New Developments in Econometrics
Lecture 16: Quantile Estimation

Jeff Wooldridge

Cemmap Lectures, UCL, June 2009

1. Review of Means, Medians, and Quantiles
2. Some Useful Asymptotic Results
3. Quantile Regression with Endogenous Explanatory Variables
4. Quantile Regression for Panel Data
5. Quantile Methods for “Censored” Data

1. Review of Means, Medians, and Quantiles

- Linear Population Model, where $\boldsymbol{\beta}$ is $K \times 1$:

$$y = \alpha + \mathbf{x}\boldsymbol{\beta} + u. \tag{1}$$

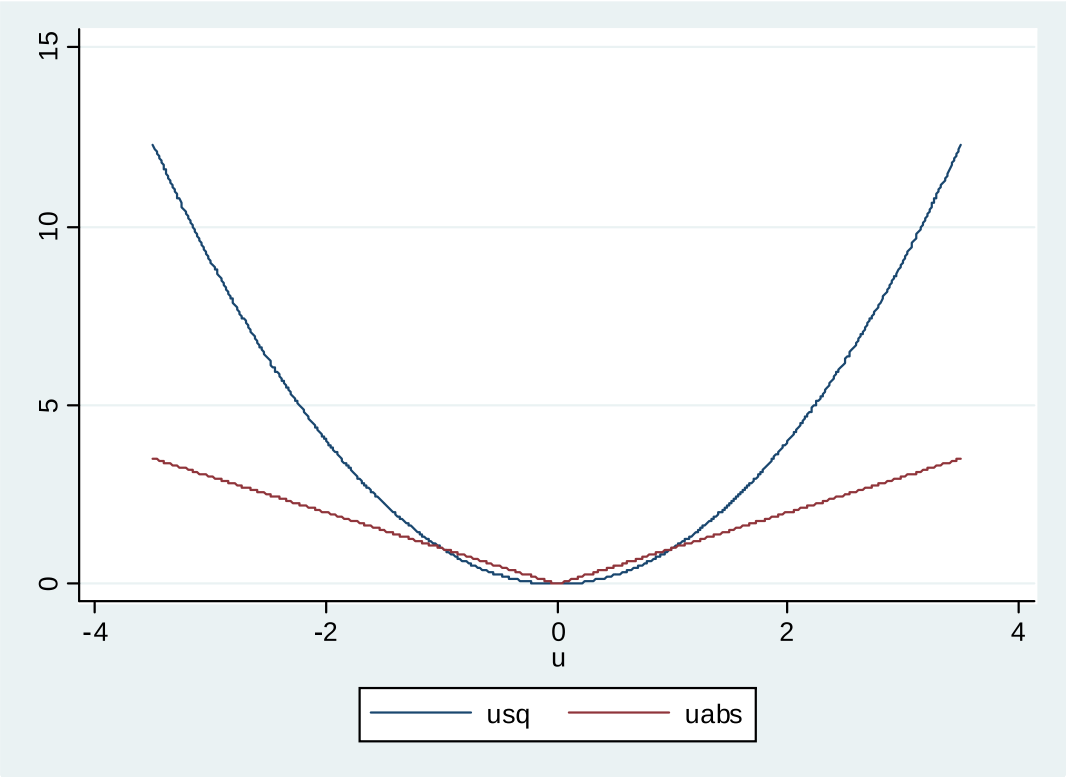
Assume $E(u^2) < \infty$, so that the distribution of u is not too spread out.

Ordinary Least Squares (OLS):

$$\min_{a,\mathbf{b}} \sum_{i=1}^{N} (y_i - a - \mathbf{x}_i\mathbf{b})^2. \tag{2}$$

Least Absolute Deviations (LAD):

$$\min_{a,\mathbf{b}} \sum_{i=1}^{N} |y_i - a - \mathbf{x}_i\mathbf{b}|. \tag{3}$$



Squared and Absolute Residual Functions

• With a large random sample, when should we expect the slope estimates to be similar? Two important cases. (i) If

$$D(u|\mathbf{x}) \text{ is symmetric about zero} \tag{4}$$

then OLS and LAD both consistently estimate α and $\boldsymbol{\beta}$. (ii) If

$$u \text{ is independent of } \mathbf{x} \text{ with } E(u) = 0, \tag{5}$$

where $E(u) = 0$ is the normalization that identifies α, then OLS and LAD both consistently estimate the slopes, $\boldsymbol{\beta}$. If u has an asymmetric distribution, then $Med(u) \equiv \eta \neq 0$, and $\hat{\alpha}_{LAD}$ converges to $\alpha + \eta$ because $Med(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + Med(u|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + \eta$.

