QUANTILE REGRESSION

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Why Quantile Regression?
2. Review of Means, Medians, and Quantiles
3. Asymptotic Results
4. Quantile Regression with Endogenous Explanatory Variables
5. Quantile Regression for Panel Data
6. Quantile Regression for Corner Solutions

1. WHY QUANTILE REGRESSION?

- Often want to know the effect of changing a covariate – such as a policy intervention – on features of the distribution other than the mean.
- For example, how does eligibility in a particular kind of pension plan affect total wealth at different quantiles of the wealth distribution? The mean effect, while useful, may mask very different effects in different parts of the wealth distribution.

- Another reason to focus on the median (and quantile more generally) is that sometimes we can estimate parameters in underlying models under weaker assumptions using zero conditional median restrictions, rather than zero conditional mean restrictions. An important case is data censoring, which we cover later.
- Manipulations with medians are useful for certain corner solution models, too.

- When we apply different estimation methods (say, ordinary least squares and least absolute deviations) to the same linear model, we must remember that these methods generally identify different quantities (mean versus median in this case).
- In the statistics literature the focus on LAD has often been on its resistance to outliers (which it certainly has). But there are other reasons OLS and LAD can give different results; it need have nothing to do with extreme data points.

2. REVIEW OF MEANS, MEDIANS, AND QUANTILES

- Start with a 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. (So, for example, we rule out a Cauchy distribution for u, or a t_2 distribution.)
- We call this equation a "linear model." There are many different ways to estimate the parameters of this model, and the goal is to evaluate the quality of the estimation procedures under different assumptions.

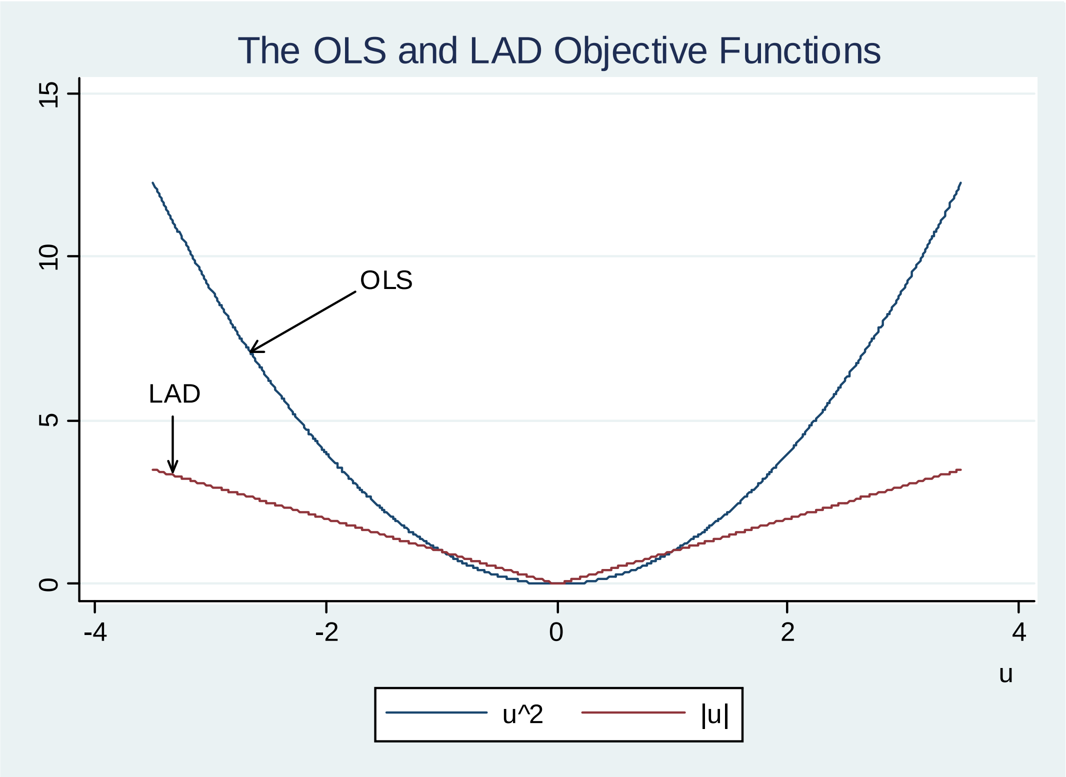
- Ordinary Least Squares (OLS) Estimation:

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

- Least Absolute Deviations (LAD) Estimation:

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

- We should **not** now refer to equation (1) as an "OLS model" or an "LAD model." OLS and LAD are *estimation methods*, not "models."



• 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 $\boldsymbol{\beta}$ because, under (4), $E(u|\mathbf{x}) = Med(u|\mathbf{x}) = 0$, and so

$$\begin{aligned} E(y|\mathbf{x}) &= \alpha + \mathbf{x}\boldsymbol{\beta} \\ Med(y|\mathbf{x}) &= \alpha + \mathbf{x}\boldsymbol{\beta} \end{aligned}$$

As we know, OLS consistently estimates the parameters in a conditional mean and LAD consistently estimates the parameters in a conditional median.

