.42	331.32

TABLE 6.1. Tests of the Location-ScaleShift and Location Shift Hypothesis

Also reported in Table 6.1 are the corresponding test statistics for the pure location-shift hypothesis. Not surprisingly, we find that the more restrictive hypothesis of constant $\beta_i(\tau)$ effects is considerably less plausible than the location scale hypothesis. The joint test statisitic is now, 331.32, with .01 critical value of 16.00, and almost all of the reported covariates effects are significant at level .05.

7. Conclusion

What should we conclude from this exercise? The linear location shift and location-scale shift models are very elegant and convenient abstractions for many statistical purposes. However, they also clearly place very stringent restrictions on the way that covariates are permitted to influence the conditional distribution of the response variable. In our unemployment duration application the location-scale shift hypothesis may be viewed as a generalized form of the familiar accelerated failure time model in which the scale of the response distribution responds linearly to the covariates. This specification is decisively rejected by the data from the Pennsylvania experiments. Not only the treatment effect of the bonus payment, but many other of the covariates appear to affect the conditional distribution of unemployment duration in ways that are not adequately represented either by pure location and/or scale shifts. One consequence of the proposed methods of inference, it may be hoped, would be a greater willingness to explore more flexible models for covariate effects in a wide variety of econometric models.

Appendix A. Proofs

Proof of Theorem 1 Note that

$$R\hat{\beta}(\tau) - r - \Psi(\tau) = R\left[\hat{\beta}(\tau) - \beta(\tau)\right] + R\beta(\tau) - r - \Psi(\tau).$$

Under Assumption A.3, $R\beta(\tau) - r - \Psi(\tau) = \zeta(\tau)/\sqrt{n}$, thus

$$R\hat{\beta}(\tau) - r - \Psi(\tau) = R\left[\hat{\beta}(\tau) - \beta(\tau)\right] + \zeta(\tau)/\sqrt{n}.$$

Under Assumptions A.1 and A.2, by Theorem 1 of Gutenbrunner and Jurečková(1992), we have, uniformly for $\tau \in \mathcal{T}$,

$$\sqrt{n}\left[\hat{\beta}(\tau) - \beta(\tau)\right] \Rightarrow \frac{1}{\varphi_0(\tau)} H_0^{-1} J_0^{1/2} v_0(\tau)$$

where $v_0(\tau)$ is a standardized p-dimensional Brownian bridge process, and $\varphi_0(\tau) = f_0(F_0^{-1}(\tau))$. Thus

$$\begin{aligned} v_n(\tau) &= \sqrt{n}\varphi_0(\tau)[R\Omega R^\top]^{-1/2}[R\hat{\beta}(\tau) - r - \Psi(\tau)] \\ &= \varphi_0(\tau)[R\Omega R^\top]^{-1/2} R\sqrt{n}\left[\hat{\beta}(\tau) - \beta(\tau)\right] + \varphi_0(\tau)[R\Omega R^\top]^{-1/2}\zeta(\tau) \\ &\Rightarrow v_0(\tau) + \eta(\tau). \end{aligned}$$

■

Proof of Corollary 1 We have,

$$\begin{aligned} v_n(\tau) &= \sqrt{n}\varphi_n(\tau)[R\Omega_n R^\top]^{-1/2}[R\hat{\beta}(\tau) - r - \Psi(\tau)] \\ &= \sqrt{n}\varphi_0(\tau)[R\Omega R^\top]^{-1/2}[R\hat{\beta}(\tau) - r - \Psi(\tau)] \\ &\quad + [\varphi_n(\tau) - \varphi_0(\tau)]\,[R\Omega_n R^\top]^{-1/2}\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)] \\ &\quad + \varphi_0(\tau)\left[[R\Omega_n R^\top]^{-1/2} - [R\Omega R^\top]^{-1/2}\right]\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)]. \end{aligned}$$

Note that

$$[R\Omega_n R^\top]^{-1/2} - [R\Omega R^\top]^{-1/2} = [R\Omega_n R^\top]^{-1/2}\left\{[R\Omega R^\top]^{1/2} - [R\Omega_n R^\top]^{1/2}\right\}[R\Omega R^\top]^{-1/2},$$

and $[R\Omega_n R^\top]^{1/2} = R\hat{H}_0^{-1}J_0^{1/2}$,

$$[R\Omega R^\top]^{1/2} - [R\Omega_n R^\top]^{1/2} = R[H_0^{-1} - \hat{H}_0^{-1}]J^{1/2} = R\hat{H}_0^{-1}[\hat{H}_0 - H_0]H_0^{-1}J^{1/2}.$$