- In many applications, neither (4) nor (5) is likely to be true. For example, y may be a measure of wealth, in which case the error distribution is probably asymmetric and $Var(u|\mathbf{x})$ not constant.
- It is important to remember that if $D(u|\mathbf{x})$ is asymmetric and changes with $\mathbf{x}$, then we should not expect OLS and LAD to deliver similar estimates of $\boldsymbol{\beta}$, even for "thin-tailed" distributions.
- Of course, LAD is much more resilient to changes in extreme values because, as a measure of central tendency, the median is much less sensitive than the mean to changes in extreme values. But it does not follow that a large difference in OLS and LAD estimates means something is "wrong" with OLS.

• Advantage for median over mean: median passes through monotonic functions. If $\log(y) = \alpha + \mathbf{x}\boldsymbol{\beta} + u$ and $Med(u|\mathbf{x}) = 0$, then $Med(y|\mathbf{x}) = \exp(Med[\log(y)|\mathbf{x}]) = \exp(\alpha + \mathbf{x}\boldsymbol{\beta})$. By contrast, we cannot generally find $E(y|\mathbf{x}) = \exp(\alpha + \mathbf{x}\boldsymbol{\beta})E[\exp(u)|\mathbf{x}]$.

• But expectation has useful properties that the median does not: linearity and the law of iterated expectations. If

$$y_i = a_i + \mathbf{x}_i\mathbf{b}_i \tag{6}$$

and $(a_i, \mathbf{b}_i)$ is independent of $\mathbf{x}_i$, then

$$E(y_i|\mathbf{x}_i) = E(a_i|\mathbf{x}_i) + \mathbf{x}_i E(\mathbf{b}_i|\mathbf{x}_i) \equiv \alpha + \mathbf{x}_i\boldsymbol{\beta}, \tag{7}$$

where $\alpha = E(a_i)$ and $\boldsymbol{\beta} = E(\mathbf{b}_i)$. OLS is consistent for α and $\boldsymbol{\beta}$.

• What can we add so that LAD estimates something of interest in (7)? If $\mathbf{u}_i$ is a vector, then its distribution conditional on $\mathbf{x}_i$ is *centrally symmetric* if $D(\mathbf{u}_i|\mathbf{x}_i) = D(-\mathbf{u}_i|\mathbf{x}_i)$, which implies that, if $\mathbf{g}_i$ is any vector function of $\mathbf{x}_i$, $D(\mathbf{g}_i'\mathbf{u}_i|\mathbf{x}_i)$ has a univariate distribution that is symmetric about zero. This implies $E(\mathbf{u}_i|\mathbf{x}_i) = \mathbf{0}$.

• Write

$$y_i = \alpha + \mathbf{x}_i\boldsymbol{\beta} + (a_i - \alpha) + \mathbf{x}_i(\mathbf{b}_i - \boldsymbol{\beta}). \tag{8}$$

If $\mathbf{c}_i = (a_i, \mathbf{b}_i)$ given $\mathbf{x}_i$ is centrally symmetric then LAD applied to the usual model $y_i = \alpha + \mathbf{x}_i\boldsymbol{\beta} + u_i$ consistently estimates α and $\boldsymbol{\beta}$.

