Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure

Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val

December 21, 2004

Abstract

Quantile regression (QR) fits a linear model for conditional quantiles, just as ordinary least squares (OLS) fits a linear model for conditional means. An attractive feature of OLS is that it gives the minimum mean square error linear approximation to the conditional expectation function even when the linear model is misspecified. Empirical research using quantile regression with discrete covariates suggests that QR may have a similar property, but the exact nature of the linear approximation has remained elusive. In this paper, we show that QR minimizes a weighted mean-squared error loss function for specification error. The weighting function is an average density of the dependent variable near the true conditional quantile. The weighted least squares interpretation of QR is used to derive an omitted variables bias formula and a partial quantile regression concept, similar to the relationship between partial regression and OLS. We also present asymptotic theory for the QR process under misspecification of the conditional quantile function. The approximation properties of QR are illustrated using wage data from the US census. These results point to major changes in inequality from 1990-2000.

Acknowledgment. We thank David Autor, Gary Chamberlain, George Deltas, Bernd Fitzenberger, Jinyong Hahn, Jerry Hausman, Frank Kleibergen, Roger Koenker, Rafael Lalive, Tony Lancaster, Art Lewbel, and Whitney Newey for helpful discussions, and seminar participants at Berkeley, BYU, Brown, Duke, the University of Michigan, Michigan State University, the Harvard-MIT Econometrics Workshop, the University of Toronto, the University of Illinois at Urbana-Champaign, and the 2001 and 2004 Winter Econometric Society Meetings for comments.

1 Introduction

The Quantile Regression (QR) estimator, introduced by Koenker and Bassett (1978), is an increasingly important empirical tool, allowing researchers to fit parsimonious models to an entire conditional distribution. Part of the appeal of quantile regression derives from a natural parallel with conventional ordinary least squares (OLS) or mean regression. Just as OLS regression coefficients offer convenient summary statistics for conditional expectation functions, quantile regression coefficients can be used to make easily interpreted statements about conditional distributions. Moreover, unlike OLS coefficients, QR estimates capture changes in distribution shape and spread, as well as changes in location.

An especially attractive feature of OLS regression estimates is their robustness and interpretability under misspecification of the conditional expectation function (CEF). In addition to consistently estimating a linear conditional expectation function, OLS estimates provide the minimum mean square error linear approximation to a conditional expectation function of any shape. The approximation properties of OLS have been emphasized by White (1980), Chamberlain (1984), Goldberger (1991), and Angrist and Krueger (1999). The fact that OLS provides a meaningful and well-understood summary statistic for conditional expectations under almost all circumstances undoubtedly contributes to the primacy of OLS regression as an empirical tool. In view of the possibility of interpretation under misspecification, modern theoretical research on regression inference also expressly allows for misspecification of the regression function when deriving limiting distributions (White, 1980).

While QR estimates are as easy to compute as OLS regression coefficients, an important difference between OLS and QR is that most of the theoretical and applied work on QR postulates a correctly specified linear model for conditional quantiles. This raises the question of whether and how QR estimates can be interpreted when the linear model for conditional quantiles is misspecified (for example, QR estimates at different quantiles may imply conditional quantile functions that cross). One interpretation for QR under misspecification is that it provides the best linear predictor for a response variable under asymmetric loss. This interpretation is not very satisfying, however, since prediction under asymmetric loss is typically not the object of interest in empirical work.[1] Empirical research on quantile regression with discrete covariates suggests that QR may have an approximation property similar to that of OLS, but the exact nature of the linear approximation has remained an important unresolved question (Chamberlain, 1994, p. 181).

[1]An exception is the forecasting literature; see, e.g., Giacomini and Komunjer (2003).

The first contribution of this paper is to show that QR is the best linear approximation to the conditional quantile function using a weighted mean-squared error loss function, much as OLS regression provides a minimum mean-squared loss fit to the conditional expectation function. The implied QR weighting function can be used to understand which, if any, parts of the distribution of regressors contribute disproportionately to a particular set of QR estimates. We also show how this approximation property can be used to interpret multivariate QR coefficients as partial regression coefficients and to develop an omitted variables bias formula for QR. A second contribution is to present a distribution theory for the QR process that accounts for possible misspecification of the conditional quantile function. We present the main inference results only, with proofs available in a supplementary appendix. The approximation theorems and inference results in the paper are illustrated with an analysis of wage data from recent U.S. censuses.[2] The results show a sharp change in the quantile process of schooling coefficients in the 2000 census, and an increase in conditional inequality in the upper half of the wage distribution from 1990-2000.

The paper is organized as follows. Section 2 introduces assumptions and notation and presents the main approximation theorems. Section 3 presents inference theory for QR processes under misspecification. Section 4 illustrates QR approximation properties with U.S. census data. Section 5 concludes.

2 Interpreting QR Under Misspecification

2.1 Notation and Framework

Given a continuous response variable Y and a $d \times 1$ regressor vector X, we are interested in the conditional quantile function (CQF) of Y given X. The conditional quantile function is defined as:

$$Q_\tau(Y|X) := \inf \{y : F_Y(y|X) \geq \tau\},$$

where $F_Y(y|X)$ is the distribution function for Y conditional on X, which is assumed to have conditional density $f_Y(y|X)$. The CQF is also known to be a solution to the following minimization problem, assuming integrability:

$$Q_\tau(Y|X) \in \arg \min_{q(X)} E\left[\rho_\tau(Y - q(X))\right], \tag{1}$$

[2] Quantile regression has been widely used to model changes in the wage distribution; see, e.g., Buchinsky (1994), Abadie (1997), Gosling, Machin, and Meghir (2000), Autor, Katz, and Kearney (2004).

where $\rho_\tau(u) = (\tau - 1(u \leq 0))u$ and the minimum is over the set of measurable functions of X. This is a potentially infinite-dimensional problem if covariates are continuous, and can be high-dimensional even with discrete X. It may nevertheless be possible to capture important features of the CQF using a linear model. This motivates linear quantile regression.

The linear quantile regression (QR), introduced by Koenker and Bassett (1978), solves the following minimization problem in the population, assuming integrability and uniqueness of the solution:

$$\beta(\tau) := \arg \min_{\beta \in \mathbb{R}^d} E\left[\rho_\tau(Y - X'\beta)\right]. \tag{2}$$

If $q(X)$ is in fact linear, the QR minimand will find it, just as if the conditional expectation function is linear, OLS will find it. More generally, QR provides the best linear predictor for Y under the asymmetric loss function, ρ_τ. As noted in the introduction, however, prediction under asymmetric loss is rarely the object of empirical work. Rather, the conditional quantile function is usually of intrinsic interest. For example, labor economists are often interested in comparisons of conditional deciles as a measure of how the spread of a wage distribution changes conditional on covariates, as in Katz and Murphy (1992), Juhn, Murphy, and Pierce (1993), and Buchinsky (1994). Thus, our first goal is to establish the nature of the approximation to conditional quantiles that QR provides.

2.2 QR Approximation Properties

Our principal theoretical result is that the population QR vector minimizes a weighted sum of squared specification errors. This is easiest to show using notation for a quantile-specific specification error and for a quantile-specific residual. For any quantile index $\tau \in (0,1)$, we define the QR specification error as:

$$\Delta_\tau(X, \beta) := X'\beta - Q_\tau(Y|X).$$

Similarly, let ϵ_τ be a quantile-specific residual, defined as the deviation of the response variable from the conditional quantile of interest:

$$\epsilon_\tau := Y - Q_\tau(Y|X),$$

with conditional density $f_{\epsilon_\tau}(e|X)$ at $\epsilon_\tau = e$. The following theorem shows that QR is a weighted least squares approximation to the unknown CQF.

Theorem 1 (Approximation Property) *Suppose that (i) the conditional density $f_Y(y|X)$ exists a.s., (ii) $E[Y]$, $E[Q_\tau(Y|X)]$, and $E\|X\|$ are finite, and (iii) $\beta(\tau)$ uniquely solves (2).*

Then

$$\beta(\tau) = \arg \min_{\beta\in\mathbb{R}^d} E\left[w_\tau(X,\beta)\cdot\Delta_\tau^2(X,\beta)\right],$$

where

$$\begin{aligned} w_\tau(X,\beta) &= \int_0^1 (1-u)\cdot f_{\epsilon_\tau}\left(u\Delta_\tau(X,\beta)|X\right)\ du \\ &= \int_0^1 (1-u)\cdot f_Y\left(u\cdot X'\beta+(1-u)\cdot Q_\tau(Y|X)|X\right) du \geq 0. \end{aligned}$$

PROOF: We have that $\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} E[\rho_\tau(\epsilon_\tau - \Delta_\tau(X,\beta))]$, or equivalently, since $E[\rho_\tau(\epsilon_\tau)]$ does not depend on β and is finite by condition (ii),

$$\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d}\left\{E\left[\rho_\tau\left(\epsilon_\tau-\Delta_\tau(X,\beta)\right)\right]-E\left[\rho_\tau\left(\epsilon_\tau\right)\right]\right\}. \tag{3}$$

By definition of ρ_τ and the law of iterated expectations, it follows further that

$$\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d}\left\{E[\mathcal{A}(X,\beta)]-E[\mathcal{B}(X,\beta)]\right\},$$

where

$$\begin{aligned} \mathcal{A}(X,\beta) &= E\left[\left(1\{\epsilon_\tau<\Delta_\tau(X,\beta)\}-\tau\right)\Delta_\tau(X,\beta)\ |X\right], \\ \mathcal{B}(X,\beta) &= E\left[\left(1\{\epsilon_\tau<\Delta_\tau(X,\beta)\}-1\{\epsilon_\tau<0\}\right)\epsilon_\tau|X\right]. \end{aligned}$$

The conclusion of the theorem can then be obtained by showing that

$$\mathcal{A}(X,\beta) = \left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du\right)\cdot\Delta_\tau^2(X,\beta), \tag{4}$$

$$\mathcal{B}(X,\beta) = \left(\int_0^1 u f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du\right)\cdot\Delta_\tau^2(X,\beta), \tag{5}$$

establishing that both components are density-weighted quadratic specification errors.

