What's New in Econometrics?
Lecture 14
Quantile Methods

Jeff Wooldridge
NBER Summer Institute, 2007

1. Reminders About 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. Reminders About Means, Medians, and Quantiles

- Consider the standard linear model in a population, with intercept α and $K \times 1$ slopes $\boldsymbol{\beta}$:

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

Assume $E(u^2) < \infty$, so that the distribution of u is not too spread out. Given a large random sample, when should we expect ordinary least squares, which solves

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

and least absolute deviations (LAD), which solves

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

to provide similar parameter estimates? There are two important cases. If

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

then OLS and LAD both consistently estimate α and $\boldsymbol{\beta}$. 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.
- Therefore, 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. It is important to separate discussions of resiliency to outliers from the

different quantities identified by least squares ($E(y|\mathbf{x})$) and least absolution deviations ($Med(y|\mathbf{x})$).

- 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.

- Big advantage for median over mean: the median passes through monotonic functions. For example, if $\log(y) = \alpha + \mathbf{x}\boldsymbol{\beta} + u$ and $Med(u|\mathbf{x}) = 0$, then $\mathrm{Med}(y|\mathbf{x}) = \exp(\mathrm{Med}[\log(y)|\mathbf{x}]) = \exp(\alpha + \mathbf{x}\boldsymbol{\beta})$. By contrast, we cannot generally find $\mathrm{E}(y|\mathbf{x}) = \exp(\alpha + \mathbf{x}\boldsymbol{\beta})\mathrm{E}[\exp(u)|\mathbf{x}]$.

● But the expectation operator has useful properties that the median does not: linearity and the law of iterated expectations. Suppose we begin with a random coefficient model

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

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

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

where $\alpha = \mathrm{E}(a_i)$ and $\boldsymbol{\beta} = \mathrm{E}(\mathbf{b}_i)$. So OLS consistently estimates α and $\boldsymbol{\beta}$. By contrast, no way to derive $\mathrm{Med}(y_i|\mathbf{x}_i)$ without imposing more restrictions.

● 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 $\mathrm{D}(\mathbf{u}_i|\mathbf{x}_i) = \mathrm{D}(-\mathbf{u}_i|\mathbf{x}_i)$, which implies that, if $\mathbf{g}_i$ is any vector function of $\mathbf{x}_i$, $\mathrm{D}(\mathbf{g}_i'\mathbf{u}_i|\mathbf{x}_i)$ has a univariate distribution that is symmetric about zero. This implies $E(\mathbf{u}_i|\mathbf{x}_i) = \mathbf{0}$.

- Apply central symmetry to random coefficient model by writing $\mathbf{c}_i = (a_i, \mathbf{b}_i)$ with $\boldsymbol{\gamma} = E(\mathbf{c}_i)$, and let $\mathbf{d}_i = \mathbf{c}_i - \boldsymbol{\gamma}$. Then

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

with $\mathbf{g}_i = (1, \mathbf{x}_i)$. If $\mathbf{c}_i$ given $\mathbf{x}_i$ is centrally symmetric about $\boldsymbol{\gamma}$, then $Med(\mathbf{g}_i'(\mathbf{c}_i - \boldsymbol{\gamma})|\mathbf{x}_i) = 0$, and 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 $\mathrm{P}(y_i \leq q(\tau)) \geq \tau$ and $\mathrm{P}(y_i \geq q(\tau)) \geq 1 - \tau$.

- Usually, we are interested in how covariates affect quantiles (of which the median is the special case with $\tau = 1/2$). Under linearity,

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

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." Consistency is relatively easy to establish because $(\alpha(\tau), \boldsymbol{\beta}(\tau))$ are known to minimize $E[c_\tau(y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta})]$ (for example, Manski (1988)). Asymptotic normality is more difficult because any sensible definition of the Hessian of the objective function, away from the nondifferentiable kink, is identically zero. But it has been worked out under a variety of conditions; see Koenker (2005) for a recent treatment.

2. Some Useful Asymptotic Results

What Happens if the Quantile Function is Misspecified?

- 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 **x,** β_j^* can be equal to or proportionl to average partial effects.

- Linear quantile formulation has been viewed by several authors as an approximation (Buchinsky (1991), Chamberlain (1991), Abadie, Angrist, Imbens (2002)). 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

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

In other words, $\boldsymbol{\beta}(\tau)$ is the best weighted mean square approximation to the true quantile function, where the weights depend on average of the conditional density of y_i over a line from $\mathbf{x}\boldsymbol{\beta}$, to the true quantile function, $q_\tau(\mathbf{x})$.

