Quantile methods

Class Notes

Manuel Arellano

December 1, 2009

Revised: February 4, 2021

1 Unconditional quantiles

Let $F(r) = \Pr(Y \leq r)$. For $\tau \in (0,1)$, the τth population quantile of Y is defined to be

$$Q_\tau(Y) \equiv q_\tau \equiv F^{-1}(\tau) = \inf\{r : F(r) \geq \tau\}.$$

$F^{-1}(\tau)$ is a generalized inverse function. It is a left-continuous function with range equal to the support of F and hence often unbounded.

A simple example Suppose that Y is discrete with pmf $\Pr(Y = s) = 0.2$ for $s \in \{1,2,3,4,5\}$. For $\tau = 0.25, 0.5, 0.75$, we have

$$\begin{aligned}
\{r : F(r) \geq 0.25\} &= \{r : r \geq 2\} \Rightarrow q_{0.25} = F^{-1}(0.25) = 2 \\
\{r : F(r) \geq 0.50\} &= \{r : r \geq 3\} \Rightarrow q_{0.5} = F^{-1}(0.50) = 3 \\
\{r : F(r) \geq 0.75\} &= \{r : r \geq 4\} \Rightarrow q_{0.75} = F^{-1}(0.75) = 4.
\end{aligned}$$

Asymmetric absolute loss Let us define the "check" function (or asymmetric absolute loss function). For $\tau \in (0,1)$

$$\rho_\tau(u) = [\tau \mathbf{1}(u \geq 0) + (1-\tau)\mathbf{1}(u < 0)] \times |u| = [\tau - \mathbf{1}(u < 0)]\, u.$$

Note that $\rho_\tau(u)$ is a continuous piecewise linear function, but nondifferentiable at $u = 0$. We should think of u as an individual error $u = y - r$ and $\rho_\tau(u)$ as the loss associated with u.[1]

Using $\rho_\tau(u)$ as a specification of loss, it is well known that q_τ minimizes expected loss:

$$s_0(r) \equiv E[\rho_\tau(Y - r)] = \tau \int_r^\infty (y - r)\, dF(y) - (1-\tau) \int_{-\infty}^r (y - r)\, dF(y).$$

Any element of $\{r : F(r) = \tau\}$ minimizes expected loss. If the solution is unique, it coincides with q_τ as defined above. If not, we have an interval of τth quantiles and the smallest element is chosen so that the quantile function is left-continuous (by convention).

In decision theory the situation is as follows: we need a predictor or point estimate for a random variable with posterior cdf F. It turns out that the τth quantile is the optimal predictor that minimizes expected loss when loss is described by the τth check function.

[1] An alternative shorter notation is $\rho_\tau(u) = \tau u^+ + (1-\tau)u^-$ where $u^+ = \mathbf{1}(u \geq 0)|u|$ and $u^- = \mathbf{1}(u < 0)|u|$.

To illustrate the minimization property let us consider the median $M = q_{0.5}$. Take $r < M$, then

$$
\begin{aligned}
E\left|Y-r\right| &= \int_{-\infty}^{r}(r-y)\,dF+\int_{r}^{\infty}(y-r)\,dF \\
&= \left[\int_{-\infty}^{M}(r-y)\,dF-\int_{r}^{M}(r-y)\,dF\right]+\left[\int_{r}^{M}(y-r)\,dF+\int_{M}^{\infty}(y-r)\,dF\right] \\
&= \int_{-\infty}^{M}(r-y)\,dF+2\int_{r}^{M}(y-r)\,dF+\int_{M}^{\infty}(y-r)\,dF \\
&= \int_{-\infty}^{M}(M-M+r-y)\,dF+2\int_{r}^{M}(y-r)\,dF+\int_{M}^{\infty}(y-r-M+M)\,dF \\
&= \left[\int_{-\infty}^{M}(M-y)\,dF+\int_{M}^{\infty}(y-M)\,dF\right]+2\int_{r}^{M}(y-r)\,dF+(r-M)\int_{-\infty}^{M}dF+(M-r)\int_{M}^{\infty}dF \\
&= E\left|Y-M\right|+2\int_{r}^{M}(y-r)\,dF+(r-M)\frac{1}{2}+(M-r)\frac{1}{2} \\
&= E\left|Y-M\right|+2\int_{r}^{M}(y-r)\,dF\geq E\left|Y-M\right|,
\end{aligned}
$$

which is minimized at $r = M$. To complete the argument we proceed along similar lines for $r > M$.