Consider $\mathcal{A}(X,\beta)$ first. Observe that

$$\begin{aligned} \mathcal{A}(X,\beta) &= \left[F_{\epsilon_\tau}(\Delta_\tau(X,\beta)|X)-F_{\epsilon_\tau}(0|X)\right]\Delta_\tau(X,\beta) \\ &= \left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)\Delta_\tau(X,\beta)du\right)\Delta_\tau(X,\beta), \end{aligned} \tag{6}$$

where the first statement follows by the definition of conditional expectation and noting that $E[1\{\epsilon_\tau\leq 0\}|X] = F_{\epsilon_\tau}(0|X) = \tau$ and the second follows from the fundamental theorem of calculus (for Lebesgue integrals). This verifies (4). Turning to $\mathcal{B}(X,\beta)$, suppose first that

$\Delta_\tau(X,\beta) > 0$. Then, setting $u_\tau = \epsilon_\tau/\Delta_\tau(X,\beta)$, we have

$$\begin{aligned}\mathcal{B}(X,\beta) &= E\ \left[1\{\epsilon_\tau \in [0,\Delta_\tau(X,\beta)]\}\cdot \epsilon_\tau \Big| X\right]\\ &= E\left[1\{u_\tau \in [0,1]\}\cdot u_\tau \cdot \Delta_\tau(X,\beta)\Big| X\right]\\ &= \left(\int_0^1 u f_{u_\tau}(u|X)du\right)\Delta_\tau(X,\beta)\\ &= \left(\int_0^1 u f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)\Delta_\tau(X,\beta)du\right)\cdot \Delta_\tau(X,\beta),\end{aligned} \tag{7}$$

which verifies (5). A similar argument shows that (5) also holds if $\Delta_\tau(X,\beta) < 0$. Finally, if $\Delta_\tau(X,\beta) = 0$, then $\mathcal{B}(X,\beta) = 0$, so that (5) holds in this case too. *Q.E.D.*

Theorem 1 states that the population QR coefficient vector $\beta(\tau)$ minimizes the expected weighted mean squared approximation error, i.e., the square of the difference between the true CQF and a linear approximation, with weighting function $w_\tau(X,\beta)$.[3] The weights are given by the average density of the response variable over a line from the point of approximation, $X'\beta$, to the true conditional quantile, $Q_\tau(Y|X)$. Pre-multiplication by the term $(1-u)$ in the integral results in more weight being applied at points on the line closer to the true CQF.

We refer to the function $w_\tau(X,\beta)$ as defining *importance weights*, since this function determines the importance the QR minimand gives to points in the support of X for a given distribution of X.[4] In addition to the importance weights, the probability distribution of X also determines the ultimate weight given to different values of X in the least squares problem. To see this, note that we can also write the QR minimand as

$$\beta(\tau) = \arg\ \min_{\beta\in\mathbb{R}^d} \int \Delta_\tau^2(x,\beta)\ w_\tau(x,\beta)\ d\Pi(x),$$

where $\Pi(x)$ is the distribution function of X with associated probability mass or density function $\pi(x)$. Thus, the overall weight varies in the distribution of X according to

$$w_\tau(x,\beta)\cdot\pi(x).$$

A natural question is what determines the shape of the importance weights. This can be understood using the following approximation. When Y has a smooth conditional density, we have for β in the neighborhood of $\beta(\tau)$:

$$w_\tau(X,\beta) = 1/2\cdot f_Y\left(Q_\tau(Y|X)|X\right) + \varrho_\tau(X), \quad |\varrho_\tau(X)| \le 1/6\cdot|\Delta_\tau(X,\beta)|\cdot \bar{f}'(X). \tag{8}$$

[3]Note that if we define $\beta(\tau)$ via (3), then integrability of Y is not required in Theorem 1.

[4]This terminology should not to be confused with similar terminology from Bayesian statistics.

Here, $\varrho_\tau(X)$ is a remainder term and the density $f_Y(y|X)$ is assumed to have a first derivative in y bounded in absolute value by $\bar{f}'(X)$ a.s.[5] Hence in many cases the *density weights* $1/2 \cdot f_Y(Q_\tau(Y|X)|X)$ are the primary determinants of the importance weights, a point we illustrate in Section 4. It is also of interest to note that $f_Y(Q_\tau(Y|X)|X)$ is constant across X in location models, and inversely proportional to the conditional standard deviation in location-scale models.[6]

QR has a second approximation property closely related to the first. This second property is particularly well-suited to the development of a partial regression decomposition and the derivation of an omitted variables bias formula for QR.

Theorem 2 (Iterative Approximation Property) *Suppose that (i) the conditional density $f_Y(y|X)$ exists and is bounded a.s., (ii) $E[Y]$, $E[Q_\tau(Y|X)^2]$, and $E\|X\|^2$ are finite, and (iii) $\beta(\tau)$ uniquely solves (2). Then $\bar{\beta}(\tau) = \beta(\tau)$ uniquely solves the equation*

$$\bar{\beta}(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E\left[\bar{w}_\tau(X, \bar{\beta}(\tau)) \cdot \Delta_\tau^2(X, \beta)\right], \tag{9}$$

where

$$\begin{aligned} \bar{w}_\tau(X, \bar{\beta}(\tau)) &= \frac{1}{2}\int_0^1 f_{\epsilon_\tau}\left(u \cdot \Delta_\tau(X, \bar{\beta}(\tau))|X\right)\, du \\ &= \frac{1}{2}\int_0^1 f_Y\left(u \cdot X'\bar{\beta}(\tau) + (1-u)\cdot Q_\tau(Y|X)|X\right)\, du \geq 0. \end{aligned}$$

Proof: We want show that

$$\beta(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E[\rho_\tau(Y - X'\beta)], \tag{10}$$

is equivalent to the fixed point $\bar{\beta}(\tau)$ that uniquely solves

$$\bar{\beta}(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E\left[\bar{w}_\tau(X, \bar{\beta}(\tau)) \cdot \Delta_\tau^2(X, \beta)\right], \tag{11}$$

where the former and the latter objective functions are finite by conditions (i) and (ii).

By convexity of (11) in β, any fixed point $\beta = \bar{\beta}(\tau)$ solves the first order condition:

$$\mathcal{F}(\beta) := \; 2 \cdot E\left[\bar{w}_\tau(X, \beta)\ \Delta_\tau(X, \beta)\ X\right] = 0.$$

By convexity of (10) in β, the quantile regression vector $\beta = \beta(\tau)$ solves the first order condition:

$$\mathcal{D}(\beta) := E\left[\mathcal{D}(X, \beta)\right] = 0,$$

[5]The remainder term $\varrho_\tau(X) = w_\tau(X, \beta) - (1/2) \cdot f_{\epsilon_\tau}(0|X)$ is bounded as $|\varrho_\tau(X)| = |\int_0^1 (1-u)(f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \beta)|X) - f_{\epsilon_\tau}(0|X))du| \leq |\Delta_\tau(X, \beta)| \cdot \bar{f}'(X) \cdot \int_0^1 (1-u)\cdot u \cdot du = (1/6) \cdot |\Delta_\tau(X, \beta)| \cdot \bar{f}'(X)$.

[6]A location-scale model is any model of the form $Y = \mu(X) + \sigma(X) \cdot e$, where e is independent of X. The location model results from setting $\sigma(X) = \sigma$.

where

$$\mathcal{D}(X,\beta) := E\left[(1\{\epsilon_\tau < \Delta_\tau(X,\beta)\} - \tau)\ X|X\right].$$

An argument similar to that used to establish equation (6) yields

$$\begin{aligned}
\mathcal{D}(X,\beta) &= (F_{\epsilon_\tau}(\Delta_\tau(X,\beta)|X) - F_{\epsilon_\tau}(0|X)) \cdot X \\
&= \left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du\right) \cdot \Delta_\tau(X,\beta) \cdot X \\
&= 2 \cdot \bar{w}_\tau(X,\beta) \cdot \Delta_\tau(X,\beta) \cdot X,
\end{aligned}$$

where we also use the definition of $\bar{w}_\tau(X;\beta)$. The functions $\mathcal{F}(\beta)$ and $\mathcal{D}(\beta)$ are therefore identical. Since $\beta = \beta(\tau)$ uniquely satisfies $\mathcal{D}(\beta) = 0$, it also uniquely satisfies $\mathcal{F}(\beta) = 0$. As a result, $\beta = \beta(\tau) = \bar{\beta}(\tau)$ is the unique solution to both (10) and (11). *Q.E.D.*

Theorem 2 differs from Theorem 1 in that it characterizes the QR coefficient as a fixed point to an iterated minimum distance approximation. Consequently, the importance weights $\bar{w}(X,\beta(\tau))$ in this approximation are defined using the QR vector $\beta(\tau)$ itself. The weighting function $\bar{w}_\tau(X,\beta(\tau))$ is also related to the conditional density of the dependent variable. In particular, when the response variable has a smooth conditional density around the relevant quantile, we have by a Taylor approximation

$$\bar{w}_\tau(X,\beta(\tau)) = 1/2 \cdot f_Y\left(Q_\tau(Y|X)|X\right) + \bar{\varrho}_\tau(X), \quad |\bar{\varrho}_\tau(X)| \leq 1/4 \cdot |\Delta_\tau(X,\beta(\tau))| \cdot \bar{f}'(X),$$

where $\bar{\varrho}_\tau(X)$ is a remainder term, and the density $f_Y\left(y|X\right)$ is assumed to have a first derivative in y bounded in absolute value by $\bar{f}'(X)$ a.s. When either $\Delta_\tau(X,\beta(\tau))$ or $\bar{f}'(X)$ is small, we then have

$$\bar{w}_\tau(X,\beta(\tau)) \approx w_\tau(X,\beta(\tau)) \approx \frac{1}{2} f_Y(Q_\tau(Y|X)|X).$$

The approximate weighting function is therefore the same as derived using Theorem 1.

