Lecture 3

Methods of Nonlinear Estimation

(revised, 2010)

Read

- Wooldridge ch12, ch13, ch14 (also, p640-659, 661-670)
- Green ch11.8.5, 16.5.2
- Hayashi ch5
- CT 5.7, 5.8, 6.5, 9

Ec 671

Lee

Extremum estimator

An extremum estimator is one which is obtained as the optimizer of a criterion function $q(\theta \mid \text{data})$.

eg) $\left\{ \begin{array}{lll} \text{LS} & \hat{\theta} = \arg\max(-\hat{e}'\hat{e}) & \text{where } \hat{e}_i = y_i - h(x_i, \hat{\theta}) \\ & & \text{(linear or nonlinear } h) \\ \text{MLE} & \hat{\theta} = \arg\max \frac{1}{n}\sum \ln f(y_i \mid x_i, \theta) & \end{array} \right. \leftarrow M$

$\text{GMM} \quad \hat{\theta} = \arg\max(-g(y_i, x_i, \theta)' \cdot W \cdot g(y_i, x_i, \theta))$ ← GMM MDE

That is, there is a scalar objective function $Q_n(\theta)$.

Three classes of extremum estimators

i) M-estimators

the obj. function is a sample average

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(y_i, x_i, \theta)$$

eg) MLE, LS (nonlinear LS is also included)

ii) GMM

the obj. function is not a sample mean

$$Q_n(\theta) = -\frac{1}{2} g_n(\theta)' \cdot W \cdot g_n(\theta)$$

where $\underset{(k\times 1)}{g_n(\theta)} = \frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i, \theta)$

W $(k\times k)$ is a pd matrix that defines the distance of $g_n(\theta)$ from zero.

- Max $Q_n(\theta)$ $\underset{\text{same}}{\Leftrightarrow}$ min the distance $g_n(\theta)' \cdot W \cdot g_n(\theta)$
- Here, $g_n(\theta)$ is a sample mean of $g(y_i, x_i, \theta)$

iii) Minimum distance estimator

Max $Q_n(\theta) = -\frac{1}{2} g_n(\theta)' W \cdot g_n(\theta)$

but [$g_n(\theta)$ is <u>not</u> a sample mean.] ☆
[W is <u>ANY</u> p.d. weighting matrix.]

(thus, GMM is a special case of the minimum distance estimator.

- $g_n(\theta)$ is a sample mean
- $W = Var(g_n(\theta))^{-1}$ is used)

<u>Minimum distance estimator (and GMM)</u>

$$\text{Min } Q_n(\theta) = \frac{1}{2} g_n(\theta)' W g_n(\theta) \; ; \quad \begin{cases} \theta = p \times 1 \\ g(\theta) = L \times 1 \end{cases}$$

where W is <u>ANY</u> p.d. (weight) matrix

ex) choose $g_n(\theta) = z'e$ for an example with a linear model.

$$\text{Min } Q(\theta) = \frac{1}{2} (z'e)' W (z'e)$$

where $y = X\beta + e$, $z = IV$ for X

$\Phi = Var(z'e) = E[z'ee'z]$ = any form: $\begin{cases} \sigma^2 z'z & \text{homoskedas} \\ z'\Omega z & \text{heterosk} \\ \text{HC variance} \end{cases}$

then

$$Q = \frac{1}{2} (\beta' x' - y') z \cdot W \cdot z' (y - X\beta)$$

$$\frac{\partial Q}{\partial \beta} = x'zWz'y - x'zWz'X\beta \overset{\text{set}}{\Rightarrow} 0$$

$$\hat{\beta} = (x'z\hat{W}z'x)^{-1} x'z\hat{W}z'y$$

$$Var(\hat{\beta}) = E(x'zWz'x)^{-1} x'zWz'\underbrace{ee'}z Wz'x (x'zWz'x)^{-1}$$

$$= (x'zWz'x)^{-1} x'zW \Phi Wz'x (x'zWz'x)^{-1} \qquad (*)$$

Note Suppose $g_n(\theta)$ is nonlinear.