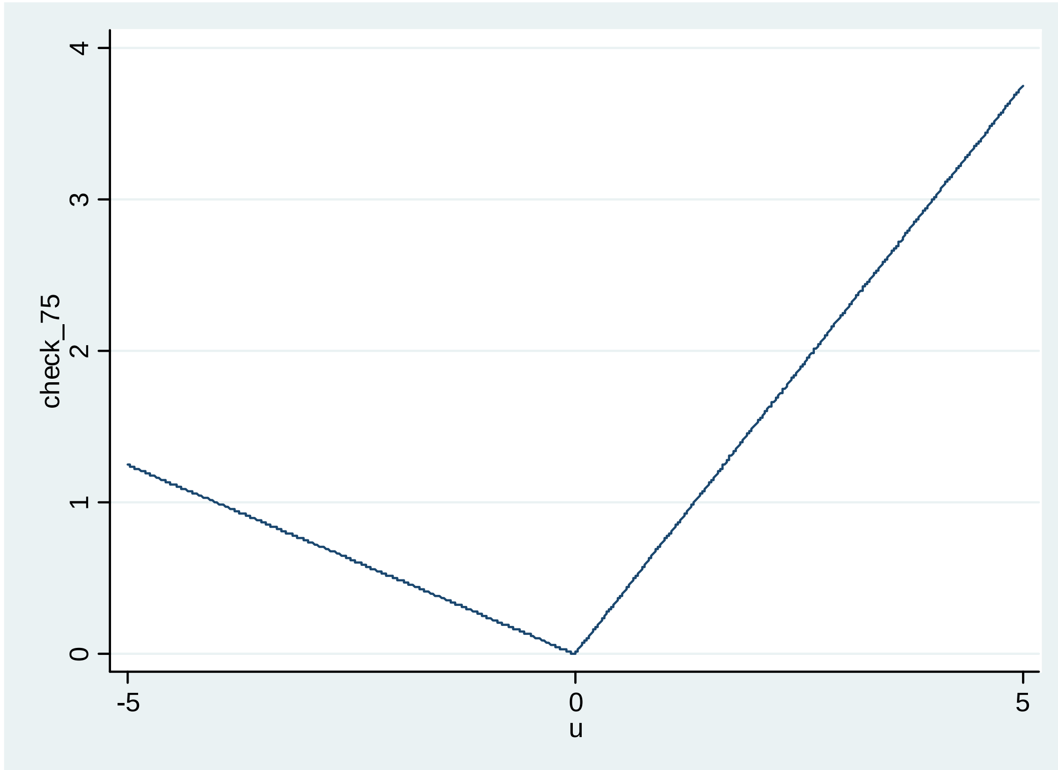
- For $0 < \tau < 1$, $q(\tau)$ is the τ^{th} quantile of y_i if $P(y_i \leq q(\tau)) \geq \tau$ and $P(y_i \geq q(\tau)) \geq 1 - \tau$.
- Let covariates affect quantiles. Under linearity,

$$Quant_\tau(y_i|\mathbf{x}_i) = \alpha(\tau) + \mathbf{x}_i\boldsymbol{\beta}(\tau). \tag{9}$$

Under (9), consistent estimators of $\alpha(\tau)$ and $\boldsymbol{\beta}(\tau)$ are obtained by minimizing the “check” function:

$$\min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^K} \sum_{i=1}^{N} c_\tau(y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}), \tag{10}$$

where $c_\tau(u) = (\tau 1[u \geq 0] + (1-\tau)1[u < 0])|u| = (\tau - 1[u < 0])u$ and $1[\bullet]$ is the “indicator function.”



The Check Function for $\tau = .75$.

2. Some Useful Asymptotic Results

What Happens if the Quantile Function is Misspecified?

- Recall property of OLS: if α^* and $\boldsymbol{\beta}^*$ are the plims from the OLS regression y_i on $1, \mathbf{x}_i$ then these provide the smallest mean squared error approximation to $E(y|\mathbf{x}) = \mu(\mathbf{x})$ in that $(\alpha^*, \boldsymbol{\beta}^*)$ solve

$$\min_{\alpha,\boldsymbol{\beta}} E\{[\mu(\mathbf{x}) - \alpha - \mathbf{x}\boldsymbol{\beta}]^2\}. \tag{11}$$

Under restrictive assumptions on distribution of $\mathbf{x}$, β_j^* can be equal to or proportional to average partial effects.