Under Assumption A.4,

$$\begin{aligned}
&[\varphi_n(\tau) - \varphi_0(\tau)]\,[R\Omega_n R^\top]^{-1/2}\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)] = o_p(1),\\
&\varphi_0(\tau)\left[[R\Omega_n R^\top]^{-1/2} - [R\Omega R^\top]^{-1/2}\right]\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)] = o_p(1),
\end{aligned}$$

thus

$$\begin{aligned}
v_n(\tau) &= \sqrt{n}\varphi_n(\tau)[R\Omega_n R^\top]^{-1/2}[R\hat{\beta}(\tau) - r - \Psi(\tau)]\\
&= \sqrt{n}\varphi_0(\tau)[R\Omega R^\top]^{-1/2}[R\hat{\beta}(\tau) - r - \Psi(\tau)] + o_p(1)\\
&\Rightarrow v_0(\tau) + \eta(\tau).
\end{aligned}$$

■

Proof of Theorem 2 We may write,

$$\begin{aligned}
\hat{v}_n(\tau) &= \sqrt{n}\varphi_0(\tau)[R_n\Omega R_n^\top]^{-1/2}[R_n\hat{\beta}(\tau) - r_n - \Psi(\tau)]\\
&= \varphi_0(\tau)[R_n\Omega R_n^\top]^{-1/2}\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)]\\
&\quad + \varphi_0(\tau)[R_n\Omega R_n^\top]^{-1/2}\sqrt{n}[r_n - r] + \varphi_0(\tau)[R_n\Omega R_n^\top]^{-1/2}\sqrt{n}[R_n - R]\hat{\beta}(\tau)\\
&= \varphi_0(\tau)[R\Omega R^\top]^{-1/2}\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)]\\
&\quad + \varphi_0(\tau)[R\Omega R^\top]^{-1/2}\sqrt{n}[r_n - r] + \varphi_0(\tau)[R\Omega R^\top]^{-1/2}\sqrt{n}[R_n - R]\beta(\tau)\\
&\quad + o_p(1)
\end{aligned}$$

Since $\beta(\tau) = \alpha + \gamma F^{-1}(\tau)$,

$$\begin{aligned}
\hat{v}_n(\tau) &= \varphi_0(\tau)[R\Omega R^\top]^{-1/2}\sqrt{n}[R\hat{\beta}(\tau) - r - \Psi(\tau)]\\
&\quad + \varphi_0(\tau)\left\{[R\Omega R^\top]^{-1/2}\sqrt{n}[r_n - r] + [R\Omega R^\top]^{-1/2}\sqrt{n}[R_n - R]\alpha\right\}\\
&\quad + \varphi_0(\tau)F^{-1}(\tau)[R\Omega R^\top]^{-1/2}\sqrt{n}[R_n - R]\gamma\\
&\quad + o_p(1)\\
&= v_n(\tau) + Z_n^\top\xi(\tau) + o_p(1)
\end{aligned}$$

where $\xi(\tau) = (\varphi_0(\tau),\ \varphi_0(\tau)F^{-1}(\tau))^\top$ and

$$Z_n = \left[\ [R\Omega R^\top]^{-1/2}\sqrt{n}[r_n - r] + [R\Omega R^\top]^{-1/2}\sqrt{n}[R_n - R]\alpha,\quad [R\Omega R^\top]^{-1/2}\sqrt{n}[R_n - R]\gamma\ \right]^\top = O_p(1).$$

And thus by Theorem 1,

$$\hat{v}_n(\tau) - Z_n^\top\xi(\tau) \Rightarrow v_0(\tau) + \eta(\tau).$$

■

Proof of Corollary 2 Similar to that of Corollary 1. ■

Proof of Theorem 3 By Theorem 2,

$$\hat{v}_n(\tau) = v_0(\tau) + Z_n^\top\xi(\tau) + \eta(\tau) + o_p(1).$$

Denote the transformation based on $\dot{g}$ as

$$Q_g(h(\tau)) = h(\tau) - \int_0^\tau \left[\dot{g}(s)^\top C(s)^{-1} \int_s^1 \dot{g}(r) dh(r) \right] ds,$$

