Inference on the Quantile Regression Process

Roger Koenker
University of Illinois at Urbana-Champaign

NAKE Workshop
Groningen
December, 2003

There is more to econometric life than is dreamt of in the conventional regression philosophies of location-scale shift models.
Outline

• Introduction and Motivation
  ★ Reemployment Bonus Experiments
  ★ The Lehmann Quantile Treatment Effect, Again
  ★ The Location-Scale Shift Hypothesis
  ★ Quantile Regression Models for Durations
• Kolmogorov-Smirnov Tests and the Durbin Problem
• Khmaladze's Martingalization Approach to the Durbin Problem
• Inference on the Quantile Regression Process
• Application to the Pennsylvania Bonus Experiment
Reemployment Bonus Experiments

Can the durations of insured unemployment spells be shortened by offering cash bonuses to recipients for early reemployment?

- 1988-89 Experiment in Pennsylvania
- 6 Treatments + Control Group
  - Two levels of bonus payment
  - Two settings of the qualification period
- Randomized Assignment to Groups
- 13,913 Participants
Some Post-Modern Econometrics

The mean deconstructed into the quantiles:

$$\mu = \int_{-\infty}^{\infty} x dF(x) = \int_{0}^{1} F^{-1}(t) dt$$

The mean treatment effect deconstructed into the quantile treatment effect:

$$\delta = \mu(G) - \mu(F) = \int_{0}^{1} (G^{-1}(t) - F^{-1}(t)) dt$$

The regression mean effect deconstructed into regression quantiles:

$$E(Y|x) = \int_{0}^{1} Q_Y(\tau|x) d\tau$$
Regression is Demeaning

'De mean is 'de meaning.

Regression is demeaning.

Regression is de-meaning.
Transformation Models for Durations

Suppose
\[ G^{-1}(S(t|x)) = h(t) - x^\top \beta \]
where \( S(t|x) \) is the conditional survival function. For \( h \) monotone,
\[
P(h(T) > t|x) = P(T > h^{-1}(t)|x)
= S(h^{-1}(t)|x)
= G(t - x^\top \beta).
\]

We have the transformation model
\[ h(T) = x^\top \beta + u \]
where \( u \) is iid from \( G \).
Example: Cox Proportional Hazard Model

For the Cox model

$$\log \Lambda_0(T) = x^\top \beta + u$$

with $G(u) = 1 - \exp(-\exp(u))$. For $\Lambda_0$ Weibull,

$$\log \Lambda_0(t) = \gamma \log t - \alpha,$$

we obtain the accelerated failure time model,

$$\log T = x^\top \beta + u.$$
Quantile Regression Transformation Models

Given the transformation model the conditional quantile functions of $h(T)$, for $0 < \tau < 1$, are

$$Q_{h(T)}(\tau|x) = x^\top \beta + F_u^{-1}(\tau)$$

Since $P(h(T) \leq t) = P(T \leq h^{-1}(t))$, (monotone equivariance!)

$$Q_T(\tau|x) = h^{-1}(x^\top \beta + F_u^{-1}(\tau)).$$

Instead, we will consider,

$$Q_{h(T)}(\tau|x) = x^\top \beta(\tau),$$

for example, consider the location-scale shift model,

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

with $u_i$ iid from $F$. In this model we have a linear family of conditional quantile functions

$$Q_{h(T)}(\tau|x) = x^\top \alpha + (x^\top \gamma)F_u^{-1}(\tau) = x^\top \beta(\tau)$$

This is considerably more flexible.
An Inference Problem

We would like to test whether covariates have a pure location shift effect on the response, a location-scale shift effect, or if they have some more general effect on the response distribution:

- Location Shift Hypothesis:
  \[ H_0 : \beta_i(\tau) = \alpha_i \quad i = 2, \ldots, p. \]

- Location-Scale Shift Hypothesis:
  \[ H_0 : \beta_i(\tau) = \alpha_i + \gamma_i \beta_1(\tau) \quad i = 2, \ldots, p. \]

Tests of the Kolmogorov-Smirnov type based on the whole quantile regression process will be considered.
The Kolmogorov-Smirnov Test

Suppose \( \{Y_1, \ldots, Y_n\} \) are iid from df \( F \). We would like to test,

\[
H_0 : F = F_0.
\]