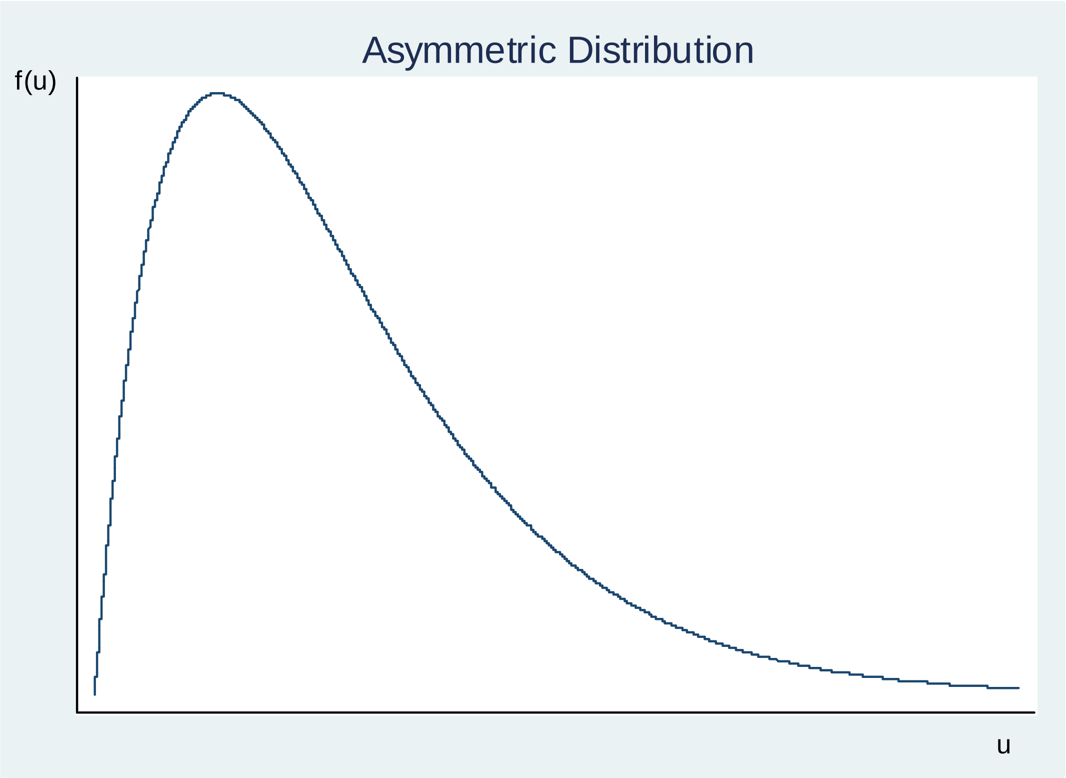
(ii) If

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

where $E(u) = 0$ is the normalization that identifies α, then OLS and LAD both consistently estimate the slopes, $\boldsymbol{\beta}$. By (5), $E(u|\mathbf{x}) = 0$ and so we still have $E(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta}$.

- If u has an asymmetric distribution, then $Med(u) \equiv \eta \neq 0$, and $\hat{\alpha}_{LAD}$ converges to $\alpha + \eta$ because $Med(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + Med(u|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + \eta = (\alpha + \eta) + \mathbf{x}\boldsymbol{\beta}$. But the slopes $\boldsymbol{\beta}$ are identified by LAD.

• 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. Therefore, claims that substantive differences between OLS and LAD estimators must be due to “outliers” may be unwarranted.

- 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.
- OLS consistently estimates the parameters the provide the best mean-square approximation to $E(y|\mathbf{x})$. Unfortunately, characterizing LAD under misspecification is harder but possible (later).
- One case where LAD is clearly preferred is when $E[|u|] < \infty$ but $E(u^2) = \infty$.

- Advantage for median over mean: median passes through monotonic functions. For example, suppose $\log(y) = \alpha + \mathbf{x}\boldsymbol{\beta} + u$ and $Med(u|\mathbf{x}) = 0$. Therefore,

$$Med[\log(y)|\mathbf{x}] = \alpha + \mathbf{x}\boldsymbol{\beta}$$

Write $y = \exp(\alpha + \mathbf{x}\boldsymbol{\beta} + u)$. Then

$$\begin{aligned} Med(y|\mathbf{x}) &= Med[\exp(\alpha + \mathbf{x}\boldsymbol{\beta} + u)|\mathbf{x}] = \exp[\alpha + \mathbf{x}\boldsymbol{\beta} + Med(u|\mathbf{x})] \\ &= \exp(\alpha + \mathbf{x}\boldsymbol{\beta}). \end{aligned}$$

- By contrast, we cannot generally find $E(y|\mathbf{x}) = \exp(\alpha + \mathbf{x}\boldsymbol{\beta})E[\exp(u)|\mathbf{x}]$ because $E[\exp(u)|\mathbf{x}]$ is an unknown function of $\mathbf{x}$ even if we assume $E(u|\mathbf{x}) = 0$ or $Med(u|\mathbf{x}) = 0$.
- As we will see, being able to pass the median through monotonic functions is very useful when data have been censored.