Since Q_g is a linear operator, we have

$$\widetilde{v}_n(\tau)^\top = Q_g \hat{v}_n(\tau)^\top = Q_g v_0(\tau)^\top + Q_g \xi(\tau)^\top Z_n + Q_g \eta(\tau)^\top + o_p(1).$$

By construction, $Q_g(\xi(\tau)^\top) = 0$, and by Khmaladze (1981), $Q_g v_0(\tau)^\top \Rightarrow w_0(\tau)^\top$, where w_0 is a q -variate standard Brownian motion. Thus

$$\widetilde{v}_n(\tau) \Rightarrow w_0(\tau) + \widetilde{\eta}(\tau).$$

Under the null hypothesis,

$$\sup_{\tau \in \mathcal{T}} \|\widetilde{v}_n(\tau)\| \Rightarrow \sup_{\tau \in \mathcal{T}} \|w_0(\tau)\|.$$

■

Proof of Corollary 3 We denote the transformation based on $\dot{g}_n$ as

$$Q_{g_n}(\hat{v}_n(\tau)^\top) = \hat{v}_n(\tau)^\top - \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \dot{g}_n(r) d\hat{v}_n(\tau)^\top \right] ds.$$

Noticing that

$$\hat{v}_n(\tau) = \sqrt{n} \varphi_n(\tau)[R_n \Omega_n R_n^\top]^{-1/2}[R_n \hat{\beta}(\tau) - r_n - \Psi(\tau)] = v_n(\tau) + Z_n^\top \xi_n(\tau) + o_p(1)$$

where Z_n is an $O_p(1)$ quantity independent of τ, and by construction, $Q_{g_n}(g_n) = 0$. Thus we have

$$\begin{aligned} & \hat{v}_n(\tau)^\top - \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \dot{g}_n(r) d\hat{v}_n(r)^\top \right] ds \\ = \; & v_n(\tau)^\top - \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \dot{g}_n(r) dv_n(r)^\top \right] ds + o_p(1). \end{aligned}$$

Because $\dot{g}_n(r)$ is a consistent estimator of $\dot{g}(r)$ uniformly on $r \in \mathcal{T} = [\varepsilon, 1-\varepsilon]$, we have, for all $s \in \mathcal{T}$

$$\left\| C(s)^{-1} \right\| = \left\| \left[\int_s^1 \dot{g}(v)\dot{g}(v)^\top dv \right]^{-1} \right\| \leq \left\| \left[\int_{1-\varepsilon}^1 \dot{g}(v)\dot{g}(v)^\top dv \right]^{-1} \right\| < \infty, \tag{A.1}$$

and

$$\begin{aligned} \left\| C_n(s)^{-1} \right\| &= \left\| \left[\int_s^1 \dot{g}_n(v)\dot{g}_n(v)^\top dv \right]^{-1} \right\| \\ &\leq \left\| \left[\int_{1-\varepsilon}^1 \dot{g}_n(v)\dot{g}_n(v)^\top dv \right]^{-1} \right\| \\ &= \left\| \left[\int_{1-\varepsilon}^1 \dot{g}(v)\dot{g}(v)^\top dv \right]^{-1} \right\| + o_p(1) < \infty. \end{aligned} \tag{A.2}$$

By assumption A.7, (A.1), and (A.2), we have

$$\begin{aligned} \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \left[\dot{g}_n(r) - \dot{g}(r) \right] dv_n(r)^\top \right] ds &= o_p(1), \\ \int_0^\tau \left[\left[\dot{g}_n(s)^\top - \dot{g}(s)^\top \right] C(s)^{-1} \int_s^1 \dot{g}(r) dv_n(r)^\top \right] ds &= o_p(1). \end{aligned}$$