Computing Standard Errors

- For given τ, write

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

and let $\hat{\boldsymbol{\theta}}$ be the quantile estimator. Define quantile residuals $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\theta}}$. Under weak conditions (see, for example, Koenker

(2005)), $\sqrt{N}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal with asymptotic variance $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, where

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

and

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

When we assume 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 (1986, 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 $N \to \infty$ with $\sqrt{N} h_N \to \infty$. For example, $h_N = aN^{-1/3}$ for any $a > 0$. Might use a smoothed version so that all residuals contribute.

- Works for reasons similar to heteroskedasticity-robust standard errors.
- If u_i and $\mathbf{x}_i$ are independent,

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

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

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

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, even the “robust” form of the variance matrix, based on the estimate in (20), is not valid. In the generalized linear models literature, the distinction is sometimes made between a “fully robust” variance estimator and a “semi-robust” variance estimator. If mean is correctly specified and estimator allows unspecified variance, it is semi-robust. If the mean is allowed to be misspecified, fully robust.

- 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 generally 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.

- First, LAD. Amemiya's (1982) two-stage LAD estimator adds a reduced form for y_2, say

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

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 $\mathbf{z}_{i1}, \hat{y}_{i2}$. The 2SLAD estimator 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

$$u_1 = \rho_1 v_2 + e_1, \tag{25}$$

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.

- Abadie (2003) and Abadie, Angrist, and Imbens (2002) define and estimate policy parameters with a 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 $D_1 > D_0$. The assumptions are

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

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

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

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

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 so treatment effects can be defined based on $D(Y|X, D, D_1 > D_0)$, where Y is the observed outcome. 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 for a linear quantile estimation. 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

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

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

- Two-step estimator solves

$$\min_{\delta} \sum_{i=1}^{N} 1[\hat{\kappa}_v(U_i) \geq 0]\hat{\kappa}_v(U_i)c_\tau(Y_i - W_i\delta). \tag{35}$$

where $W_i = (D_i, X_i)$ and δ contains α and β. The indicator function $1[\hat{\kappa}_v(U_i) \geq 0]$ ensures that only observations with nonnegative weights are used. Can use flexible parametric models (series) estimators for $\hat{v}(u)$ and $\hat{\pi}(x)$.

- Chernozhukov and Hansen (2005, 2006) consider identification and estimation of QTEs in a model with endogenous treatment. Let $q(d, x, \tau)$ denote the τ^{th} quantile function for treatment level $D = d$ and covariates x. In the binary case, CH define the QTE as

$$QTE_\tau(x) = q(1, x, \tau) - q(0, x, \tau). \tag{36}$$

- CH use the representation that Y_d, conditional on $X = x$, can be expressed as

$$Y_d = q(d, x, U_d) \tag{37}$$

where

$$U_d | Z \sim \text{Uniform}(0, 1), \tag{38}$$

and Z is the instrumental variable for treatment assignment, D. Key assumptions are that $q(d, x, u)$ is strictly increasing in u and a "rank invariance" condition, whose simplest form is conditional on $X = x$ and $Z = z$, U_d does not depend on d. CH show that, with the observed Y defined as $Y = q(D, X, U_D)$,

$$P[Y \leq q(D,X,\tau)|X,Z] = P[Y < q(D,X,\tau)|X,Z] = \tau. \quad (39)$$

If we could take $Z = D$, (39) would define the quantile $\text{Quant}_{\tau}(Y|D,X)$. Generally, it defines conditional moment conditions

$$E(\{1[Y \leq q(D,X,\tau)] - \tau\}|X,Z) = 0, \quad (40)$$

which is analogous to conditional moment conditions in models with additive errors.

● Chernozhukov and Hansen (2006) assume a linear functional form and obtain the *quantile regression instrumental variables estimator*.

4. Quantile Regression for Panel Data

● Without unobserved effects, easy to use quantile regression methods on panel data:

$$\text{Quant}_\tau(y_{it}|\mathbf{x}_{it}) = \mathbf{x}_{it}\boldsymbol{\theta},\ t = 1,\ldots,T. \tag{41}$$

Use pooled quantile regression. But need to generally account for serial correlation in the “scores,

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

Use

$$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\mathbf{s}_{it}(\hat{\boldsymbol{\theta}})\mathbf{s}_{ir}(\hat{\boldsymbol{\theta}})' \tag{42}$$

and then

$$\hat{\mathbf{A}} = (2Nh_N)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}1[|\hat{u}_{it}| \leq h_N]\mathbf{x}_{it}'\mathbf{x}_{it}. \tag{43}$$

- Explicitly allowing unobserved effects is harder.

$$\text{Quant}_\tau(y_{it}|\mathbf{x}_i, c_i) = \text{Quant}_\tau(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\theta} + c_i. \tag{44}$$