• Linear quantile formulation has been viewed by several authors as an approximation. Recently, Angrist, Chernozhukov, and Fernandez-Val (2006) characterized the probability limit of the quantile regression estimator. Absorb the intercept into $\mathbf{x}$ and let $\boldsymbol{\beta}(\tau)$ be the solution to the population quantile regression problem. ACF show that $\boldsymbol{\beta}(\tau)$ solves

$$\min_{\boldsymbol{\beta}} E\{w_\tau(\mathbf{x},\boldsymbol{\beta})[q_\tau(\mathbf{x}) - \mathbf{x}\boldsymbol{\beta})]^2\}, \tag{12}$$

where the weight function $w_\tau(\mathbf{x},\boldsymbol{\beta})$ is

$$w_\tau(\mathbf{x},\boldsymbol{\beta}) = \int_0^1 (1-u) f_{y|x}(u\mathbf{x}\boldsymbol{\beta} + (1-u)q_\tau(\mathbf{x})|\mathbf{x})du. \tag{13}$$

Computing Standard Errors

- For given τ, write

$$y_i = \mathbf{x}_i\boldsymbol{\theta} + u_i,\ Quant_\tau(u_i|\mathbf{x}_i) = 0, \tag{14}$$

and let $\hat{\boldsymbol{\theta}}$ be the quantile estimator. Define quantile residuals $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\theta}}$. Generally, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal with asymptotic variance $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, where

$$\mathbf{A} \equiv E[f_u(0|\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i] \tag{15}$$

and

$$\mathbf{B} \equiv \tau(1-\tau)E(\mathbf{x}_i'\mathbf{x}_i). \tag{16}$$

- If the quantile function is actually linear, a consistent estimator of **B** is

$$\hat{\mathbf{B}} = \tau(1-\tau)\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right). \tag{17}$$

Generally, a consistent estimator of **A** is (Powell (1991))

$$\hat{\mathbf{A}} = (2Nh_N)^{-1}\sum_{i=1}^{N}1[|\hat{u}_i| \leq h_N]\mathbf{x}_i'\mathbf{x}_i, \tag{18}$$

where $\{h_N > 0\}$ is a nonrandom sequence shrinking to zero as $N \rightarrow \infty$ with $\sqrt{N}h_N \rightarrow \infty$. For example, $h_N = aN^{-1/3}$ for any $a > 0$. Might use a smoothed version so that all residuals contribute.

- If u_i and $\mathbf{x}_i$ are independent,

$$Avar\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{\tau(1-\tau)}{[f_u(0)]^2}[E(\mathbf{x}_i'\mathbf{x}_i)]^{-1}, \tag{19}$$

and $Avar(\hat{\boldsymbol{\theta}})$ is estimated as

$$\widehat{Avar(\hat{\boldsymbol{\theta}})} = \frac{\tau(1-\tau)}{[\hat{f}_u(0)]^2}\left(\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}, \tag{20}$$

where, say, $\hat{f}_u(0)$ is the histogram estimator

$$\hat{f}_u(0) = (2Nh_N)^{-1}\sum_{i=1}^{N} 1[|\hat{u}_i| \leq h_N]. \tag{21}$$

Estimate in (20) is commonly reported (by, say, Stata).

• If the quantile function is misspecified, the "robust" form based on (20), is not valid. In the generalized linear models literature, distinction between "fully robust" variance estimator (mean correctly specified) and a "semi-robust" estimator (mean might be misspecified).

• For quantile regression, a fully robust variance requires a different estimator of **B**. Kim and White (2002) and Angrist, Chernozhukov, and Fernández-Val (2006) show

$$\hat{\mathbf{B}} = \left(N^{-1} \sum_{i=1}^{N} (\tau - 1[\hat{u}_i < 0])^2 \mathbf{x}_i' \mathbf{x}_i \right) \tag{22}$$

is consistent, and then $\widehat{Avar(\hat{\boldsymbol{\theta}})} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}$ with $\hat{\mathbf{A}}$ given by (18).