• Aside: Suppose $u \sim Normal(0, \sigma^2)$ and define $w = \exp(u)$, so w has a lognormal distribution. Then $Med(w) = \exp[Med(u)] \ = \exp(0) = 1$ and $E[\exp(u)] \ = \exp(\sigma^2/2) > 1$. So

$$E(w) > Med(w),$$

as is often the case for asymmetric distributions in economics.

- The previous derivation for finding the conditional median of y given the conditional median of $\log y$) is just a special case of

$$Med[g(y)|\mathbf{x}] = g[Med(y|\mathbf{x})]$$

where $g(\bullet)$ is monotonically increasing or decreasing (and need not be strictly increasing or decreasing) on the support of y.

- And we also know that for nonlinear functions $g(\bullet)$,

$$E[g(y)|\mathbf{x}] = g[E(y|\mathbf{x})].$$

- The expectation cannot be passed through monotonic functions, but it has useful properties that the median does not, particularly linearity and the law of iterated expectations. (The median operator is not linear; in particular, $Med(w+y) \neq Med(w) + Med(y)$. Also, there is no "law of iterated medians.")
- Suppose

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

where $(a_i, \mathbf{b}_i)$ is independent of $\mathbf{x}_i$. Define the population averages $\alpha = E(a_i)$ and $\boldsymbol{\beta} = E(\mathbf{b}_i)$ (so the β_j are average partial effects).

• $E(y_i|\mathbf{x}_i)$ is easy to find:

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

• Equation (7) immediatly shows OLS is consistent for α and $\boldsymbol{\beta}$ because OLS is consistent for the parameters in a conditional mean linear in those parameters.

• Generally, we cannot find $Med(y_i|\mathbf{x}_i)$.

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

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

If $\mathbf{r}_i = (a_i - \alpha, \mathbf{b}_i - \boldsymbol{\beta})$ 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 $\boldsymbol{\beta}$.

- Therefore, we can only be guaranteed that LAD consistently estimates an interesting set of parameters under assumptions that imply OLS would consistently estimate those same parameters. (Again, we are ruling out case of very fat-tailed distributions.)
- Generally, the problem of what LAD estimates when we deviate from the model with a single source of heterogeneity appears unsolved, unless we impose strong assumptions.

Quantiles

• For $0 < \tau < 1$, $q(\tau)$ is the τ^{th} quantile of y_i if $P[y_i \leq q(\tau)] \geq \tau$ and $P[y_i \geq q(\tau)] \geq 1 - \tau$.

• In general, a quantile need not be unique. (Special case: a median need not be unique.)

• In the common case where y_i is continuous with a strictly increasing cdf, $q(\tau)$ is the unique value such that

$$\begin{aligned} P[y_i \leq q(\tau)] &= \tau \\ P[y_i \geq q(\tau)] &= P[y_i > q(\tau)] = 1 - \tau. \end{aligned}$$

- Let covariates affect quantiles. Index the parameters by τ. Under linearity,

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

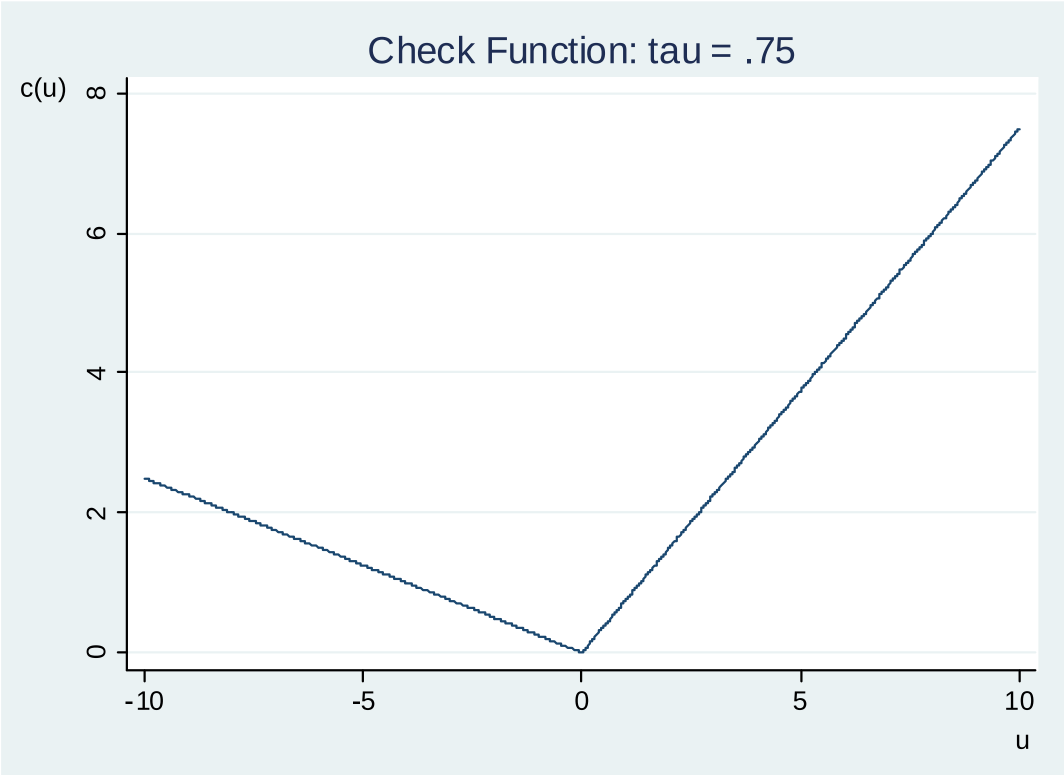
Under (9), consistent estimators of $\alpha(\tau)$ and $\boldsymbol{\beta}(\tau)$ are obtained by minimizing the "check" function:

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

where

$$c_{\tau}(u) = (\tau 1[u \geq 0] + (1-\tau)1[u < 0])|u| = (\tau - 1[u < 0])u$$

and $1[\cdot]$ is the indicator function.