- "Fixed effects" approach, where do not restrict $\text{D}(c_i|\mathbf{x}_i)$, is attractive. From Honoré (1992) applied to the uncensored case, LAD on the first differences is consistent when $\{u_{it} : t = 1, \ldots, T\}$ is an iid. sequence conditional on $(\mathbf{x}_i, c_i)$, even if the common distribution is not symmetric. But this is a fairly strong assumption. When $T = 2$, applying LAD on the first differences is equivalent to estimating the c_i along with $\boldsymbol{\theta}$. Generally, an incidental parameters problem with small T.
- Alternative suggested by Abrevaya and Dahl (2006) for $T = 2$. In

Chamberlain's correlated random effects linear model,

$$E(y_t|\mathbf{x}_1,\mathbf{x}_2) = \psi_t + \mathbf{x}_t\boldsymbol{\beta} + \mathbf{x}_1\boldsymbol{\xi}_1 + \mathbf{x}_2\boldsymbol{\xi}_2, t = 1, \tag{45}$$

$$\boldsymbol{\beta} = \frac{\partial E(y_1|\mathbf{x})}{\partial \mathbf{x}_1} - \frac{\partial E(y_2|\mathbf{x})}{\partial \mathbf{x}_1}. \tag{46}$$

Abrevaya and Dahl suggest modeling $\text{Quant}_\tau(y_t|\mathbf{x}_1,\mathbf{x}_2)$ as in (46) and then defining the partial effect as

$$\boldsymbol{\beta}_\tau = \frac{\partial \text{Quant}_\tau(y_1|\mathbf{x})}{\partial \mathbf{x}_1} - \frac{\partial \text{Quant}_\tau(y_2|\mathbf{x})}{\partial \mathbf{x}_1}. \tag{47}$$

- Generally, correlated random effects approaches are hampered because finding quantiles of sums of random variables is difficult. Suppose we write $c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i$ and then

$$y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it}. \quad (48)$$

Generally, $v_{it} = a_i + u_{it}$ will not have zero conditional quantile. Could just estimate (48) by pooled quantile regression for different quantiles and use the ACF results on approximating quantiles.

- A little more flexibility if we start with median,

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\theta} + c_i + u_{it}, \text{ Med}(u_{it}|\mathbf{x}_i, c_i) = 0, \quad (49)$$

and make symmetry assumptions. If $\mathrm{D}(\mathbf{u}_i|\mathbf{x}_i) = \mathrm{D}(-\mathbf{u}_i|\mathbf{x}_i)$ then all linear combinations of the errors have a symmetric distribution, and so we 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 as in (48), we can get by with central symmetry of $\mathrm{D}(a_i, u_{it}|\mathbf{x}_i)$ has a symmetric distribution around zero then $\mathrm{D}(a_i + u_{it}|\mathbf{x}_i)$ 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.

5. Quantile Methods for "Censored" Data

● Censored LAD applicable to data censoring and and corner solutions. Very useful for true data censoring, where parameters of underlying linear model are of interest. w_i is the response variable (say, wealth or log of a duration) following

$$w_i = \mathbf{x}_i\boldsymbol{\beta} + u_i, \tag{50}$$

but it is top coded or right censored at r_i, then we can estimate $\boldsymbol{\beta}$ under the assumption

$$Med(u_i|\mathbf{x}_i, r_i) = 0 \tag{51}$$

because $Med(y_i|\mathbf{x}_i, r_i) = \min(\mathbf{x}_i\boldsymbol{\beta}, r_i)$ where $y_i = \min(y_i^*, r_i)$. Leads to 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:

$$y = \max(0, \mathbf{x}\boldsymbol{\beta} + u),\ Med(u|\mathbf{x}) = 0. \tag{52}$$

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 (52) is

$$y = a \cdot \exp(\mathbf{x}\boldsymbol{\beta}),\ E(a|\mathbf{x}) = 1, \tag{53}$$

in which case $E(y|\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})$ is identified. Allows for corner because

$P(a = 0|\mathbf{x}) > 0$ is allowed.

- How to interpret panel data applications of CLAD for corner solutions?

$$Med(y_{it}|\mathbf{x}_i, c_i) = \max(0, \mathbf{x}_{it}\boldsymbol{\beta} + c_i). \tag{54}$$

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

$$\theta_{tj}(\mathbf{x}_t, c) = 1[\mathbf{x}_t\boldsymbol{\beta} + c > 0]\beta_j. \tag{55}$$

What values should we insert for c? We need to know something about $D(c_i)$. The average of (55) across the distribution of unobserved heterogeneity would be average partial effects (on the median). Again, we need to identify $D(c_i)$. The β_j give us the sign and relative effects of the APEs. If c_i has a $Normal(\mu_c, \sigma_c^2)$ distribution, then it is easy to show $E_{c_i}[\theta_{tj}(\mathbf{x}_t, c_i)] = \Phi[(\mu_c - \mathbf{x}_t\boldsymbol{\beta})/\sigma_c]\beta_j$.