Equivariance of quantiles under monotone transformations This is an interesting property of quantiles not shared by expectations. Let $g(.)$ be a nondecreasing function. Then, for any random variable Y

$$Q_{\tau}\left[g\left(Y\right)\right]=g\left[Q_{\tau}\left(Y\right)\right].$$

Thus, the quantiles of $g(Y)$ coincide with the transformed quantiles of Y. To see this in the case of a monotonic transformation note that

$$\Pr\left[Y\leq Q_{\tau}\left(Y\right)\right]=\tau\Rightarrow\Pr\left(g\left(Y\right)\leq g\left[Q_{\tau}\left(Y\right)\right]\right)=\tau.$$

Sample quantiles Given a random sample $\{Y_1,...,Y_N\}$ we obtain sample quantiles replacing F by the empirical cdf:

$$F_N\left(r\right)=\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(Y_i\leq r).$$

That is, we choose $\widehat{q}_{\tau}=F_N^{-1}\left(\tau\right)\equiv\inf\left\{r:F_N\left(r\right)\geq\tau\right\}$, which minimizes

$$s_N\left(r\right)=\int\rho_{\tau}\left(y-r\right)dF_N\left(y\right)=\frac{1}{N}\sum_{i=1}^{N}\rho_{\tau}\left(Y_i-r\right). \tag{1}$$

An important advantage of expressing the calculation of sample quantiles as an optimization problem, as opposed to a problem of ordering the observations, is computational (specially in the regression context). The optimization perspective is also useful for studying statistical properties.

Linear program representation An alternative presentation of the minimization of (1) is

$$\min_{r,u_i^+,u_i^-} \sum_{i=1}^{N} \left[\tau u_i^+ + (1-\tau)\, u_i^-\right]$$

subject to[2]

$$Y_i - r = u_i^+ - u_i^-, \quad u_i^+ \geq 0, u_i^- \geq 0, \quad (i = 1, ..., N)$$

where here $\left\{u_i^+, u_i^-\right\}_{i=1}^{2N}$ denote $2N$ artificial additional arguments, which allow us to represent the original problem in the form of a linear program. A linear program takes the form:[3]

$$\min_x c'x \text{ subject to } Ax \geq b, x \geq 0.$$

The simplex algorithm for numerical solution of this problem was created by George Dantzig in 1947.

Nonsmoothness in sample but smoothness in population The sample objective function $s_N(r)$ is continuous but not differentiable for all r. Moreover, the gradient or moment condition

$$b_N(r) = \frac{1}{N} \sum_{i=1}^{N} \left[\mathbf{1}(Y_i \leq r) - \tau\right]$$

is not continuous in r. Note that if each Y_i is distinct, so that we can reorder the observations to satisfy $Y_1 < Y_2 < ... < Y_N$, for all τ we have

$$|b_N(\widehat{q}_\tau)| \equiv |F_N(\widehat{q}_\tau) - \tau| \leq \frac{1}{N}.$$

Despite lack of smoothness in $s_N(r)$ or $b_N(r)$, smoothness of the distribution of the data can smooth their population counterparts. Suppose that F is differentiable at q_τ with positive derivative $f(q_\tau)$, then $s_0(r)$ is twice continuously differentiable with derivatives:[4]

$$\frac{d}{dr} E\left[\rho_\tau(Y - r)\right] = -\tau\left[1 - F(r)\right] + (1-\tau) F(r) = F(r) - \tau \equiv E\left[\mathbf{1}(Y \leq r) - \tau\right]$$
$$\frac{d^2}{dr^2} E\left[\rho_\tau(Y - r)\right] = f(r).$$

---

[2] Note that

$$u^+ - u^- = \mathbf{1}(u \geq 0)|u| - \mathbf{1}(u < 0)|u| = \mathbf{1}(u \geq 0)\, u + \mathbf{1}(u < 0)\, u = u.$$

[3] See Koenker, 2005, section 6.1, for an introduction oriented to quantiles.