- Hahn (1995, 1997) shows that the nonparametric bootstrap and the Bayesian bootstrap generally provide consistent estimates of the fully robust variance without claims about the conditional quantile being correct. Bootstrap does not provide "asymptotic refinements" for testing and confidence intervals.
- ACF provide the covariance function for the process $\{\hat{\theta}(\tau) : \varepsilon \leq \tau \leq 1 - \varepsilon\}$ for some $\varepsilon > 0$, which can be used to test hypotheses jointly across multiple quantiles (including all quantiles at once).
- Example using Abadie (2003). These are nonrobust standard errors. *nettfa* is net total financial assets.

Dependent Variable:	*nettfa*			
Explanatory Variable	Mean (OLS)	.25 Quantile	Median (LAD)	.75 Quantile
inc	.783	.0713	.324	.798
	(.104)	(.0072)	(.012)	(.025)
age	−1.568	.0336	−.244	−1.386
	(1.076)	(.0955)	(.146)	(.287)
age^2	.0284	.0004	.0048	.0242
	(.0138)	(.0011)	(.0017)	(.0034)
*e*401*k*	6.837	1.281	2.598	4.460
	(2.173)	(.263)	(.404)	(.801)
N	2,017	2,017	2,017	2,017

3. Quantile Regression with Endogenous Explanatory Variables

- Suppose

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + u_1, \tag{23}$$

where $\mathbf{z}$ is exogenous and y_2 is endogenous – whatever that means in the context of quantile regression.

- Amemiya's (1982) two-stage LAD estimator: reduced form for y_2,

$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2. \tag{24}$$

First step applies OLS or LAD to (24), and gets fitted values, $y_{i2} = \mathbf{z}_i\hat{\boldsymbol{\pi}}_2$. These are inserted for y_{i2} to give LAD of y_{i1} on $\mathbf{z}_{i1},\hat{y}_{i2}$. 2SLAD relies on symmetry of the composite error $\alpha_1 v_2 + u_1$ given $\mathbf{z}$.

• If $D(u_1, v_2|\mathbf{z})$ is centrally symmetric, can use a control function approach. Write

$$u_1 = \rho_1 v_2 + e_1, \qquad (25)$$

where e_1 given $\mathbf{z}$ would have a symmetric distribution. Get LAD residuals $\hat{v}_{i2} = y_{i2} - \mathbf{z}_i \hat{\boldsymbol{\pi}}_2$ and do LAD of y_{i1} on $\mathbf{z}_{i1}, y_{i2}, \hat{v}_{i2}$. Use t test on $\hat{v}_{i2}$ to test null that y_2 is exogenous.

• Interpretation of LAD in context of omitted variables is difficult unless lots of symmetry assumed.

• Chesher (2003) shows how to identify partial derivatives of structural functions without functional form restrictions, but with monotonicity restrictions. As an example, consider the system

$$\begin{aligned} y_1 &= g_1(y_2, \mathbf{z}_1, f_1, a_2) \\ y_2 &= g_2(\mathbf{z}, a_2) \end{aligned}$$

where the functions are strictly increasing in the unobserved heterogeneity, f_1 and a_2.

• Under quantile restrictions and exclusion restrictions, the derivative of $g_1(\cdot)$ with respect to y_2 can be identified at particular values of the observables and quantiles of the unobservables.

• Allows very flexible functional forms in observables and weak assumptions about the dependence between (f_1, a_2) and $\mathbf{z}$ (conditional quantile restrictions).

• Compared with Blundell and Powell (2003), Chesher restricts the way heterogeneity can appear in the structural equation:

$$y_1 = f_1 y_2 + \mathbf{z}_1 \mathbf{b}_1 + a_2$$
$$y_2 = \mathbf{z}\mathbf{b}_2 + \psi_2 a_2$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are random slopes on the exogenous variables.

• But BP require an additive error in the reduced form.

