28 May 2009
16:50

$y_i = g(x_i, \beta_0, \varepsilon_i)$ $\quad \Pr\{\varepsilon_i \geq 0 \mid x_i\} \geq \tau$

$\Pr\{\varepsilon_i \leq 0 \mid x_i\} \geq 1-\tau$

$g(\ )$ increasing in ε (nondecreasing)

$\varepsilon \geq 0 \Rightarrow g(x_i, \beta_0, \varepsilon) \geq g(x_i, \beta_0, 0)$

$\Pr\{\cdot \mid x_i\} \geq \tau$

Quantile minimizes $E[\rho_\tau(y - g(x, \beta, 0))]$

$\rho_\tau(u) = |u|\,|\tau - 1(u \leq 0)|$

$\tau = 1/2 \qquad \mathrm{med}\{\varepsilon_i \mid x_i\} = 0$

$\beta_0 = \arg\min E[\,|y_i - g(x_i, \beta, 0)|\,]$

If $\Pr\{\varepsilon_i \leq \alpha_0 \mid x_i\} = 1/2$

Binary Response: $\quad y_i = 1(x_i'\beta_0 + \varepsilon_i \geq 0)$

$\Pr\{\varepsilon_i \geq 0 \mid x_i\} = 1/2$

$$\hat\beta_{MS} = \arg\max \sum_{i=1}^{N} \{ y_i \cdot 1(x_i'\beta > 0) + (1-y_i)\cdot 1(x_i'\beta \leq 0)\}$$

$$= \arg\min \sum_{i=1}^{N} |y_i - 1(x_i'\beta > 0)|$$

Conditions: (1) $\Pr\{\mathrm{sgn}(x_i'b) \neq \mathrm{sgn}(x_i'\beta_0)\} > 0$ if $\beta \neq \beta_0$

$\Rightarrow \|\beta_0\| = 1$ or $\beta_0 = \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix}$

$x_i'\beta$ cts near zero, $\beta \in B$

② $\varepsilon_i | x_i \sim f_{\varepsilon|x}(u|x_i)\,du \qquad f_{\varepsilon|x}(0|x_i) \geq c_0 > 0$

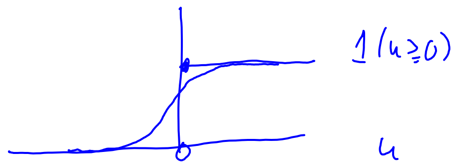
$$\hat{\beta}_{MS} \xrightarrow{p} \beta_0$$

Horowitz "SMOOTHED" MAX SCORE:

$$\hat{\beta}_{sms} = \underset{\beta \in B}{\arg\max} \sum_{i=1}^{N} y_i H\left(\frac{x_i'\beta}{h}\right) + (1-y_i)\left(1 - H\left(\frac{x_i'\beta}{h}\right)\right)$$

$$H(u) = \int_{-\infty}^{u} K(v)\,dv \quad , \quad \int_{-\infty}^{\infty} K(v)\,dv = 1$$

$$h = h_N \to 0 \quad \text{as} \quad N \to \infty$$



F.O.C. $\quad \beta = \begin{pmatrix} 1 \\ \gamma \end{pmatrix} \quad x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}$

$$0 = \sum_{i=1}^{N} (1 - 2y_i)\, K\left(\frac{x_i'\hat{\beta}}{h}\right) x_{i2}$$

rate of convergence is $\sqrt{Nh}$

Censored Regression: $\quad y_i = \max\{0, x_i'\beta_0 + \varepsilon_i\}$

$$\text{med}\{\varepsilon_i | x_i\} = 0 \Rightarrow \text{med}\{y_i | x_i\} = \max\{0, x_i'\beta_0\}$$

$$\hat{\beta}_{CLAD} = \underset{\beta \in B}{\arg\min} \sum_{i=1}^{N} \left| y_i - \max\{0, x_i'\beta\} \right|$$

Approx F.O.C.: $\quad 0 \cong \sum_{i=1}^{N} 1(x_i'\hat{\beta} > 0)\ \text{sgn}(y_i - x_i'\hat{\beta}) \cdot x_i$

CONDITIONS INCLUDE

$$C_0 = E\left[f_{\varepsilon|x}(0|x_i)\, x_i x_i' \cdot 1(x_i'\beta_0 > 0) \right] \quad \text{p.d.}$$

$$V_0 = E\left[1(x_i'\beta_0 > 0)\, x_i x_i' \right]$$

$$\sqrt{N}\left(\hat{\beta}_{CLAD} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \tfrac{1}{4} C_0^{-1} V_0 C_0^{-1'}\right)$$

$$\uparrow \pi(1-\pi) \qquad \pi = 1/2$$

$$\hat{V} = \frac{1}{N}\sum_i 1(x_i'\hat{\beta} > 0)\, x_i x_i'$$

$$\hat{C} = \frac{1}{N}\sum_i 1(x_i'\hat{\beta} > 0)\, \underbrace{1(x_i'\hat{\beta} < y_i < x_i'\hat{\beta} + h)}_{h}\, x_i x_i'$$

$$\hat{V} \xrightarrow{p} V_0, \quad \hat{C} \xrightarrow{p} C_0 \qquad \sqrt{N} h_N \to \infty$$