We want to consider the K-S statistic,

\[
K_n = \sup_{x \in \mathbb{R}} \sqrt{n} |F_n(x) - F_0(x)|
\]

where \( F_n(x) = n^{-1} \sum I(Y_i \leq x) \).
KS Test is ADF

Classically, from Doob (1949), we know

$$U_n(x) = \sqrt{n}(F_n(x) - F_0(x))$$

or, changing variables $x \rightarrow F_n^{-1}(\tau)$,

$$u_n(\tau) = \sqrt{n}(\tau - F_0(F_n^{-1}(\tau)))$$

converges weakly under $H_0$ to a Brownian Bridge process, i.e., a Gaussian process, $u_0$, with mean zero and covariance function $Cov(u_0(\tau_1), u_0(\tau_2)) = \tau_1 \wedge \tau_2 - \tau_1 \tau_2$. So the test is asymptotically distribution free (ADF).
The Durbin Problem

Now suppose \( F_0 \) is known only up to parameters, e.g., \( F_0(x, \theta_0) = \Phi((x - \mu_0)/\sigma_0) \), but \( \theta_0 = (\mu_0, \sigma_0) \) is unknown. We are tempted to consider the process,

\[
\hat{U}_n(x) = \sqrt{n}(F_n(x) - F_0(x, \hat{\theta}_n))
\]

and again changing variables, so \( \tau = F_0(x, \theta_0) \), setting \( G(\tau, \theta_0) = \tau \),

\[
\hat{u}_n(\tau) = \sqrt{n}(G_n(\tau) - G(\tau, \hat{\theta}_n))
\]

Like \( u_n(\tau) \), \( \hat{u}_n(\tau) \) converges weakly to zero mean Gaussian process, say, \( \hat{u}_n(\tau) \Rightarrow \hat{u}_0(\tau) \), but now for the mle \( \hat{\theta}_n \),

\[
E(\hat{u}_0(\tau_1)\hat{u}_0(\tau_2)) = \tau_1 \wedge \tau_2 - \tau_1\tau_2 - g_0(\tau_1)^\top J^{-1} g_0(\tau_2)
\]

where \( g_0(\tau) = \partial F_0(y, \theta_0)/\partial \theta|_{y=F_0^{-1}(\tau, \theta_0)} \), and \( J \) is the Fisher information about \( \theta \) in model \( F_0 \). Now \( \hat{K}_n = \sup |\hat{u}_n(\tau)| \) depends on \( F_0 \); this is the Durbin Problem.
The Doob-Meyer Decomposition

The process $G_n(\tau) = F_0(F_n^{-1}(\tau))$ is Markov:

$$n \Delta G_n(\tau) = n[G_n(\tau + \Delta\tau) - G_n(\tau)] \sim \text{Bin}(n(1 - G_n(\tau)), \Delta\tau/(1 - \tau)).$$

So,

$$E[\Delta G_n(\tau) | \mathcal{F}_\tau^{G_n}] = \frac{1 - G_n(\tau)}{1 - \tau} \Delta\tau$$

and this suggests the representation,

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

where $m_n(t)$ is a martingale. Now substituting from $u_n(t) = \sqrt{n}(G_n(t) - t)$ we have

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

where $w_n(t) = \sqrt{nm_n(t)} \Rightarrow w_0(\tau)$, is standard Brownian motion.
“Marmalade” in a Martingale

Etymology: a. Fr. martingale of obscure etymology. [First found in Rabelais in *chausses a la martingale*, men’s socks that fastened at the back of the leg. This is commonly supposed to mean literally ‘hose after the fashion of Martigues’ (in Provence).
Doob-Meyer as Recursive OLS

Let \( g(t) = (t, g_1(t), \ldots, g_p(t))^\top \) be a \((p + 1)\)-vector of real-valued functions on \([0, 1]\). Suppose \( \dot{g}(t) = dg(t)/dt \) are linearly independent, so

\[
C(t) = \int_t^1 \dot{g}(s)\dot{g}(s)^\top ds
\]

is nonsingular, and consider the transformation,

\[
w_n(t) = v_n(t) - \int_0^t \dot{g}(s)^\top C^{-1}(s) \int_s^1 \dot{g}(r)dv_n(r)ds
\]