- Abadie, Angrist, and Imbens (2002) consider binary endogenous treatment, say D, and binary instrumental variable, say Z. The potential outcomes are Y_d, $d = 0, 1$ – that is, without treatment and with treatment, respectively. The counterfactuals for treatment are D_z, $z = 0, 1$. Observed are $X, Z, D = (1 - Z)D_0 + ZD_1$, and $Y = (1 - D)Y_0 + DY_1$. AAI study treatment effects for *compliers*, that is, the (unobserved) subpopulation with $D_1 > D_0$.

- Assumptions:

$$(Y_1, Y_0, D_1, D_0) \text{ independent of } Z \text{ conditional on } X \tag{26}$$

$$0 < P(Z = 1|X) < 1 \tag{27}$$

$$P(D_1 = 1|X) \neq P(D_0 = 1|X) \tag{28}$$

$$P(D_1 \geq D_0|X) = 1. \tag{29}$$

Under these assumptions, treatment is unconfounded for compliers:

$$D(Y_0, Y_1|D, X, D_1 > D_0) = D(Y_0, Y_1|X, D_1 > D_0) \tag{30}$$

and treatment effects can be defined based on $D(Y|X, D, D_1 > D_0)$.

• AAI focus on *quantile treatment effects* (Abadie looks at other distributional features):

$$Quant_\tau(Y|X, D, D_1 > D_0) = \alpha_\tau D + X\beta_\tau. \tag{31}$$

(This results in estimated differences for the quantiles of Y_1 and Y_0, not the quantile of the difference $Y_1 - Y_0$.)

• If the dummy variable $C = 1[D_1 > D_0]$ could be observed, problem would be straightforward. Would like to use linear quantile estimation for the subpopulation $C = 1$ because the parameters solve

$$\min_{\alpha,\boldsymbol{\beta}} E[C \cdot g(Y, X, D, \alpha, \boldsymbol{\beta})] \tag{32}$$

where $g(Y, X, D, \alpha, \boldsymbol{\beta}) = c_\tau(Y - \alpha D - X\beta)$ is the check function.

- Instead, can solve

$$\min_{\alpha,\boldsymbol{\beta}} E[\kappa(U) \cdot g(Y, X, D, \alpha, \boldsymbol{\beta})], \tag{33}$$

where $U = (Y, X, D)$ and $\kappa(U) = P(C = 1|U)$. AAI show

$$\kappa_\nu(U) = 1 - \frac{D(1 - \nu(U))}{1 - \pi(X)} - \frac{(1 - D)\nu(U)}{\pi(X)}, \tag{34}$$

where $\nu(U) = P(Z = 1|U)$, and $\pi(X) = P(Z = 1|X)$, which can both be estimated using observed data.

- Two-step estimator solves

$$\min_{\alpha,\beta} \sum_{i=1}^{N} 1[\hat{\kappa}_\nu(U_i) \geq 0]\hat{\kappa}_\nu(U_i) c_\tau(Y_i - \alpha D_i - X_i\beta). \tag{35}$$

• Chernozhukov and Hansen (2005) show how to identify quantile functions without restricting the functional form. But monotonicity in the unosbervable (a scalar) is key.

• If $q(d, x, \tau)$ is the τ^{th} quantile conditional on x for treatment level $D = d$. Under the assumption that the unobservables are independent of the instruments, they show

$$P[Y \leq q(D, X, \tau)|X, Z] = \tau.$$

• This defines moment conditions

$$E(\{1[Y \leq q(D, X, \tau)] - \tau\}|X, Z) = 0.$$

• Chernozhukov and Hansen (2006) consider estimation.

4. Quantile Regression for Panel Data

- Without unobserved effects, QR easy on panel data:

$$Quant_\tau(y_{it}|\mathbf{x}_{it}) = \mathbf{x}_{it}\boldsymbol{\theta},\ t = 1,\ldots,T. \tag{36}$$

Pooled QR, but account for serial correlation in

$$\mathbf{s}_{it}(\boldsymbol{\theta}) = -\mathbf{x}_{it}'\{\tau 1[y_{it} - \mathbf{x}_{it}\boldsymbol{\theta} \geq 0] - (1-\tau)1[y_{it} - \mathbf{x}_{it}\boldsymbol{\theta} < 0]\}.$$