[4] The required derivatives are:

$$\frac{d}{dr} \int_{-\infty}^{r} (y - r) f(y)\, dy = \frac{d}{dr} \int_{-\infty}^{r} y f(y)\, dy - \frac{d}{dr}\left[rF(r)\right] = rf(r) - \left[F(r) + rf(r)\right] = -F(r)$$

and

$$\begin{aligned} \frac{d}{dr} \int_{r}^{\infty} (y - r) f(y)\, dy &= \frac{d}{dr} \int_{r}^{\infty} y f(y)\, dy - \frac{d}{dr}\left\{r\left[1 - F(r)\right]\right\} = \frac{d}{dr}\left[\int_{-\infty}^{\infty} y f(y)\, dy - \int_{-\infty}^{r} y f(y)\, dy\right] - \frac{d}{dr}\left\{r\left[1 - F(r)\right]\right\} \\ &= -rf(r) - \left\{\left[1 - F(r)\right] - rf(r)\right\} = -\left[1 - F(r)\right]. \end{aligned}$$

Consistency Consistency of sample quantiles follows from the theorem by Newey and McFadden (1994) that we discussed in a previous class note. This theorem relies on continuity of the limiting objective function and uniform convergence. The quantile sample objective function $s_N(r)$ is continuous and convex in r. Suppose that F is such that $s_0(r)$ is uniquely maximized at q_τ. By the law of large numbers $s_N(r)$ converges pointwise to $s_0(r)$. Then use the fact that pointwise convergence of convex functions implies uniform convergence on compact sets.[5]

Asymptotic normality The asymptotic normality of sample quantiles cannot be established in the standard way because of the nondifferentiability of the objective function. However, it has long been known that under suitable conditions sample quantiles are asymptotically normal and there are direct approaches to establish the result.[6] Here we just re-state the asymptotic normality result for unconditional quantiles following the discussion in the class note on nonsmooth GMM around Newey and McFadden's theorems. The general idea is that as long as the limiting objective function is differentiable the familiar approach for differentiable problems is possible if a stochastic equicontinuity assumption holds.

Fix $0 < \tau < 1$. If F is differentiable at q_τ with positive derivative $f(q_\tau)$, then

$$\sqrt{N}\left(\widehat{q}_\tau - q_\tau\right) = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{1}\left(Y_i \leq q_\tau\right) - \tau}{f\left(q_\tau\right)} + o_p(1).$$

Consequently,

$$\sqrt{N}\left(\widehat{q}_\tau - q_\tau\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{\left[f\left(q_\tau\right)\right]^2}\right).$$

The term $\tau(1-\tau)$ in the numerator of the asymptotic variance tends to make $\widehat{q}_\tau$ more precise in the tails, whereas the density term in the denominator tends to make $\widehat{q}_\tau$ less precise in regions of low density. Typically the latter effect will dominate so that quantiles closer to the extremes will be estimated with less precision.

Computing standard errors The asymptotic normality result justifies the large N approximation

$$\frac{\widehat{f}\left(\widehat{q}_\tau\right)}{\sqrt{\tau(1-\tau)}}\sqrt{N}\left(\widehat{q}_\tau - q_\tau\right) \approx \mathcal{N}(0,1)$$

where $\widehat{f}\left(\widehat{q}_\tau\right)$ is a consistent estimator of $f\left(q_\tau\right)$.[7] Since

$$f(r) = \lim_{h\to 0}\frac{F(r+h) - F(r-h)}{2h} \equiv \lim_{h\to 0}\frac{1}{2h}E\left[\mathbf{1}(|Y-r| \leq h)\right],$$

[5] See Amemiya (1985, p. 150) for a proof of consistency of the median, and Koenker (2005, p. 117–119) for conditional and unconditional quantiles.

[6] See for example the proofs in Cox and Hinkley (1974, p. 468) and Amemiya (1985, p. 148–150).

