ECO 519. Handout on Powell, Censored LAD (1984)

The basic setup is a censored-dependent variable situation, where for each observation $i = 1, \ldots, n$ we observe $y_i = \max\{0, x_i'\beta_0 + u_i\}$, where the error term u_i is continuously distributed, independent of x_i and has **median equal to zero**[1]. Let the cdf of u_i be $F(\cdot)$ with corresponding density function $f(\cdot)$. We will assume throughout that $f(0) > 0$, so the median of u is unique[2]. Given this setup, is is easy to see that conditional on x_i:

$$\Pr(y_i < y|x_i) = \begin{cases} 0 & \text{if } y < 0 \\ F\big(y - x_i'\beta_0\big) & \text{if } y \geq 0 \end{cases}$$

Therefore, the median of y (conditional on x_i) is determined by $\min\Big\{F(0), F\big(-x_i'\beta_0\big)\Big\}$. We have:

$$\text{median}(y_i|x_i) = \max\big\{0, x_i'\beta_0\big\}.$$

So, it was easy to characterize the conditional median of y_i in this censored-dependent variable context. How can we exploit this information in order to estimate β? It is a well known result that for any scalar random variable y, the loss function $E\big|y - b\big|$ is minimized at $b = \text{median}(y)$. Following this principle, the censored-LAD estimator $\widehat{\beta}$ minimizes the objective function $S_n(\beta)$ given by:

$$S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Big|y_i - \max\{0, x_i'\beta\}\Big|. \tag{1}$$

over a compact parameter space $\mathrm{B} \subset \mathbb{R}^k$.

1 Consistency of $\widehat{\beta}$

If we had $y < 0$ w.p.1, then the median of y would not depend on β. In this case, β would fail to be point-identified; any β for which $x'\beta \leq 0$ w.p.1 would yield the same median of y

[1]More generally, we assume that conditional on x_i, the error term u_i has median equal to zero. This allows for heteroskedasticity.

[2]This is not completely accurate. Remember from our discussion about Almost Sure Representation theorems that the median of y is formally defined as $\inf\{t : \Pr(y \geq t) \geq 1/2\}$, which is always unique. We need $f(0) > 0$ here in order to ensure asymptotic normality of our estimator.

(zero). So first of all, we need to rule out the case in which $u \leq -x'\beta_0$ w.p.1. In addition to this, we also need to rule out the case $x'\beta_0 < 0$ w.p.1, because in that case —once again— there is no longer a one-to-one relationship between the median of y and β_0: If we change β_0 slightly, the median of y would not change. Finally, we also need to rule out perfect multicollinearity of x. These concerns are captured in Assumption R.1 of Powell:

R.1 The regressors x_i are independently distributed[3] with $E\|x_i\|^3 \leq K_0$ for all i and some K_0, and for some $\epsilon_0 > 0$ and some $n_0 \in \mathbb{N}$, the smallest characteristic root of the matrix

$$E\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i'\beta_0 \geq \epsilon_0\} x_i x_i'\right]$$

is bounded below by some ν_0 for all $n \geq n_0$

the smallest eigenvalue is involved because eigenvalues can be used to characterize lower bounds for the norm of the matrix. This assumption ensures that the distribution of the covariates has enough mass in the set $\{x_i'\beta_0 > 0\}$. The paper also assumes a compact parameter space B, in addition to the assumptions stated above regarding the $F(\cdot)$ and $f(\cdot)$ (the cdf and density of u respectively). Assumption R.1 yields the following result: For any $\epsilon > 0$, $\exists \delta > 0$ such that

$$\inf_{\|\beta-\beta_0\|>\epsilon} E\big[S_n(\beta)\big] \geq E\big[S_n(\beta_0)\big] + \delta$$

This condition and the previous assumptions yield strong consistency: $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s} \boldsymbol{\beta_0}$. See especially Equations (A.10) and (A.11) in the appendix. Right now, we are more interested in using this result (strong consistency) in order to see how the paper uses Huber's Lemma to establish asymptotic normality (next).

2 Asymptotic normality of $\widehat{\beta}$ using Huber's conditions.

Due to the nature of the objective function, we cannot assume that $\widehat{\beta}$ satisfies some smooth first-order conditions, which would enable us to find the asymptotic distribution of $\widehat{\beta}$ via a

[3]Throughout these notes, we will act as if the x_i's were in fact also identically distributed, which changes nothing fundamentally. The assumption of independence is never relaxed in the paper. See footnote #3 in the paper.

