QUANTILE REGRESSION

PAUL SCHRIMPF
NOVEMBER 15, 2011
UNIVERSITY OF BRITISH COLUMBIA
ECONOMICS 628: TOPICS IN ECONOMETRICS

[1]

1. INTRODUCTION

1.1. **Unconditional quantiles.** Let Y be a random variable with cdf F. The τth quantile of Y is

$$Q_Y(\tau) \equiv \inf_y \{y : F(y) \geq \tau\}. \tag{1}$$

If F is continuous and strictly monotonic (i.e. Y is continuously distributed), then the quantile and distribution functions are inverses of one another.

$$Q_Y(F(y)) = y$$

and

$$F(Q_Y(\tau)) = \tau.$$

Quantiles are often associated with sorting. If we estimate quantiles by using an estimate of the cdf in (1), we have

$$\begin{aligned}\hat{Q}^s_Y(\tau) =& \inf_y \{y : \hat{F}(y) \geq \tau\} \\ =& \inf_y \{y : \frac{1}{n}\sum_{i=1}^n 1[y \leq y_i] \geq \tau\} \\ =& \inf_y \{y : \sum_{i=1}^n 1[y \leq y_i] \geq \tau n\}\end{aligned}$$

If we let $y_{i(1)}$ denote $\min_{i\in\{1,\ldots,n\}} y_i$, $y_{i(2)}$ denote $\min_{i\in\{1,\ldots,n\}\setminus\{i(1)\}} y_i$, etc, so that $y_{i(j)}$ is the jth smallest y_i, then we get

$$\hat{Q}^s_Y(\tau) = y_{i(\lceil \tau n \rceil)} \tag{2}$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Using the $\lceil n\tau \rceil$ instead of $\lfloor n\tau \rfloor$, or the nearest integer to $n\tau$ is to be consistent with (1), but this convention is of no practical importance.

We would like to be able to estimate not just unconditional quantile functions, but also conditional quantile functions,

$$Q_Y(\tau|x) = \inf_y \{y : F(y|x) \geq \tau\}.$$

[1]**References**: these notes draw largely on Chernozhukov (2007a) and Chernozhukov (2007b). Sections 2 and 4.1.2 were written when I was a TA for that course. Koenker (2005) is a very good reference. Other sources are cited in the sections of the text where they are used.

If x is continuously distributed, observing any $x_i = x$ is a zero probability event, so we cannot hope to sort conditional on x. Fortunately, another way to characterize quantiles is as a solution to a minimization problem. Let

$$\begin{aligned}\rho_\tau(z) \equiv& z\left(\tau - 1\{z < 0\}\right)\\ =&\tau z 1\{z \geq 0\} + (1-\tau) z 1\{z < 0\}\\ =&\tau z^+ + (1-\tau) z^-,\end{aligned}$$

where the three lines are just slightly different ways of writing the same expression. ρ_τ is called the check function. When $\tau = 0.5$, $\rho_{0.5}(z) = |z|$. Figure, 1 shows plots of the check function for various values of τ. We can show that $Q_Y(\tau)$ minimizes the expectation of the check function.

Lemma 1.1.

$$Q_Y(\tau) = \arg\min_{\theta \in \mathbb{R}} \mathrm{E}[\rho_\tau(y - \theta)]$$

Proof. ρ_τ is convex, so $\rho_\tau(y-\theta)$ is a convex function of θ. $\mathrm{E}[\rho_\tau(y-\theta)]$ is a linear transformation of $\rho_\tau(y-\theta)$, so $\mathrm{E}[\rho_\tau(y-\theta)]$ is convex as a function of θ. Therefore, if we can find a θ^* that satisfies the first order condition, then θ^* is a minimizer. The first order condition is

$$\begin{aligned}0 =& \frac{d}{d\theta}\mathrm{E}[\rho_\tau(y-\theta)] = \mathrm{E}\left[-\left(\tau - 1\{y - \theta < 0\}\right)\right]\\ \tau =& F(\theta)\\ \theta =& Q_Y(\tau)\end{aligned}$$

□

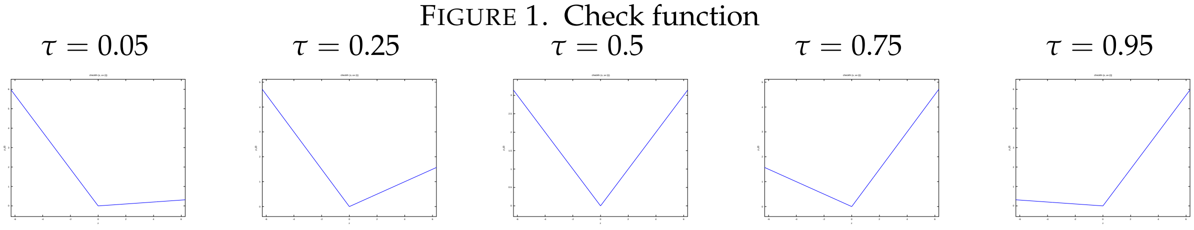
Thus, we could estimate the τth quantile by

$$\hat{Q}_Y(\tau) = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}_n[\rho_\tau(y_i - \theta)]$$

A good exercise would be to verify that this $\hat{Q}_Y(\tau)$ is the same as the estimate of the τth quantile from sorting as in (2) (up to rounding). That is,

$$y_{i(\lfloor n\tau \rfloor)} \leq \hat{Q}_Y(\tau) \leq y_{i(\lceil n\tau \rceil)}.$$

FIGURE 1. Check function



1.2. **Conditional quantiles.** To begin our study of conditional quantiles assume that the true conditional quantile function is linear in x,

$$Q_Y(\tau|x) = x'\beta(\tau)$$

where $x \in \mathbb{R}^d$, and $\beta(\tau) \in \mathbb{R}^d$. Later, we will go over what happens if make this assumption, but the true conditional quantile function is not linear. As in the unconditional case, we can show that $\beta(\tau)$ minimizes the expectation of the check function.

Lemma 1.2.

$$\beta(\tau) = \arg\min_{b \in \mathbb{R}^d} \mathrm{E}[\rho_\tau(y - x'b)]$$

Proof. The proof is nearly identical to 1.1, and is left as an exercise. □

Although the linearity assumption is somewhat restrictive, a linear quantile model can capture much richer forms of variation than a linear mean regression.

Example 1.1 (Location model)**.** The classic regression setup is

$$y_i = x_i' B + \epsilon_i$$

with ϵ_i independent of x_i. In this case, if the first component of x is a constant, then the conditional quantile coefficients are

$$\beta_1(\tau) = B_1 + Q_\epsilon(\tau)$$

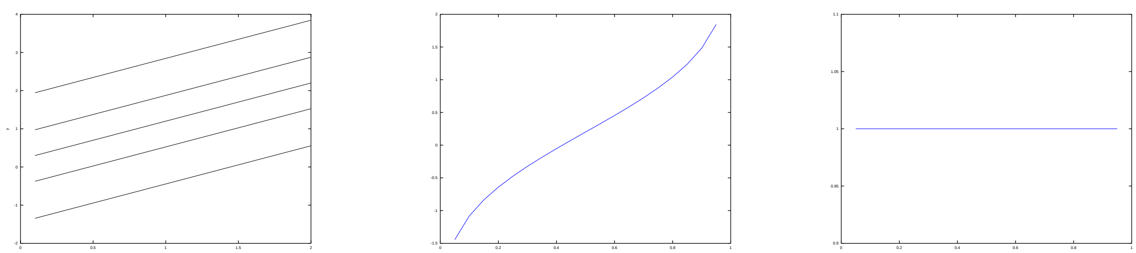
and

$$\beta_j(\tau) = B_j$$

for $j > 1$.

The figures below illustrate this for a model with two x variables—a constant, and another variable. The left panel shows $Q_Y(\tau|x)$ as a function of x for various τ, the next two panels show the intercept, $\beta_1(\tau)$, and slope, $\beta_2(\tau)$ as functions of τ.

Conditional quantile functions $\beta_1(\tau)$ $\beta_2(\tau)$



Example 1.2 (Location-scale (heteroskadesticity) model)**.** If

$$y_i = x_i' B + x_i' \gamma \epsilon_i$$

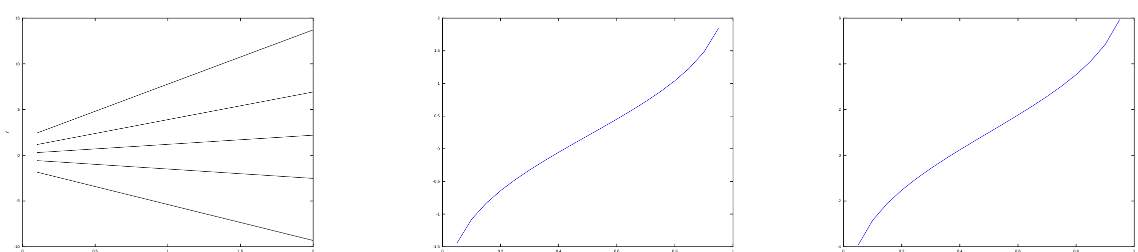
then

$$\beta(\tau) = B + \gamma Q_\epsilon(\tau)$$

In particular, $\beta_j(\tau)$ is monotone in τ for each j, and $\beta(\tau_1)$ is a linear function of $\beta(\tau_2)$ for any τ_1, and τ_2.

The figures below illustrate this for a model with two x variables—a constant, and another variable. The left panel shows $Q_Y(\tau|x)$ as a function of x for various τ, the next two panels show the intercept, $\beta_1(\tau)$, and slope, $\beta_2(\tau)$ as functions of τ.

Conditional quantile functions $\beta_1(\tau)$ $\beta_2(\tau)$



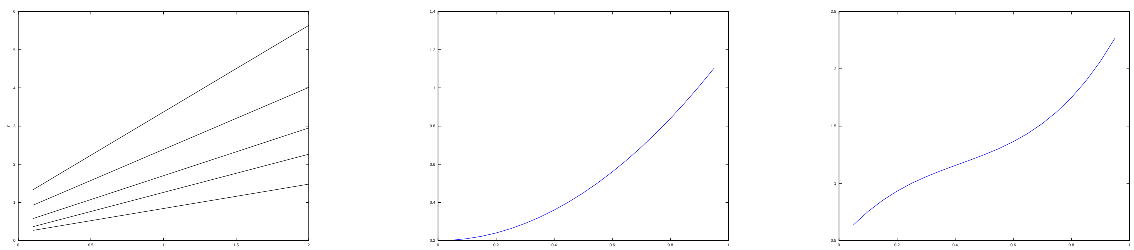
Example 1.3 (General model)**.** In general, we could think of y coming from a correlated random coefficients model,

$$y_i = x'\beta(u_i)$$

with $u_i|x \sim U(0,1)$. The conditional quantile functions will still be linear in x, but their slopes can be related in any way (provided they do not cross).

The figures below illustrate this for a model with two x variables—a constant, and another variable. The left panel shows $Q_Y(\tau|x)$ as a function of x for various τ, the next two panels show the intercept, $\beta_1(\tau)$, and slope, $\beta_2(\tau)$ as functions of τ.

Conditional quantile functions $\beta_1(\tau)$ $\beta_2(\tau)$



2. Inference

This section goes over the asymptotic behavior of quantile regression. It is based on Koenker (2005).

2.1. Setup.

Let $\{Y_i\}$ be independent random variables with distributions $\{F_i\}$. Suppose that the τth quantile of Y_i given x_i is linear in x:

$$Q_{Y_i}(\tau|x_i) = x_i'\beta(\tau) \tag{3}$$

By definition, we have

$$F_i^{-1}(\tau|x_i) = Q_{Y_i}(\tau|x_i) \equiv \xi_i(\tau) \tag{4}$$

We will consider the behavior of the quantile regression estimator:

$$\hat{\beta}_n(\tau) = \arg\min_{b \in \Re^p} \sum \rho_\tau(y_i - x_i'b) \tag{5}$$

where $\rho_\tau(u) = u\,(\tau - \mathbf{1}(u < 0))$ is the check function.

2.2. Asymptotic distribution.

Assume:

A1 $\{F_i\}$ are uniformly continuous with $f_i(\xi)$ uniformly bounded away from 0 and ∞ at $\{\xi_i(\tau)\}$.

A2 There exist positive definite D_0 and $D_1(\tau)$ such that

(a) $\lim_{n\to\infty} n^{-1} \sum x_i x_i' = D_0$

(b) $\lim_{n\to\infty} n^{-1} \sum f_i(\xi_i(\tau)) x_i x_i' = D_1(\tau)$

(c) $\max ||x_i||/\sqrt{n} \to 0$

Theorem 2.1 (Asymptotic Normality)**.** *Under these assumptions,*

$$\sqrt{n}\left(\hat{\beta}_n(\tau) - \beta(\tau)\right) \xrightarrow{d} N\left(0, \tau(1-\tau) D_1(\tau)^{-1} D_0 D_1(\tau)^{-1}\right)$$

Proof. We show this result by finding the limit distribution of the objective function. Convexity then implies that the limit distribution of the estimator is the distribution of the minimizer limiting objective function.