[7] Alternatively we can use the density $f_{U_\tau}(r)$ of the error $U_\tau = Y - q_\tau$ noting that $f(q_\tau) = f_{U_\tau}(0)$.

an obvious possibility is to use the histogram estimator

$$\begin{aligned}\widehat{f}(r) &= \frac{F_N(r+h_N)-F_N(r-h_N)}{2h_N} = \frac{1}{2Nh_N}\sum_{i=1}^{N}\left[\mathbf{1}(Y_i \leq r+h_N) - \mathbf{1}(Y_i \leq r-h_N)\right] \\ &= \frac{1}{2Nh_N}\sum_{i=1}^{N}\mathbf{1}(|Y_i - r| \leq h_N)\end{aligned}$$

for some $h_N > 0$ sequence such that $h_N \to 0$ as $N \to \infty$.[8] Thus,

$$\widehat{f}(\widehat{q}_\tau) = \frac{1}{2Nh_N}\sum_{i=1}^{N}\mathbf{1}(|Y_i - \widehat{q}_\tau| \leq h_N).$$

Other alternatives are kernel estimators for $f(q_\tau)$, the bootstrap, or directly obtain an approximate confidence interval using the normal approximation to the binomial distribution (Chamberlain, 1994; Koenker, 2005, p. 73).

2 Conditional quantiles

Consider the conditional distribution of Y given X:

$$\Pr(Y \leq r \mid X) = F(r; X)$$

and denote the τth quantile of Y given X as

$$Q_\tau(Y \mid X) \equiv q_\tau(X) \equiv F^{-1}(\tau; X).$$

Now quantiles minimize expected asymmetric absolute loss in a conditional sense:

$$q_\tau(X) = \arg\min_c E\left[\rho_\tau(Y-c) \mid X\right].$$

Suppose that $q_\tau(X)$ satisfies a parametric model $q_\tau(X) = g(X, \beta_\tau)$, then

$$\beta_\tau = \arg\min_b E\left[\rho_\tau(Y - g(X, b))\right].$$

Also, since in general

$$\Pr(Y \leq q_\tau(X) \mid X) = \tau \quad \text{or} \quad E\left[1(Y \leq q_\tau(X)) - \tau \mid X\right] = 0,$$

it turns out that β_τ solves moment conditions of the form

$$E\left\{h(X)\left[1(Y \leq g(X, \beta_\tau)) - \tau\right]\right\} = 0.$$

[8] A sufficient condition for consistency is $\sqrt{N}h_N \to \infty$. One possibility is $h_N = aN^{-1/3}$ for some $a > 0$.

Conditional quantiles in a location-scale model The standardized variable in a location-scale model of $Y \mid X$ has a distribution that is independent of X. Namely, letting $E(Y \mid X) = \mu(X)$ and $Var(Y \mid X) = \sigma^2(X)$, the variable

$$V = \frac{Y - \mu(X)}{\sigma(X)}$$

is distributed independently of X according to some cdf G. Thus, in a location scale model all dependence of Y on X occurs through mean translations and variance re-scaling.

An example is the classical normal regression model:

$$Y \mid X \sim \mathcal{N}\left(X'\beta, \sigma^2\right).$$

In the location-scale model:

$$\Pr(Y \leq r \mid X) = \Pr\left(\frac{Y - \mu(X)}{\sigma(X)} \leq \frac{r - \mu(X)}{\sigma(X)} \mid X\right) = G\left(\frac{r - \mu(X)}{\sigma(X)}\right)$$

and

$$G\left(\frac{Q_\tau(Y \mid X) - \mu(X)}{\sigma(X)}\right) = \tau$$

or

$$Q_\tau(Y \mid X) = \mu(X) + \sigma(X) G^{-1}(\tau)$$

so that

$$\frac{\partial Q_\tau(Y \mid X)}{\partial X_j} = \frac{\partial \mu(X)}{\partial X_j} + \frac{\partial \sigma(X)}{\partial X_j} G^{-1}(\tau).$$

Under homoskedasticity, $\partial Q_\tau(Y \mid X) / \partial X_j$ is the same at all quantiles since they only differ by a constant term. More generally, in a location-scale model the relative change between two quantiles $\partial \ln\left[Q_{\tau_1}(Y \mid X) - Q_{\tau_2}(Y \mid X)\right] / \partial X_j$ is the same for any pair (τ_1, τ_2). These assumptions have been found to be too restrictive in studies of the distribution of individual earnings conditioned on education and labor market experience.

In the classical normal regression model

$$Q_\tau(Y \mid X) = X'\beta + \sigma \Phi^{-1}(\tau).$$

3 Quantile regression

A linear regression is an optimal linear predictor that minimizes average quadratic loss. Given data $\{Y_i, X_i\}_{i=1}^N$ OLS sample coefficients are given by

$$\widehat{\beta}_{OLS} = \arg\min_b \sum_{i=1}^N \left(Y_i - X_i'b\right)^2.$$

If $E(Y \mid X)$ is linear it coincides with the least squares population predictor, so that $\widehat{\beta}_{OLS}$ consistently estimates $\partial E(Y \mid X)/\partial X$.