Check Function: tau = .75
c(u)
8
6
4
2
0
-10
-5
0
5
10
u

• The check function identifies $\alpha(\tau)$ and $\boldsymbol{\beta}(\tau)$ in the sense that these parameters solve the population problem

$$\min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^K} E[c_\tau(y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta})],$$

and then we have to assume uniqueness.

• The proof proceeds by showing that, for any $\mathbf{x}_i$, $\alpha(\tau)$ and $\boldsymbol{\beta}(\tau)$ actually solve

$$\min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^K} E[c_\tau(y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta})|\mathbf{x}_i].$$

• Manski (1988, *Analog Estimation Methods in Econometrics*) contains a proof.

3. 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 **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 Fernández-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

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

Computing Standard Errors

- For given τ, write

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

where $x_1 \equiv 1$, 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

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

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 simply

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

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 $N \to \infty$ 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.

- 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

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

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

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

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

- If the quantile function is misspecified, the “robust” form based on (18) (with $\hat{\mathbf{B}}$ as in (17)), is not valid. In the generalized linear models literature, distinction between “semi-robust” variance estimator (mean correctly specified) and a “ fully robust” estimator (mean might be misspecified).
- For quantile regression, a fully robust variance estimator, which allows the quantile function to be misspecfied, requires a different estimator of $\mathbf{B}$.

• 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 – that is, just resample all variables – generally provides 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.
- Example using Abadie (2003). These are nonrobust standard errors. *nettfa* is net total financial assets.

- Stata Output:

```
. use 401ksubs

. keep if fsize == 1
(7258 observations deleted)

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       e401k |      2017    .3604363    .4802461          0          1
         inc |      2017    29.44618    16.67356     10.008    143.067
        marr |      2017    .0183441    .1342256          0          1
        male |      2017    .5418939    .4983654          0          1
         age |      2017    39.27814    10.82328         25         64
-------------+--------------------------------------------------------
       fsize |      2017           1           0          1          1
      nettfa |      2017    13.59498    47.59058     -143.5   1134.098
       p401k |      2017    .2429351    .4289625          0          1
        pira |      2017    .2141795    .4103536          0          1
       incsq |      2017    1144.947    1581.761   100.1601   20468.17
-------------+--------------------------------------------------------
       agesq |      2017    1659.857    922.5799        625       4096
```

```
. count if nettfa <= 0
  706

. tab e401k

      =1 if |
eligble for |
    401(k) |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,290       63.96       63.96
          1 |        727       36.04      100.00
------------+-----------------------------------
      Total |      2,017      100.00
```

```
. reg nettfa inc age agesq e401k, robust

Linear regression                                      Number of obs =    2017
                                                       F(  4,  2012) =   30.66
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1273
                                                       Root MSE      =  44.502

------------------------------------------------------------------------------
             |               Robust
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .7825921   .1041046     7.52   0.000     .5784281    .9867562
         age |  -1.567659   1.075848    -1.46   0.145    -3.677551    .5422324
       agesq |   .0283926   .0138173     2.05   0.040      .001295    .0554902
       e401k |   6.836563   2.173342     3.15   0.002     2.574328     11.0988
       _cons |   2.533552   19.26135     0.13   0.895    -35.24072    40.30782
------------------------------------------------------------------------------
```

```
. qreg nettfa inc age agesq e401k
Iteration  1:  WLS sum of weighted deviations =  34159.391

Iteration  1: sum of abs. weighted deviations =  34187.253

Iteration 26: sum of abs. weighted deviations =  30905.329

Median regression                                    Number of obs =      2017
  Raw sum of deviations 33151.39 (about 1.4)
  Min sum of deviations 30905.33                     Pseudo R2     =    0.0678

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .3239283   .0116808    27.73   0.000     .3010205    .3468361
         age |  -.2443716   .1458544    -1.68   0.094     -.530413    .0416699
       agesq |   .0047983     .00171     2.81   0.005     .0014448    .0081518
       e401k |   2.597726   .4038806     6.43   0.000     1.805658    3.389794
       _cons |  -3.572832   2.897657    -1.23   0.218    -9.255555     2.10989
------------------------------------------------------------------------------
```

```
. qreg nettfa inc age agesq e401k, q(.25)
Iteration  1:  WLS sum of weighted deviations =  29542.707

Iteration  1: sum of abs. weighted deviations =  29403.746

Iteration 21: sum of abs. weighted deviations =  19568.944

.25 Quantile regression                             Number of obs =      2017
  Raw sum of deviations 19760.67 (about -.15000001)
  Min sum of deviations 19568.94                    Pseudo R2     =    0.0097

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .0712858   .0072275     9.86   0.000     .0571118    .0854599
         age |   .0336287   .0954666     0.35   0.725     -.153595    .2208524
       agesq |    .000372    .001113     0.33   0.738    -.0018107    .0025547
       e401k |   1.281012   .2627072     4.88   0.000     .7658052    1.796218
       _cons |  -4.372772   1.895672    -2.31   0.021    -8.090457   -.6550865
------------------------------------------------------------------------------
```

```
. qreg nettfa inc age agesq e401k, q(.75)
Iteration  1:  WLS sum of weighted deviations =  35270.543

Iteration  1: sum of abs. weighted deviations =   35277.14

Iteration 22: sum of abs. weighted deviations =  33600.122

.75 Quantile regression                             Number of obs =      2017
  Raw sum of deviations 41098.57 (about 13.2)
  Min sum of deviations 33600.12                    Pseudo R2     =    0.1825

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |    .797724   .0252319    31.62   0.000     .7482406    .8472075
         age |  -1.385644   .2865236    -4.84   0.000    -1.947558   -.8237297
       agesq |    .024192   .0033797     7.16   0.000     .0175639    .0308202
       e401k |   4.460003   .8006231     5.57   0.000     2.889866     6.03014
       _cons |   7.538962   5.732206     1.32   0.189    -3.702718    18.78064
------------------------------------------------------------------------------
```