CAN DO OTHER MONOTONIC MODELS, BOX-COX + CENSOR

<u>Symmetry Given x_i</u>: $\quad \varepsilon_i | x_i \sim -\varepsilon_i | x_i$

$y_i = g(x_i, \beta_0, \varepsilon_i)$ if $G(y_i; x_i, \beta_0)$

$= G(g(x_i, \beta_0, \varepsilon_i); x_i, \beta_0)$

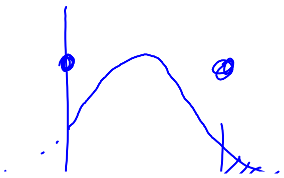
is an odd function of ε

$G(g(\cdot, \varepsilon)) = -G(g(\cdot, -\varepsilon))$, then

$$E\left[G(y_i; x_i, \beta_0) \mid x_i\right] = 0$$

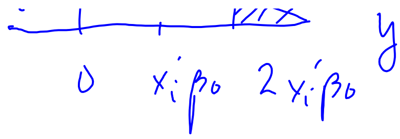
<u>Ex</u>: $y_i = x_i'\beta_0 + \varepsilon_i$, $\quad G(y_i, x_i, \beta_0) = y_i - x_i'\beta_0$

If $y_i = \max\{0, x_i'\beta_0 + \varepsilon_i\}$



$\psi(y_i, x_i, \beta_0) = 1(x_i'\beta_0 > 0) \cdot$

$\min(y_i, 2x_i'\beta_0) - x_i'\beta_0$

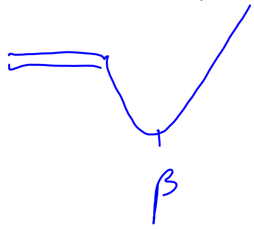


$\hat\beta_{SCLS}$ solves $\frac{1}{N}\sum_{i=1}^{N} 1(x_i'\hat\beta > 0)\left(\min\{y_i, 2x_i'\hat\beta\} - x_i'\hat\beta\right) x_i$

$= 0$

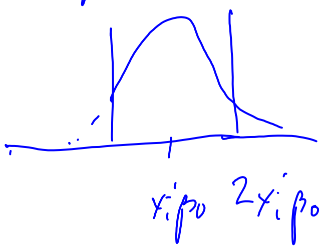
$\hat\beta_{SCLS} = \arg\min_{\beta} S_N(\beta)$ for some $S_N(\beta)$

$S_N(\beta) \xrightarrow{P} \bar S(\beta) > \bar S(\beta_0)$ if $\beta \neq \beta_0$



Truncated Regression $y_i = x_i'\beta_0 + \varepsilon_i$ conditional on $y_i > 0$

ie $y_i = x_i'\beta_0 + v_i$, $v_i \sim \varepsilon_i \mid \varepsilon_i > -x_i'\beta_0$



$$0 = \frac{1}{N}\sum_i 1(0 < y_i < 2x_i'\hat\beta)(y_i - x_i'\hat\beta)x_i$$

$$\hat\beta = \arg\min \sum_i \left(y_i - \max\{\tfrac{y_i}{2}, x_i'\beta\}\right)^2$$

Honoré-Kyriazidou-Udry:

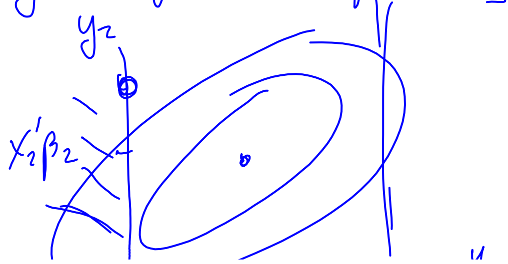
$y_{i1} = \max\{0, x_{i1}'\beta_1 + \varepsilon_{i1}\}$

$y_{i2} = 1(y_{i1} > 0)\cdot[x_{i2}'\beta_2 + \varepsilon_{i2}]$

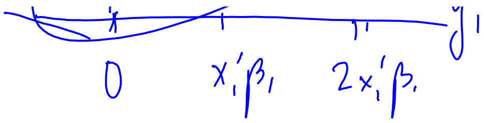
$y_{i1} = \max\{0, y_{i1}^*\}$

$y_{i2} = 1(y_{i1}^* > 0)\cdot y_{i2}^*$

$(\varepsilon_{i1}\ ) \mid x_i \sim (\varepsilon_{i1}\ ) \mid x_i$



$\left(e_{i2}\right)$ | ... $\left(\varepsilon_{i2}\right)$ | ...



$$0 = \frac{1}{N}\sum_{i=1}^{N} 1\left(0 < y_{i1} < 2x_{i1}'\hat{\beta}_1\right)\left(y_{i2} - x_{i2}'\hat{\beta}_2\right)X_{i2}$$

Independence: $\varepsilon_i \perp x_i \qquad \varepsilon_i$ iid

$$\varepsilon_i - \varepsilon_j \mid x_i, x_j \sim \varepsilon_j - \varepsilon_i \mid x_i, x_j$$

Binary Response: $y_i = 1\left(x_i'\beta_0 + \varepsilon_i \geq 0\right)$

Avg Log Likelihood: $\mathcal{L}_N(\beta) = \frac{1}{N}\sum_i y_i \log F(x_i'\beta_0) + (1-y_i)\log\left(1-F(x_i'\beta_0)\right)$

$\Pr\{\varepsilon_i \leq \lambda\} = F(\lambda)$

$\mathcal{L}_N(\beta) = \mathcal{L}_N(\beta; F(\cdot)) \qquad \beta \in \mathcal{B} = \{\beta : \|\beta\| = 1\}$

$F(\cdot) \in \mathcal{F} = \{H(u) :$ $H(u)$ nondecreasing, $\lim_{u\to-\infty} H(u) = 0$, $\lim_{u\to\infty} H(u) = 1\}$