Also notice that, under Assumption A.7, for all $s \in \mathcal{T}$,

$$C(s) - C_n(s) = \int_s^1 \left[\dot{g}(v)\dot{g}(v)^\top - \dot{g}_n(v)\dot{g}_n(v)^\top\right] dv = o_p(1), \tag{A.3}$$

thus, by (A.3), (A.1), and (A.2),

$$\begin{aligned}
&\int_0^\tau \left[\dot{g}_n(s)^\top \left[C_n(s)^{-1} - C(s)^{-1}\right] \int_s^1 \dot{g}(r) dv_n(r)^\top\right] ds \\
=& \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \left[C(s) - C_n(s)\right] C(s)^{-1} \int_s^1 \dot{g}(r) dv_n(r)^\top\right] ds \\
=& \ o_p(1).
\end{aligned}$$

Thus

$$\begin{aligned}
&\int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \dot{g}_n(r) dv_n(r)\right] ds - \int_0^\tau \left[\dot{g}(s)^\top C(s)^{-1} \int_s^1 \dot{g}(r) dv_n(r)^\top\right] ds \\
=& \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \left[\dot{g}_n(r) - \dot{g}(r)\right] dv_n(r)^\top\right] ds \\
&+ \int_0^\tau \left[\dot{g}_n(s)^\top \left[C_n(s)^{-1} - C(s)^{-1}\right] \int_s^1 \dot{g}(r) dv_n(r)^\top\right] ds \\
&+ \int_0^\tau \left[\left[\dot{g}_n(s)^\top - \dot{g}(s)^\top\right] C(s)^{-1} \int_s^1 \dot{g}(r) dv_n(r)^\top\right] ds \\
=& \ o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
v_n(\tau)^\top \quad - & \int_0^\tau \left[\dot{g}_n(s)^\top C_n(s)^{-1} \int_s^1 \dot{g}_n(r) dv_n(r)^\top\right] ds \\
= & \ v_n(\tau)^\top - \int_0^\tau \left[g(s)^\top C(s)^{-1} \int_s^1 g(r) dv_n(r)^\top\right] ds + o_p(1),
\end{aligned}$$

and the result of Corollary 3 follows immediately. ■

APPENDIX B. ASYMPTOTIC CRITICAL VALUES

Like many other Kolmogorov-Smirnov type tests (see, e.g. Andrews (1993), the limiting distribution $\sup_{\tau\in\mathcal{T}} \|w_0(\tau)\|$ is dependent on the norm $\|\cdot\|$, the pre-specified $\mathcal{T}$ and the dimension parameter q. Notice that the transformation is generally unstable in the extreme right tails, and the uniform convergency of existing estimators of the density and score $(f(F^{-1}(s))$ and $f'/f(F^{-1}(s)))$ usually requires that $\mathcal{T}$ be bounded away from zero and one, we consider a subset of $[0,1]$ whose closure lies in $(0,1)$.

We calculated the 1%, 5%, and 10% critical values for the test statistic $\sup_{\tau\in\mathcal{T}} \|\widetilde{v}_n(\tau)\|$ based on simulations where the Brownian motion was approximated by a Gaussian random walk, using a sample size $n = 2000$ and $20,000$ replications. For the norm $\|\cdot\|$, we use the ℓ_1 norm for a q-dimensional vector x, $\|x\| = \sum_{j=1}^q |x_j|$. Table 1 covers $\mathcal{T} = [\varepsilon, 1-\varepsilon]$ for $\varepsilon = 0.05, 0.1, 0.15, 0.2, 0.25$, 0.3, and $q = 1, 2, ..., 20$. Although conventionally we consider symmetric intervals $\mathcal{T} = [\varepsilon, 1-\varepsilon]$ for some small numbers ε, a much wider range of intervals $\mathcal{T}$ may be considered for the proposed tests. Critical values based other choices of the interval $\mathcal{T}$ and the dimension parameter q can be similarly calculated. Gauss programs are available from the authors upon request.