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

• Can use the bootstrap to get the fully robust standard errors (valid with or with independence between u_i and $\mathbf{x}_i$). Below is for LAD, and the standard errors are substantially larger than the nonrobust ones.

```
. bsqreg nettfa inc age agesq e401k, reps(500)
(fitting base model)

Median regression, bootstrap(500) SEs                  Number of obs =       2017
  Raw sum of deviations 33151.39 (about 1.4)
  Min sum of deviations 30905.33                       Pseudo R2     =     0.0678

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .3239283   .0396347     8.17   0.000      .246199    .4016576
         age |  -.2443716   .1997378    -1.22   0.221    -.6360862     .147343
       agesq |   .0047983   .0025729     1.86   0.062    -.0002475    .0098441
       e401k |   2.597726   .5752944     4.52   0.000     1.469491    3.725961
       _cons |  -3.572832   3.819017    -0.94   0.350    -11.06247    3.916809
------------------------------------------------------------------------------
```

Efficiency Calculations

- Suppose in the model

$$y = \mathbf{x}\boldsymbol{\beta} + u,$$

where $\mathbf{x}$ includes unity, u is independent of $\mathbf{x}$ with a symmetric distribution about zero. Also, $f_u(0) > 0$ and $E(u^2) < \infty$.

- Then OLS and LAD are both consistent for $\boldsymbol{\beta}$ and $\sqrt{N}$-asymptotically normal:

$$Avar\sqrt{N}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) = \sigma^2[E(\mathbf{x}'\mathbf{x})]^{-1}$$
$$Avar\sqrt{N}(\hat{\boldsymbol{\beta}}_{LAD} - \boldsymbol{\beta}) = \frac{1}{4[f_u(0)]^2}[E(\mathbf{x}'\mathbf{x})]^{-1}.$$

• We can compare asymptotic efficiency by comparing the scalars out front: σ^2 versus $1/\{4[f_u(0)]^2\}$.

• For example, suppose the population distribution is $Normal(0,\sigma^2)$. Then $f_u(0) = 1/\sqrt{2\pi\sigma^2}$ and so

$$\frac{1}{4[f_u(0)]^2} = \frac{2\pi\sigma^2}{4} = \left(\frac{\pi}{2}\right)\sigma^2 \approx 1.57\sigma^2$$

• This shows that LAD is asymptotically inefficient – 57% less efficient – compared with OLS when the u_i come from a normal distribution.

• Now suppose that u_i has a double exponential distribution with mean zero, which has density

$$f(u) = \frac{1}{2\kappa} \exp(-|u|/\kappa),$$

where $Var(u_i) = 2\kappa^2$.

• Can show the MLE of $\boldsymbol{\beta}$ in this case is the LAD estimator. The log likelihood is

$$\mathcal{L}_N(\mu, \kappa) = -\log(\kappa) - \kappa^{-1} \sum_{i=1}^{N} |y_i - \mathbf{x}_i \boldsymbol{\beta}|.$$

- Further, $f(0) = 1/(2\kappa)$ and so

$$\frac{1}{4[f_u(0)]^2} = 4\kappa^2/4 = \kappa^2$$

whereas

$$\sigma^2 = 2\kappa^2$$

- The asymptotic variance of the sample average is twice that of the sample median.

- For general asymmetric distributions it makes no sense to discuss asymptotic efficiency of the two estimators unless u is independent of $\mathbf{x}$, in which case we can compare the slopes.
- If $D(u|\mathbf{x})$ is asymmetric and depends on $\mathbf{x}$, OLS and LAD generally estimate different parameters.