As is well known the median may be preferable to the mean if the distribution is long-tailed. The median lacks the sensitivity to extreme values of the mean and may represent the position of an asymmetric distribution better than the mean. For similar reasons in the regression context one may be interested in median regression. That is, an optimal predictor that minimizes average absolute loss:

$$\widehat{\beta}_{LAD} = \arg\min_{b} \sum_{i=1}^{N} \left| Y_i - X_i' b \right|.$$

If $med(Y \mid X)$ is linear it coincides with the least absolute deviation (LAD) population predictor, so that $\widehat{\beta}_{LAD}$ consistently estimates $\partial med(Y \mid X)/\partial X$.

The idea can be generalized to quantiles other than $\tau = 0.5$ by considering optimal predictors that minimize average asymmetric absolute loss:

$$\widehat{\beta}(\tau) = \arg\min_{b} \sum_{i=1}^{N} \rho_\tau \left(Y_i - X_i' b \right).$$

As before if $Q_\tau(Y \mid X)$ is linear $\widehat{\beta}(\tau)$ consistently estimates $\partial Q_\tau(Y \mid X)/\partial X$. Clearly, $\widehat{\beta}_{LAD} = \widehat{\beta}(0.5)$.

Structural representation Define U such that

$$F(Y; X) = U.$$

It turns out that U is uniformly distributed independently of X between 0 and 1.[9] Also

$$Y = F^{-1}(U; X) \text{ with } U \mid X \sim \mathcal{U}(0,1).$$

This is sometimes called the Skorohod representation. For example, the Skorohod representation of the Gaussian linear regression model is $Y = X'\beta + \sigma V$ with $V = \Phi^{-1}(U)$, so that $V \mid X \sim \mathcal{N}(0,1)$.

Linear quantile model A semiparametric alternative to the normal linear regression model is the linear quantile regression

$$Y = X'\beta(U) \qquad U \mid X \sim \mathcal{U}(0,1).$$

where $\beta(u)$ is a nonparametric function, such that $x'\beta(u)$ is strictly increasing in u for each value of x in the support of X. Thus, it is a semiparametric one-factor random coefficients model.

This model nests linear regression as a special case and allows for interactions between observable and unobservable determinants of Y. Partitioning $\beta(U)$ into intercept and slope components $\beta(U) =$

[9] Note that if $\Pr(Y \leq r \mid X) = F(r; X)$ then $\Pr(F(Y; X) \leq F(r; X) \mid X) = F(r; X)$ or $\Pr(U \leq s \mid X) = s$.

$\left[\beta_0(U), \beta_1(U)'\right]'$, the normal linear regression arises as a particular case of the linear quantile model with $\beta_1(U) = \beta_1$ and $\beta_0(U) = \beta_0 + \sigma\Phi^{-1}(U)$.

The error U can be understood as the rank of a particular unit in the population. For example of ability, in a situation where Y denotes log earnings and X contains education and labor market experience.

The practical usefulness of this model is that for given $\tau \in (0,1)$ estimation of $\beta(\tau)$ can be easily obtained as the τ-th quantile linear regression coefficient, since $Q_\tau(Y \mid X) = X'\beta(\tau)$.

4 Asymptotic inference for quantile regression

Using our earlier results, the first and second derivatives of the limiting objective function can be obtained as

$$\begin{aligned}
\frac{\partial}{\partial b} E\left[\rho_\tau\left(Y - X'b\right)\right] &= E\left\{X\left[1\left(Y \leq X'b\right) - \tau\right]\right\} \\
\frac{\partial^2}{\partial b \partial b'} E\left[\rho_\tau\left(Y - X'b\right)\right] &= E\left[f\left(X'b \mid X\right) X X'\right] = H(b)
\end{aligned}$$