Taylor approximation. Note that $S_n(\beta)$ is not differentiable whenever $y_i - \max\{0, x_i'\beta\} = 0$, but it has well-defined left and right derivatives everywhere. This would not hold if $S_n(\cdot)$ were not continuous for example. For each one of the individual terms of $S_n(\beta)$, the right-partial derivative with respect to β_j is given by

$$\begin{cases} -x_{ij} & \text{if } x_i'\beta > 0 \text{ and } y_i \geq x_i'\beta \\ x_{ij} & \text{if } x_i'\beta > 0 \text{ and } y_i < x_i'\beta \\ 0 & \text{otherwise} \end{cases}$$

the left and right derivatives are the same everywhere except when $y_i - \max\{0, x_i'\beta\} = 0$, which is when the function fails to be differentiable.

Before proceeding, we need additional assumptions to those that ensured strong consistency of $\widehat{\beta}$. We need the conditional median of y to be well-behaved as a function of β at least inside some neighborhood containing θ_0. This conditional median will not be well-behaved as a function of β for realizations of x_i for which $x_i'\beta = 0$, since $\max\{0, x_i'\beta\}$ is not differentiable in β whenever $x_i'\beta = 0$. This is achieved by introducing the following assumption (in addition to those that yielded strong consistency)

R.2 For $r, z \in \mathbb{R}$ define

$$G_i(z, \beta, r) = E\Big[\mathbb{1}\big\{|x_i'\beta| \leq \|x_i\| \cdot z\big\}\|x_i\|^r\Big]$$

Then there exists ξ_0 and K_1 such that for any $0 \leq z < \xi_0$ and $r = 0, 1, 2$:

$$\sup_{\substack{1 \leq i \leq n \\ \|\beta - \beta_0\| \leq \xi_0}} G_i(z, \beta, r) \leq K_1 \cdot z$$

The first consequence of this assumption is that $\Pr\big[x_i'\beta = 0 \text{ for some } 1 \leq i \leq n\big] = 0$ for all $\|\beta - \beta_0\| \leq \xi_0$, but it also ensures that $\|x_i\|$ has uniformly bounded 1st and second moments for values of β near β_0 such that $x_i'\beta \approx 0$. Now let us quickly see how the paper applies Huber's results.

2.1 Asymptotic "First-order conditions"

Suppress right now the term $1/n$ in $S_n(\beta)$. The right-partial derivative of $S_n(\beta)$ with respect to β_j, $j = 1, \ldots, k$ is given by:

$$\begin{aligned}\psi_j(\beta) &= \sum_{i=1}^{n}\Big[-x_{ij}\mathbb{1}\{y_i \geq x_i'\beta\} + x_{ij}\mathbb{1}\{y_i < x_i'\beta\}\Big]\mathbb{1}\{x_i'\beta > 0\} \\ &= 2\sum_{i=1}^{n} x_{ij}\left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\}\end{aligned}$$

Let e_j be the k-dimensional vector that places 1 in the j^{th} place and zero otherwise. Now comes a key fact: Since $\widehat{\beta}$ minimizes (1), the following must be true for each $j = 1, \ldots, k$ and any $\delta > 0$:

$$\psi_j(\widehat{\beta} - \delta e_j) \leq 0, \quad \psi_j(\widehat{\beta} + \delta e_j) \geq 0$$

$$\psi_j(\widehat{\beta} - \delta e_j) \leq \psi_j(\widehat{\beta}) \leq \psi_j(\widehat{\beta} + \delta e_j)$$

Therefore $\left|\psi_j(\widehat{\beta})\right| \leq \psi_j(\widehat{\beta}+\delta e_j) - \psi_j(\widehat{\beta}-\delta e_j)$. Therefore, for any $\delta > 0$ and all $j = 1, \ldots, k$:

$$\left|\psi_j(\widehat{\beta})\right| \leq$$
$$2\sum_{i=1}^{n}\left(x_{ij}\left[\mathbb{1}\{y_i < x_i'\widehat{\beta} + \delta x_{ij}\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\widehat{\beta} + \delta x_{ij} > 0\} - x_{ij}\left[\mathbb{1}\{y_i < x_i'\widehat{\beta} - \delta x_{ij}\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\widehat{\beta} - \delta x_{ij} > 0\}\right)$$