Asymptotic Critical Values

	$\varepsilon = 0.05$			$\varepsilon = 0.1$			$\varepsilon = 0.15$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$p = 1$	2.721	2.140	1.872	2.640	2.102	1.833	2.573	2.048	1.772
$p = 2$	4.119	3.393	3.011	4.034	3.287	2.946	3.908	3.199	2.866
$p = 3$	5.350	4.523	4.091	5.267	4.384	3.984	5.074	4.269	3.871
$p = 4$	6.548	5.560	5.104	6.340	5.430	4.971	6.148	5.284	4.838
$p = 5$	7.644	6.642	6.089	7.421	6.465	5.931	7.247	6.264	5.758
$p = 6$	8.736	7.624	7.047	8.559	7.412	6.852	8.355	7.197	6.673
$p = 7$	9.876	8.578	7.950	9.573	8.368	7.770	9.335	8.125	7.536
$p = 8$	10.79	9.552	8.890	10.53	9.287	8.662	10.35	9.044	8.412
$p = 9$	11.81	10.53	9.820	11.55	10.26	9.571	11.22	9.963	9.303
$p = 10$	12.91	11.46	10.72	12.54	11.17	10.43	12.19	10.85	10.14
$p = 11$	14.03	12.41	11.59	13.58	12.10	11.29	13.27	11.77	10.98
$p = 12$	15.00	13.34	12.52	14.65	13.00	12.20	14.26	12.61	11.86
$p = 13$	15.93	14.32	13.37	15.59	13.90	13.03	15.22	13.48	12.69
$p = 14$	16.92	15.14	14.28	16.52	14.73	13.89	16.12	14.34	13.48
$p = 15$	17.93	16.11	15.19	17.53	15.67	14.76	17.01	15.24	14.36
$p = 16$	18.85	16.98	16.06	18.46	16.56	15.65	17.88	16.06	15.22
$p = 17$	19.68	17.90	16.97	19.24	17.44	16.53	18.78	16.93	16.02
$p = 18$	20.63	18.83	17.84	20.21	18.32	17.38	19.70	17.80	16.86
$p = 19$	21.59	19.72	18.73	21.06	19.24	18.24	20.53	18.68	17.70
$p = 20$	22.54	20.58	19.62	22.02	20.11	19.11	21.42	19.52	18.52

	$\varepsilon = 0.2$			$\varepsilon = 0.25$			$\varepsilon = 0.3$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$p = 1$	2.483	1.986	1.730	2.420	1.923	1.664	2.320	1.849	1.602
$p = 2$	3.742	3.100	2.781	3.633	3.000	2.693	3.529	2.904	2.602
$p = 3$	4.893	4.133	3.749	4.737	4.018	3.632	4.599	3.883	3.529
$p = 4$	6.023	5.091	4.684	5.818	4.948	4.525	5.599	4.807	4.365
$p = 5$	6.985	6.070	5.594	6.791	5.853	5.406	6.577	5.654	5.217
$p = 6$	8.147	6.985	6.464	7.922	6.760	6.241	7.579	6.539	6.024
$p = 7$	9.094	7.887	7.299	8.856	7.611	7.064	8.542	7.357	6.832
$p = 8$	10.03	8.775	8.169	9.685	8.510	7.894	9.413	8.211	7.633
$p = 9$	10.90	9.672	9.018	10.61	9.346	8.737	10.27	9.007	8.400
$p = 10$	11.89	10.52	9.843	11.48	10.17	9.517	11.15	9.832	9.192
$p = 11$	12.85	11.35	10.66	12.48	10.99	10.28	12.06	10.62	9.929
$p = 12$	13.95	12.22	11.48	13.54	11.82	11.11	12.96	11.43	10.74
$p = 13$	14.86	13.09	12.31	14.34	12.66	11.93	13.82	12.24	11.51
$p = 14$	15.69	13.92	13.11	15.26	13.46	12.67	14.64	13.03	12.28
$p = 15$	16.55	14.77	13.91	16.00	14.33	13.47	15.46	13.85	13.05
$p = 16$	17.41	15.58	14.74	16.81	15.09	14.26	16.25	14.61	13.78
$p = 17$	18.19	16.43	15.58	17.59	15.95	15.06	17.04	15.39	14.54
$p = 18$	19.05	17.30	16.37	18.49	16.78	15.83	17.85	16.14	15.30
$p = 19$	19.96	18.09	17.17	19.40	17.50	16.64	18.78	16.94	16.05
$p = 20$	20.81	18.95	17.97	20.14	18.30	17.38	19.48	17.74	16.79

REFERENCES

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

BAI, J. (1998): "Testing parametric conditional distributions of dynamic models," preprint.

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.

BICKEL, P. J. (1982): "On Adaptive Estimation," *AnlsStat*, 10, 647–671.