Moreover, under some regularity conditions we can use Newey and McFadden's asymptotic normality theorem, leading to

$$\sqrt{N}\left[\widehat{\beta}(\tau) - \beta(\tau)\right] = -H_0{}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \left\{1\left[Y_i \leq X_i'\beta(\tau)\right] - \tau\right\} + o_p(1).$$

where $H_0 = H(\beta(\tau))$ is the Hessian of the limit objective function at the truth, and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \left\{1\left[Y_i \leq X_i'\beta(\tau)\right] - \tau\right\} \xrightarrow{d} \mathcal{N}(0, V_0)$$

where

$$V_0 = E\left(\left\{1\left[Y_i \leq X_i'\beta(\tau)\right] - \tau\right\}^2 X_i X_i'\right) = \tau(1-\tau) E\left(X_i X_i'\right).$$

Note that the last equality follows under the assumption of linearity of conditional quantiles.

Thus,

$$\sqrt{N}\left[\widehat{\beta}(\tau) - \beta(\tau)\right] \xrightarrow{d} \mathcal{N}(0, W_0)$$

where

$$W_0 = H_0{}^{-1} V_0 H_0{}^{-1}.$$

To get a consistent estimate of W_0 we need consistent estimates of H_0 and V_0. A simple estimator of H_0 suggested in Powell (1984, 1986), which mimics the histogram estimator discussed above, is as follows:

$$\widehat{H} = \frac{1}{2Nh_N} \sum_{i=1}^{N} \mathbf{1}\left(\left| Y_i - X_i' \widehat{\beta}(\tau) \right| \leq h_N \right) X_i X_i'.$$

This estimator is motivated by the following iterated expectations argument:

$$\begin{aligned} H_0 &= E\left[f\left(X'\beta(\tau) \mid X \right) XX' \right] \equiv \lim_{h \to 0} \frac{1}{2h} E\left\{ E\left[\mathbf{1}(\left| Y - X'\beta(\tau) \right| \leq h) \mid X \right] XX' \right\} \\ &= \lim_{h \to 0} \frac{1}{2h} E\left[\mathbf{1}(\left| Y - X'\beta(\tau) \right| \leq h) XX' \right]. \end{aligned}$$

If the quantile function is correctly specified a consistent estimate of V_0 is

$$\widehat{V} = \tau(1-\tau) \frac{1}{N} \sum_{i=1}^{N} X_i X_i'.$$

Otherwise, a fully robust estimator can be obtained using

$$\widetilde{V} = \frac{1}{N} \sum_{i=1}^{N} \left\{ 1\left[Y_i \leq X_i' \widehat{\beta}(\tau) \right] - \tau \right\}^2 X_i X_i'$$

Finally, if $U_\tau = Y - X'\beta(\tau)$ is independent of X (as in the location model) it turns out that

$$H_0 = f_{U_\tau}(0) E\left(X_i X_i' \right)$$

so that

$$W_0 = \frac{\tau(1-\tau)}{\left[f_{U_\tau}(0) \right]^2} \left[E\left(X_i X_i' \right) \right]^{-1},$$

which can be consistently estimated as

$$\widehat{W}_{NR} = \frac{\tau(1-\tau)}{\left[\widehat{f}_{U_\tau}(0) \right]^2} \left(\frac{1}{N} \sum_{i=1}^{N} X_i X_i' \right)^{-1}$$

where

$$\widehat{f}_{U_\tau}(0) = \frac{1}{2Nh_N} \sum_{i=1}^{N} \mathbf{1}(\left| Y_i - X_i' \widehat{\beta}(\tau) \right| \leq h_N).$$

In summary, we have considered three different alternative estimators for standard errors: A non-robust variance matrix estimator under independence $\widehat{W}_{NR}$, a robust estimator under correct specification: $\widehat{W}_R = \widehat{H}^{-1} \widehat{V} \widehat{H}^{-1}$, and a fully robust estimator under misspecification: $\widehat{W}_{FR} = \widehat{H}^{-1} \widetilde{V} \widehat{H}^{-1}$.

5 Further topics

5.1 Flexible QR

- Linearity is restrictive. It may also be at odds with the monotonicity requirement of $q(x,u)$ in u for every value of x.

- Linear QR may be interpreted as an approximation to the true quantile function.

- An approach to nonparametric QR is to use series methods:

$$q(x,u) = \theta_0(u) + \theta_1(u) g_1(x) + ... + \theta_P(u) g_P(x).$$

- The g's are anonymous functions without an economic interpretation. Objects of interest are derivative effects and summary measures of them.