Rewrite the objetive function as:

$$Q_n(\hat{\beta}_n(\tau)) = \sum \rho_\tau(y_i - x_i'\hat{\beta}_n(\tau)) \tag{6}$$
$$= \sum \rho_\tau(y_i - x_i'\beta(\tau) - x_i'(\sqrt{n}(\hat{\beta}_n(\tau) - \beta(\tau)))/\sqrt{n} \tag{7}$$

Let $u_i = y_i - x_i'\beta(\tau)$ and $\delta = \sqrt{n}(\hat{\beta}_n(\tau) - \beta(\tau))$. We can add a constant to the objetive function without changing our estimator. Consider

$$Z_n(\delta) = \sum \rho_\tau(u_i - x_i'\delta/\sqrt{n}) - \rho_\tau(u_i) \tag{8}$$

As in Knight (1998), consider the identity:

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (\mathbf{1}(u \leq s) - \mathbf{1}(u \leq 0))ds \tag{9}$$

where $\psi_\tau(u) = \tau - \mathbf{1}(u < 0)$[2] Using this identity, we can rewrite (8) as:

$$Z_n(\delta) = -\sum x_i'\delta/\sqrt{n}\psi_\tau(u_i) + \sum \int_0^{x_i'\delta/\sqrt{n}} \mathbf{1}(u_i \leq s) - \mathbf{1}(u_i \leq 0)ds \tag{10}$$

We can deal with these two terms separately. Let $Z_{1n}(\delta) = -\sum x_i'\delta/\sqrt{n}\psi_\tau(u_i)$ and $Z_{2n}(\delta) = \sum \int_0^{x_i'\delta/\sqrt{n}} \mathbf{1}(u_i \leq s) - \mathbf{1}(u_i \leq 0)ds$.

A standard CLT applies to $Z_{1n}(\delta)$. It is a sum of independent terms with expectation 0 (because $E[\psi_\tau(u)|x_i] = 0$) and variance

$$E[x_i'\delta\psi_\tau(u_i)^2\delta'x_i] = E[x_i'\delta E[\psi_\tau(u_i)^2|x_i]\delta'x_i]$$
$$= \tau(1-\tau)\delta'E[x_i'x_i]\delta \tag{11}$$

so,

$$Z_{1n}(\delta) \rightsquigarrow -\delta'W \text{ where } W \sim N(0, \tau(1-\tau)D_0) \tag{12}$$

Now, let $Z_{2ni}(\delta) = \int_0^{x_i'\delta/\sqrt{n}} \mathbf{1}(u_i \leq s) - \mathbf{1}(u_i \leq 0)ds$. Note that $P(u_i < s) = F_i(\xi_i + s)$ ($u_i < s$ means that the difference between y_i and its τth quantile is less than s, i.e. $y_i - \xi_i < s$ or $y_i < \xi_i + s$).

[2]This identity is easily verified by plugging in the definitions of the various functions. Begin with the right side:

$$-v\psi_\tau(u) + \int_0^v (\mathbf{1}(u \leq s) - \mathbf{1}(u \leq 0))ds = -v(\tau - \mathbf{1}(u < 0)) - v\mathbf{1}(u < 0) + \int_0^v \mathbf{1}(u \leq s)ds$$
$$= -v\tau + \begin{cases} v & u < 0,\ v > u \\ 0 & u < 0,\ v < u \\ (v-u) & u > 0,\ v > u \\ 0 & u > 0,\ v < u \end{cases}$$
$$= -v\tau + (v-u)\mathbf{1}(v > u) + u\mathbf{1}(u < 0)$$
$$= (u-v)(\tau - \mathbf{1}(u - v < 0)) - u(\tau - \mathbf{1}(u < 0))$$
$$= \rho_\tau(u-v) - \rho_\tau(u)$$

Therefore,

$$\begin{aligned}
\sum EZ_{2ni}(\delta) =& \sum \int_0^{x_i'\delta/\sqrt{n}} F_i(\xi_i + s) - F_i(\xi_i) ds \\
=& \frac{1}{n} \sum \int_0^{x_i'\delta} \sqrt{n} \left(F_i(\xi_i + t/\sqrt{n}) - F_i(\xi_i) \right) dt \\
=& \frac{1}{n} \sum \int_0^{x_i'\delta} f_i(\xi_i) t dt + o(1) \\
=& \frac{1}{2n} \sum f_i(\xi_i) \delta' x_i x_i' \delta + o(1) && (13) \\
\to& \frac{1}{2} \delta' D_1(\tau) \delta && (14)
\end{aligned}$$

Furthermore, the variance of $(Z_{2n}(\delta))$ is bounded by:

$$V(Z_{2n}(\delta)) \leq \frac{1}{\sqrt{n}} \max_i |x_i'\delta| \sum EZ_{2ni}(\delta) \quad (15)$$

By assumption 0c, $V(Z_{2n}(\delta)) \to 0$, so $Z_{2n}(\delta) \xrightarrow{p} EZ_{2n}(\delta)$ and we can conlude that

$$Z_n(\delta) \rightsquigarrow Z_0(\delta) = -\delta' W + \frac{1}{2} \delta' D_1 \delta \quad (16)$$

Finally, we obtain a limiting distribution for $\hat{\beta}(\tau)$ by noting that (ignoring some details)

$$\begin{aligned}
\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) =& \hat{\delta}_n = \arg\min Z_n(\delta) \\
& \rightsquigarrow \arg\min Z_0(\delta) = \hat{\delta}_0 = D_1^{-1} W \sim N\left(0, D_1^{-1} D_0 D_1^{-1} \tau(1-\tau)\right) && (17)
\end{aligned}$$

□

2.3. **Inference in practice.** We can apply the above result to perform Wald or t-tests. The primary difficulty is that we must estimate the inverse of the conditional density (aka the sparsity) at the τth quantile, $f_i(\xi_i(\tau))^{-1} = s(\tau)$. One option would be to use a standard kernel density estimator. Powell suggested using

$$\hat{D}_1(\tau) \frac{1}{nh_n} \sum K(\hat{u}_i(\tau)/h_n) x_i x_i'$$

Another approach recognizes that $\frac{d}{dt} F^{-1}(t|x) = x' \frac{d}{dt} \beta(t) = s(t)$. Then we can use, e.g. (as in Siddiqui (1960))

$$\hat{s}_n(t) = \frac{x'(\hat{\beta}(t + h_n) - \hat{\beta}(t - h_n))}{2h_n}$$

where $h_n \to 0$ as $n \to \infty$. More complicated ways of approximating the derivative are also possible.

2.3.1. *Bootstrap.* To avoid having to estimate the sparsity function, one can also use the bootstrap for inference. Note that if we do not estimate the sparsity function, we will be bootstrapping a statistic that is not asymptotically pivotal. This has led to a number of papers about variants of the bootstrap and their rates. The residual bootstrap converges slower than standard asymptotic distribution. Smoothed variants of the bootstrap do as well as the standard asymptotic distribution. Another approach based on resampling of the subgradient condition has attractive computational properties, especially for non-convex problems such as censored quantile regression. See Koenker (2005)for more information.

2.3.2. *Inference on $\beta(\cdot)$.* Some interesting hypotheses depend on the entire function $\beta(\cdot)$ instead of just $\beta(\tau)$ at some fixed τ. For example, we might want to test whether x has an effect at any quantile, $H_0 : \beta(\tau) = 0 \forall \tau$, or whether x has a constant effect, $H_0 : \beta(\tau) = \beta(0.5) \forall \tau$. To test hypotheses of this form, we need to derive the limit distribution of $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$ viewed as a function of τ. From the result above we know that for a finite set of points, $\sqrt{n} D_0^{-1/2} D_1 \begin{pmatrix} \hat{\beta}(\tau_1) - \beta(\tau_1) \\ \vdots \\ \hat{\beta}(\tau_k) - \beta(\tau_k) \end{pmatrix}$, is asymptotically normal with variances $\tau_j(1-\tau_j)$ and covariances given by:

$$\begin{aligned} E\psi_{\tau_j}(u_i)\psi_{\tau_k}(u_i) =& E\left[(\tau_j - \mathbf{1}(y_i - x\beta(\tau_j) \le 0))(\tau_k - \mathbf{1}(y_i - x\beta(\tau_k) \le 0))\right] \\ =& (\tau_j \wedge \tau_k) - \tau_j \tau_k \end{aligned}$$

In fact, this convergence is true for all τ. $\sqrt{n} D_0^{-1/2} D_1(\hat{\beta}(\tau) - \beta(\tau))$ converges to a random function, $\nu(\tau)$, which is normally distributed at any finite set of points with the variance above. This sort of random function is called a Brownian bridge. The sense in which this convergence occurs is called weak convergence, and is often denoted by $\rightsquigarrow$. The weak convergence is that for all bound continuous functions , $f : T \to \mathbb{R}$ (where T is the space in which $\nu(\cdot)$ lies, in this case, $T = \{g : [\epsilon, 1-\epsilon] \to \mathbb{R}^k, g \text{ continuous}\}$ with the ℓ^∞ metric), $Ef(\sqrt{n} D_0^{-1/2} D_1(\hat{\beta}(\cdot) - \beta(\cdot))) \to Ef(\nu(\cdot))$. In fact, we already relied on this sort of convergence when we said the limit distribution of $\hat{\beta}(\tau)$ is the distribution of the minimum of the limiting objective function. For the problem at hand, weak convergence implies that

$$\sup_\tau n(\hat{\beta}(\tau) - \beta(\tau))' D_1' D_0^{-1} D_1(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} \sup_\tau \nu(\tau)'\nu(\tau)$$

The distribution of the later can be simulated to obtain critical values. This result can be used to test the hypothesis of no effect, $H_0 : \beta(\tau) = 0 \forall \tau$. Testing the hypothesis of constant effect is slightly more complicated since it involves the value of the constant effect as a nuisance parameter.

2.4. **Misspecification.** You are likely familiar with the fact that mean regression estimates the best (in least squares) linear approximation to the conditional expectation function, regardless of whether the conditional expectation function is linear. That is, OLS estimates

$$\beta_0 \arg\min \mathrm{E}[(\mathrm{E}[y|x] - x'\beta)^2].$$

Analogously, we should expect that quantile regression estimates the "best" linear approximation to the true conditional quantile function. However, "best" in what sense? For the median ($\tau = 0.5$), we might guess that it would be the best approximation in $\mathcal{L}^1$ norm, i.e.

$$\min_\beta \mathrm{E}\left[\left|Q_Y(0.5|x) - x'\beta\right|\right]$$

This guess is incorrect. In fact, quantile regression estimates the best linear approximation in a weighted $\mathcal{L}^2$ norm. Angrist, Chernozhukov, and Fernández-Val (2006) show the following.

Theorem 2.2. *Suppose that (i) the conditional density of y given x, $f_Y(y|x)$ exists almost surely, (ii)*$\mathrm{E}[Y]$, $\mathrm{E}[Q_Y(\tau|x)^2]$, *and* $E[\|X\|^2]$ *are finite, and (iii) $\beta(\tau)$ uniquely solves*

$$\beta(\tau) = \arg\min_{b \in \mathbb{R}^d} \mathrm{E}[\rho_\tau(y - x'b)]$$

Then,

$$\beta(\tau) = \arg\min_{\beta \in \mathbb{R}^d} \mathrm{E}\left[w_\tau(x,\beta)(Q_Y(\tau|x) - x'\beta)^2\right]$$

where
$$w_\tau(x,\beta) = \int_0^1 (1-u) f_Y\left(ux'\beta + (1-u)Q_Y(\tau|x)|x\right) du$$
Additionally,
$$\beta(\tau) = \arg\min_{b\in\mathbb{R}^d} \mathrm{E}\left[w_\tau(x,\beta(\tau))(Q_Y(\tau|x) - x'b)^2\right]$$
where now the weights are being help constant.

A useful corollary of this theorem is that you can get an omitted variable bias formula for quantile regression somewhat similar to what you get for OLS.

Corollary 2.1. *Let $\beta_1(\tau)$ and $\beta_2(\tau)$ be the coefficients from the population quantile regression of y on x_1 and x_2. Let $\gamma_1(\tau)$ be the coefficient from the quantile regression of y on x_1 only. Then,*
$$\gamma_1(\tau) = \beta_1(\tau) + \mathrm{E}[\tilde{w}_\tau(x,\gamma_1(\tau))x_1'x_1]^{-1}\mathrm{E}\left[\tilde{w}_\tau(x,\gamma_1(\tau))x_1'\left(Q_Y(\tau|x) - x_1'\beta_1(\tau)\right)\right]$$
where
$$\tilde{w}_\tau(x,\gamma) = \int_0^1 (1-u) f_Y\left(ux_1'\gamma + (1-u)Q_Y(\tau|x)|x\right) du.$$

Proof. The same argument used to prove theorem 2.2 (see Angrist, Chernozhukov, and Fernández-Val (2006) for this proof) also shows that
$$\gamma_1(\tau) = \arg\min_{\gamma} \mathrm{E}\left[\tilde{w}_\tau(x,\gamma_1(\tau))(Q_Y(\tau|x) - x_1'\gamma)^2\right]$$
From the usual least squares first order condition, we get that
$$\gamma_1(\tau) = \mathrm{E}[\tilde{w}_\tau(x,\gamma_1(\tau))x_1'x_1]^{-1}\mathrm{E}\left[\tilde{w}_\tau(x,\gamma_1(\tau))x_1'Q_Y(\tau|x)\right]$$
Substituting $Q_Y(\tau|x) = x_1\beta_1(\tau) + Q_Y(\tau|x) - x_1\beta_1(\tau)$ gives the conclusion. □

If the true conditional quantile function is linear, then $Q_Y(\tau|x) - x_1\beta_1(\tau) = x_2\beta_2(\tau)$. It seems plausible to believe that $\mathrm{E}\left[x_1'(x_2'\beta_2(\tau))\tilde{w}_\tau(x,\gamma_1(\tau))\right]$ has the same sign as $\mathrm{E}[x_1'(x_2'\beta_2)]$. In this case, the omitted variable bias in quantile regression and the omitted variable bias in mean regression have the same sign.