2.3 Partial Quantile Regression and Omitted Variable Bias

Partial quantile regression is defined with regard to a partition of the regressor vector X into a variable, X_1, and the remaining variables X_2, along with the corresponding partition of QR coefficients $\beta(\tau)$ into $\beta_1(\tau)$ and $\beta_2(\tau)$. We can now decompose $Q_\tau(Y|X)$ and X_1 using orthogonal projections onto X_2 weighted by $\bar{w}_\tau(X) := \bar{w}(X;\beta(\tau))$ defined in Theorem 2:

$$\begin{aligned}
Q_\tau(Y|X) &= X_2'\pi_Q + q_\tau(Y|X), &&\text{where } E[\bar{w}_\tau(X) \cdot X_2 \cdot q_\tau(Y|X)] = 0, \\
X_1 &= X_2'\pi_1 + V_1, &&\text{where } E[\bar{w}_\tau(X) \cdot X_2 \cdot V_1] = 0.
\end{aligned}$$

In this decomposition, $q_\tau(Y|X)$ and V_1 are residuals created by a weighted linear projection of $Q_\tau(Y|X)$ and X_1 on X_2, respectively, using $\bar{w}_\tau(X)$ as the weight.[7] Standard least squares algebra then gives

$$\beta_1(\tau) = \arg\min_{\beta_1} E\left[\bar{w}_\tau(X)\ (q_\tau(Y|X) - V_1\beta_1)^2\right]$$

and also $\beta_1(\tau) = \arg\min_{\beta_1} E\left[\bar{w}_\tau(X)\ (Q_\tau(Y|X) - V_1\beta_1)^2\right]$. This shows that $\beta_1(\tau)$ is a partial quantile regression coefficient in the sense that it can be obtained from a weighted least squares regression of $Q_\tau(Y|X)$ on X_1, once we have partialled out the effect of X_2. Both the first-step and second-step regressions are weighted by $\bar{w}_\tau(X)$.

We can similarly derive an omitted variables bias formula for QR. In particular, suppose we are interested in a quantile regression with explanatory variables $X = [X_1', X_2']'$, but X_2 is not available, e.g., a measure of ability or family background in a wage equation. We run QR on X_1 only, obtaining the coefficient vector $\gamma_1(\tau) = \arg\min_{\gamma_1} E[\rho_\tau(Y - X_1'\gamma_1)]$. The long regression coefficient vectors are given by $(\beta_1(\tau)', \beta_2(\tau)')' = \arg\min_{\beta_1,\beta_2} E[\rho_\tau(Y - X_1'\beta_1 - X_2'\beta_2)]$. Then,

$$\gamma_1(\tau) = \beta_1(\tau) + (E[\tilde{w}_\tau(X) \cdot X_1X_1'])^{-1}E[\tilde{w}_\tau(X) \cdot X_1R_\tau(X)],$$

where $R_\tau(X) := Q_\tau(Y|X) - X_1'\beta_1(\tau)$, $\tilde{w}_\tau(X) := \int_0^1 f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \gamma_1(\tau))|X)du/2$, $\Delta_\tau(X, \gamma_1) := X_1'\gamma_1 - Q_\tau(Y|X)$, and $\epsilon_\tau := Y - Q_\tau(Y|X)$.[8] Here $R_\tau(X)$ is the part of the CQF not explained by the linear function of X_1 in the long QR. If the CQF is linear, then $R_\tau(X) = X_2'\beta_2(\tau)$. The proof of this result is similar to the previous arguments and therefore omitted.

As with OLS short and long calculations, the omitted variables formula in this case shows the short QR coefficients to be equal to the corresponding long QR coefficients plus the coefficients in a weighted projection of omitted effects on included variables. While the parallel with OLS seems clear, there are two complications in the QR case. First, the effect of omitted variables appears through the remainder term, $R_\tau(X)$. In practice, it seems reasonable to think of this as being approximated by the omitted linear part, $X_2'\beta_2(\tau)$. Second, the regression of omitted variables on included variables is weighted by $\tilde{w}_\tau(X)$, while for OLS it is unweighted.[9]

3 Sampling Properties of QR Under Misspecification

Parallelling the interest in robust inference methods for OLS, it is also of interest to know how specification error affects inference for QR. In this case, inference under misspecification means

[7] Thus, $\pi_Q = E\ [\bar{w}_\tau(X)X_2X_2']^{-1}E\ [\bar{w}_\tau(X)X_2Q_\tau(Y|X)]$ and $\pi_1 = E\ [\bar{w}_\tau(X)X_2X_2']^{-1}E\ [\bar{w}_\tau(X)X_2X_1]$.

[8] Note that the weights in this case depend on how the regressor vector is partitioned.

[9] The formula obtained above can be used to determine the bias from measurement error in regressors, by setting the error to be the omitted variable. This suggests that classical measurement error is likely to generate an attenuation bias in QR as well as OLS estimates. We thank Arthur Lewbel for pointing this out.

distribution theory for quantile regressions in large samples without imposing the restriction that the CQF is linear. While not consistent for the true nonlinear CQF, quantile regression consistently estimates the approximations to the CQF given in Theorems 1 and 2. We would therefore like to quantify the sampling uncertainty in estimates of these approximations. This question can be compactly and exhaustively addressed by obtaining the large sample distribution of the sample quantile regression process, which is defined by taking all or many sample quantile regressions.

As in Koenker and Xiao (2001), the entire QR process is of interest here because we would like to either test global hypotheses about (approximations to) conditional distributions or make comparisons across different quantiles. Therefore our interest is in the QR process, and is not confined to a specific quantile. The second motivation for studying the process comes from the fact that formal statistical comparisons across quantiles, often of interest in empirical work, require the construction of simultaneous (joint) confidence regions. Process methods provide a natural and simple way of constructing these regions.

The QR process $\hat{\beta}(\cdot)$ is formally defined as

$$\hat{\beta}(\tau) \in \arg\min_{\beta\in\mathbb{R}^d} \; n^{-1}\sum_{i=1}^{n}\rho_\tau(Y_i - X_i'\beta),\; \tau \in \mathcal{T} := \text{ a closed subset of } [\epsilon, 1-\epsilon] \text{ for } \epsilon > 0. \quad (12)$$

Koenker and Machado (1999) and Koenker and Xiao (2001) previously focused on QR process inference in correctly specified models, while earlier treatments of specification error discussed only pointwise inference for a single quantile coefficient (Hahn, 1997). As it turns out, the empirical results in the next section show misspecification has a larger effect on process inference than on pointwise inference. Our main theoretical result on inference is as follows:

Theorem 3 *Suppose that (i) $(Y_i, X_i, i \leq n)$ are iid on the probability space $(\Omega, \mathcal{F}, P)$ for each n, (ii) the conditional density $f_Y(y|X=x)$ exists, and is bounded and uniformly continuous in y, uniformly in x over the support of X, (iii) $J(\tau) := E[f_Y(X'\beta(\tau)|X)XX']$ is positive definite for all $\tau \in \mathcal{T}$, and (iv) $E\ \|X\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then, the quantile regression process is uniformly consistent, $\sup_{\tau\in\mathcal{T}}\|\hat{\beta}(\tau) - \beta(\tau)\| = o_p(1)$, and $J(\cdot)\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ converges in distribution to a zero mean Gaussian process $z(\cdot)$, where $z(\cdot)$ is defined by its covariance function $\Sigma(\tau,\tau') := E\ [z(\tau)z(\tau')']$, with*

$$\Sigma(\tau,\tau') = E\ \left[\left(\tau - 1\left\{Y < X'\beta(\tau)\right\}\right)\left(\tau' - 1\left\{Y < X'\beta(\tau')\right\}\right)XX'\right]. \quad (13)$$

If the model is correctly specified, i.e. $Q_\tau(Y|X) = X'\beta(\tau)$ a.s., then $\Sigma(\tau,\tau')$ simplifies to

$$\Sigma_0(\tau,\tau') := [\min(\tau,\tau') - \tau\tau'] \cdot E\ [XX']. \quad (14)$$

Theorem 3 establishes joint asymptotic normality for the entire QR process.[10] The proof of this theorem appears in the supplementary appendix. Theorem 3 allows for misspecification and imposes little structure on the underlying conditional quantile function (e.g., smoothness of $Q_\tau(Y|X)$ in X, needed for a fully nonparametric approach, is not needed here). The result states that the limiting distribution of the QR process (and of any single QR coefficient) will in general be affected by misspecification. The covariance function that describes the limiting distribution is generally different from the covariance function that arises under correct specification.