Use “cluster robust” variance matrix estimate:

$$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\mathbf{s}_{it}(\hat{\boldsymbol{\theta}})\mathbf{s}_{ir}(\hat{\boldsymbol{\theta}})' \tag{37}$$

$$\hat{\mathbf{A}} = (2Nh_N)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} 1[|\hat{u}_{it}| \leq h_N] \mathbf{x}_{it}' \mathbf{x}_{it}. \tag{38}$$

- Explicitly allowing unobserved effects is harder.

$$Quant_\tau(y_{it}|\mathbf{x}_i, c_i) = Quant_\tau(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\theta} + c_i. \tag{39}$$

- "Fixed effects" approach, where $D(c_i|\mathbf{x}_i)$ unrestricted, is attractive. Honoré (1992) applied to the uncensored case: LAD on the first differences consistent when $\{u_{it} : t = 1, \ldots, T\}$ is an iid. sequence conditional on $(\mathbf{x}_i, c_i)$ (symmetry not required). When $T = 2$, LAD on the first differences is equivalent to estimating the c_i along with $\boldsymbol{\theta}$, but not with general T.

- Alternative suggested by Abrevaya and Dahl (2006) for $T = 2$. In Chamberlain's correlated random effects linear model,

$$E(y_t|\mathbf{x}_1, \mathbf{x}_2) = \psi_t + \mathbf{x}_t\boldsymbol{\beta} + \mathbf{x}_1\boldsymbol{\xi}_1 + \mathbf{x}_2\boldsymbol{\xi}_2, t = 1, \tag{40}$$

$$\boldsymbol{\beta} = \frac{\partial E(y_1|\mathbf{x})}{\partial \mathbf{x}_1} - \frac{\partial E(y_2|\mathbf{x})}{\partial \mathbf{x}_1}. \tag{41}$$

Abrevaya and Dahl suggest modeling $Quant_\tau(y_t|\mathbf{x}_1, \mathbf{x}_2)$ as in (41) and then defining the partial effect as

$$\boldsymbol{\beta}_\tau = \frac{\partial Quant_\tau(y_1|\mathbf{x})}{\partial \mathbf{x}_1} - \frac{\partial Quant_\tau(y_2|\mathbf{x})}{\partial \mathbf{x}_1}. \tag{42}$$

• Correlated RE approaches difficult: quantiles of sums not sums of quantiles. If $c_i = \psi + \bar{\mathbf{x}}_i \boldsymbol{\xi} + a_i$,

$$y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\theta} + \bar{\mathbf{x}}_i \boldsymbol{\xi} + a_i + u_{it}. \tag{43}$$

Generally, $v_{it} = a_i + u_{it}$ will not have zero conditional quantile. Might estimate (43) by pooled quantile regression for different quantiles.

• More flexibility if we start with median,

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\theta} + c_i + u_{it}, \; Med(u_{it}|\mathbf{x}_i, c_i) = 0, \tag{44}$$

and make symmetry assumptions. Can apply LAD to the time-demeaned equation $\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\theta} + \ddot{u}_{it}$, being sure to obtain fully robust standard errors for pooled LAD.

- If we impose the Chamberlain-Mundlak device, $y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it}$, we can get by with central symmetry of $D(a_i, u_{it}|\mathbf{x}_i)$, so that $D(a_i + u_{it}|\mathbf{x}_i)$ is symmetric about zero, and, if this holds for each t, pooled LAD of y_{it} on $1, \mathbf{x}_{it}$, and $\bar{\mathbf{x}}_i$ consistently estimates $(\psi_t, \boldsymbol{\theta}, \boldsymbol{\xi})$. (If we use pooled OLS with $\bar{\mathbf{x}}_i$ included, we obtain the FE estimate.) Should use robust inference.
- We might even just apply quantile regression directly to

$$y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + v_{it}$$

and interpret $\hat{\boldsymbol{\theta}}_\tau$ as approximating the partial effects of $\mathbf{x}_t$ on the τ^{th} quantile.