In the Doob-Meyer case, we set \( g(t) = t \) so \( \dot{g}(t) = 1 \), \( C(s) = 1 - s \), and noting that,

\[
\int_s^1 \dot{g}(r)dv_n(r) = v_n(1) - v_n(s) = -v_n(s)
\]

we obtain the Doob-Meyer decomposition.
Khmaladze’s Martingalization

Ingredients:

\[ G(\tau, \hat{\theta}_n) = \tau + (\hat{\theta} - \theta_0)^\top g(\tau, \theta^*) \]

\[ \sqrt{n}(\hat{\theta} - \theta_0) = \int_0^1 h(s, \theta_0) du_n(s) + o_p(1) \]

\[ \hat{u}_n(\tau) = \sqrt{n}(G_n(\tau) - \tau + \tau - G(\tau, \hat{\theta}_n)) \]

Combine and stir:

\[ \hat{u}_n(\tau) = u_n(\tau) - g(\tau, \theta_0)^\top \int_0^1 h(s, \theta_0) du_n(s) + o_p(1) \]  \hspace{1cm} (1)

\[ \Rightarrow u_0(\tau) - g(\tau, \theta_0)^\top \int_0^1 h(s, \theta_0) du_0(s) \]  \hspace{1cm} (2)

but,

\[ \tilde{u}_n(\tau) = \hat{u}_n(\tau) - \int_0^\tau \dot{g}(s)^\top C^{-1}(s) \int_s^1 \dot{g}(r) d\hat{u}_n(r) ds \]  \hspace{1cm} (3)

\[ \Rightarrow w_0(\tau) \]  \hspace{1cm} (4)

Martingalization annihilates the \( g(\tau, \theta_0) \) term and restores ADF property of KS-test!


Khmaladze for the Quantile Process

Let

\[ \hat{\alpha}(\tau) = \arg\min_{a \in \mathbb{R}} \sum \rho_\tau(y_i - a) \]

where \( \{y_i\} \) are iid from \( F_0((y - \mu)/\sigma) \). Consider

\[ H_0 : \quad \alpha(\tau) = F_y^{-1}(\tau) = \mu + \sigma F_0^{-1}(\tau) \]

under \( H_0 \),

\[ v_n(\tau) = \sqrt{n} \varphi_0(\tau)(\hat{\alpha}(\tau) - \alpha(\tau))/\sigma \Rightarrow v_0(\tau) \]

where \( \varphi_0(\tau) = f_0(F_0^{-1}(\tau)) \) and \( v_0(\tau) \) is the Brownian Bridge process. To test \( H_0 \), set \( \tilde{\alpha}(\tau) = \xi(\tau)^\top \tilde{\theta} = (1, F_0^{-1}(\tau))\tilde{\theta} \), then

\[ \hat{v}_n(t) = \sqrt{n} \varphi_0(t)(\hat{\alpha}(t) - \tilde{\alpha}(t))/\sigma \]

\[ = \sqrt{n} \varphi_0(t)(\hat{\alpha}(t) - \alpha(t) - (\tilde{\alpha}(t) - \alpha(t)))/\sigma \]

\[ = v_n(t) - \sqrt{n} \varphi_0(\tau)(\tilde{\theta} - \theta_0)^\top \xi(t)/\sigma \]

Now we apply martingalization as before.
Testing for Normality

In the typical case that $\theta_0$ consists of a location and scale parameter we have,

$$g(\tau) = (\tau, \varphi_0(\tau)\xi(\tau)^T)^T$$

so,

$$\dot{g}(\tau) = (1, \dot{f}/f, 1 - F_0^{-1}(\tau)\dot{f}/f)^T$$

where $\dot{f}/f$ is evaluated at $F^{-1}(\tau)$. In the Gaussian case, $F_0 = \Phi$, we have

$$\dot{g}(\tau) = (1, -\Phi^{-1}(\tau), 1 - \Phi^{-1}(\tau)^2)^T$$
Inference for Quantile Regression

Now consider the quantile regression process,

$$\hat{\beta}(\tau) = \arg\min_{b \in \mathbb{R}^p} \sum \rho_\tau(y_i - x_i^\top b)$$

The analogue of the location scale model is

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

with \( \{u_i\} \) iid from \( F_0 \). This implies the null hypothesis,

$$H_0 : \beta_i(\tau) = \alpha_i + \gamma_i F_0^{-1}(\tau) \quad i = 1, \ldots, p.$$  