Fully Parametric Approaches

- If we have fully specified a model for $D(y|\mathbf{x})$ then we can learn anything about $D(y|\mathbf{x})$, including any quantile we want.
- If we start with

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

then specifying $D(u|\mathbf{x})$ implies a model for $D(y|\mathbf{x})$.

- As an example, suppose

$$D(u|\mathbf{x}) = Normal(0, \sigma^2 \exp(2\mathbf{x}\boldsymbol{\gamma}))$$

- Then $E(y|\mathbf{x}) = Med(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta}$ and, for any quantile τ,

$$Quant_\tau(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + Quant_\tau(u|\mathbf{x}).$$

- Now $Quant_\tau(u|\mathbf{x})$ is the value $q_\tau(\mathbf{x})$ such that

$$P[u \leq q_\tau(\mathbf{x})|\mathbf{x}] = \tau$$

or

$$P\left[\frac{u}{\sigma \exp(\mathbf{x}\boldsymbol{\gamma})} \leq \frac{q_\tau(\mathbf{x})}{\sigma \exp(\mathbf{x}\boldsymbol{\gamma})} \right] = \tau$$

• But $e \equiv u/[\sigma \exp(\mathbf{x}\boldsymbol{\gamma})]$ is independent of $\mathbf{x}$ with a standard normal distribution, and so we must have

$$\frac{q_\tau(\mathbf{x})}{\sigma \exp(\mathbf{x}\boldsymbol{\gamma})} = a_\tau$$

where a_τ is the τ^{th} quantile of the standard normal distribution (which we can easily find). Therefore,

$$q_\tau(\mathbf{x}) = a_\tau \sigma \exp(\mathbf{x}\boldsymbol{\gamma})$$

and so

$$Quant_\tau(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + a_\tau \sigma \exp(\mathbf{x}\boldsymbol{\gamma}).$$

- If $\boldsymbol{\gamma} \neq \mathbf{0}$, this quantile function is nonlinear in $\mathbf{x}$ except when $\tau = .5$, in which case $a_\tau = 0$.
- When $\boldsymbol{\gamma} = \mathbf{0}$ (homskedasticity), $Quant_\tau(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + a_\tau\sigma$ for any τ. (u is independent of $\mathbf{x}$ in this case, and so all quantile functions have the same slopes, $\boldsymbol{\beta}$, but with intercepts $\alpha + a_\tau\sigma$.)
- The parameters α, $\boldsymbol{\beta}, \sigma^2$, and $\boldsymbol{\gamma}$ can be estimated by MLE using the normal density that recognizes the mean and variance both depend on $\mathbf{x}$. One might try using only $Quant_\tau(y|\mathbf{x}) = \alpha + \mathbf{x}\boldsymbol{\beta} + a_\tau\sigma\exp(\mathbf{x}\boldsymbol{\gamma})$, but there is an identification problem when $\boldsymbol{\gamma} = \mathbf{0}$.

4. 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. Specify a reduced form for y_2,

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

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

- 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 $(\mathbf{z}, v_2)$ has 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 joint symmetry assumed.

- Little has been done on y_2 binary (except in the special case of treatment effects). Clearly cannot just plug in, say, probit fitted values, and then use LAD. Similar comments hold for other discrete y_2.
- Control function approaches with “generalized residuals” may provide good approximations.

5. QUANTILE REGRESSION FOR PANEL DATA

- Without unobserved effects, QR easy on panel data:

$$Quant_{\tau}(y_{it}|\mathbf{x}_{it}) = \mathbf{x}_{it}\boldsymbol{\theta},\ t = 1,\ldots,T. \tag{26}$$

Pooled QR, but account for serial correlation in the score,

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

Use “cluster robust” variance matrix estimate for **B**:

$$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\mathbf{s}_{it}(\hat{\boldsymbol{\theta}})\mathbf{s}_{ir}(\hat{\boldsymbol{\theta}})' \tag{27}$$

$$\hat{\mathbf{A}} = (2Nh_N)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} 1[|\hat{u}_{it}| \leq h_N] \mathbf{x}_{it}' \mathbf{x}_{it}. \tag{28}$$

- Explicitly allowing unobserved effects is harder.

$$Quant_\tau(y_{it}|\mathbf{x}_i, c_i) = Quant_\tau(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\theta} + c_i. \tag{29}$$

- "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,

$$E(y_t|\mathbf{x}_1, \mathbf{x}_2) = \psi_t + \mathbf{x}_t\boldsymbol{\beta} + \mathbf{x}_1\boldsymbol{\xi}_1 + \mathbf{x}_2\boldsymbol{\xi}_2, t = 1, 2 \tag{30}$$

$$\boldsymbol{\beta} = \frac{\partial E(y_1|\mathbf{x})}{\partial \mathbf{x}_1} - \frac{\partial E(y_2|\mathbf{x})}{\partial \mathbf{x}_1} \tag{31}$$

Abrevaya and Dahl suggest modeling $Quant_\tau(y_t|\mathbf{x}_1, \mathbf{x}_2)$ as in (30) and then *defining* the partial effects as

$$\boldsymbol{\beta}_\tau = \frac{\partial Quant_\tau(y_1|\mathbf{x})}{\partial \mathbf{x}_1} - \frac{\partial Quant_\tau(y_2|\mathbf{x})}{\partial \mathbf{x}_1} \tag{32}$$

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

$$y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\theta} + \bar{\mathbf{x}}_i \boldsymbol{\xi} + a_i + u_{it}. \tag{33}$$

Generally, $v_{it} = a_i + u_{it}$ will not have zero conditional quantile. Nevertheless, might estimate (33) by pooled quantile regression for different quantiles.