let's examine the components of this sum and take the limit as $\delta \to 0$

$$\lim_{\delta\to 0}\sum_{i=1}^{n} x_{ij}\left[\mathbb{1}\{x_i'\widehat{\beta} - \delta x_{ij} > 0\} - \mathbb{1}\{x_i'\widehat{\beta} + \delta x_{ij} > 0\}\right] = \sum_{i=1}^{n} x_{ij}\mathbb{1}\{x_i'\widehat{\beta} = 0\}$$

and since $y_i \geq 0$ for all i (by the censored-nature of the problem)

$$\begin{aligned}&\lim_{\delta\to 0}\sum_{i=1}^{n} x_{ij}\left[\mathbb{1}\{y_i < x_i'\widehat{\beta} + \delta x_{ij}\}\mathbb{1}\{x_i'\widehat{\beta} + \delta x_{ij} > 0\} - \mathbb{1}\{y_i < x_i'\widehat{\beta} - \delta x_{ij}\}\mathbb{1}\{x_i'\widehat{\beta} - \delta x_{ij} > 0\}\right] \\ &= \lim_{\delta\to 0}\sum_{i=1}^{n} x_{ij}\left[\mathbb{1}\{y_i < x_i'\widehat{\beta} + \delta x_{ij}\} - \mathbb{1}\{y_i < x_i'\widehat{\beta} - \delta x_{ij}\}\right] = \sum_{i=1}^{n} x_{ij}\mathbb{1}\{y_i = x_i'\widehat{\beta}\}\end{aligned}$$

Therefore,

$$\left|\psi_j(\widehat{\beta})\right| \leq 2\sum_{i=1}^{n} x_{ij}\left[\mathbb{1}\{y_i = x_i'\widehat{\beta}\} + \frac{1}{2}\mathbb{1}\{x_i'\widehat{\beta} = 0\}\right] \leq 2\sum_{i=1}^{n}|x_{ij}|\left[\mathbb{1}\{y_i = x_i'\widehat{\beta}\} + \frac{1}{2}\mathbb{1}\{x_i'\widehat{\beta} = 0\}\right] \tag{2}$$

now let $\psi(\widehat{\beta}) = \big(\psi_1(\widehat{\beta}), \ldots, \psi_k(\widehat{\beta})\big)'$ be the entire vector of right-hand derivatives. So

$$\psi(\beta) = 2\sum_{i=1}^{n} x_i \left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\}$$

Using the bound in Equation (2), we have

$$\left\|\psi(\widehat{\beta})\right\| \leq 2\max_{1\leq i\leq n}\|x_i\| \times \sum_{i=1}^{n}\left[\mathbb{1}\{y_i = x_i'\widehat{\beta}\} + \frac{1}{2}\mathbb{1}\{x_i'\widehat{\beta} = 0\}\right].$$

Given Assumption (R2) and the strong consistency of $\widehat{\beta}$, we have $\Pr\big[x_i'\widehat{\beta} = 0$ for at least one $i]$ goes to zero as $n \to \infty$. Assumption (R2) and the continuously distributed nature of u_i also imply that $\Pr\big[y_i = x_i'\widehat{\beta}\big]$ for at least one i goes to zero as $n \to \infty$. The paper's assumptions also ensure that $\max_{1\leq i\leq n}\|x_i\| = O_p(1)$. All these facts together imply that

$$\left\|\frac{1}{\sqrt{n}}\psi(\widehat{\beta})\right\| \leq \frac{2}{\sqrt{n}}\max_{1\leq i\leq n}\|x_i\| \times \sum_{i=1}^{n}\left[\mathbb{1}\{y_i = x_i'\widehat{\beta}\} + \frac{1}{2}\mathbb{1}\{x_i'\widehat{\beta} = 0\}\right] = o_p(1).$$