2.4.1. *Inference under misspecification.* Misspecification changes the asymptotic distribution of the quantile regression coefficients slightly. Angrist, Chernozhukov, and Fernández-Val (2006) show the following.

Theorem 2.3. *Suppose that (i) $(Y_i, X_i, i \leq n)$ are independent and identically distributed on the probability space (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) $J(\tau) := \mathrm{E}[f_Y(X\beta(\tau)|X)XX']$ is positive definite for all $\tau \in T = [\delta, 1-\delta]$, and (iv) $\mathrm{E}[X^{2+\epsilon}] < \infty$ for some $\epsilon > 0$.*

Then the quantile regression process is uniformly consistent,
$$\sup_{\tau\in T} \left\|\beta(\tau) - \hat{\beta}(\tau)\right\| = o_p(1),$$
and $J(\cdot)^{-1}\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ converges in distribution to a zero mean Gaussian process $z(\cdot)$, where $z(\cdot)$ has covariance function
$$\begin{aligned}\Sigma(\tau,\tau') :=& \mathrm{E}[z(\tau)z(\tau')] \\ =& \mathrm{E}\left[(\tau - 1\{Y < X\beta(\tau)\})(\tau' - 1\{Y < X\beta(\tau')\})XX'\right]\end{aligned} \tag{18}$$
If the model is correctly specified, i.e., $Q_Y(\tau|x) = x'\beta(\tau)$ a.s., then $\Sigma(\tau,\tau')$ simplifies to
$$\Sigma_0(\tau,\tau') = \left(\min\{\tau,\tau'\} - \tau\tau'\right)\mathrm{E}[XX']$$

2.5. **Exact finite sample inference.** Consider the problem of estimating an unconditional quantile of Y_i. The estimate solves

$$\hat{q} = \arg\min \mathbb{E}_n[\rho_\tau(y - q)]$$

$\mathbb{E}_n \rho_\tau$ is not differentiable, but it is differentiable from the left and the right. In this case, the usual first order condition of the derivative being zero becomes the derivatives from the left and from the right must both be non-negative. That

$$\mathbb{E}_n[\tau - 1\{y < \hat{q}]\} \geq 0$$

and

$$\mathbb{E}_n[1\{y < \hat{q}\} - \tau] \geq 0.$$

Note that $\mathbb{E}_n[1\{y < \hat{q}\} - \tau]$ is monotonically increasing in $\hat{q}$. Therefore, the event $\hat{q} > q$ is the same as

$$\mathbb{E}_n[1\{y < q\} - \tau] < 0.$$

If y_i is i.i.d. with distribution F, then this is the probability of n events, each with probability $F(q)$ happening fewer than $n\tau$ times. The probability of this happening is

$$F_{\hat{q}}(q) = \sum_{i=1}^{\lfloor n\tau \rfloor} \binom{n}{i} F(q)^i (1 - F(q))^{n-i}$$

Suppose $q = Q_Y(\tau)$. Then,

$$F_{\hat{q}}\left(Q_Y(\tau)\right) = \sum_{i=1}^{\lfloor n\tau \rfloor} \binom{n}{i} \tau^i (1 - \tau)^{n-i}.$$

This distribution is pivotal. It does not depend on any unknown nuisance parameters. Therefore, it can be used to construct an exact finite sample test of the null hypothesis that $q = Q_Y(\tau)$.

Chernozhukov, Hansen, and Jansson (2009) showed that a similar approach applies to quantile regression and IV quantile regression (discussed below). In those cases, the gradient of the objective function is

$$\mathbb{E}_n\left[(\tau - 1\{y < x'\beta(\tau)\})z\right]$$

where $z = x$ for quantile regression and z is the instrument in IV quantile regression. If you assume correct specification, then

$$\mathrm{P}(y < x'\beta(\tau)) = \tau$$

and the distribution of

$$\mathbb{E}_n\left[(\tau - 1\{y < x'\beta(\tau)\})z\right]$$

conditional on $Z_1, ..., Z_n$ is the same as

$$\mathbb{E}_n\left[(\tau - B)z\right]$$

where $B_i \sim \text{Bernoulli}(\tau)$. Thus, you can construct tests and confidence intervals for $\beta(\tau)$ as follows.

(1) Let
$$\mathcal{L}_n = \frac{1}{2}\left(\sqrt{n}\mathbb{E}_n[(\tau - B)z]\right)' W_n \left(\sqrt{n}\mathbb{E}_n[(\tau - B)z]\right)$$
where W_n is some weighting matrix that can depend on x and z.

(2) Simulate the distribution of $\mathcal{L}_n$ by repeatedly drawing B_i and computing $\mathcal{L}_n$. For a test with size α or confidence region with coverage $1 - \alpha$, find the $1 - \alpha$ quantile of $\mathcal{L}_n$ and call it $c_n(\alpha)$. This will be the critical value.

(3) Let

$$L_n(\beta) = \frac{1}{2}\left(\sqrt{n}\mathbb{E}_n[(\tau - 1\{y < x'\beta\})z]\right)' W_n \left(\sqrt{n}\mathbb{E}_n[(\tau - 1\{y < x'\beta\})z]\right).$$

The test does not reject $\beta_0(\tau) = \beta$ (or the confidence region includes β) if $L_n(\beta) \leq c_n(\alpha)$. Note that the scaling of $\mathcal{L}_n$ and L_n by $\sqrt{n}$ is not needed here, but is done for comparability with the asymptotic results. Asymptotically, $\mathcal{L}_n$ and $L_n(\beta_0)$ are each χ^2. Also, these are score statistics, and, as usual, have favorable asymptotic power properties. See Chernozhukov, Hansen, and Jansson (2009) for details. An additional advantage of this approach to inference is that it is valid even under weak and partial identification.

3. Quantile regression with endogeneity

Just like OLS always estimates the best linear approximation to the conditional expectation function, quantile regression always estimates the (weighted) best linear approximation to the conditional quantile function. To meaningfully talk about endogeneity, we must have some model in mind where the function we want to estimate is not the conditional quantile function we observe in the data. This model could simply come from a thinking about an ideal experiment and the causal effects it would reveal, or it could come from a more structured economic model. In any case, suppose our model implies that

$$\begin{aligned} y_i =& x_i'\beta(\tau) + z_{1i}\gamma(\tau) + u_i \\ x_i =& z_i'\pi(\tau_r) + v_i \end{aligned}$$

where x_i are some endogenous regressors, $z_i = (z_{1i}, z_{2i})$ are exogenous, and we are primarily interested in estimating $\beta(\tau)$ and $\gamma(\tau)$.

Recall that if we were doing mean regression, there are three essentially equivalent ways to proceed.[3]

(1) 2SLS: regress x_i on z_i, form $\hat{x}_i = z_i\hat{\pi}$, then regress y_i on $\hat{x}_i$ and z_{1i}.
(2) Control function: regress x_i on z_i, estimate the residuals $\hat{v}_i$, and then regress y_i on x_i, z_{1i}, and $\hat{v}_i$.
(3) IV: form instruments $w_i = \Phi(z_i)$ do GMM using the moment conditions $\mathrm{E}[(y_i - x_i'\beta - z_{1i}\gamma)w_i] = 0$.

Two-stage least square and the control function approach produce identical estimates. The IV estimate generally differs, but all three approaches are consistent under the assumption that $\mathrm{E}[u_i|z_i] = 0$[4]. If you do not remember this result, it would be a good review exercise to show it. There are quantile analogs of 2SLS, control functions, and IV, but interestingly, they each require different assumptions about u_i and v_i.

3.1. Two stage quantile regression.

Two stage quantile regression was first studied by Amemiya (1982) and Powell (1983). These papers specifically focused on the median and called the estimator two-stage least absolute deviations, but the same sort of analysis applies to any quantile. In the first stage, we estimate $\hat{\pi}(\tau_r)$ by quantile regression. Note that the quantile in the first stage need not equal the quantile in the second stage, but it is hard to imagine why they would differ. In the second stage, we perform quantile regression on

$$y_i = (z_i\hat{\pi}(\tau_r) + \hat{v}_i)\beta(\tau) + z_{1i}\gamma(\tau) + u_i.$$

Under the assumptions in 2 $\hat{\pi}(\tau_r) \xrightarrow{p} \pi(\tau_r)$ and $\hat{v}_i \xrightarrow{p} v_i$. Then, the second stage estimates will converge to the weighted best linear approximation to the τth conditional quantile of

$$(z_i\pi(\tau_r) + v_i)'\beta(\tau) + z_{1i}\gamma(\tau) + u_i = (z_i\pi(\tau_r))'\beta(\tau) + z_{1i}\gamma(\tau) + u_i + v_i'\beta(\tau).$$

If the τ quantile of $u_i + v_i'\beta(\tau)$ conditional on z does not depend on z, then it is clear that $\beta(\tau)$ and $\gamma(\tau)$ will be consistently estimated (except the intercept). In contrast to mean regression, this is assumption about $u_i + v_i'\beta$ instead of u_i. Conditional expectations are linear, so with mean regression we have

$$\mathrm{E}[u_i + v_i'\beta(\tau)|z] = \mathrm{E}[u_i|z] + \mathrm{E}[v_i|z]'\beta$$

[3]This comparison of the three approaches to endogeneity is based largely on Lee (2007) and Imbens and Wooldridge (2007)

[4]2SLS and control function just need $\mathrm{E}[u_i z_i] = 0$. IV works with this assumption as well, but only for $\Phi(z_i) = z_i$

and $\mathrm{E}[v_i|z] = 0$ because of the properties of projections. Conditional quantiles do not have this linearity property, and $Q_v(\tau|z)$ need not equal zero unless $\tau = \tau_r$. The required assumption for 2SQR to be consistent, $Q_{u+v'\beta}(\tau|z)$ does not depend on z, is difficult to interpret. Consequently, there have not been many applications of this approach.

3.2. **Control function approach to quantile regression.** Lee (2007) analyzes the control function approach to quantile regression with endogeneity. Blundell and Powell (2007) study the control function approach to quantile regression with endogeneity and censoring. To motivate the control function, first observe that, as long as $x = f(z, \nu)$ for some function that is invertible in ν, it will always be true that

$$Q_u(\tau|x,z) = Q_u(\tau|v,z).$$

Now, suppose that (v, u) is independent of z. This is the strongest sense in which z could be assumed exogenous. Then,

$$Q_u(\tau|v,z) = Q_u(\tau|v). \tag{19}$$

In this case,

$$Q_y(\tau|x,z,v) = x'\beta(\tau) + z_1'\gamma(\tau) + Q_u(\tau|v).$$

In fact, to get this equation, we just need to assume (19) instead of independence.

This sort of model, where want to estimate a function that is linear in some variables plus an unrestricted function of some other variables is called a partially linear model. If we observed v, we could estimate it be doing a quantile regression of y_i on x_i, z_{1i}, and a series or kernel of v_i. v is not observed, but we can estimate it. We begin by estimating $\hat{\pi}(\tau_r)$ using quantile regression. We then form an estimate of $\hat{v}_i = x_i - z_i\hat{\phi}(\tau_r)$. Finally, we can estimate β, γ, and $Q_u(\tau|v)$ by performing a partially linear quantile regression. See Lee (2007) or Blundell and Powell (2007) for details.

3.3. **IV quantile regression.** Instrumental variable quantile regression assumes that

$$Q_u(\tau|z) = 0.$$

It is immediate that

$$Q_{y-x'\beta(\tau)-z_1'\gamma(\tau)}(\tau|z) = 0.$$

In terms of a minimization problem, this means that

$$0 \in \arg\min_{f\in\mathcal{F}} \mathrm{E}\left[\rho_\tau\left(y - x'\beta(\tau) - z_1'\gamma(\tau) - f(z)\right)\right].$$

This suggests estimating β and γ as follows. Choose $w_i = \Phi(z_i)$. Define $\hat{\alpha}(\beta,\tau)$ and $\hat{\gamma}(\beta,\tau)$ as

$$(\hat{\gamma}(\beta,\tau), \hat{\alpha}(\beta,\tau))) = \arg\min \mathbb{E}_n\left[\rho_\tau\left(y - x'\beta - z_1'\gamma - w'\alpha\right)\right].$$

In other words, $\hat{\gamma}(\beta,\tau)$ and $\hat{\alpha}(\beta,\tau)$ are the coefficients from a quantile regression of $y - x'\beta$ on z_1 and w. At the true β and γ, we should get $\alpha = 0$. Therefore, estimate β by

$$\hat{\beta}(\tau) = \arg\min_{\beta} \|\hat{\alpha}(\beta,\tau)\|,$$

and set $\hat{\gamma}(\tau) = \hat{\gamma}(\hat{\beta}(\tau),\tau)$. Chernozhukov and Hansen (2006) discuss inference for IV quantile regression.