Inference on the QR process is useful for testing basic hypotheses of the form:

$$R(\tau)'\beta(\tau) = r(\tau) \text{ for all } \tau \in \mathcal{T}. \tag{15}$$

For example, we may be interested in whether a variable or a subset of variables $j \in \{k+1, ..., d\}$ enter the regression equations at all quantiles with zero coefficients, i.e. whether $\beta_j(\tau) = 0$ for all $\tau \in \mathcal{T}$ and $j \in \{k+1, ..., d\}$. This corresponds to $R(\tau) = \left[0_{(d-k)\times k}\ I_{d-k}\right]$ and $r(\tau) = 0_{d-k}$. Similarly, we may want to construct simultaneous (uniform) confidence intervals for linear functions of parameters

$$R(\tau)'\beta(\tau) - r(\tau) \quad \text{for all } \tau \in \mathcal{T}.$$

Theorem 3 has a direct consequence for these confidence intervals and hypothesis testing, since it implies that $(EXX')^{-1}\Sigma(\tau, \tau') \neq [\min(\tau, \tau') - \tau\tau'] \cdot I_d$. That is, the covariance function under misspecification is not proportional to the covariance function of the standard d-dimensional Brownian bridge arising in the correctly specified case. Hence, unlike in the correctly specified case, the critical values for confidence regions and tests are not distribution-free and can not be obtained from standard tabulations based on the Brownian bridge. However, the following corollaries facilitate both testing and the construction of confidence intervals under misspecification:

Corollary 1 *Define* $V(\tau) := R(\tau)'J(\tau)^{-1}\ \Sigma(\tau,\tau)\ J(\tau)^{-1}R(\tau)$ *and* $|x| := \max_j |x_j|$. *Under the conditions of Theorem 3, the Kolmogorov statistic* $\mathcal{K}_n := \sup_{\tau\in\mathcal{T}} |V(\tau)^{-1/2}\sqrt{n}(R(\tau)'\hat{\beta}(\tau) - r(\tau))|$ *for testing (15) converges in distribution to variable* $\mathcal{K} := \sup_{\tau\in\mathcal{T}} |V(\tau)^{-1/2}R(\tau)'J(\tau)^{-1}z(\tau)|$ *with an absolutely continuous distribution. The result is not affected by replacing* $J(\tau)$ *and* $\Sigma(\tau,\tau)$ *with estimates that are consistent uniformly in* $\tau \in \mathcal{T}$.

[10]A simple corollary is that any finite collection of $\sqrt{n}(\hat{\beta}(\tau_k) - \beta(\tau_k))$, $k = 1, 2, ...$, are asymptotically jointly normal, with asymptotic covariance between the k-th and l-th subsets equal to $J(\tau_k)^{-1}\Sigma(\tau_k, \tau_l)J(\tau_l)^{-1}$. Hahn (1997) previously derived this corollary for a single quantile.

Thus, Kolmogorov-type statistics have a well-behaved limit distribution.[11] Unlike in the correctly specified case, however, this distribution is non-standard. Nevertheless, critical values and simultaneous confidence regions can be obtained as follows:

Corollary 2 *For $\kappa(\alpha)$ denoting the α-quantile of $\mathcal{K}$ and $\hat{\kappa}(\alpha)$ any consistent estimate of it, for instance the estimate defined below, $\lim_{n\to\infty} P\{\sqrt{n}(R(\tau)'\beta(\tau) - r(\tau)) \in \widehat{I}_n(\tau),\ \text{for all } \tau \in \mathcal{T}\} = \alpha$, where $\widehat{I}_n(\tau) = [u(\tau) : |V(\tau)^{-1/2}\sqrt{n}(R(\tau)'\hat{\beta}(\tau) - r(\tau) - u(\tau))| \leq \hat{\kappa}(\alpha)]$. If $R(\tau)'\beta(\tau) - r(\tau)$ is scalar, the simultaneous confidence interval is $\widehat{I}_n(\tau) = [R(\tau)'\hat{\beta}(\tau) - r(\tau) \pm \hat{\kappa}(\alpha) \cdot V(\tau)^{1/2}]$. This result is not affected by replacing $V(\tau)$ with an estimate that is consistent uniformly in $\tau \in \mathcal{T}$.*

A consistent estimate of the critical value, $\hat{\kappa}(\alpha)$, can be obtained by subsampling. Let $j = 1, ..., B$ index B randomly chosen subsamples of $((Y_i, X_i), i \leq n)$ of size b, where $b \to \infty$, $b/n \to 0$, $B \to \infty$ as $n \to \infty$. Compute the test statistic for each subsample as $K_j = \sup_{\tau\in\mathcal{T}} |\hat{V}(\tau)^{-1/2}\sqrt{b}R(\tau)'(\widehat{\beta}_j(\tau) - \widehat{\beta}(\tau))|$, where $\widehat{\beta}_j(\tau)$ is the QR estimate using j-th subsample. Then, set $\widehat{\kappa}(\alpha)$ to be the α-quantile of $\{K_1, ..., K_B\}$.

Finally, the inference procedure above requires estimators of $\Sigma(\tau, \tau')$ and $J(\tau)$ that are uniformly consistent in $(\tau, \tau') \in \mathcal{T} \times \mathcal{T}$. These are given by:

$$\begin{aligned}
\hat{\Sigma}(\tau, \tau') &= n^{-1}\sum_{i=1}^{n}(\tau - 1\{Y_i \leq X_i'\hat{\beta}(\tau)\})(\tau' - 1\{Y_i \leq X_i'\hat{\beta}(\tau')\}) \cdot X_i X_i', \\
\hat{J}(\tau) &= (2nh_n)^{-1}\sum_{i=1}^{n} 1\{|Y_i - X_i'\hat{\beta}(\tau)| \leq h_n\} \cdot X_i X_i',
\end{aligned}$$

where $\hat{\Sigma}(\tau, \tau')$ differs from its usual counterpart $\hat{\Sigma}_0(\tau, \tau') = [\min(\tau, \tau') - \tau\tau'] \cdot n^{-1}\sum_{i=1}^{n} X_i X_i'$ used in the correctly specified case; and $\hat{J}(\tau)$ is Powell's (1986) estimator of the Jacobian, with h_n such that $h_n \to 0$ and $h_n^2 n \to \infty$. Koenker (1994) suggests $h_n = C \cdot n^{-1/3}$ and provides specific choices of C. The supplementary appendix shows that these estimates are consistent uniformly in (τ, τ').

4 Application to U.S. Wage Data

In this section we study the approximation properties of QR in widely used U.S. Census micro data sets.[12] The main purpose of this section is to show that linear QR indeed provides a useful

[11] In practice, by stochastic equicontinuity of QR process, in the definition of $\mathcal{K}_n$-statistics we can replace any continuum of quantile indices $\mathcal{T}$ by a finite-grid $\mathcal{T}_{K_n}$, where the distance between adjacent grid points goes to zero as $n \to \infty$.

[12] The data were drawn from the 1% self-weighted 1980 and 1990 samples, and the 1% weighted 2000 sample, all from the IPUMS website (Ruggles *et al.,* 2003). The sample consists of US-born black and white men of age

minimum distance approximation to the conditional distribution of wages, accurately capturing changes in the wage distribution from 1980 to 2000. We also report new substantive empirical findings arising from the juxtaposition of data from the 2000 census with earlier years. The inference methods derived in the previous section facilitate this presentation. In our analysis, Y is the real log weekly wage for U.S. born men aged 40-49, calculated as the log of reported annual income from work divided by weeks worked in the previous year, and the regressor X consists of a years-of-schooling variable and other basic controls.[13]

The nature of the QR approximation property is illustrated in Figure 1. Panels A-C plot a nonparametric estimate of the conditional quantile function, $Q_\tau(Y|X)$, along with the linear QR fit for the 0.10, 0.50, and 0.90 quantiles, where X includes only the schooling variable. Here we take advantage of the discreteness of the schooling variable and the large census sample to compare QR fits to the nonlinear CQFs computed at each point in the support of X. We focus on the 1980 data for this figure because the 1980 Census has a true highest grade completed variable, while for more recent years this must be imputed. It should be noted, however, that the approximation results for the 1990 and 2000 censuses are similar.

Our theorems establish that QR implicitly provides a weighted minimum distance approximation to the true nonlinear CQF. It is therefore useful to compare the QR fit to an explicit minimum distance (MD) fit similar to that discussed by Chamberlain (1994).[14] The MD estimator for QR is the sample analog of the vector $\tilde{\beta}(\tau)$ solving

$$\tilde{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E\left[(Q_\tau(Y|X) - X'\beta)^2\right] = \arg \min_{\beta \in \mathbb{R}^d} E\left[\Delta_\tau^2(X,\beta)\right].$$

In other words, $\tilde{\beta}(\tau)$ is the slope of the linear regression of $Q_\tau(Y|X)$ on X, weighted only by the probability mass function of X, $\pi(x)$. In contrast to QR, this MD estimator relies on the ability to estimate $Q_\tau(Y|X)$ in a nonparametric first step, which, as noted by Chamberlain (1994), may be feasible only when X is low dimensional, the sample size is large, and sufficient smoothness of $Q_\tau(Y|X)$ is assumed.

Figure 1 plots this MD fit with a dashed line. The QR and MD regression lines are close, as predicted by our approximation theorems, but they are not identical because the additional

40-49 with 5 or more years of education, with positive annual earnings and hours worked in the year preceding the census. Individuals with imputed values for age, education, earnings or weeks worked were also excluded from the sample. The resulting sample sizes were 65,023, 86,785, and 97,397 for 1980, 1990, and 2000.

[13]Annual income is expressed in 1989 dollars using the Personal Consumption Expenditures Price Index. The schooling variable for 1980 corresponds to the highest grade of school completed. The categorical schooling variables in the 1990 and 2000 Census were converted to years of schooling using essentially the same coding scheme as in Angrist and Krueger (1999). See Angrist, Chernozhukov, and Fernandez-Val (2004) for details.