- More flexibility if we start with median,

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\theta} + c_i + u_{it},\ Med(u_{it}|\mathbf{x}_i, c_i) = 0, \tag{34}$$

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 $(\psi_t, \boldsymbol{\theta}, \boldsymbol{\xi})$. (Remember, if we use pooled OLS with $\bar{\mathbf{x}}_i$ included along with $\mathbf{x}_{it}$, we obtain the FE estimates.)
- Should use serial-correlation-robust inference.

6. QUANTILE REGRESSION FOR CORNER SOLUTIONS

- Suppose that y is a corner solution response with a corner at zero. We know that a general model that captures this feature is

$$y = \max(0, \mathbf{x}\boldsymbol{\beta} + u). \tag{35}$$

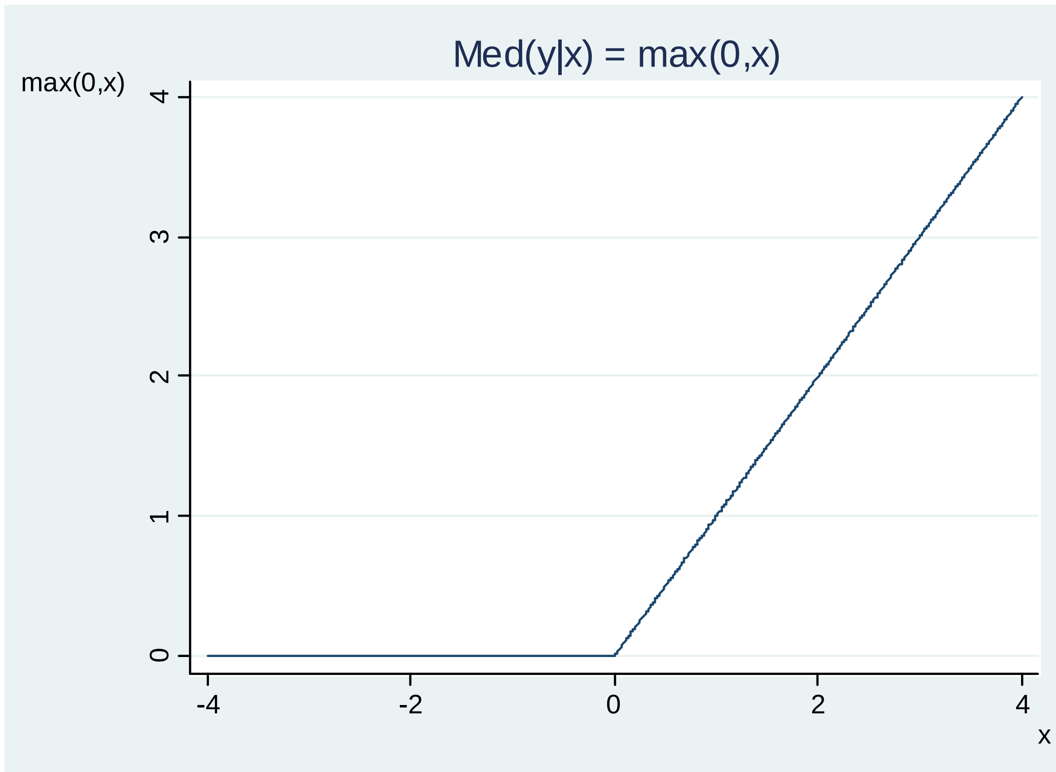
- If we assume

$$Med(u|\mathbf{x}) = 0 \tag{36}$$

then

$$Med(y|\mathbf{x}) = \max(0, \mathbf{x}\boldsymbol{\beta}). \tag{37}$$

• In other words, the zero conditional median restriction on u identifies one feature of $D(y|\mathbf{x})$, namely, $Med(y|\mathbf{x})$.



• The β_j measure the partial effects on $Med(y|\mathbf{x})$ once $Med(y|\mathbf{x}) > 0$. For example, if x_j is continuous,

$$\frac{\partial Med(y|\mathbf{x})}{\partial x_j} = \beta_j 1[\mathbf{x}\boldsymbol{\beta} > 0]. \tag{38}$$

• As Honoré (2008) has recently shown, a simple estimator of the average of these effects (across the distribution of $\mathbf{x}$) is easily estimated: $\hat{\eta}\hat{\beta}_j$ where $\hat{\eta}$ is the fraction of strictly positive y_i.

• The so-called *censored least absolute deviations (CLAD)* estimator solves

$$\min_{\mathbf{b}\in\mathbb{R}^K} \sum_{i=1}^{n} |y_i - \max(0, \mathbf{x}_i\mathbf{b})| \tag{39}$$

• The objective function is continuous in the parameters, so consistency is relatively straightforward (under identification).