BILIAS, Y., S. CHEN, AND Z. YING (1999): "Simple resampling methods for censored regression quantiles," preprint.

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

CORSON, W., P. DECKER, S. DUNSTAN, AND S. KERANSKY (1992): "Pennsylvania 'reemployment bonus' demonstration: Final Report," *Unemployment Insurance Occasional Paper*, 92.

COX, D. D. (1985): "A Penalty Method for Nonparametric Estimation of the Logarithmic Derivative of a Density Function," *AnInStMa*, 37, 271–288.

DARLING, D. (1955): "The Cramér-Smirnov test in the parametric case," *AnlsStat*, 26, 1–20.

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

DOOB, J. (1949): "A Heuristic Approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, 20, 393–403.

DURBIN, J. (1973): *Distribution theory for tests based on the sample distribution function.* SIAM.

FLEMING, T., AND D. HARRINGTON (1991): *Counting Processes and Survival Processes.* Wiley.

GUTENBRUNNER, C., AND J. JUREČKOVÁ (1992): "Regression quantile and regression rank score process in the linear model and derived statistics," *Ann. Statist.*, 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 Statistics*, 2, 307–33.

HAHN, J. (1995): "Bootstrapping quantile regression models," *Econometric Theory*, 11, 105–121.

HE, X., AND F. HU (1999): "Markov chain marginal bootstrap," preprint.

HECKMAN, J., J. SMITH, AND N. CLEMENTS (1997): "Making the most our of Programme evaluation and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64, 487–535.

HOROWITZ, J. (1998): "Bootstrap methods for median regression models," *Econometrica*, 66, 1327–1352.

HSIEH, D. A., AND C. F. MANSKI (1987): "Monte Carlo Evidence on Adaptive Maximum Likelihood Estimation of a Regression," *AnlsStat*, 15, 541–551.

KHMALADZE, E. V. (1981): "Martingale Approach in the Theory of Goodness-of-fit Tests," *Theory of Prob. and its Apps*, 26, 240–257.

KIEFER, J. (1959): "K-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests," *Ann. Math. Statist*, 30, 420–447.

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

——— (1982): "Robust tests for heteroscedasticity based on regression quantiles," *Econometrica*, 50, 43–61.

KOENKER, R., AND Y. BILIAS (1999): "Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments," *Empirical Economics*, forthcoming.

KOENKER, R., AND J. MACHADO (1999): "Goodness of fit and related inference processes for quantile regression," *J. of Am. Stat. Assoc.*, pp. 1296–1310.

KOENKER, R., AND S. PORTNOY (1987): "L-estimation for Linear Models," *J. of Am. Stat. Assoc.*, 82, 851–857.

Koul, H., and W. Stute (1999): "Nonparametric Model Checks for Time Series," *Ann. of Statist.*, 27, 204–236.

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

Meyer, B. (1995): "Lessons from the U.S. unemployment insurance experiments," *Journal of Economic Literature*, 33, 191–131.

——— (1996): "What have we learned from the Illinois reemployment bonus experiment," *Journal of Labor Economics*, 14, 26–51.

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

Oja, H. (1981): "On Location, Scale, Skewness and Kurtosis of Univariate Distributions," *Scand. J. of Stat.*, 8, 154–168.

Portnoy, S., and R. Koenker (1989): "Adaptive L-estimation of linear models," *Ann. Statist,*, 17, 362–381.

Prakasa Rao, B. (1999): *Semimartingales and their Statistical Inference.* Chapman-Hall, London.

Serfling, R. (1980): *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Sheather, S., and J. Maritz (1983): "An estimate of the asymptotic standard error of the sample medi an," *Australian J. of Statistics*, 25, 109–122.

Shorack, G., and J. Wellner (1986): *Empirical Processes with Applications to Statistics.* Wiley.

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

Stigler, S. (1980): "Stigler's Law of Eponymy," *Transactions of the New York Academy of Sciences*, 39, 147–157.

Stute, W., H. Thies, and L. X. Zhu (1998): "Model Checks for Regression: An Innovation Process Approach," *Ann. of Statist.*, 26, 1916–1934.

van der Vaart, A. W. (1998): *Asymptotic Statistics.* Cambridge U. Press.

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