[14]See Ferguson (1958) and Rothenberg (1971) for general discussions of MD. Buchinsky (1994) and Bassett, Knight, and Tam (2002) present other applications of MD to quantile problems.

weighting by $w_\tau(X, \beta)$ in the QR fit accentuates quality of the fit at values of X where Y is more densely distributed near true quantiles. To further investigate the QR weighting function, panels D-F in Figure 1 plot the overall QR weights, $w_\tau(X, \beta(\tau)) \cdot \pi(X)$, against the regressor X. The panels also show estimates of the importance weights from Theorem 1, $w_\tau(X, \beta(\tau))$, and their density approximations, $f_Y(Q_\tau(Y|X)|X)$.[15] The importance weights and the actual density weights are fairly close. The importance weights are stable across X and tend to accentuate the middle of the distribution a bit more than other parts. The overall weighting function ends up placing the highest weight on 12 years of schooling, implying that the linear QR fit should be the best in the middle of the design.

Also of interest is the ability of QR to track changes in quantile-based measures of conditional inequality. The column labeled CQ in panel A of Table 1 shows nonparametric estimates of the average 90-10 quantile spread conditional on schooling, potential experience, and race. This spread increased from 1.2 to about 1.35 from 1980 to 1990, and then to about 1.43 from 1990 to 2000. QR estimates match this almost perfectly, not surprisingly since an implication of our theorems is that QR should fit (weighted) average quantiles exactly. The fit is not as good, however, when averages are calculated for specific schooling groups, as reported in panels B and C of the table. These results highlight the fact that QR is only an approximation. Table 1 also documents two important substantive findings, apparent in both the CQ and QR estimates. First, the table shows conditional inequality increasing in both the upper and lower halves of the wage distribution from 1980 to 1990, but in the top half only from 1990 to 2000. Second, the increase in conditional inequality since 1990 has been much larger for college graduates than for high school graduates.

Figure 2 provides a useful complement to, and a partial explanation for, the patterns and changes in Table 1. In particular, Panel A of the figure shows estimates of the schooling coefficient quantile process, along with robust simultaneous 95% confidence intervals. These estimates are from quantile regressions of log-earnings on schooling, race and a quadratic function of experience, using data from the 1980, 1990 and 2000 censuses.[16] The robust simultaneous

[15]The importance weights defined in Theorem 1 are estimated at $\beta = \hat{\beta}(\tau)$ as follows:

$$\widehat{w}_\tau(X, \widehat{\beta}(\tau)) = (1/U) \cdot \sum_{u=1}^{U} [(1 - u/U) \cdot \hat{f}_Y((u/U) \cdot X'\widehat{\beta}(\tau) + (1 - u/U) \cdot \widehat{Q}_\tau(Y|X)|X)], \quad (16)$$

where U is set to 100; $\hat{f}_Y(y|X)$ is a kernel density estimate of $f_Y(y|X)$, which employs a Gaussian kernel and Silverman's rule for bandwidth; $\widehat{Q}_\tau(Y|X)$ is a non-parametric estimate of $Q_\tau(Y|X)$, for each cell of the covariates X; and $X'\widehat{\beta}(\tau)$ is the QR estimate. Approximate weights are calculated similarly. Weights based on Theorem 2 are similar and therefore not shown.

[16]The simultaneous bands were obtained by subsampling using 500 repetitions with subsample size $b = 5n^{2/5}$ and a grid of quantiles $\mathcal{T}_n = \{.10, .11, ..., .90\}$.

confidence intervals allow us to asses of the significance of changes in schooling coefficients across quantiles and across years. The horizontal lines in the figure indicate the corresponding OLS estimates.

The figure suggests the returns to schooling were low and essentially constant across quantiles in 1980, a finding similar to Buchinsky's (1994) using Current Population Surveys for this period. On the other hand, the returns increased sharply and became much more heterogeneous in 1990 and especially in 2000, a result we also confirmed in Current Population Survey data. Since the simultaneous confidence bands do not contain a horizontal line, we reject the hypothesis of constant returns to schooling for 1990 and 2000. The fact that there are quantile segments where the simultaneous bands do not overlap indicates statistically significant differences across years at those segments. For instance, the 1990 band does not overlap with the 1980 band, suggesting a marked and statistically significant change in the relationship between schooling and the conditional wage distribution in this period.[17] The apparent twist in the schooling coefficient process explains why inequality increased for college graduates from 1990 to 2000. In the 2000 census, higher education was associated with increased wage dispersion to a much greater extent than in earlier years.

Another view of the stylized facts laid out in Table 1 is given in Figure 2B. This figure plots changes in the approximate conditional quantiles, based on a QR fit, with covariates evaluated at their mean values for each year. The figure also shows simultaneous 95% confidence bands. This figure provides a visual representation of the finding that between 1990 and 2000 conditional wage inequality increased more in the upper half of the wage distribution than in the lower half, while between 1980 and 1990 the increase in inequality occurred in both tails. Changes in schooling coefficients across quantiles and years, sharper above the median than below, clearly contributed to the fact that recent (conditional) inequality growth has been mostly confined to the upper half of the wage distribution.

Finally, it is worth noting that the simultaneous bands differ from the corresponding pointwise bands (the latter are not plotted). Moreover, the simultaneous bands allow multiple comparisons across quantiles without compromising confidence levels. Even more importantly in our context, accounting for misspecification substantially affects the width of simultaneous confidence intervals in this application. Uniform bands calculated assuming correct specification can be constructed using the critical values for the Kolmogorov statistic $\mathcal{K}$ reported in Andrews

[17] Due to independence of samples across Census years, the test that looks for overlapping in two 95% confidence bands has a significance level of about 10%, namely $1 - .95^2$. Alternately, an α-level test can be based on a simultaneous α-level confidence band for the difference in quantile coefficients across years, again constructed using Theorem 3.

(1993). In this case, the resulting bands for the schooling coefficient quantile process are 26%, 23%, and 32% narrower than the robust intervals plotted in Figure 2A for 1980, 1990, and 2000.[18]

5 Summary and conclusions

We have shown how linear quantile regression provides a weighted least squares approximation to an unknown and potentially nonlinear conditional quantile function, much as OLS provides a least squares approximation to a nonlinear CEF. The QR approximation property leads to partial quantile regression relationships and an omitted variables bias formula analogous to those for OLS. While misspecification of the CQF functional form does not affect the usefulness of QR, it does have implications for inference. We also present a misspecification-robust distribution theory for the QR process. This provides a foundation for simultaneous confidence intervals and a basis for global tests of hypotheses about distributions.

An illustration using US. census data shows the sense in which QR fits the CQF. The empirical example also shows that QR accurately captures changes in the wage distribution from 1980 to 2000. An important substantive finding is the sharp twist in schooling coefficients across quantiles in the 2000 census. We use simultaneous confidence bands robust to misspecification, to show that this pattern is highly significant. A related finding is that most inequality growth after 1990 has been in the upper part of the wage distribution.

[18]The simultaneous bands take the form of $\hat{\beta}(\tau) \pm \hat{\kappa}(\alpha) \cdot$ robust std. error$(\hat{\beta}(\tau))$. Using the procedure described in Corollary 2, we obtain estimates for $\hat{\kappa}(0.05)$ of 3.78, 3.70, and 3.99 for 1980, 1990, and 2000. The simultaneous bands that impose correct specification take the form $\hat{\beta}(\tau) \pm \kappa_0(\alpha) \cdot$ std. error$(\hat{\beta}(\tau))$, where $\kappa_0(\alpha)$ is the α-quantile of the supremum of (the absolute value of) a standardized tied-down Bessel process of order 1. For example, $\kappa_0(0.05) = (9.31)^{1/2} = 3.05$ from Table I in Andrews (1993).

References

[1] Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, pp. 821856.

[2] Abadie, A. (1997): "Changes in Spanish Labor Income Structure during the 1980's: A Quantile Regression Approach," *Investigaciones Economicas* XXI(2), pp. 253-272.

[3] Angrist, J., and A. Krueger (1999): "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3. Amsterdam. Elsevier Science.

[4] Autor, D., L.F. Katz, and M.S. Kearney (2004): "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists," MIT Department of Economics, mimeo, August 2004.

[5] Bassett, G.W., M.-Y. S. Tam, and K. Knight (2002): "Quantile Models and Estimators for Data Analysis," *Metrika* 55, pp. 17-26.

[6] Buchinsky, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, pp. 405-458.

[7] Chamberlain, G. (1984): "Panel Data," in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Volume 2. North-Holland. Amsterdam.

[8] Chamberlain, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics, Sixth World Congress,* Volume 1. Cambridge University Press. Cambridge.

[9] Chernozhukov, V. (2002): "Inference on the Quantile Regression Process, An Alternative," Unpublished Working Paper 02-12 (February), MIT (www.ssrn.com).

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

[11] Davydov, Yu. A., M. A. Lifshits, and N. V. Smorodina (1998): *Local Properties of Distributions of Stochastic Functionals*. Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. Translations of Mathematical Monographs, 173. American Mathematical Society, Providence, RI.

[12] Ferguson, T. S. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *The Annals of Mathematical Statistics* 29(4), pp. 1046-1062.

[13] Giacomini, R., and I. Komunjer (2003): "Evaluation and Combination of Conditional Quantile Forecasts," Working Paper 571, Boston College and California Institute of Technology, 06/2003.

[14] Goldberger, A. S. (1991): *A Course in Econometrics*. Harvard University Press. Cambridge, MA.

[15] Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the U.K.," *Review of Economic Studies* 67, pp. 635-666.

[16] Gutenbrunner, C., and J. Jurečková (1992): "Regression Quantile and Regression Rank Score Process in the Linear Model and Derived Statistics," *Annals of Statistics* 20, pp. 305-330.

[17] Hahn, J. (1997): "Bayesian Bootstrap of the Quantile Regression Estimator: A Large Sample Study," *International Economic Review* 38(4), pp. 795-808.

[18] Hansen, L.-P. , J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14(3), pp. 262-280.