- In practice one may use orthogonal polynomials, wavelets or splines.

- This type of specification may be seen as an approximating model that becomes more accurate as P increases, or simply as a parametric flexible model of the quantile function.

- From the point of view of computation the model is still a linear QR, but the regressors are now functions of X instead of the Xs themselves.

5.2 Decompositions

- Basic idea of decomposition:

$$F_M(y) - F_F(y) = \int F_M(y \mid x) f_M(x) dx - \int F_F(y \mid x) f_F(x) dx$$

$$= \int [F_M(y \mid x) - F_F(y \mid x)] f_M(x) dx + \int F_M(y \mid x) [f_M(x) - f_F(x)] dx.$$

- We have added and subtracted the counterfactual cdf:

$$F_F^C(y) = \int F_F(y \mid x) f_M(x) dx.$$

- The indices (M,F) could be male/female gender gaps, a pair of countries or two different periods.

- The decomposition can be done for *cdf*s (as shown) or for other distributional characteristics such as quantiles or moments.

- When done for differences in means using linear regression they are called Oaxaca decompositions (after the work of Ronald Oaxaca):

$$\overline{y}_M - \overline{y}_F = \overline{x}'_M \beta_M - \overline{x}'_F \beta_F = \overline{x}'_F (\beta_M - \beta_F) + (\overline{x}_M - \overline{x}_F)' \beta_M.$$

Machado-Mata method To do decompositions for quantiles or for distributions based on QR models, we need to be able to calculate the marginal distribution of the outcome implied by the QR.

To do so we can use the following simulation method (Machado and Mata 2005):

1) Generate $u_1, ..., u_m \sim iid\, \mathcal{U}(0,1)$.

2) Get $\widehat{\beta}(u_1), ..., \widehat{\beta}(u_m)$ by running QRs from the actual data $\{y_i, x_i\}_{i=1}^n$.

3) Get a random sample of size m of the covariates: $x_1^*, ..., x_m^*$.

4) Compute $y_j^* = x_j^* \widehat{\beta}(u_j)$ for $j = 1, 2, ..., m$.

5) The sample quantiles of $y_1^*, ..., y_m^*$ are consistent estimates of the marginal quantiles of y_i for large n and m.

5.3 Other topics

- Crossings and rearrangements.
- Functional inference.
- Quantile regression under misspecification (Angrist et al, 2006).
- Asymptotic efficiency: GLS and optimal instrument arguments.
- Instrumental variable models
  - Chernozhukov and Hansen (2006) estimators.
  - Chesher (2003) and Ma and Koenker (2006).
  - Treatment effect perspectives.
- Censored regression quantiles
  - Powell's estimators.
  - Chamberlain's minimum distance approach.
  - Honoré (1992) panel data approaches.
- Quantile selection models (Arellano and Bonhomme 2017).

References

[1] Amemiya, T. (1985): *Advanced Econometrics*, Blackwell.

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

[3] Arellano, M. and S. Bonhomme (2017): "Quantile Selection Models with an Application to Understanding Changes in Wage Inequality", *Econometrica*, 85, 1–28.

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

[5] Chernozhukov, V. and C. Hansen (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models", *Journal of Econometrics*, 132, 491–525.

[6] Chernozhukov, V., I. Fernández-Val and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

[7] Chesher, A. (2003): "Identification in Nonseparable Models", *Econometrica*, 71, 1405–1441.

[8] Cox, D. R. and D. V. Hinkley (1974): *Theoretical Statistics*, Chapman and Hall, London.

[9] Honoré, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects", *Econometrica*, 60, 533–565.

[10] Koenker, R. and G. Basset (1978): "Regression Quantiles", *Econometrica*, 46, 33–50.

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

[12] Koenker, R. (2017): "Quantile Regression: 40 Years On" *Annual Review of Economics* 9, 155–176.

[13] Ma, L. and R. Koenker (2006): "Quantile Regression Methods for Recursive Structural Equation Models", *Journal of Econometrics*, 134, 471–506.

[14] Machado, J. and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression", *Journal of Applied Econometrics*, 20, 445–465.

[15] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier.

[16] Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model", *Journal of Econometrics*, 25, 303-25.

[17] Powell, J. L. (1986): "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.