3.3.1. *Quantile treatment effects.* Chernozhukov and Hansen (2005) show how the assumption

$$Q_u(\tau|z) = 0 \tag{20}$$

arises from a model of quantile treatment effects. As in the previous notes on treatment effects, suppose we have some treatment d. For each possible value of d there is a potential outcome Y_d. We have some exogenous covariates x and instruments z. Y_d is given by

$$Y_d = q(d, x, U_d)$$

where $U_d \sim U(0,1)$. q is called the quantile treatment response, and we are interested in quantile treatment effects defined as

$$q(d_1, x, \tau) - q(d_0, x, \tau).$$

These represent the change in the τth quantile of outcomes conditional on x if everyone is changed from treatment d_0 to treatment d_1. Note that we can write the average treatment effect in terms of quantile treatment effects by integrating, $\int_0^1 q(d_1, x, \tau) - q(d_0, x, \tau) d\tau$.

Chernozhukov and Hansen (2005) show that the following assumptions imply (20).

A1 (Potential outcomes)**.** *Conditional on $X = x$, for each d,*

$$Y_d = q(d, x, U_d)$$

where $U_d \sim U(0,1)$ and q is strictly increasing in its third argument.

A2 (Independence)**.** *Conditional on $X = x$, $\{U_d\}$ are independent of Z.*

A3 (Selection)**.**

$$D = \delta(Z, X, V)$$

for some unknown function δ and random vector V.

A4 (Rank similarity)**.** *Conditional on $X = x$ and $Z = z$, $\{U_d\}$ are identically distributed, conditional on V.*

A5 (Observed variables)**.** *$Y = q(D, X, U_D)$, D, X, and Z are observed.*

Theorem 3.1. *If assumptions A1-A5 hold, then for all $\tau \in (0,1)$, a.s.*

$$\mathrm{P}\left(Y \leq q(D, X, \tau)|X, Z\right) = \tau \tag{21}$$

and $U_D \sim U(0,1)$ conditional on Z and X.

Proof. See Chernozhukov and Hansen (2005). □

Given the result of theorem 3.1 and the definition of conditional quantiles, an immediate implication is that

$$Q_{Y-q(D,X,\tau)}(\tau|X, Z) = 0.$$

If we additionally assume that $q(D, X, \tau) = D\beta(\tau) + X\gamma(\tau)$ we have exactly the same condition as in instrumental variables quantile regression.

Theorem 3.1 gives us an estimating equation for $q(d, x, \tau)$, but it does not guarantee that $q(d, x, \tau)$ is the only function satisfying (21). We need an additional assumption to guarantee that the solution to (21) is unique. This assumption is the analog of the rank condition for mean IV regression. Suppose the treatment and the instrument are both binary. Conditional on $X = x$, $q(D, X, \tau)$ is just a two-dimensional vector, $q_0 \equiv q(0,X,\tau)$ and $q_1 \equiv q(1,X,\tau)$. Then for each possible $\tilde{q} = \tilde{q}_0, \tilde{q}_1$, (21) can be written,

$$\Pi(\tilde{q}) \equiv \begin{pmatrix} \mathrm{P}(Y \leq \tilde{q}_0(1-D) + D\tilde{q}_1 | X, Z=0) - \tau \\ \mathrm{P}(Y \leq \tilde{q}_0(1-D) + D\tilde{q}_1 | X, Z=1) - \tau \end{pmatrix} = 0$$

We want to know whether $\tilde{q} = q$ uniquely solves this equations. We know that this equation has a locally unique solution if its Jacobian has rank two. The Jacobian can be written

$$\Pi'(\tilde{q}) = \begin{pmatrix} f_Y(\tilde{q}_0|X, D=0, Z=0)\mathrm{P}(D=0|X,Z=0) & f_Y(\tilde{q}_1|X, D=1, Z=0)\mathrm{P}(D=1|X,Z=0) \\ f_Y(\tilde{q}_0|X, D=0, Z=1)\mathrm{P}(D=0|X,Z=1) & f_Y(\tilde{q}_1|X, D=1, Z=1)\mathrm{P}(D=1|X,Z=1) \end{pmatrix}$$
$$= \begin{pmatrix} f_{Y,D}(\tilde{q}_0, 0|X, Z=0) & f_{Y,D}(\tilde{q}_1, 1|X, Z=0) \\ f_{Y,D}(\tilde{q}_0, 0|X, Z=1) & f_{Y,D}(\tilde{q}_1, 1|X, Z=1) \end{pmatrix}$$

For this to have full rank when $\tilde{q} = q$, we need

$$f_{Y,D}(q_0, 0, |X, Z=0) f_{Y,D}(q_1, 1|X, Z=1) - f_{Y,D}(q_1, 1|X, Z=0) f_{Y,D}(q_0, 0|X, Z=1) > 0$$
$$\frac{f_{Y,D}(q_1, 1|X, Z=1)}{f_{Y,D}(q_0, 1|X, Z=1)} > \frac{f_{Y,D}(q_1, 0|X, Z=0)}{f_{Y,D}(q_0, 0, |X, Z=0)}$$

It would be a good exercise to try to think of a simpler assumption that implies this condition. See Chernozhukov and Hansen (2005) for one such condition. Similar identification conditions can be stated when the treatment and/or instrument are continuous, see Chernozhukov and Hansen (2005).

Assumptions A1-A5 are fairly natural if we have a randomized experiment, or if we think of the instrument as inducing a natural experiment. However, they also make sense if we think about some other models. Chernozhukov and Hansen (2006) give two examples: a Roy model of education, and supply and demand. Homework 5 goes over the supply and demand example.

Example 3.1 (Roy model of education)**.** Let $d \in \mathcal{D} := \{0, 1, ..., \bar{d}\}$ be possible levels of education. Suppose that potential earnings are

$$Y_d = q(d, x, U)$$

where the rank variable, U, is determined by ability and other factors that do not vary with d. d is chosen to maximize utility,

$$D = \arg\max_{d \in \mathcal{D}} \mathrm{E}[W(Y_d, d, X)|X, Z, \nu]$$

where W is an unobserved utility function, and ν includes unobserved information that is correlated with U, and other shocks that affect the education decision.

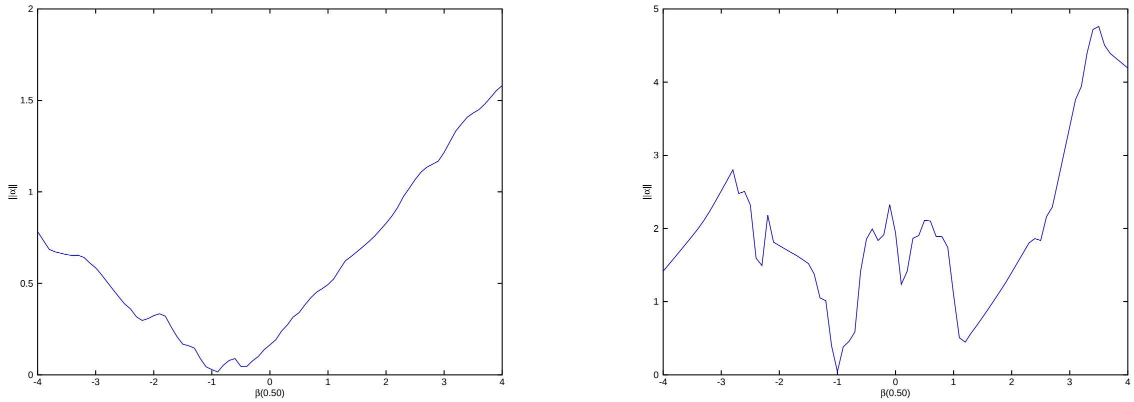
3.3.2. *Comparison to control function quantile regression.* Control function quantile regression requires that the first stage error, v, and the endogenous variable, X, have the same type[5] of support so that

$$Q_u(\tau|x, z) = Q_u(\tau|v, z).$$

In particular, if v is continuous, x, must also be continuous. This rules out the common cases where the endogenous variable is binary and generated from a latent index model. Additionally, the assumption that $Q_u(\tau|v, z) = Q_u(\tau|v)$ required for the control function approach is not the same as the assumption that $Q_u(\tau|z) = 0$ required for IV quantile regression. Neither assumption implies the other. It would be interesting to come up with some examples where one assumption holds but not the other.

[5]More formally, their supports need to have the same topology.

FIGURE 2. $\|\hat{\alpha}(\beta,\tau)\|$



3.3.3. *Computation.* As described above, IV quantile regression can be computed by solving

$$\hat{\beta}(\tau) = \arg\min_{\beta} \|\hat{\alpha}(\beta,\tau)\|, \tag{22}$$

where

$$(\hat{\gamma}(\beta,\tau), \hat{\alpha}(\beta,\tau))) = \arg\min \mathbb{E}_n\left[\rho_\tau\left(y - x'\beta - z_1'\gamma - w'\alpha\right)\right].$$

Given β, $\hat{\gamma}(\beta,\tau)$ and $\hat{\alpha}(\beta,\tau)$ are given by a standard quantile regression. These are fast and easy to compute. However, $\|\hat{\alpha}(\beta,\tau)\|$ need not be a well-behaved function. If the identification condition holds, then $\|\alpha(\beta,\tau)\|$ has a unique minimum. However, $\|\hat{\alpha}(\beta,\tau)\|$ need not have a unique minimum in finite samples. Additionally, $\|\hat{\alpha}(\beta,\tau)\|$ tends to have many local minima. Figure 2 shows two examples of $\|\hat{\alpha}(\beta,\tau)\|$ as function of β. These were generated using the data for homework 5. The left panel uses mixed and stormy as instruments, the right panel uses mixed, stormy, and wind speed as instruments. Both have $\tau = 0.5$. As shown, $\|\hat{\alpha}(\beta,\tau)\|$ can be very badly behaved. When β is only one or two dimensional, this can be overcome by doing an exhaustive grid search for minimization. However, as the dimension of β increases, the number of points in a grid needed to achieve a fixed precision increases exponentially.

When the dimension of β is high, the MCMC approach of Chernozhukov and Hong (2003) can be used to compute IV quantile regression. Let

$$\hat{g}(\beta,\gamma) = \mathbb{E}_n\left[(\tau - 1\{Y < X'\beta + Z_1'\gamma\})Z\right],$$

$$W_n = \frac{1}{\tau(1-\tau)}\mathbb{E}_n[ZZ'],$$

and

$$\hat{Q}(\beta,\gamma) = n\frac{1}{2}\hat{g}(\beta,\gamma)'W_n\hat{g}(\beta,\gamma).$$

Then Chernozhukov and Hong (2003) show that the quasi-posterior mean,

$$(\hat{beta}^{MCMC}, \hat{\gamma}^{MCMC}) \int (\beta,\gamma)\frac{e^{-\hat{Q}(\beta,\gamma)}}{\int e^{-\hat{Q}(\beta,\gamma)}d(\beta,\gamma)}d(\beta,\gamma)$$

converges to β. Furthermore, $\hat{\beta}^{MCMC}, \hat{\gamma}^{MCMC}$ have the same asymptotic distribution as the estimates you would get from minimizing $\|\hat{\alpha}(\beta,\tau)\|$ (when the norm is chosen appropriately). Finally, the quasi-distribution,

$$\frac{e^{-\hat{Q}(\beta,\gamma)}}{\int e^{-\hat{Q}(\beta,\gamma)} d(\beta,\gamma)}$$

can be used for inference, so, for example, the interval from the 2.5%tile to the 97.5%tile of this distribution is a valid confidence interval for β.

Finding posterior distributions is central to Bayesian statistics. As a result, there are many methods for simulating from posterior distributions. Collectively, these methods are called Markov Chain Monte Carlo or MCMC. They construct a Markov Chain with transition densities that can be easily simulated and with stationary distribution equal to the desired posterior. One general purpose MCMC method is the Metropolis-Hastings algorithm. Let $\theta = (\beta, \gamma)$. Suppose we have some conditional density that can easily be sampled, $q(\theta'|\theta)$. In applications, $q(\theta'|\theta)$ is often a random walk, such as

$$\theta'|\theta \sim N(\theta, \sigma^2)$$

or

$$\theta'|\theta \sim U(\theta - \sigma, \theta + \sigma).$$

In the Metropolis Hasting algorithm you then

(1) Choose a starting value $\theta^{(0)}$.
(2) Draw ξ from $q(\xi|\theta^{(j)})$.
(3) Set

$$\theta^{(j+1)} = \begin{cases} \xi & \text{with probability } \rho(\theta^{(j)}, \xi) \\ \theta^{(j)} & \text{with probability } 1 - \rho(\theta^{(j)}, \xi) \end{cases}$$

where

$$\rho(\theta, \xi) = \min\{\frac{e^{-\hat{Q}(\xi)} q(\theta|\xi)}{e^{-\hat{Q}(\theta)} q(\xi|\theta)}, 1\}.$$

Note that this algorithm is more likely to accept ξ when $e^{-\hat{Q}(\xi)}$ is relatively high. This ensures that the stationary distribution of the chain is proportional to $e^{-\hat{Q}(\xi)}$. If we use either a normal or uniform random walk for $q(\theta'|\theta)$, then we face a tradeoff when choosing σ. If σ is low, we will accept many draws of ξ, but the values of $\theta^{(j)}$ will be close together. This may lead us to need many draws for $\theta^{(j)}$ to adequately explore the support of θ. On the other hand if σ is high, accepted draws of $\theta^{(j)}$ will be further apart, but the probability of acceptance is lower. See the references in Chernozhukov and Hong (2003) for more information. It will take some time for the draws of $\theta^{(j)}$ to converge to their stationary distribution, so the first B_0 draws should be discarded. Then, the average of the next B_1 draws can be used as estimate of $\hat{\beta}^{MCMC}$, and confidence regions can be constructed from the quantiles of these draws.