Therefore $\left\|\frac{1}{\sqrt{n}}\psi(\widehat{\beta})\right\| = o_p(1)$. **These are the asymptotic "first-order conditions" that will fit into Huber's Lemma to establish asymptotic normality of $\widehat{\boldsymbol{\beta}}$.** Using Huber's notation, let $\lambda(\beta) = E\big[\psi(\beta)/n\big]$. Note that

$$\lambda(\beta) = \frac{2}{n}\sum_{i=1}^{n} E_{x_i}\left[x_i\left(F(x_i'(\beta - \beta_0)) - \frac{1}{2}\right)\mathbb{1}\{x_i'\beta > 0\}\right]$$

and therefore $\lambda(\beta_0) = 0$. To simplify matters, and as I mentioned before, we will assume that the x_i's are iid and not just independent. (Lemma A.3 in Powell points out that Huber's Lemma 3 extends immediately and automatically to the case in which the covariates are only independent, but not identically distributed without further assumptions simply by generalizing $\lambda(\beta)$ to $(1/n)\sum_{i=1}^{n} E[\psi(x_i, \beta)]$.). So $\lambda(\beta)$ simply becomes

$$\lambda(\beta) = 2E_x\left[x\left(F(x'(\beta - \beta_0)) - \frac{1}{2}\right)\mathbb{1}\{x'\beta > 0\}\right]$$

Asymptotic normality will be established by showing that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \beta_0) + \sqrt{n}\lambda(\widehat{\beta}) = o_p(1)$$

which is achieved by showing that the assumptions made by the paper are sufficient for the conditions in Lemma 3 of Huber to hold.

Inside the neighborhood characterized in Assumption (R2), we have

$$\nabla_\beta \lambda(\beta) \equiv \Lambda(\beta) = 2E_x\Big[xx'f(x'(\beta-\beta_0))\mathbb{1}\{x'\beta>0\}\Big] \quad \text{if the covariates are iid}$$
$$= \frac{2}{n}\sum_{i=1}^n E_{x_i}\Big[x_i x_i' f(x_i'(\beta-\beta_0))\mathbb{1}\{x_i'\beta>0\}\Big] \quad \text{if the covariates are only independent (Powell's setting)}$$

Dominated convergence and (R2) yield: $\Lambda(\widehat{\beta}) \xrightarrow{p} \Lambda(\beta_0) = f(0)E_x\Big[xx'\mathbb{1}\{x'\beta_0>0\}\Big]$. We already know the punchline: If the remaining conditions in Huber, Lemma 3 are satisfied then we will have

$$\sqrt{n}(\widehat{\beta}-\beta_0)$$
$$= \Big(f(0)E_x\Big[xx'\mathbb{1}\{x'\beta>0\}\Big]\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i\Big[\mathbb{1}\{y_i<x_i'\beta_0\}-\frac{1}{2}\Big]\mathbb{1}\{x_i'\beta_0>0\}+o_p(1)$$

(the number 2 we have been carrying around drops out here) the final version of this handout will provide further details of how Assumptions (N-3)(i-iii) in Huber are satisfied here.

Before proceeding, let us state the rest of the assumptions in the paper (in addition to those stated above already)

- P.2 β_0 is an interior point of B.

- E.2 Summary of assumptions about $f(\cdot)$ (the density of the error term u): $F(0)=1/2$, $f(0)>0$, $f(u)<f_0<\infty$ for all u; $\big|f(u_1)-f(u_2)\big| \le L_0\big|u_1-u_2\big|$ for all u_1,u_2 and some $L_0>0$ ($f(\cdot)$ is bounded and Lipschitz continuous everywhere)

By Assumption R.2, the following second-order Taylor approximation is valid in a neighborhood around β_0:

$$\lambda(\beta) = \underbrace{\lambda(\beta_0)}_{=0} + \Lambda(\beta_0)(\beta-\beta_0) + \big\|\beta-\beta_0\big\|^2\underbrace{O(1)}_{\text{by R.2}}$$

where

$$\Lambda(\beta_0) = \begin{cases} f(0)E_x\Big[xx'\mathbb{1}\{x'\beta_0>0\}\Big] & \text{if covariates are iid} \\ f(0)\frac{1}{n}\sum_{i=1}^n E_{x_i}\Big[x_i x_i'\mathbb{1}\{x_i'\beta_0>0\}\Big] & \text{if covariates are only independent} \end{cases}$$

Now, assumption R.1 kicks in, since (combined with R.2 and the assumption $f(0) > 0$) it implies that there exists a strictly positive constant c_0 such that $\|\Lambda(\beta_0)\| > c_0$. Therefore, in a neighborhood of β_0, we have: $\|\lambda(\beta)\| > c_0\|\beta - \beta_0\|$. **This satisfies condition (N-3)(i) in Huber**. Next we have to verify if there exists a d_0 b and c such that for all $d \geq 0$:

$$E\left[\sup_{\|\beta-\tau\|+d\leq d_0}\left\|x_i\left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\} - x_i\left[\mathbb{1}\{y_i < x_i'\tau\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\tau > 0\}\right\|\right] \leq b\cdot d$$
$$E\left[\sup_{\|\beta-\tau\|+d\leq d_0}\left\|x_i\left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\} - x_i\left[\mathbb{1}\{y_i < x_i'\tau\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\tau > 0\}\right\|^2\right] \leq c\cdot d \tag{3}$$