5. Quantile Methods for “Censored” Data

• Censored LAD applicable to data censoring and and corner solutions. For true data censoring, let w_i be the underling response (say, wealth or log of a duration) following

$$w_i = \mathbf{x}_i\boldsymbol{\beta} + u_i, \tag{45}$$

but it is top coded or right censored at r_i. Can estimate $\boldsymbol{\beta}$ if

$$Med(u_i|\mathbf{x}_i, r_i) = 0 \tag{46}$$

because $Med(y_i|\mathbf{x}_i, r_i) = \min(\mathbf{x}_i\boldsymbol{\beta}, r_i)$ where $y_i = \min(w_i, r_i)$. Powell’s (1986) CLAD estimator. (Need to always observe r_i; see Honoré, Khan, and Powell (2002) to relax.)

- Less clear that CLAD is “better” than parametric models for corner solution responses. CLAD identifies a single feature of $D(y|\mathbf{x})$, namely, $Med(y|\mathbf{x})$. Models such as Tobit assume more but deliver more. Not just enough to estimate parameters. Common model for corner at zero:

$$y = \max(0, \mathbf{x}\boldsymbol{\beta} + u),\ Med(u|\mathbf{x}) = 0. \tag{47}$$

β_j measures the partial effects on $Med(y|\mathbf{x}) = \max(0, \mathbf{x}\boldsymbol{\beta})$ once $Med(y|\mathbf{x}) > 0$.

- A model no more or less restrictive than (47) is

$$y = a \bullet \exp(\mathbf{x}\boldsymbol{\beta}),\ E(a|\mathbf{x}) = 1, \tag{48}$$

and $E(y|\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})$ is identified. Can have $P(a = 0|\mathbf{x}) > 0$.

• How to interpret applications of CLAD for corner solutions?

$$Med(y_{it}|\mathbf{x}_i, c_i) = \max(0, \mathbf{x}_{it}\boldsymbol{\beta} + c_i). \quad (49)$$

Honoré (1992), Honoré and Hu (2004) show how to estimate $\boldsymbol{\beta}$ under exchangeability assumptions on the idiosyncratic errors in the latent variable model. The partial effect of x_{tj} on $Med(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t, c_i = c)$ is

$$\theta_{tj}(\mathbf{x}_t, c) = 1[\mathbf{x}_t\boldsymbol{\beta} + c > 0]\beta_j. \quad (50)$$

What values should we insert for c? Average of (50) across $D(c_i)$ would be average partial effects (on the median). The β_j give the relative effects of the APEs on the median. If c_i has a $Normal(\mu_c, \sigma_c^2)$ distribution, $E_{c_i}[\theta_{tj}(\mathbf{x}_t, c_i)] = \Phi[(\mu_c - \mathbf{x}_t\boldsymbol{\beta})/\sigma_c]\beta_j$.

- Honoré (2008) has argued that partial effects averaged across the entire distribution for continuous elements of $\mathbf{x}_t$ can be obtained. Without any particular structure on the error, write $y_t = \max(0, \mathbf{x}_t\boldsymbol{\beta} + v_t)$. Then, if v_t has a continuous distribution, the probability of being at the kink is zero. So

$$\frac{\partial y_t}{\partial x_{tj}}(\mathbf{x}_t, v_t) = 1[\mathbf{x}_t\boldsymbol{\beta} + v_t > 0]\beta_j \tag{51}$$

and averaging out across the joint distribution of $(\mathbf{x}_t, v_t)$ gives

$$E_{(\mathbf{x}_t, v_t)}\left[\frac{\partial y_t}{\partial x_{tj}}(\mathbf{x}_t, v_t)\right] = P(y_t > 0)\beta_j \tag{52}$$

- Given $\hat{\beta}_j$ – available using methods by Honoré (2008) and co-authors – (51) is easily estimated by multiplying by the fraction of positive y_t in the sample.
- But we cannot obtain partial effects as a function of $\mathbf{x}_t$ – a leading reason for estimating models with nonlinearity. That would require knowing the distribution of v_t (just as in the case where c is explicitly introduced).
- Unclear what to do about discrete covariates or discrete changes more generally.