3.3.4. *Local QTE.* Abadie, Angrist, and Imbens (2002) develop an approach to quantile treatment effects that is similar to LATE. As in LATE, suppose there is a binary treatment d and a binary instrument z. Let Y_d be the potential outcomes and D_z be the potential treatments. As in LATE, assume the following,

LQTE–A6. (1) Independence*:* $(Y_1, Y_0, D_1, D_0) \perp\!\!\!\perp Z|X$
(2) Nontrivial assignment*:* $\mathrm{P}(Z = 1|X) \in (0, 1)$
(3) First-stage*:* $\mathrm{E}[D_1|X] \neq \mathrm{E}[D_0|X]$
(4) Monotonicity*:* $\mathrm{P}(D_1 \geq D_0|X) = 1$.

As in LATE, with these assumptions, we can identify the quantile treatment effect for compliers ($D_1 > D_0$). For simplicity, assume linearity.

LQTE–A7.

$$Q_Y(\tau|X, D, D_1 > D_0) = \alpha(\tau)D + X'\beta(\tau)$$

It follows that

$$(\alpha(\tau), \beta(\tau)) = \arg\min_{\alpha,\beta} \mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)|D_1 > D_0\right]$$

The event $D_1 > D_0$ is unobserved. However, it can be shown that

$$\begin{aligned}\mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)|D_1 > D_0\right] \mathrm{P}(D_1 > D_0) =& \mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)\left(1 - \frac{D(1-Z)}{1-\mathrm{P}(Z=1|X)} - \frac{(1-D)Z}{\mathrm{P}(Z=1|X)}\right)\right] \\ =& \mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)\kappa\right]\end{aligned}$$

with $\kappa \equiv 1 - \frac{D(1-Z)}{1-\mathrm{P}(Z=1|X)} - \frac{(1-D)Z}{\mathrm{P}(Z=1|X)}$. κ is an estimable function of observable variables, so

$$(\alpha(\tau), \beta(\tau)) = \arg\min_{\alpha,\beta} \mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)\kappa\right]$$

can be used for estimating. Since the population objective function is equal to the conditional expectation of the check function, it must be convex. However, in finite samples κ can be both positive and negative, so the sample minimization problem,

$$(\hat{\alpha}(\tau), \hat{\beta}(\tau)) = \arg\min_{\alpha,\beta} \mathbb{E}_n\left[\rho_\tau(Y - \alpha D - X'\beta)\kappa\right],$$

need not be convex. However, by iterated expectations,

$$\mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)\kappa\right] = \mathrm{E}\left[\rho_\tau(Y - \alpha D - X'\beta)\mathrm{E}[\kappa|Y, D, X]\right]$$

and as Abadie, Angrist, and Imbens (2002) show, $\mathrm{E}[\kappa|Y, D, X] = \mathrm{P}(D_1 > D_0|Y, D, X) \geq 0$. This suggests estimating by solving

$$(\hat{\alpha}(\tau), \hat{\beta}(\tau)) = \arg\min_{\alpha,\beta} \mathbb{E}_n\left[\rho_\tau(Y - \alpha D - X'\beta)\mathrm{E}[\kappa\widehat{|Y, D}, X]\right],$$

where $\mathrm{E}[\kappa\widehat{|Y, D}, X]$ is some non-negative consistent estimate of

$$\mathrm{E}[\kappa|Y, D, X] = \mathrm{E}\left[1 - \frac{D(1-\mathrm{E}[Z|Y, D, X])}{1-\mathrm{P}(Z=1|X)} - \frac{(1-D)\mathrm{E}[Z|Y, D, X]}{\mathrm{P}(Z=1|X)}\right].$$

Abadie, Angrist, and Imbens (2002) propose estimating $\mathrm{P}(Z = 1|X)$ and $\mathrm{E}[Z|Y, D, X]$ by series regression, and then setting

$$(\hat{\alpha}(\tau), \hat{\beta}(\tau)) = \arg\min_{\alpha,\beta} \mathbb{E}_n\left[\rho_\tau(Y - \alpha D - X'\beta)\mathrm{E}[\kappa\widehat{|Y, D}, X]1\{\mathrm{E}[\kappa\widehat{|Y, D}, X] \geq 0\}\right].$$

The indicator function here is just to ensure convexity of the sample objective function. Asymptotically, $\mathrm{P}(\mathrm{E}[\kappa\widehat{|Y, D}, X] \geq 0) \to 1$, so the indicator function does not affect asymptotic behavior of the objective function. Abadie, Angrist, and Imbens (2002) give conditions for $\hat{\alpha}$ and $\hat{\beta}$ to be $\sqrt{n}$ asymptotically normal and give the asymptotic variance.

4. APPLICATIONS

4.1. **Quantile regression.**

4.1.1. *Subjective wine quality and physical wine characteristics.* Figure 3 shows quantile regression estimates from regressing wine quality (measured by the median of three reviewers' assessments on a scale from 1 to 10) on various chemical characteristics of the wine. The data comes from Cortez, Cerdeira, Almeida, Matos, and Reis (2009) and is available at the UCI machine learning repository, `http://archive.ics.uci.edu/ml/datasets/Wine+Quality`. Each of the covariates have been standardized to have mean zero and standard deviation one. The solid red line is the OLS estimate, and the dashed lines form a 95% confidence interval. The dotted black line are the quantile regression estimates as a function of τ, and the gray region is a 95% confidence interval. Figure 4 shows the same thing for white wine. Figure 5 shows a scatter plot of quality as a function of alcohol, and the fitted quantile regression lines for the bivariate quantile regression of quality on alcohol.

FIGURE 3. Quantile regression of red wine quality on physical characteristics

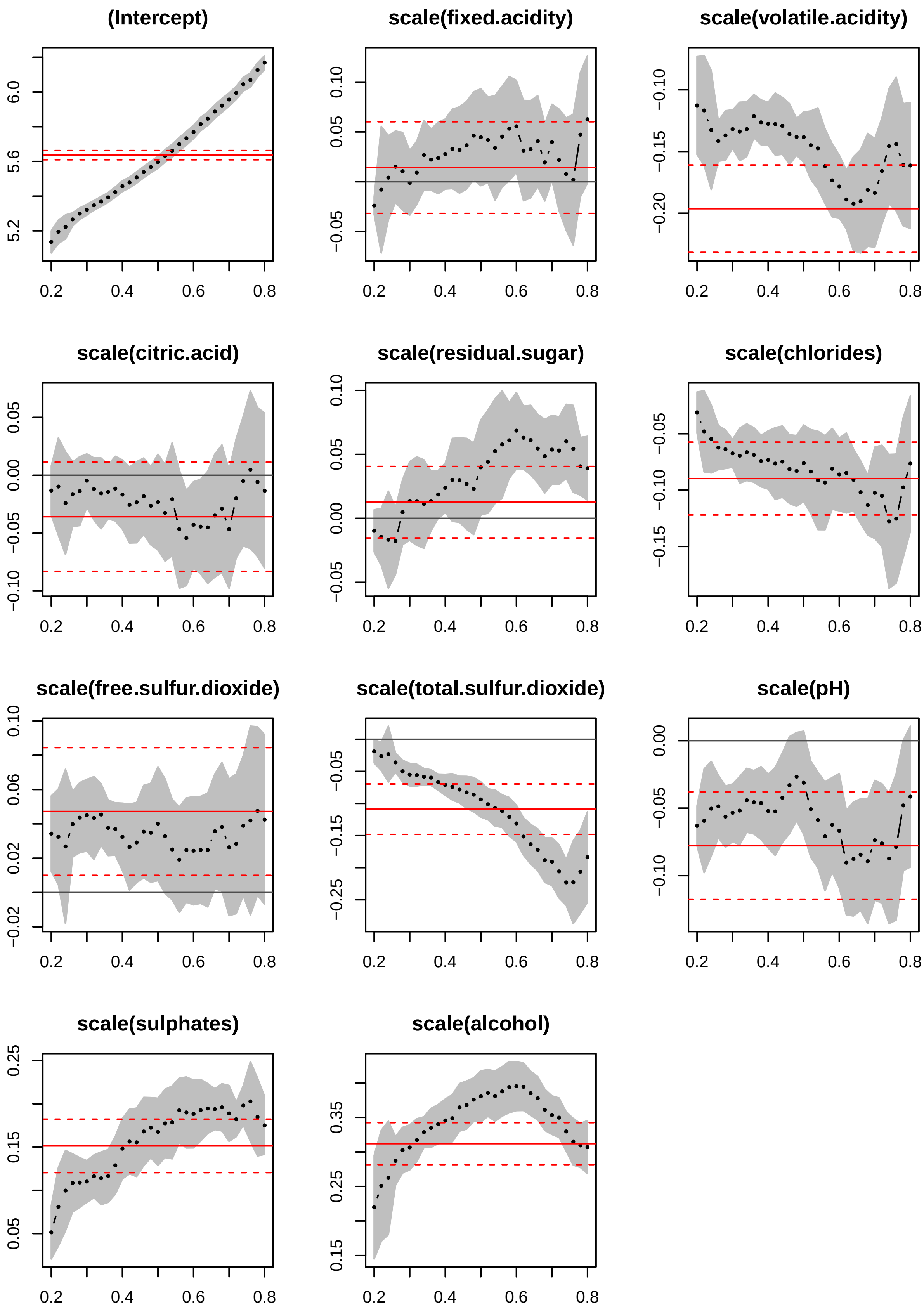


FIGURE 4. Quantile regression of white wine quality on physical characteristics

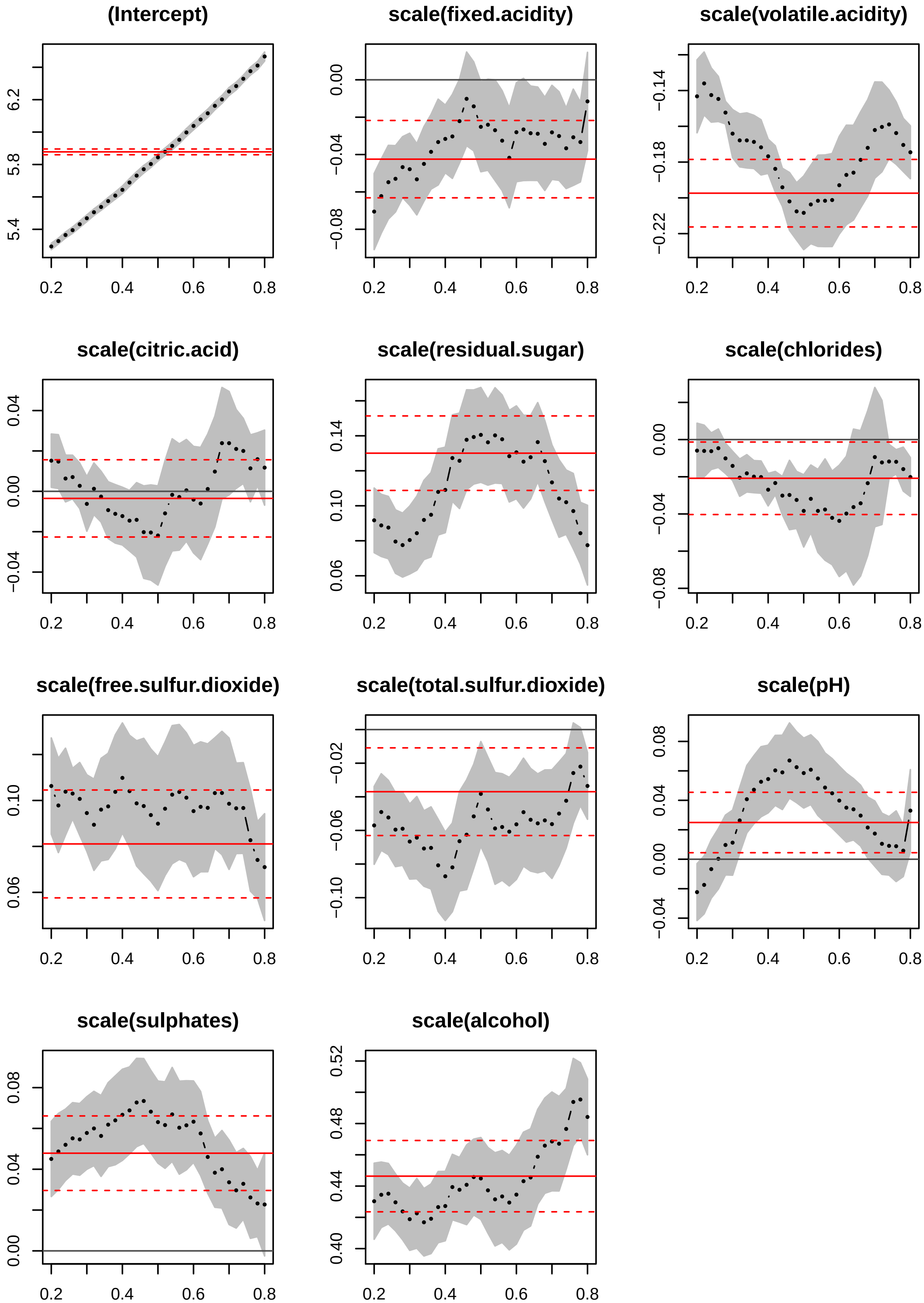
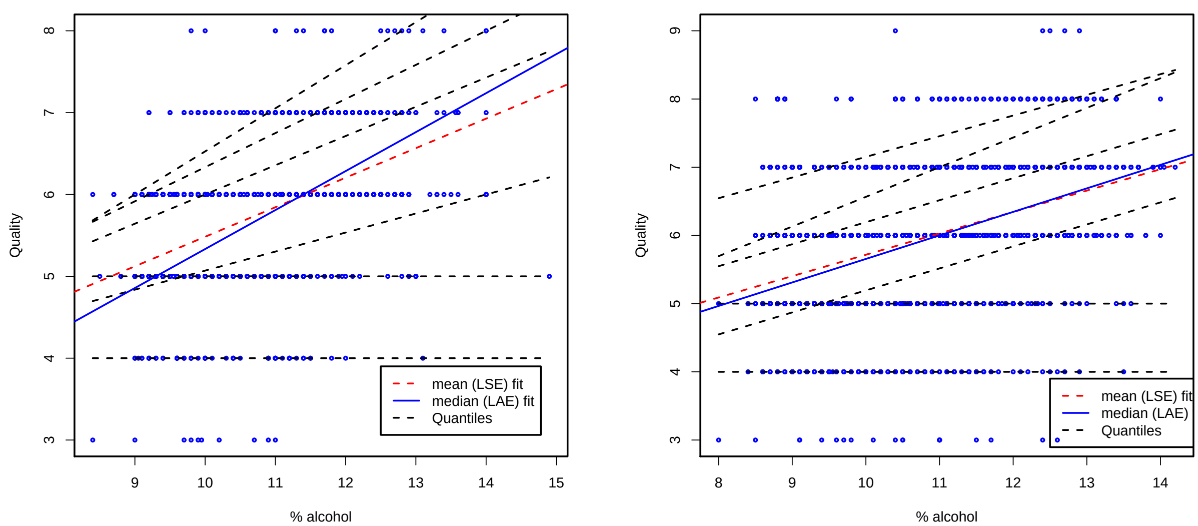