[19] Juhn, C., K. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Return to Skill," *Journal of Political Economy* 101, pp. 410-422.

[20] Katz, L., and D. Autor (1999): "Changes in the Wage Structure and Earnings Inequality," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3A. Elsevier Science. Amsterdam.

[21] Katz, L., and K. Murphy (1992): "Changes in the Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics* 107, pp. 35-78.

[22] Kim T.H., and H. White (2002): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions," in *Advances in Econometrics*, forthcoming.

[23] Koenker, R. (1994): "Confidence Intervals for Regression Quantiles," in M.P. and M. Hušková (eds.), *Asymptotic Statistics: Proceeding of the 5th Prague Symposium on Asymptotic Statistics*. Physica-Verlag.

[24] Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33-50.

[25] Koenker, R., and J. A. Machado (1999): "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association* 94(448), pp. 1296-1310.

[26] Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica* 70, no. 4, pp. 1583–1612.

[27] Lemieux, T. (2003): "Residual Wage Inequality: A Re-examination," Mimeo. University of British Columbia.

[28] Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, no. 2, 186–199.

[29] Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* Springer-Verlag. New York.

[30] Portnoy, S. (1991): "Asymptotic Behavior of Regression Quantiles in Nonstationary, Dependent Cases" *Journal of Multivariate Analysis* 38, no. 1, pp. 100-113.

[31] Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics* 32, no. 1, pp. 143-155.

[32] Powell, J. L. (1994): "Estimation of Semiparametric Models," in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Volume IV. Amsterdam. Elsevier Science.

[33] Rothenberg, T. J. (1973): "Comparing Efficient Estimation with A Priori Information," Cowles Foundation Monograph 23. Yale University.

[34] Ruggles, S., and M. Sobek *et al.* (2003): *Integrated Public Use Microdata Series: Version 3.0.* Minneapolis: Historical Census Project. University of Minesota.

[35] Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag. New York.

[36] White, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review* 21(1), pp. 149-170.

SUPPLEMENTARY APPENDIX.

A Proof of Theorems 3 and its Corollaries

The proof has two steps.[19] The first step establishes uniform consistency of the sample QR process. The second step establishes asymptotic Gaussianity of the sample QR process.[20] For $W = (Y, X)$, let $\mathbb{E}_n[f(W)]$ denote $n^{-1}\sum_{i=1}^n f(W_i)$ and $\mathbb{G}_n[f(W)]$ denote $n^{-1/2}\sum_{i=1}^n (f(W_i) - E[f(W_i)])$. If $\widehat{f}$ is an estimated function, $\mathbb{G}_n[\widehat{f}(W)]$ denotes $n^{-1/2}\sum_{i=1}^n (f(W_i) - E[f(W_i)])_{f=\widehat{f}}$.

A.1 Uniform consistency of $\hat{\beta}(\cdot)$

For each τ in $\mathcal{T}$, $\hat{\beta}(\tau)$ minimizes $Q_n(\tau, \beta) := \mathbb{E}_n[\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]$. Define $Q_\infty(\tau, \beta) := E[\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]$. It is easy to show that $E\|X\| < \infty$ implies that $E|\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))| < \infty$. Therefore, $Q_\infty(\tau, \beta)$ is finite, and by the stated assumptions, it is uniquely minimized at $\beta(\tau)$ for each τ in $\mathcal{T}$.

We first show the uniform convergence, namely for any compact set $\mathcal{B}$, $Q_n(\tau, \beta) = Q_\infty(\tau, \beta) + o_{p^*}(1)$, uniformly in $(\tau, \beta) \in \mathcal{T} \times \mathcal{B}$. This statement holds pointwise by the Khinchine law of large numbers. The uniform convergence follows because $|Q_n(\tau', \beta') - Q_n(\tau'', \beta'')| \le C_{1n} \cdot |\tau' - \tau''| + C_{2n} \cdot \|\beta' - \beta''\|$, where $C_{1n} = 2 \cdot \mathbb{E}_n\|X\| \cdot \sup_{\beta \in \mathcal{B}} \|\beta\| = O_p(1)$ and $C_{2n} = 2 \cdot \mathbb{E}_n\|X\| = O_p(1)$. Hence the empirical process $(\tau, \beta) \mapsto Q_n(\tau, \beta)$ is stochastically equicontinuous, which implies the uniform convergence.

Next, we show uniform consistency. Consider a collection of closed balls $B_M(\beta(\tau))$ of radius M and center $\beta(\tau)$, and let $\beta_M(\tau) = \beta(\tau) + \delta_M(\tau) \cdot v(\tau)$, where $v(\tau)$ is a direction vector with unity norm $\|v(\tau)\| = 1$ and $\delta_M(\tau)$ is a positive scalar such that $\delta_M(\tau) \ge M$. Then uniformly in $\tau \in \mathcal{T}$, $(M/\delta_M(\tau)) \cdot (Q_n(\tau, \beta_M(\tau)) - Q_n(\tau, \beta(\tau))) \overset{(a)}{\ge} Q_n(\tau, \beta_M^*(\tau)) - Q_n(\tau, \beta(\tau)) \overset{(b)}{\ge} Q_\infty(\tau, \beta_M^*(\tau)) - Q_\infty(\tau, \beta(\tau)) + o_{p^*}(1) \overset{(c)}{>} \epsilon_M + o_{p^*}(1)$, for some $\epsilon_M > 0$; where (a) follows by convexity in β, for $\beta_M^*(\tau)$ the point of the boundary of $B_M(\beta(\tau))$ on the line connecting $\beta_M(\tau)$ and $\beta(\tau)$; (b) follows by the uniform convergence established above; and (c) follows since $\beta(\tau)$ is the unique minimizer of $Q_\infty(\beta, \tau)$ uniformly in $\tau \in \mathcal{T}$, by convexity and assumption (iii). Hence for any $M > 0$, the minimizer $\hat{\beta}(\tau)$ must be within M from $\beta(\tau)$ uniformly for all $\tau \in \mathcal{T}$, with probability approaching one.

A.2 Asymptotic Gaussianity of $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$

First, by the computational properties of $\hat{\beta}(\tau)$, for all $\tau \in \mathcal{T}$, cf. Theorem 3.3 in Koenker and Bassett (1978), we have that $\|\mathbb{E}_n[\varphi_\tau(Y - X'\hat{\beta}(\tau))X]\| \le \text{const} \cdot \sup_{i \le n} \|X_i\|/n$, where $\varphi_\tau(u) = \tau - 1\{u \le 0\}$. Note that $E\|X_i\|^{2+\varepsilon} < \infty$ implies $\sup_{i \le n} \|X_i\| = o_{p^*}(n^{1/2})$, since $P\left(\sup_{i \le n} \|X_i\| > n^{1/2}\right) \le nP(\|X_i\| > n^{1/2}) \le nE\|X_i\|^{2+\varepsilon}/n^{\frac{2+\varepsilon}{2}} = o(1)$. Hence uniformly in $\tau \in \mathcal{T}$,

$$\sqrt{n}\mathbb{E}_n\left[\varphi_\tau(Y - X'\hat{\beta}(\tau))X\right] = o_p(1). \tag{17}$$

[19]Basic concepts used in the proof, including weak convergence in the space of bounded functions, stochastic equicontinuity, Donsker and Vapnik-Červonenkis (VC) classes, are defined as in van der Vaart and Wellner (1996).

[20]The step does not rely on Pollard's (1991) convexity argument, as it does not apply to the process case.

Second, $(\tau, \beta) \mapsto \mathbb{G}_n [\varphi_\tau (Y - X'\beta) X]$ is stochastically equicontinuous over $\mathcal{B} \times \mathcal{T}$, where $\mathcal{B}$ is any compact set, with respect to the $L_2(P)$ pseudometric

$$\rho((\tau', \beta'), (\tau'', \beta''))^2 := \max_{j \in 1,\ldots,d} E\left[\left(\varphi_{\tau'} (Y - X'\beta') X_j - \varphi_{\tau''} (Y - X'\beta'') X_j\right)^2 \right],$$

for $j \in 1, ..., d$ indexing the components of X. Note that the functional class $\{\varphi_\tau (Y - X'\beta) X, \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$ is formed as $(\mathcal{T} - \mathcal{F})X$, where $\mathcal{F} = \{1\{Y \leq X'\beta\}, \beta \in \mathcal{B}\}$ is a VC subgraph class and hence a bounded Donsker class. Hence $\mathcal{T} - \mathcal{F}$ is also bounded Donsker, and $(\mathcal{T} - \mathcal{F})X$ is therefore Donsker with a square integrable envelope $2 \cdot \max_{j \in 1,..d} |X|_j$, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). The stochastic equicontinuity then is a part of being Donsker.

Third, by stochastic equicontinuity of $(\tau, \beta) \mapsto \mathbb{G}_n [\varphi_\tau (Y - X'\beta) X]$ we have that

$$\mathbb{G}_n \left[\varphi_\tau(Y - X'\hat{\beta}(\tau))X\right] = \mathbb{G}_n [\varphi_\tau(Y - X'\beta(\tau))X] + o_{p^*}(1), \quad \text{in } \ell^\infty(\mathcal{T}), \tag{18}$$

which follows from $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| = o_{p^*}(1)$, and resulting convergence with respect to the pseudometric $\sup_{\tau \in \mathcal{T}} \rho[(\tau, \hat{\beta}(\tau)), (\tau, \beta(\tau))]^2 = o_p(1)$. The latter is immediate from $\sup_{\tau \in \mathcal{T}} \rho\left[(\tau, b(\tau)), (\tau, \beta(\tau))\right]^2 \leq C_3 \cdot \sup_{\tau \in \mathcal{T}} \|b(\tau) - \beta(\tau)\|^{\frac{\varepsilon}{(2+\varepsilon)}}$, where $C_3 = (\bar{f} \cdot (E\|X\|^2)^{1/2})^{\frac{\varepsilon}{2(2+\varepsilon)}} \cdot (E\|X\|^{2+\varepsilon})^{\frac{2}{2+\varepsilon}} < \infty$ and $\bar{f}$ is the a.s. upper bound on $f_Y(Y|X)$. (This follows by the Hölder's inequality and Taylor expansion.)