We would like to test, \( H_0 \), versus a general alternative. Note that, \( H_0 \) implies that all \( p \) coordinates of \( \beta(\cdot) \) are affine functions of a single univariate function.
Simple Nulls

When $\alpha, \gamma, F_0$ are all known we have, subject to some regularity conditions,

$$v_n(\tau) = \sqrt{n} J_n^{-1/2} H_n(\hat{\beta}(\tau) - \beta(\tau)) \Rightarrow v_0$$

where $v_0$ is now a p-variate Brownian Bridge, $J_n = n^{-1} X^\top X$, $H_n = n^{-1} X^\top \Gamma^{-1} X$, and $\Gamma = \text{diag}(x_i^\top \gamma)$.

This leads to Wald, LR and LM/rankscore tests as in Koenker and Machado (JASA, 1999), employing Bessel processes as in Kiefer(1959). But when $(\alpha, \gamma)$ are unknown, the Durbin problem arises again.
A General Linear Hypothesis

Consider the hypothesis,

\[ R\beta(\tau) - r = \Psi(\tau) \quad \tau \in \mathcal{T} \]  \hspace{1cm} (8)

where \( R \) denotes a \( q \times p \) matrix, \( q \leq p \), \( r \in \mathbb{R}^q \), and \( \Psi(\tau) \) denotes a known function \( \Psi : \mathcal{T} \to \mathbb{R}^q \). and the local alternative,

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

Test based on:

\[ v_n(\tau) = \sqrt{n} \varphi_0(\tau) (R\Omega R^\top)^{-1/2} (R\hat{\beta}(\tau) - r - \Psi(\tau)) \]

where \( \Omega = H_0^{-1} J_0 H_0^{-1} \) with \( J_0 = \lim n^{-1} \sum x_i x_i^\top \), and \( H_0 = \lim n^{-1} \sum x_i x_i^\top / \gamma^\top x_i \).
Regularity Conditions

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

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

(i) $x_{i1} \equiv 1$ \quad $i = 1, 2, \ldots$

(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 $\Gamma_n = \text{diag}(\gamma^\top x_i)$.

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

$$R\beta_n(\tau) - r - \Psi(\tau) = \frac{\zeta(\tau)}{\sqrt{n}}.$$
More Regularity Conditions

Assumption 4. There exist estimators $\varphi_n(\tau)$ and $\Omega_n$ satisfying

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

ii. $||\Omega_n - \Omega|| = o_p(1)$.

Assumption 5. The function $g(t)$ satisfies:

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

ii. $\{\dot{g}_i(t) : i = 1, \ldots, m\}$ are linearly independent in a neighborhood of 1.
Theorem 1. Let $T$ denote the closed interval $[\varepsilon, 1 - \varepsilon]$, for some $\varepsilon \in (0, 1/2)$. Under conditions A.1-3

$$v_n(\tau) = \sqrt{n} \varphi_0(\tau)(R\Omega R^\top)^{-1/2}(R\hat{\beta}(\tau) - r - \Psi(\tau))$$ \hfill (9)

$$\Rightarrow v_0(\tau) + \eta(\tau) \text{ for } \tau \in T$$ \hfill (10)

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

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

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

$$\hat{v}_n(\tau) = \sqrt{n} \varphi_0(\tau)[R_n \Omega R_n^\top]^{-1/2}(R_n\hat{\beta}(\tau) - r_n - \Psi(\tau))$$ \hfill (11)

$$\Rightarrow Z_n^\top \xi(\tau) + v_0(\tau) + \eta(\tau) \text{ for } \tau \in T$$ \hfill (12)

where $\xi(\tau) = \varphi_0(\tau)(1, F_0^{-1}(\tau))^\top$, and $Z_n = O_p(1)$, with $v_0(\tau)$ and $\eta(\tau)$ as specified in Theorem 1.
Theorem 3. Under conditions A.1 - 6, we have

$$\tilde{v}_n(\tau)^\top = \hat{v}_n(\tau)^\top - \int_0^\tau \dot{g}(s)^\top C^{-1}(s) \int_s^1 \dot{g}(r)d\hat{v}_n(r)^\top ds \tag{13}$$

$$\Rightarrow w_0(\tau) + \tilde{\eta}(\tau) \text{ for } \tau \in \mathcal{T} \tag{14}$$

where $w_0(\tau)$ denotes a $q$-variate standard Brownian motion, and under the null hypothesis, $\zeta(\tau) = 0$,