FIGURE 5. Wine quality and alcohol content

Red White



4.1.2. *Decomposition of distribution changes* . One popular application of quantile regression has been the analysis of the change in inequality. During the 80s and 90s, income inequality increased in the US and much of the rest of the world. Labor economists have been interested in the mechanism through which this happened. One way of thinking about the increase in inequality is to try to break it into (1) changes in the observed distribution of characteristics, (2) changes in the prices of worker characteristics, and (3) residual changes. Juhn, Murphy, and Pierce (1993) were the first to consider this sort of decomposition, but they had a somewhat ad-hoc method. DiNardo, Fortin, and Lemieux (1996) proposed a more complicated method based on kernel re-weighting. Many others have used similar methods. Melly (2005) and Machado and Mata (2005) use quantile regression to perform the decomposition. In a quantile regression,

$$y_{it} = x_{it}\beta_t(\tau)$$

x_{it} are the observed characteristics, β_t are the prices, and τ captures residual changes. For each year, t, we can estimate $\beta_t(\tau)$. We can use these estimates to simulate what y_{it} would have been had x's been distributed as in year s,

$$\hat{y}_{it|s} = x_{is}\hat{\beta}_t(\tau)$$

We can seperate out residual inequality by looking at the distribution of

$$\hat{u}_{it|s} = \hat{y}_{it|s} - E[\hat{y}_{it|s}|x_{is}] = x_{is}(\hat{\beta}_t(\tau) - \hat{\beta}_{1/2}(\tau))$$

Angrist, Chernozhukov, and Fernández-Val (2006) is, in part, about how to interpret this sort of quantile regression if the true conditional quantile is not linear. A main result is that the $x\beta(\tau)$ minimizes a weighted squared difference from the true conditional quantile function.

4.2. **IVQR and LQTE.**

4.2.1. *JTPA.* The Job Training Partnership Act was a large publicly-funded job training program in the US. Individuals in the treatment were randomly offered training immediately, while individuals in the control group were only offered training 18 months later. Abadie, Angrist, and

Imbens (2002) applied their local QTE estimator to the JTPA. The outcome variable is the sum of earnings in the 30 months following treatment. The treatment is receiving training, and the instrument is being offered training. The LATE assumptions are quite natural here. In particular, the monotonicity assumption seems sensible. The covariates include race, education, marital status, and age. Table II from Abadie, Angrist, and Imbens (2002) shows OLS and quantile regression estimates of the effect of the training program. Table III from Abadie, Angrist, and Imbens (2002) shows 2SLS and local QTE estimates.

TABLE II

QUANTILE REGRESSION AND OLS ESTIMATES

Dependent Variable: 30-month Earnings

		Quantile				
	OLS	0.15	0.25	0.50	0.75	0.85
A. Men						
Training	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of Training	21.2	135.6	75.2	34.5	17.2	13.4
High school or GED	4,015	339	1,280	3,665	6,045	6,224
	(571)	(186)	(305)	(618)	(1,029)	(1,170)
Black	−2,354	−134	−500	−2,084	−3,576	−3,609
	(626)	(194)	(324)	(684)	(1,087)	(1,331)
Hispanic	251	91	278	925	−877	−85
	(883)	(315)	(512)	(1,066)	(1,769)	(2,047)
Married	6,546	587	1,964	7,113	10,073	11,062
	(629)	(222)	(427)	(839)	(1,046)	(1,093)
Worked less than 13	−6,582	−1,090	−3,097	−7,610	−9,834	−9,951
weeks in past year	(566)	(190)	(339)	(665)	(1,000)	(1,099)
Constant	9,811	−216	365	6,110	14,874	21,527
	(1,541)	(468)	(765)	(1,403)	(2,134)	(3,896)
B. Women						
Training	2,215	367	1,013	2,707	2,729	2,058
	(334)	(105)	(170)	(425)	(578)	(657)
% Impact of Training	18.5	60.8	44.4	32.3	14.5	8.09
High school or GED	3,442	166	681	2,514	5,778	6,373
	(341)	(99)	(156)	(396)	(606)	(762)
Black	−544	22	−60	−129	−866	−1,446
	(397)	(115)	(188)	(451)	(679)	(869)
Hispanic	−1,151	−31	−222	−995	−1,620	−1,503
	(488)	(130)	(194)	(546)	(911)	(992)
Married	−667	−213	−392	−758	−1,048	−902
	(436)	(127)	(209)	(522)	(785)	(970)
Worked less than 13	−5,313	−1,050	−3,240	−6,872	−7,670	−6,470
weeks in past year	(370)	(137)	(289)	(522)	(672)	(787)
AFDC	−3,009	−398	−1,047	−3,389	−4,334	−3,875
	(378)	(107)	(174)	(468)	(737)	(834)
Constant	10,361	649	2,633	8,417	16,498	20,689
	(815)	(255)	(490)	(966)	(1,554)	(1,232)

Note: The table reports OLS and quantile regression estimates of the effect of training on earnings. The specification also includes indicators for service strategy recommended, age group, and second follow-up survey. Robust standard errors are reported in parentheses.

QTE estimates of the effect of training on median earnings, reported in Table III, are similar in magnitude though less precise than the benchmark 2SLS estimates. For men, the QTE estimates show a pattern very different from the quantile regression estimates, with no evidence of an impact on the .15 or .25

we computed QTE coefficient estimates by weighted quantile regression using the Barrodale-Roberts (1973) linear programming algorithm for quantile regression (see, e.g., Koenker and D'Orey (1987)). A biweight kernel was used for the estimation of standard errors.

TABLE III
QUANTILE TREATMENT EFFECTS AND 2SLS ESTIMATES
Dependent Variable: 30-month Earnings

		Quantile				
	2SLS	0.15	0.25	0.50	0.75	0.85
A. Men						
Training	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of Training	8.55	5.19	12.0	9.64	10.7	9.02
High school or GED	4,075	714	1,752	4,024	5,392	5,954
	(573)	(429)	(644)	(940)	(1,441)	(1,783)
Black	−2,349	−171	−377	−2,656	−4,182	−3,523
	(625)	(439)	(626)	(1,136)	(1,587)	(1,867)
Hispanic	335	328	1,476	1,499	379	1,023
	(888)	(757)	(1,128)	(1,390)	(2,294)	(2,427)
Married	6,647	1,564	3,190	7,683	9,509	10,185
	(627)	(596)	(865)	(1,202)	(1,430)	(1,525)
Worked less than 13	−6,575	−1,932	−4,195	−7,009	−9,289	−9,078
weeks in past year	(567)	(442)	(664)	(1,040)	(1,420)	(1,596)
Constant	10,641	−134	1,049	7,689	14,901	22,412
	(1,569)	(1,116)	(1,655)	(2,361)	(3,292)	(7,655)
B. Women						
Training	1,780	324	680	1,742	1,984	1,900
	(532)	(175)	(282)	(645)	(945)	(997)
% Impact of Training	14.6	35.5	23.1	18.4	10.1	7.39
High school or GED	3,470	262	768	2,955	5,518	5,905
	(342)	(178)	(274)	(643)	(930)	(1026)
Black	−554	0	−123	−401	−1,423	−2,119
	(397)	(204)	(318)	(724)	(949)	(1,196)
Hispanic	−1,145	−73	−138	−1,256	−1,762	−1,707
	(488)	(217)	(315)	(854)	(1,188)	(1,172)
Married	−652	−233	−532	−796	38	−109
	(437)	(221)	(352)	(846)	(1,069)	(1,147)
Worked less than 13	−5,329	−1,320	−3,516	−6,524	−6,608	−5,698
weeks in past year	(370)	(254)	(430)	(781)	(931)	(969)
AFDC	−2,997	−406	−1,240	−3,298	−3,790	−2,888
	(378)	(189)	(301)	(743)	(1,014)	(1,083)
Constant	10,538	984	3,541	9,928	15,345	20,520
	(828)	(547)	(837)	(1,696)	(2,387)	(1,687)

Note: The table reports 2SLS and QTE estimates of the effect of training on earnings. Assignment status is used as an instrument for training. The specification also includes indicators for service strategy recommended, age group, and second follow-up survey. Robust standard errors are reported in parentheses.

quantile. The estimates at low quantiles are substantially smaller than the corresponding quantile regression estimates, and they are small in absolute terms. For example, the QTE estimate (standard error) of the effect on the .15 quantile for men is $121 (475), while the corresponding quantile regression estimate is $1,187 (205). Similarly, the QTE estimate (standard error) of the effect on the .25 quantile for men is $702 (670), while the corresponding quantile regression estimate is

Chernozhukov and Hansen (2008) also analyze the JTPA using IVQR instead of local QTE. Figure 6 shows their estimates of the effect of training. The estimates are fairly similar to Abadie, Angrist, and Imbens (2002), but not identical. The LQTE estimator of Abadie, Angrist, and Imbens (2002) and the IVQR estimator of Chernozhukov and Hansen (2005) generally identify and estimate different quantities. The LQTE estimator identifies the quantile treatment effect conditional on being a complier. It only applies to binary treatments and instruments. The IVQR estimator identifies quantile treatment effects unconditional on being a complier, and applies non-binary as well as binary treatments and instruments. However, the IVQR estimator requires the perhaps stronger assumption of rank similarity. Nonetheless, IVQR and LQTE can estimate the same thing if the assumptions of both models are satisfied (mainly rank similarity and monotonicity), and the compliers are representative of the population. If these conditions are not met, then the two estimators will in general have different probability limits. Therefore, a comparison of results based on the two models can provide evidence on the plausibility of their assumptions.