Further, the following expansion is valid uniformly in $\tau \in \mathcal{T}$

$$E\left[\varphi_\tau(Y - X'\beta)X\right]\Big|_{\beta = \hat{\beta}(\tau)} = [J(\tau) + o_p(1)] \left(\hat{\beta}(\tau) - \beta(\tau)\right). \tag{19}$$

Indeed, by Taylor expansion $E\left[\varphi_\tau(Y - X'\beta)X\right]|_{\beta = \hat{\beta}(\tau)} = E\left[f_Y(X'b(\tau)|X)XX'\right]|_{b(\tau) = \beta^*(\tau)}(\hat{\beta}(\tau) - \beta(\tau))$, where $\beta^*(\tau)$ is on the line connecting $\hat{\beta}(\tau)$ and $\beta(\tau)$ for each τ, and is different for each row of the Jacobian matrix. Then, (19) follows by the uniform consistency of $\hat{\beta}(\tau)$, and the assumed uniform continuity and boundedness of the mapping $y \mapsto f_Y(y|x)$, uniformly in x over the support of X.

Fourth, since the left hand side (lhs) of (17) = lhs of $n^{1/2}$(19)+ lhs of (18), we have that

$$o_p(1) = [J(\cdot) + o_p(1)](\hat{\beta}(\cdot) - \beta(\cdot)) + \mathbb{G}_n[\varphi_\cdot(Y - X'\beta(\cdot))X]. \tag{20}$$

Therefore, using that mineig $[J(\tau)] \geq \lambda > 0$ uniformly in $\tau \in \mathcal{T}$,

$$\sup_{\tau \in \mathcal{T}} \left\|\mathbb{G}_n[\varphi_\tau(Y - X'\beta(\tau))X] + o_p(1)\right\| \geq (\sqrt{\lambda} + o_p(1)) \cdot \sup_{\tau \in \mathcal{T}} \sqrt{n}\|\hat{\beta}(\tau) - \beta(\tau)\|. \tag{21}$$

Fifth, the mapping $\tau \mapsto \beta(\tau)$ is continuous by the implicit function theorem and stated assumptions. In fact, since $\beta(\tau)$ solves $E\left[(\tau - 1\{Y \leq X'\beta\})X\right] = 0$, $d\beta(\tau)/d\tau = J(\tau)^{-1}E\left[X\right]$. Hence $\tau \mapsto \mathbb{G}_n [\varphi_\tau (Y - X'\beta(\tau)) X]$ is stochastically equicontinuous over $\mathcal{T}$ for the pseudo-metric given by $\rho(\tau', \tau'') := \rho((\tau', \beta(\tau')), (\tau'', \beta(\tau'')))$. Stochastic equicontinuity of $\tau \mapsto \mathbb{G}_n [\varphi_\tau(Y - X'\beta(\tau))X]$ and a multivariate CLT imply that

$$\mathbb{G}_n \left[\varphi_\cdot(Y - X'\beta(\cdot))X\right] \Rightarrow z(\cdot) \text{ in } \ell^\infty(\mathcal{T}), \tag{22}$$

where $z(\cdot)$ is a Gaussian process with covariance function $\Sigma(\cdot, \cdot)$ specified in the statement of Theorem 3. Therefore, the lhs of (21) is $O_p(n^{-1/2})$, implying $\sup_{\tau \in \mathcal{T}} \|\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))\| = O_{p^*}(1)$.

Finally, the latter fact and (20)-(22) imply that in $\ell^\infty(\mathcal{T})$

$$J(\cdot)\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot)) = -\mathbb{G}_n \left[\varphi_\cdot(Y - X'\beta(\cdot))\right] + o_{p^*}(1) \Rightarrow z(\cdot). \tag{23}$$

Q.E.D.

A.3 Proof of Corollaries

Proof of Corollary 1. The result follows by the continuous mapping theorem in $\ell^\infty(\mathcal{T})$. Absolute continuity of $\mathcal{K}$ follows from Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998). *Q.E.D.*

Proof of Corollary 2. The result follows by absolute continuity of $\mathcal{K}$. The consistency of subsampling estimator of $\hat{\kappa}(\alpha)$ follows from Theorem 2.2.1 and Corollary 2.4.1 in Politis, Romano and Wolf (1999), for the case when $V(\tau)$ are known. When $V(\tau)$ is estimated consistently uniformly in $\tau \in \mathcal{T}$, the result follows by an argument similar to the proof of Theorem 2.5.1 in Politis et. al. (1999). *Q.E.D.*

A.4 Uniform Consistency of $\widehat{\Sigma}(\cdot,\cdot)$ and $\widehat{J}(\cdot)$.

Here it is shown that under the conditions of Theorem 3 and the additional assumption that $E\|X\|^4 < \infty$, the estimates described in the main text are consistent uniformly in $(\tau, \tau') \in \mathcal{T} \times \mathcal{T}'$.[21]

First, recall that $\hat{J}(\tau) = [1/(2h_n)] \cdot \mathbb{E}_n[1\{|Y_i - X_i'\hat{\beta}(\tau)| \leq h_n\} \cdot X_i X_i']$. We will show that

$$\hat{J}(\tau) - J(\tau) = o_{p^*}(1) \text{ uniformly in } \tau \in \mathcal{T}. \tag{24}$$

Note that $2h_n \hat{J}(\tau) = \mathbb{E}_n[f_i(\hat{\beta}(\tau), h_n)]$, where $f_i(\beta, h) = 1\{|Y_i - X_i'\beta| \leq h\} \cdot X_i X_i'$. For any compact set B and positive constant H, the functional class $\{f_i(\beta, h), \beta \in B, h \in (0, H]\}$ is a Donsker class with a square-integrable envelope by Theorem 2.10.6 in Van der Vaart and Wellner (1996), since this is a product of a VC subgraph class $\{1\{|Y_i - X_i'\beta| \leq h\},\ \beta \in B,\ h \in (0, H]\}$ and a square integrable random matrix $X_i X_i'$ (recall $E\|X_i\|^4 < \infty$ by assumption). Therefore, $(\beta, h) \mapsto \mathbb{G}_n\left[f_i(\beta, h)\right]$ converges to a Gaussian process in $\ell^\infty(B \times (0, H])$, which implies that $\sup_{\beta \in B, 0 < h \leq H} \|\mathbb{E}_n\left[f_i(\beta, h)\right] - E\left[f_i(\beta, h)\right]\| = O_{p^*}(n^{-1/2})$. Letting B be any compact set that covers $\cup_{\tau \in \mathcal{T}} \beta(\tau)$, this implies $\sup_{\tau \in \mathcal{T}} \|\mathbb{E}_n[f_i(\hat{\beta}(\tau), h_n)] - E[f_i(\beta, h_n)]\big|_{\beta = \hat{\beta}(\tau)}\| = O_{p^*}(n^{-1/2})$. Hence (24) follows by using $2h_n \hat{J}(\tau) = \mathbb{E}_n[f_i(\hat{\beta}(\tau), h_n)]$, $1/(2h_n) \cdot E[f_i(\beta, h_n)]\big|_{\beta = \widehat{\beta}(\tau)} = J(\tau) + o_p(1)$, and the assumption $h_n^2 n \to \infty$.

Second, we can write $\hat{\Sigma}(\tau, \tau') = \mathbb{E}_n[g_i(\hat{\beta}(\tau), \hat{\beta}(\tau'), \tau, \tau') X_i X_i']$, where $g_i(\beta', \beta'', \tau', \tau'') = (\tau - 1\{Y_i \leq X_i'\beta'\})(\tau' - 1\{Y_i \leq X_i'\beta''\}) \cdot X_i X_i'$. We will show that

$$\hat{\Sigma}(\tau, \tau') - \Sigma(\tau, \tau') = o_{p^*}(1) \text{ uniformly in } (\tau, \tau') \in \mathcal{T} \times \mathcal{T}. \tag{25}$$

It is easy to verify that $\{g_i(\beta', \beta'', \tau', \tau''), (\beta', \beta'', \tau', \tau'') \in B \times B \times \mathcal{T} \times \mathcal{T}\}$ is Donsker and hence a Glivenko-Cantelli class, for any compact set B, e.g., using Theorem 2.10.6 in Van der Vaart and Wellner (1996). This implies that $\mathbb{E}_n\left[g_i(\beta', \beta'', \tau', \tau'') X_i X_i'\right] - E\left[g_i(\beta', \beta'', \tau', \tau'') X_i X_i'\right] = o_{p^*}(1)$ uniformly in $(\beta', \beta'', \tau', \tau'') \in (B \times B \times \mathcal{T} \times \mathcal{T})$. The latter and continuity of $E\left[g_i(\beta', \beta'', \tau', \tau'') X_i X_i'\right]$ in $(\beta', \beta'', \tau', \tau'')$ imply (25). *Q.E.D.*

[21]Note that the result for $\widehat{J}(\tau)$ is not covered by Powell (1986) because his proof applies only pointwise in τ, whereas we require a uniform result.