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

Table 1: Treatment Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Bonus Amount</th>
<th>Qualification Period</th>
<th>Workshop Offer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>Treatment 1</td>
<td>Low</td>
<td>Short</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>Low</td>
<td>Long</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment 3</td>
<td>High</td>
<td>Short</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment 4</td>
<td>High</td>
<td>Long</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment 5</td>
<td>Declining</td>
<td>Long</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment 6</td>
<td>High</td>
<td>Long</td>
<td>No</td>
</tr>
</tbody>
</table>

Note: The low benefit was 3 times UI weekly benefit amount, the high benefit was 6 times this amount. The declining bonus declined from 6 times the weekly benefit to zero, over a 12 week period. The short qualification period was 6 weeks, and the long period was 12 weeks.
## Sample Sizes

<table>
<thead>
<tr>
<th>Groups</th>
<th>Target $n$</th>
<th>Collected $n$</th>
<th>Analysis $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>3,000</td>
<td>3,392</td>
<td>3,354</td>
</tr>
<tr>
<td>Treatment 1</td>
<td>1,030</td>
<td>1,395</td>
<td>1,385</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>2,240</td>
<td>2,456</td>
<td>2,428</td>
</tr>
<tr>
<td>Treatment 3</td>
<td>1,740</td>
<td>1,910</td>
<td>1,885</td>
</tr>
<tr>
<td>Treatment 4</td>
<td>1,590</td>
<td>1,771</td>
<td>1,745</td>
</tr>
<tr>
<td>Treatment 5</td>
<td>1,740</td>
<td>1,860</td>
<td>1,831</td>
</tr>
<tr>
<td>Treatment 6</td>
<td>1,780</td>
<td>1,302</td>
<td>1,285</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>13,120</strong></td>
<td><strong>14,086</strong></td>
<td><strong>13,913</strong></td>
</tr>
</tbody>
</table>
Quantile Regression Process

(Intercept)

-0.25 -0.05

treatmentTRUE

-0.05 0.15

female

-0.4 -0.2 0.0

recall

-0.4 0.2 0.0

young

old
Fitted Quantile Regression Process

- treatment
- female
- black
- recall
- young
- old
Standardized Residual Quantile Regression Process
Khmaladzized Quantile Regression Process
## Test Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Location Scale Shift</th>
<th>Location Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>5.41</td>
<td>5.48</td>
</tr>
<tr>
<td>Female</td>
<td>4.47</td>
<td>4.42</td>
</tr>
<tr>
<td>Black</td>
<td>5.77</td>
<td>22.00</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2.74</td>
<td>2.00</td>
</tr>
<tr>
<td>N-Dependents</td>
<td>2.47</td>
<td>2.83</td>
</tr>
<tr>
<td>Recall Effect</td>
<td>4.45</td>
<td>16.84</td>
</tr>
<tr>
<td>Young Effect</td>
<td>3.42</td>
<td>3.90</td>
</tr>
<tr>
<td>Old Effect</td>
<td>6.81</td>
<td>7.52</td>
</tr>
<tr>
<td>Durable Effect</td>
<td>3.07</td>
<td>2.83</td>
</tr>
<tr>
<td>Lusd Effect</td>
<td>3.09</td>
<td>3.05</td>
</tr>
<tr>
<td>Joint Effect</td>
<td>112.23</td>
<td>449.83</td>
</tr>
</tbody>
</table>

Table 2: Tests of the Location-Scale and Location Shift Hypotheses: Critical values for the univariate tests are 1.92 at .05 and 2.42 at .01. For the joint tests the .01 critical value is 16.0.
Conclusions

- Quantile regression methods complement established survival analysis methods.
- By focusing on local slices of the conditional distribution, they offer a useful deconstruction of conditional mean models.
- They offer a more flexible role for covariate effects allowing them to influence location, scale and shape of the response distribution.
- The Khmaladze transformation approach offers a flexible way to handle nuisance parameter problems in semi-parametric inference for quantile regression.