Notice that for all β, τ:

$$x_i\left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\} - x_i\left[\mathbb{1}\{y_i < x_i'\tau\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\tau > 0\}$$
$$= \begin{cases} 1/2 & \text{if } x_i'\beta > 0 \text{ and } x_i'\tau \leq 0 \text{ or } x_i'\beta \leq 0 \text{ and } x_i'\tau > 0 \\ 1 & \text{if } 0 < x_i'\beta < y_i \leq x_i'\tau \text{ or } 0 < x_i'\tau < y_i \leq x_i'\beta \\ 0 & \text{otherwise} \end{cases}$$

Notice that

$$\mathbb{1}\left\{\text{if } x_i'\beta > 0 \text{ and } x_i'\tau \leq 0 \text{ or } x_i'\beta \leq 0 \text{ and } x_i'\tau > 0\right\} \leq \mathbb{1}\left\{|x_i'\beta| \leq |x_i'\beta - x_i'\tau|\right\}$$
$$\mathbb{1}\left\{\text{if } 0 < x_i'\beta < y_i \leq x_i'\tau \text{ or } 0 < x_i'\tau < y_i \leq x_i'\beta\right\} \leq \mathbb{1}\left\{|y_i - x_i'\beta| < |x_i'\beta - x_i'\tau|\right\}$$

Therefore

$$\left\|x_i\left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\} - x_i\left[\mathbb{1}\{y_i < x_i'\tau\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\tau > 0\}\right\|$$
$$\leq \|x_i\| \times \left(\frac{1}{2}\mathbb{1}\left\{|x_i'\beta| \leq |x_i'\beta - x_i'\tau|\right\} + \mathbb{1}\left\{|y_i - x_i'\beta| < |x_i'\beta - x_i'\tau|\right\}\right)$$
$$\leq \|x_i\| \times \left(\frac{1}{2}\mathbb{1}\left\{|x_i'\beta| \leq \|x_i\|\cdot\|\beta - \tau\|\right\} + \mathbb{1}\left\{|y_i - x_i'\beta| < \|x_i\|\cdot\|\beta - \tau\|\right\}\right)$$

which means that

$$\sup_{\|\beta-\tau\|+d\leq d_0}\left\|x_i\left[\mathbb{1}\{y_i < x_i'\beta\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\beta > 0\} - x_i\left[\mathbb{1}\{y_i < x_i'\tau\} - \frac{1}{2}\right]\mathbb{1}\{x_i'\tau > 0\}\right\|$$
$$\leq \|x_i\| \times \left(\frac{1}{2}\mathbb{1}\left\{|x_i'\beta| \leq \|x_i\|\cdot(d + d_0)\right\} + \mathbb{1}\left\{|y_i - x_i'\beta| < \|x_i\|\cdot(d + d_0)\right\}\right)$$

By Assumption R.2 and choosing $d + d_0 < \xi_0$, we verify that Conditions (3) are satisfied. Thus, all the assumptions of Huber are satisfied and we obtain:

$$\sqrt{n}(\widehat{\beta} - \beta_0) =$$

$$\underbrace{\Big(f(0)E_x\Big[xx'f(\mathbb{1}\{x'\beta > 0\}\Big]\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\Big[\mathbb{1}\{y_i < x_i'\beta_0\} - \frac{1}{2}\Big]\mathbb{1}\{x_i'\beta_0 > 0\} + o_p(1)}_{\text{if covariates are iid}}$$

$$\underbrace{\Big(f(0)\frac{1}{n}\sum_{i=1}^{n} E_{x_i}\Big[x_i x_i'\mathbb{1}\{x_i'\beta > 0\}\Big]\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\Big[\mathbb{1}\{y_i < x_i'\beta_0\} - \frac{1}{2}\Big]\mathbb{1}\{x_i'\beta_0 > 0\} + o_p(1)}_{\text{if covariates are only independent}}$$