FIGURE 6. Chernozhukov and Hansen (2008) figure 4

V. Chernozhukov, C. Hansen / Journal of Econometrics 142 (2008) 379–398 393

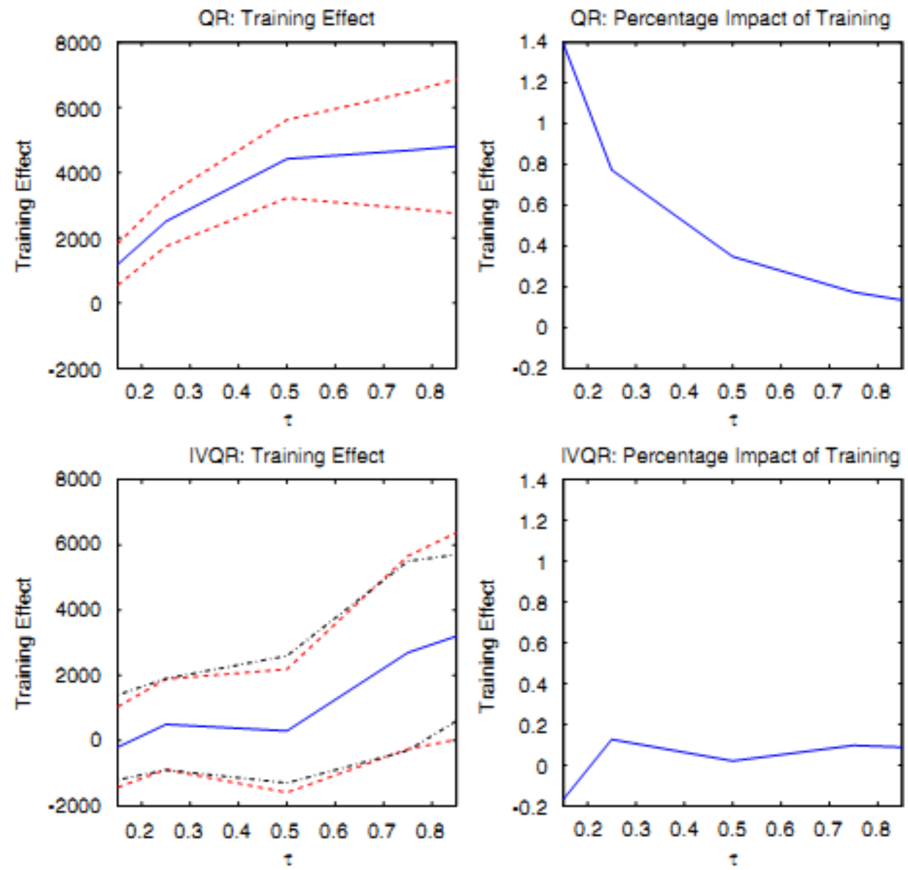
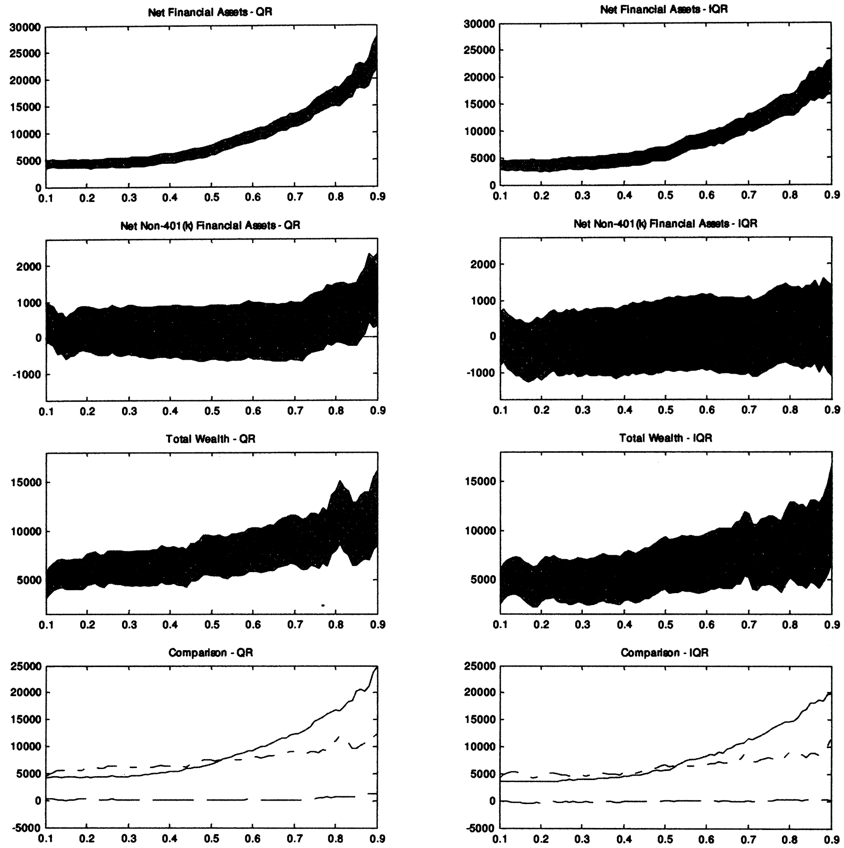


Fig. 4. Estimates of the training impact by QR and by IVQR. *Notes*: Left column. QR and IVQR estimates of the impact of a job training program on earnings for $\tau = 0.15, 0.25, 0.50, 0.75$, and 0.85. The top panel reports the QR estimate of the training impact, and the bottom panel reports the IVQR results. In each figure, the solid line represents the point estimates, and the dashed (- -) line represents the 95% confidence interval formed using the direct inference approach. For the IVQR results, the dash-dot (-.) line represents the 95% confidence bound constructed using the dual inference procedure described in the text. In both figures, the horizontal axis measures the quantile index τ, and the vertical axis is the impact of training on earning quantiles measured in dollars. Models include covariates as specified in the text, and the sample size is 5102. Right column. QR and IVQR estimates of the percentage impact of training for $\tau = 0.15, 0.25, 0.50, 0.75$, and 0.85. The top panel reports the QR estimate of the training impact, and the bottom panel reports the IVQR results. Percentage impacts are for moving from non-training to training and all other covariates are evaluated at their sample mean. In both figures, the horizontal axis measures the quantile index τ, and the vertical axis is the percentage impact of training.

4.2.2. *401(k) participation and wealth.* Chernozhukov and Hansen (2004) estimate the effect of 401(k) participation and wealth. The outcome is wealth, the treatment is 401(k) participation, and the instrument is 401(k) eligibility. Chernozhukov and Hansen (2004) discuss how rank similarity might not be a valid assumption in this context. They say, "In the context of 401(k) participation, matching practices of employers could jeopardize the validity of the similarity assumption. This is because individuals in firms with high match rates may be expected to have a higher rank in the asset distribution than workers in firms with less generous match rates.' This suggests that the distribution of U_d may be different across the treatment states." Still they argue, that rank similarity may still hold. As evidence of this, they also estimate LQTE and find that the results are similar to the IVQR estimates.

FIGURE 1.—QR AND IQR ESTIMATES OF EFFECT OF 401(K) PARTICIPATION



The sample size is 9915. The left column contains standard quantile regression estimates, and the right column contains instrumental quantile regression. Each panel is labeled with the dependent variable used in estimation of the presented results. The bottom panel in each column compares the point estimates for each wealth measure. The solid line corresponds to net financial assets, the dashed line to net non-401(k) financial assets, and the dash-dot line to total wealth. The vertical axis measures the dollar increase in the wealth measure due to 401(k) participation. The quantile of the conditional wealth distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [0.10, 0.90]$ at 0.01-unit intervals.

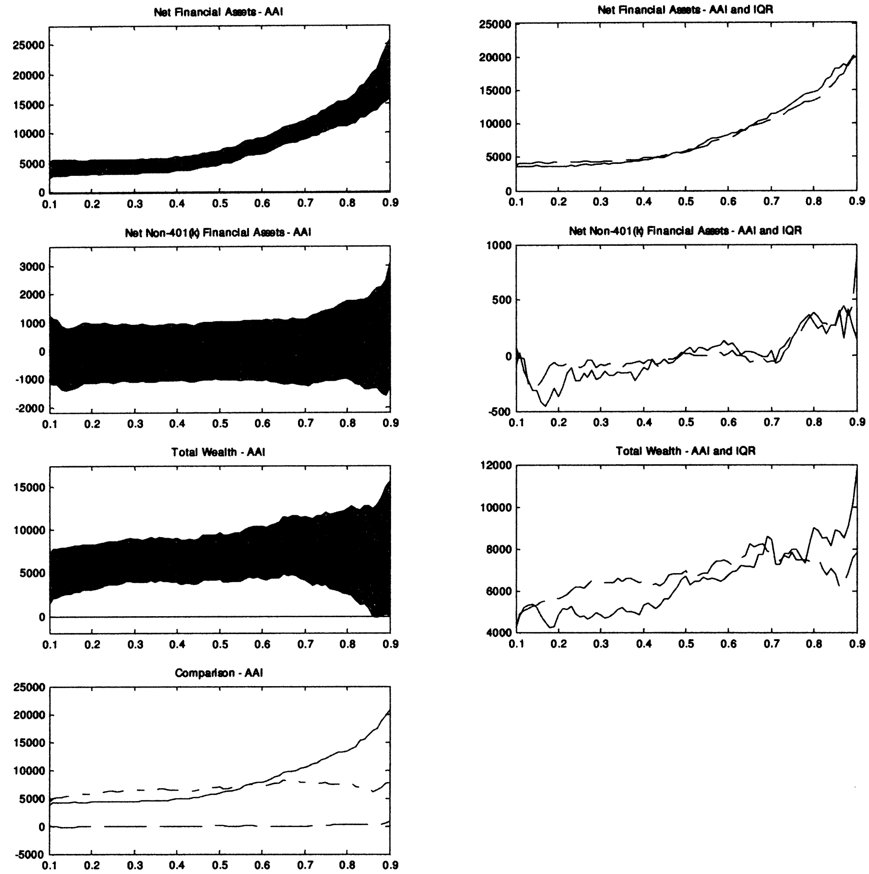
The left column of Figure 1 contains QR estimates of the effect of 401(k) participation on the wealth measures, and the right column presents the IQR estimates of the QTE. The shaded region in the first six panels represents the 95% confidence interval.[22] The last two panels plot the estimated effects for each of the dependent variables together, to provide a comparison of the magnitudes and to facilitate the discussion of substitution between the different wealth measures.

The results exhibit a number of striking features. First, the difference between the QR and IQR estimators is not dramatic. Both exhibit the same pattern of results, though there is some upward bias evident in the QR estimates. This bias is most evident in the estimates for net financial assets and net non-401(k) financial assets, but is hardly noticeable in the total wealth results.

Another interesting feature of the results is that the effect of participation on net financial assets is highly nonconstant, appearing to increase monotonically in the quantile index. This result suggests that, conditional on income and other observables, people who rank higher in the conditional wealth distribution are affected far more than those ranking lower in the conditional distribution. In addition, the effect is strongly positive across the entire distribution. Though these results correspond to our intuition, there is actually no other a priori reason to believe that net financial assets must react in this way. In particular, if people were simply

[22] Standard errors were estimated using heteroskedasticity-consistent standard errors as in Powell (1984, 1986) and Buchinsky (1995), using the methods outlined in Chernozhukov and Hansen (2001).

FIGURE 6.—COMPARISON OF AAI AND IQR



Note: The sample size is 9915. The left column contains estimates obtained using the estimator of Abadie, Angrist, and Imbens (2002) (AAI), and the right column compares them with the corresponding estimates obtained through the IQR estimator of Chernozhukov and Hansen (2001). The solid line corresponds to the IQR estimator, and the dashed line corresponds to the AAI estimator. Each panel is labeled with the dependent variable used in estimation of the presented results. The bottom panel in the left column compares the AAI point estimates for each wealth measure. The solid line corresponds to net financial assets, the dashed line to net non-401(k) financial assets, and the dash-dot line to total wealth. The vertical axis measures the dollar increase in the wealth measure due to 401(k) participation. The quantile of the conditional wealth distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [0.10, 0.90]$ at 0.01-unit intervals.

The IQR estimates suggest that the effect of 401(k) participation on net financial assets is quite heterogeneous, with the largest returns accruing to those who are in the upper tail of the assets distribution. The results also indicate that the effect of 401(k) participation on net financial assets is positive and significant over the entire range of the asset distribution and that the effect is monotonically increasing in the quantile index. Effects on total wealth and net non-401(k) financial assets, on the other hand, appear to be constant, and the effect on net non-401(k) financial assets is not significantly different from 0, whereas the effect on total wealth is positive and significant. Overall, the results suggest that participation in 401(k)s increases net financial assets across the asset distribution, but that this effect is mitigated by substitution with other forms of wealth in the upper tail of the distribution. They also demonstrate that estimates of treatment effects which focus on a single feature of the outcome distribution may fail to capture the full effect of the treatment and that examining additional features may enhance our understanding of the economic relationships involved.

REFERENCES

Abadie, Alberto, "Changes in Spanish Labor Income Structure During the 1980s: A Quantile Regression Approach," *Investigaciones Económicas* 21 (May 1997), 253–272.

4.3. **Quantile regression with censoring.** Unlike mean regression, quantile regression can deal with censoring without making any distributional assumptions. Suppose Y^* has conditional quantile function $Q_{Y^*}(\tau|X)$, but is not observed. Instead, we observe $Y_i = \max\{Y_i^*, C_i\}$. Then it is immediate that the conditional quantile function of Y is

$$Q_Y(\tau|X) = \max\{Q_{Y^*}(\tau|X), C\}.$$

If $Q_{Y^*}(\tau|X) = X'\beta(\tau)$, then we could estimate $\beta(\tau)$ by

$$\hat{\beta}(\tau) = \arg\min_{\beta} \mathbb{E}_n \left[Y - \max\{C, X'\beta(tau)\} \right].$$

This estimator was first proposed by L. and Powell (1986). Due to the max inside the check function, this objective function is not necessarily convex. Nonetheless, a number of effective computational strategies exist for finding the minimum, see the references on Koenker (2005) p253.