• The nonsmoothness makes asymptotic normality hard, but Powell (1984) shows it under general conditions. Estimation of the asymptotic variance has been coded, too, and the bootstrap is used.

- One way to view the CLAD approach is that it identfies $Med(y|\mathbf{x}) = \max(0, \mathbf{x}\boldsymbol{\beta})$ for a variety of shapes for $D(u|\mathbf{x})$. But there is a cost: other features of $D(y|\mathbf{x})$, such as the mean, are not identified. So, CLAD does not allow us to aggregate the effects of a policy or program. We can get the median effect for groups indexed by the observed covariates.

• A model no more or less restrictive than $y = \max(0, \mathbf{x}\boldsymbol{\beta} + u)$, $Med(u|\mathbf{x}) = 0$, is

$$y = a \cdot \exp(\mathbf{x}\boldsymbol{\beta}),\ E(a|\mathbf{x}) = 1, \tag{40}$$

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

• The standard Tobit model, or extensions such as allowing for heteroskedasticity in the latent error, can be restrictive, but they identify all of $D(y|\mathbf{x})$. In other words, there is a tradeoff between assumptions and how much can be learned.

- In the panel case, we can start with

$$\begin{aligned} y_{it} &= \max(0, \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}) \\ Med(u_{it}|\mathbf{x}_i, c_i) &= 0,\ t = 1, \ldots, T. \end{aligned}$$

- These imply that

$$Med(y_{it}|\mathbf{x}_i, c_i) = \max(0, \mathbf{x}_{it}\boldsymbol{\beta} + c_i). \tag{41}$$

- Notice that strict exogeneity is assumed because $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})$ appears on the left hand side.
- Honoré (1992) showed how to estimate $\boldsymbol{\beta}$ without restricting $D(c_i|\mathbf{x}_i)$ by imposing "exchangeability" assumptions on $\{u_{it} : t = 1, \ldots, T\}$; independent and identicallly distributed is sufficient but not necessary. Nonstationarity, including heteroskedasticity, is ruled out. So, it is like a "fixed effects" method for corner solutions but the c_i are not parameters to estimate. And it does impose extra assumptions on $\{u_{it} : t = 1, \ldots, T\}$.

• The partial effect of x_{tj} on $Med(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t, c_i = c)$ is

$$\theta_{tj}(\mathbf{x}_t, c) = 1[\mathbf{x}_t\boldsymbol{\beta} + c > 0]\beta_j. \tag{42}$$

• What values should we insert for c? Average of (42) across $D(c_i)$ would be average partial effects (on the median):

$$\gamma_{tj}(\mathbf{x}_t) = E_{c_i}[\theta_{tj}(\mathbf{x}_t, c_i)] \;\; = [1 - G(\mathbf{x}_t\boldsymbol{\beta})]\beta_j$$

where $G(\cdot)$ is the unconditional cdf of c_i.

• 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) can also be applied in this case. Write $y_t = \max(0, \mathbf{x}_t\boldsymbol{\beta} + v_t)$ where, in this case, we are thinking $v_t = c + u_t$. Then, if v_t has a continuous distribution, the probability of being at the kink is zero. So

$$\frac{\partial y_t}{\partial x_{tj}}(\mathbf{x}_t, v_t) = 1[\mathbf{x}_t\boldsymbol{\beta} + v_t > 0]\beta_j \tag{43}$$

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

$$E_{(\mathbf{x}_t,v_t)}\left[\frac{\partial y_t}{\partial x_{tj}}(\mathbf{x}_t, v_t)\right] = P(y_t > 0)\beta_j \tag{44}$$

• Given $\hat{\beta}_j$ – available using methods summarized by Honoré (2008) – see also Arellano and Honoré (2001, *Handbook of Econometrics*, Volume 5) – (44) is easily estimated by multiplying $\hat{\beta}_j$ by the fraction of positive y_t in the sample.

• Remember, we can also use parametric models: the Chamberlain-Mundlak version of the RE Tobit model (where $D(u_{it}|\mathbf{x}_i, c_i) = Normal(0, \sigma_u^2)$ and $c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i$, $D(a_i|\mathbf{x}_i) = Normal(0, \sigma_a^2)$). In the parametric setting, we can easily obtain APEs, and if we impose conditional independence on the $\{u_{it}\}$, also other partial effects (such as the partial effect at the average).

• Generally, using median and quantile restrictions on $D(y_{it}|\mathbf{x}_{it}, c_i)$, without restricting the distribution of c_i (either unconditionally or conditional on $\mathbf{x}_i$), we cannot obtain partial effects as a function of $\mathbf{x}_t$; that would require knowing at least the central tendency the distribution of c_i (say, the mean or median).

• With the semiparametric approach, it is unclear what to do about discrete changes. If $\mathbf{x}_t^{(1)}$ and $\mathbf{x}_t^{(0)}$ are two values of $\mathbf{x}_t$, we would like to study

$$\max(0, \mathbf{x}_t^{(1)}\boldsymbol{\beta} + c) - \max(0, \mathbf{x}_t^{(0)}\boldsymbol{\beta} + c),$$

but we do not know what to plug in for c, or how to average out across the distribution of c.