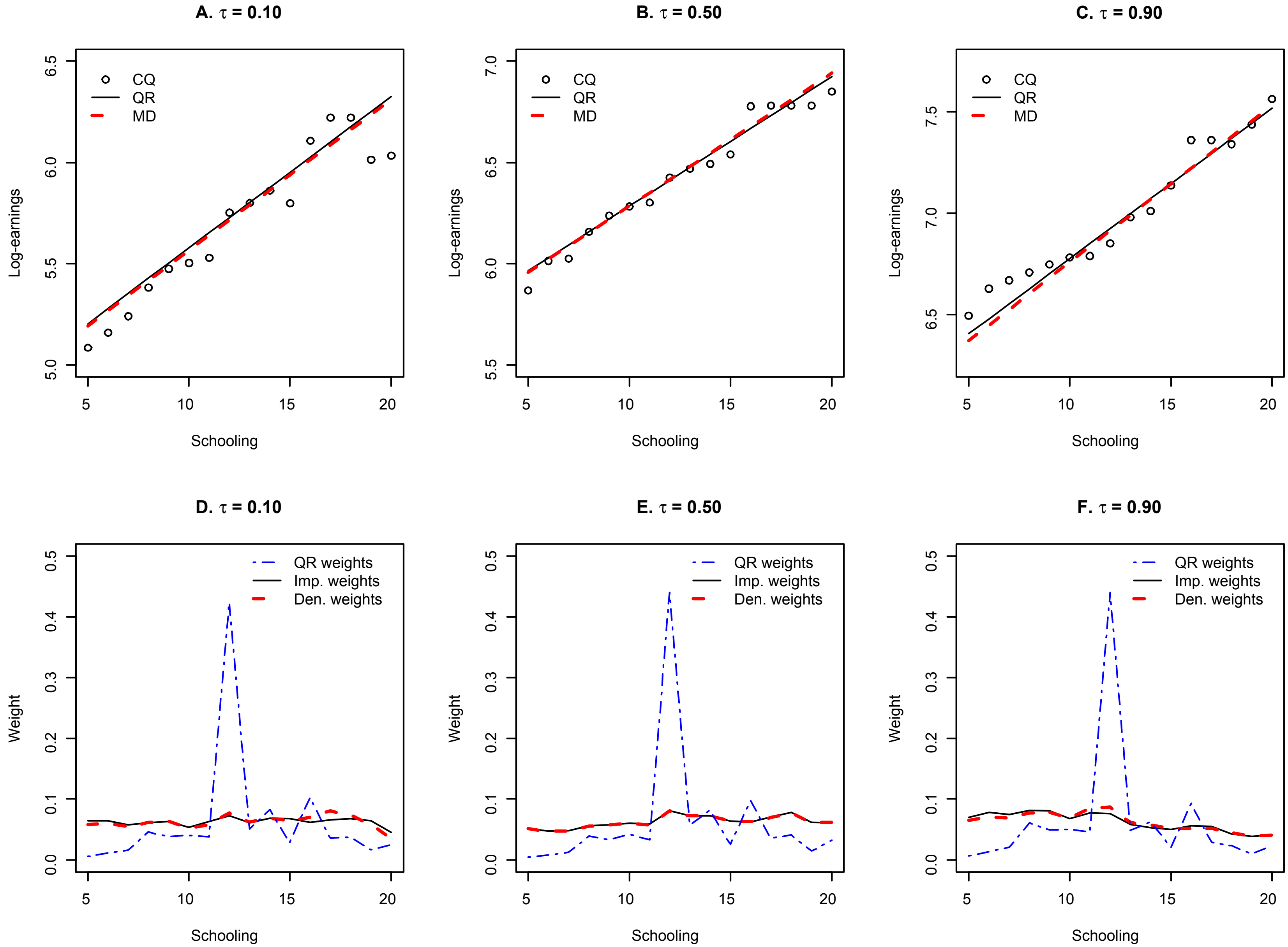

Figure 1: CQF and Weighting schemes in 1980 Census (US-born white and black men aged 40-49). Panels A - C plot the Conditional Quantile Function, Linear Quantile Regression fit, and Chamberlain's Minimum Distance fit for log-earnings given years of schooling. Panels D - F plot QR weighting function (histogram $\times$ importance weights), importance weights and density weights.

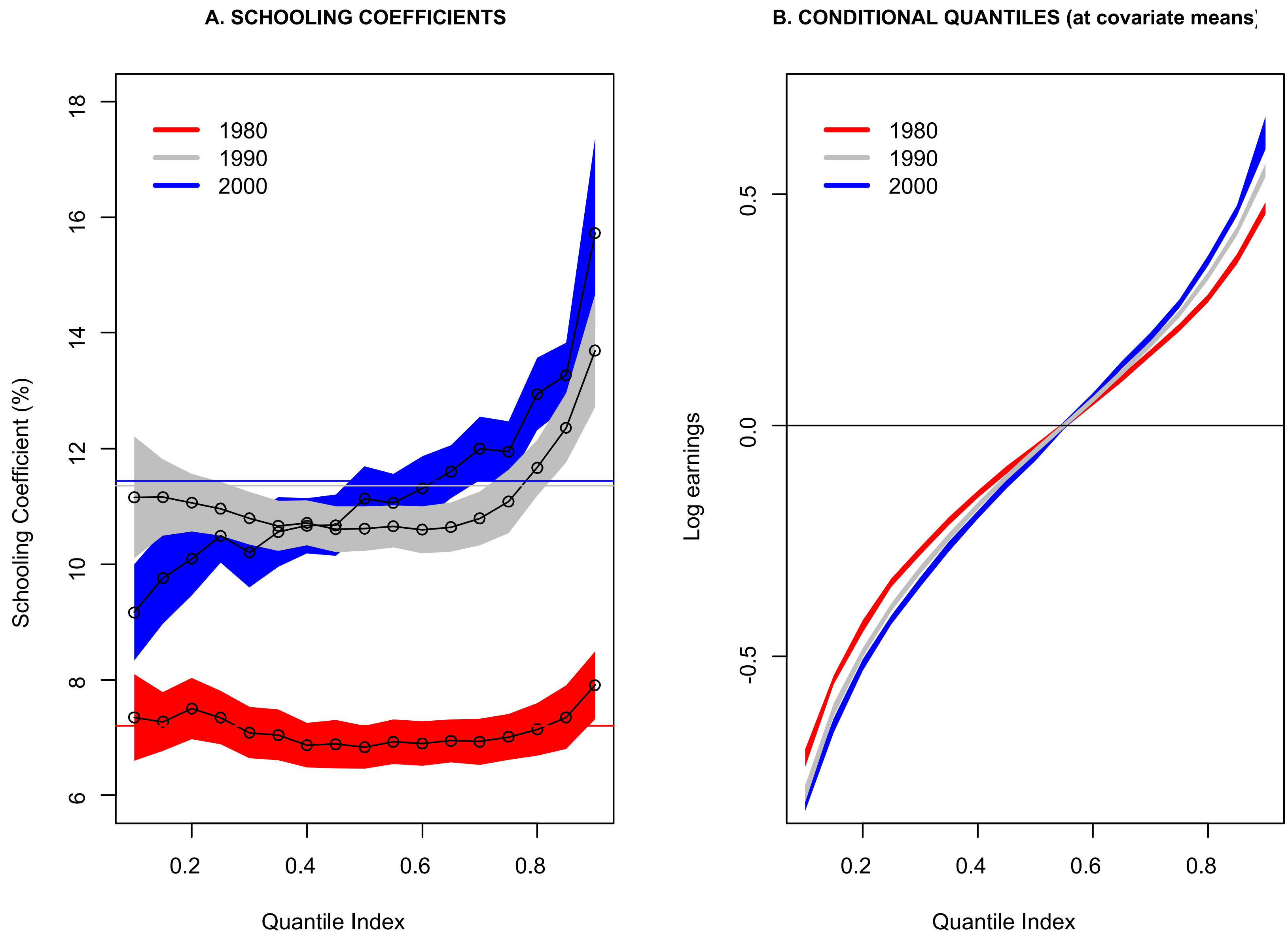

Figure 2: Schooling coefficients and conditional quantiles of log-earnings in 1980, 1990, and 2000 censuses (US-born white and black mean aged 40-49). Panel A plots the quantile process for the coefficient of schooling in the QR of log-earnings on years of schooling, race, and a quadratic function of experience; and robust simultaneous 95 % confidence bands. Panel B plots simultaneous 95 % confidence bands for the QR approximation to the conditional quantile function given schooling, race, and a quadratic function of experience. Horizontal lines correspond to OLS estimates of the schooling coefficients in Panel A. In Panel B, covariates are evaluated at sample mean values for each year, and distributions are centered at median earnings for each year (i.e., for each τ and year, $E[X]'(\widehat{\beta}(\tau) - \widehat{\beta}(.5))$ is plotted).

Table 1: Comparison of CQF and QR-based Interquantile Spreads

		Interquantile Spread					
		90-10		90-50		50-10	
Census	Obs.	CQ	QR	CQ	QR	CQ	QR
A. Overall							
1980	65,023	1.20	1.19	0.51	0.52	0.68	0.67
1990	86,785	1.35	1.35	0.60	0.61	0.75	0.74
2000	97,397	1.43	1.45	0.67	0.70	0.76	0.75
B. High School Graduates							
1980	25,020	1.09	1.17	0.44	0.50	0.65	0.67
1990	22,837	1.26	1.31	0.52	0.55	0.74	0.76
2000	25,963	1.29	1.32	0.59	0.60	0.70	0.72
C. College Graduates							
1980	7,158	1.26	1.19	0.61	0.54	0.65	0.64
1990	15,517	1.44	1.38	0.70	0.66	0.74	0.72
2000	19,388	1.55	1.57	0.75	0.80	0.80	0.78

Notes: US-born white and black men aged 40-49. Average measures calculated using the distribution of the covariates in each year. The covariates are schooling, race and a quadratic function of experience. Sampling weights used for 2000 Census.