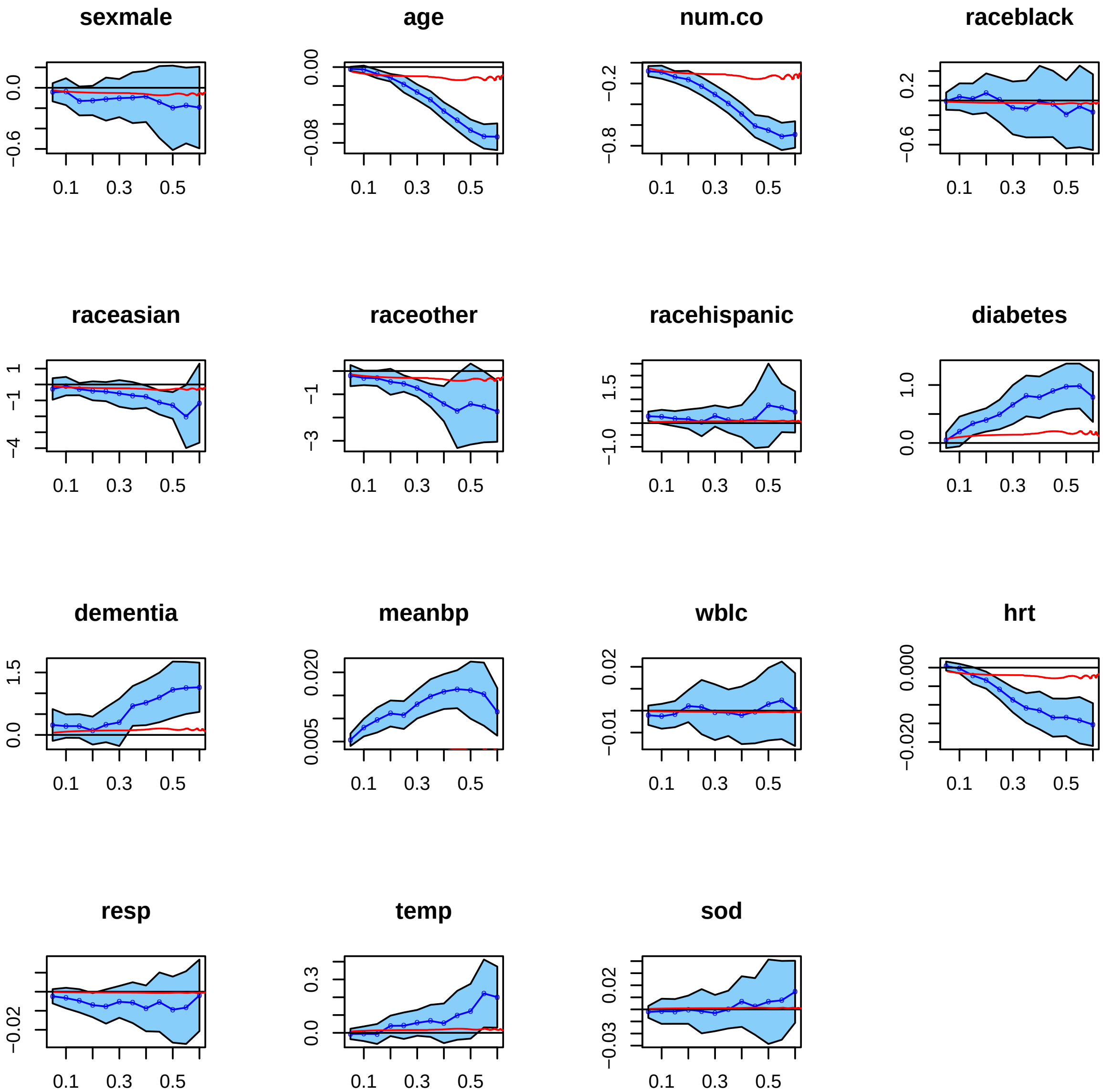
A limitation of the censored quantile regression estimate of L. and Powell (1986) is that it requires that the censoring point, C_i, be known for all observations. For some applications, this is a fine assumption. For example, if Y is spending on some good, then we know $C = 0$. However, in other situations, we do not know C. For example, if Y^* is how long someone survived after some medical treatment and C is when either the study ends or when the person drops out of the study, then we cannot know when someone would have dropped out of the study if they had not died. Portnoy (2003) and Honor, Khan, and Powell (2002) describe an estimators for such situations with random censoring. Both estimators are based on constructing an estimate of Y^* for the censored observations or C for the uncensored observations, and using this estimate to adjust the check function appropriately.

4.3.1. *Survival after hospitalization for ACF or MOSF.* Figure 7 shows Portnoy (2003) censored quantile regression estimates. The outcome is survival time after hospitalization for patients with acute renal failure or multiple organ system failure[6]. About 62% of these patients died within a year, 50% within 6 months, and 32% within one month. The median survival time of those with observed deaths is 28 days. The median censored time of those with no observed death is 918 days. The data comes from the SUPPORT[7] study `http://biostat.mc.vanderbilt.edu/wiki/Main/SupportDesc`.

[6] At least, I think that's what ACF/MOSF stands for.

[7] See the following reference for more information about the SUPPORT study (funded by the Robert Wood Johnson Foundation): Knaus WA, Harrell FE, Lynn J et al. (1995): The SUPPORT prognostic model: Objective estimates of survival for seriously ill hospitalized adults. Annals of Internal Medicine 122:191-203.

FIGURE 7. Censored quantile regression of survival time



`num.co` is the number of co-morbidities (other conditions), `meanbp` is blood pressure, `wblc` is white blood cell count, `hrt` is heart rate, `resp` is respiratory rate, `temp` is body temperature, `sod` is sodium level. Other variables should be self-explanatory. All variables were measured within three days of hospitalization. The solid blue line are the randomly censored quantile regression coefficient estimates. The shaded blue area is a 95% point-wise confidence band.

4.3.2. *Censored quantile regression with endogeneity.* Blundell and Powell (2007) and Chernozhukov, Fernndez-Val, and Kowalski (2011) estimate censored (with known censoring points) quantile regression with endogeneity using a control function approach. Kowalski (2009) uses this approach to estimate the price elasticity of medical care. She has detailed data on medical spending for families that face an annual piecewise linear cost of medical care. These families have insurance plans with a deductible, followed by a constant coinsurance rate, with a stoploss that caps total out of pocket spending. So, for the first X dollars of medical care each year, the family must pay for it completely. For the next X' dollars, the family pays $c < 1$ for each dollar of medical care. If the family has spent more than X'' dollars that year, then any additional medical care has zero marginal price. The outcome is medical care consumed by an individual. The endogenous variable is the marginal cost of medical care at the end of the year. Whether another family member was injured in the year is used as an instrument.

Table 4: 2004 and 2003 CQIV Year-End Price Coefficients

2004 and 2003 CQIV Year-End Price Coefficients
Dependent variable: Ln(Expenditure)

		Censored Quantile IV							
2004 Sample		65	70	75	80	85	90	95	Tobit IV
A. Employee									
N= 29,010	Year-end price	-4.34	-4.27	-4.46	-4.52	-4.62	-4.72	-4.58	-6.36
	lower bound	-5.15	-5.11	-5.16	-5.06	-5.21	-5.27	-4.98	-7.42
	upper bound	-3.30	-3.54	-3.59	-3.95	-4.03	-4.21	-4.17	-5.30
	[Elasticity]	-[2.17]	-[2.14]	-[2.23]	-[2.26]	-[2.31]	-[2.36]	-[2.29]	-[3.18]
B. Employee and Spouse									
N= 53,185	Year-end price	-4.71	-4.69	-4.66	-4.57	-4.51	-4.66	-4.48	-6.57
	lower bound	-5.43	-5.20	-5.10	-5.03	-4.92	-5.01	-4.77	-7.29
	upper bound	-3.93	-4.05	-4.13	-4.10	-4.09	-4.37	-4.12	-5.86
	[Elasticity]	-[2.35]	-[2.35]	-[2.33]	-[2.29]	-[2.25]	-[2.33]	-[2.24]	-[3.29]
C. Everyone									
N= 127,119	Year-end price	-4.03	-3.96	-3.96	-3.92	-3.99	-4.08	-4.13	-6.78
	lower bound	-4.35	-4.23	-4.22	-4.16	-4.24	-4.28	-4.31	-7.28
	upper bound	-3.67	-3.66	-3.74	-3.67	-3.71	-3.86	-3.92	-6.29
	[Elasticity]	-[2.01]	-[1.98]	-[1.98]	-[1.96]	-[2.00]	-[2.04]	-[2.06]	-[3.39]
2003 Sample									
D. Employee									
N= 29,886	Year-end price	-5.02	-4.87	-4.63	-4.33	-4.34	-4.32	-4.43	-7.55
	lower bound	-5.89	-5.49	-5.22	-5.09	-4.90	-4.92	-4.96	-8.56
	upper bound	-4.24	-4.11	-3.96	-3.66	-3.87	-3.81	-3.98	-6.54
	[Elasticity]	-[2.51]	-[2.43]	-[2.32]	-[2.17]	-[2.17]	-[2.16]	-[2.22]	-[3.77]
E. Employee and Spouse									
N= 54,683	Year-end price	-5.53	-5.16	-4.81	-4.51	-4.42	-4.42	-4.53	-7.83
	lower bound	-6.20	-5.72	-5.38	-4.92	-4.81	-4.75	-4.88	-8.56
	upper bound	-4.89	-4.57	-4.38	-4.10	-4.05	-4.06	-4.19	-7.10
	[Elasticity]	-[2.76]	-[2.58]	-[2.40]	-[2.26]	-[2.21]	-[2.21]	-[2.26]	-[3.91]
F. Everyone									
N= 131,815	Year-end price	-4.72	-4.43	-4.24	-4.19	-4.12	-4.08	-4.14	-7.75
	lower bound	-5.12	-4.73	-4.61	-4.49	-4.44	-4.30	-4.39	-8.28
	upper bound	-4.35	-4.08	-3.89	-3.85	-3.89	-3.88	-3.92	-7.21
	[Elasticity]	-[2.36]	-[2.21]	-[2.12]	-[2.09]	-[2.06]	-[2.04]	-[2.07]	-[3.87]

Lower and upper bounds of 95% confidence interval from 200 bootstrap replications.
Lower and upper bounds for specifications B, C, E, and F account for intra-family correlations.
Controls include: employee dummy (when applicable), spouse dummy (when applicable), male dummy, plan (saturated), census region (saturated), salary dummy (vs. hourly), spouse on policy dummy, YOB of oldest dependent, YOB of youngest dependent, family size (saturated with 8-11 as one group), count family born 1944 to 1953, count family born 1954 to 1963, count family born 1974 to 1983, count family born 1984 to 1993, count family born 1994 to 1998, count family born 1999, count family born 2000, count family born 2001, count family born 2002, count family born 2003, count family born 2004 (when applicable).

In all of the estimated quantiles, the CQIV expenditure elasticities are an order of magnitude larger than those in the literature. For example, at the .85 quantile of the expenditure distribution, the implied expenditure elasticity is -2.3, which indicates that a one percent increase in price would decrease spending at the .85 quantile of the expenditure distribution by 2.3 percent. This elasticity estimate is fairly stable

References

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

Amemiya, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50(3), pp. 689–711.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74(2), 539–563.

Blundell, R., and J. L. Powell (2007): "Censored regression quantiles with endogenous regressors," *Journal of Econometrics*, 141(1), 65 – 83.

Chernozhukov, V. (2007a): "Lecture 8: An Introduction to Quantile Methods," MIT OpenCourseWare, course materials for 14.385.

——— (2007b): "Lecture 9: Quantile Methods 2," MIT OpenCourseWare, course materials for 14.385.

Chernozhukov, V., I. Fernndez-Val, and A. E. Kowalski (2011): "Quantile Regression with Censoring and Endogeneity," Working Paper 16997, National Bureau of Economic Research.

Chernozhukov, V., and C. Hansen (2004): "The Effects of 401(k) Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis," *The Review of Economics and Statistics*, 86(3), pp. 735–751.

——— (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491 – 525.

——— (2008): "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 142(1), 379 – 398.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Finite sample inference for quantile regression models," *Journal of Econometrics*, 152(2), 93 – 103, ¡ce:title¿Nonparametric and Robust Methods in Econometrics¡/ce:title¿.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293 – 346.

Cortez, P., A. Cerdeira, F. Almeida, T. Matos, and J. Reis (2009): "Modelling wine preferences by data mining from physiocochemical properties," *Decision Support Systems*, 47(4), 547–553.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64(5), pp. 1001–1044.

Honor, B., S. Khan, and J. L. Powell (2002): "Quantile regression under random censoring," *Journal of Econometrics*, 109(1), 67 – 105.

Imbens, G., and J. Wooldridge (2007): "Quantile Methods," Discussion paper, NBER Lecture Notes 14.

Juhn, C., K. M. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101(3), pp. 410–442.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

Kowalski, A. E. (2009): "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care," Working Paper 15085, National Bureau of Economic Research.

L., J., and Powell (1986): "Censored regression quantiles," *Journal of Econometrics*, 32(1), 143 – 155.

Lee, S. (2007): "Endogeneity in quantile regression models: A control function approach," *Journal of Econometrics*, 141, 1131–1158.

Machado, J. A. F., and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4), 445–465.

Melly, B. (2005): "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 12(4), 577 – 590.

Portnoy, S. (2003): "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98(464), 1001–1012.

Powell, J. L. (1983): "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51(5), pp. 1569–1575.