let $\dfrac{\partial g_n(\theta)}{\partial \theta'} = G(\theta) \overset{\text{let}}{=} G$

then, we have the variance of min. distance estimator:

see: (Wooldridge p. 423 eq (14.10). same equation) ⇒

$$\boxed{Var(\hat{\beta}) = (G'WG)^{-1} G'W \Phi W G (G'WG)^{-1}}$$

eg) if $g(\theta) = z'e = z'(y - X\beta)$

$$G = \frac{\partial g(\theta)}{\partial \theta'} = \frac{\partial g}{\partial \beta'} = z'X \Rightarrow$$ p. 2 (*)

Note If $W = \Phi^{-1}$ where $\Phi = Var(g(\theta)) = Var(z'e)$

(ie. optimal IV: GMM) $= z'\Omega z$

$$Var(\hat{\beta}) = (G'\Phi^{-1}G)^{-1} G' \cancel{\Phi^{-1}} \cancel{\Phi} \Phi^{-1} G (G'\Phi^{-1}G)^{-1}$$

$$\boxed{= (G'\Phi^{-1}G)^{-1}}$$

... see lecture 2, p. 21
$Var(\hat{\beta}) = (D'W^{-1}D)^{-1}$ (different notations)
(GMM estimator)

eg) if $g(\theta) = z'e = z'(y - X\beta)$

$$G = \frac{\partial g(\theta)}{\partial \theta'} = z'X \quad (\text{where } \theta = \beta)$$

$\Phi = var(g(\theta)) = var(z'e) = \sigma^2(z'z)$ under homoskedasticity

$$Var(\hat{\beta}) = [(z'X)'\sigma^{-2}(z'z)^{-1}(z'X)]^{-1}$$
$$= \sigma^2 (X'z(z'z)^{-1}z'X)^{-1}$$
$$= \sigma^2 (X'P_z X)^{-1} \quad \text{as in 2SLS}$$

Note If $L = K$ (# of moment conditions = # of parameters),

the optimal $W = I$.

Note which is more efficient?

$$\begin{cases} \text{any } W \Rightarrow var(\hat{\beta}) = (G'WG)^{-1} G'W \Phi WG (G'WG)^{-1} \\ W = \Phi^{-1} \Rightarrow var(\hat{\beta}) = (G'\Phi^{-1}G)^{-1} \end{cases}$$

where $G = \frac{\partial g(\theta)}{\partial \theta'}$, $\Phi = var(g(\theta))$

Exercise) Wooldridge ex 8.5 (p. 206) H/W

(also see Hayashi, p. 213)

Show that

$$(G'\Phi^{-1}G) - (G'WG)(G'W\Phi WG)^{-1}(G'WG) \text{ is psd.}$$

$[A^{-1} - B^{-1} \text{ is psd} \Leftrightarrow B - A \text{ is psd}]$

Hint $G'\Phi^{-\frac{1}{2}}[I_L - D(D'D)^{-1}D']\Phi^{-\frac{1}{2}}G$

where $D = \Phi^{\frac{1}{2}}WG$

Thus, using $W = \Phi^{-1}$ (weighting matrix $= var(g(\theta))^{-1}$) leads to the GMM estimator, which is more efficient than other estimators using any p.d. W.

GMM

Thus, GMM = Min. distance estimator which uses:

(i) $W = \Phi^{-1}$

(ii) $g(\theta)$ is a sample mean.

Choice of Φ^{-1} ($var(g(\theta))^{-1}$)

(a) $\Phi^{-1} = (\sigma^2(Z'Z))^{-1}$ linear, homoskedasticity

(b) $\Phi^{-1} = (Z'\Omega Z)^{-1}$ where $var(e) = \Omega$ under heteroskedasticity

GLS type

(c) $\Phi^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i \hat{u}_i \hat{u}_i' z_i'\right)^{-1}$ Heteroskedasticity-consistent variance

- Over-identifying restriction tests

$$\widehat{\delta_n}(\theta) = \left(\frac{1}{\sqrt{n}}\sum_i^n g_i(\hat{\theta})\right)' \hat{\Phi}^{-1} \left(\frac{1}{\sqrt{n}}\sum_i^n g_i(\hat{\theta})\right)$$

$\sim \chi^2_{L-K}$: Hansen's over-identifying rest. test (J-statistic)

this expression is used to test validity of the moment conditions. (Read Hayashi, p.218 for weakness)

Note the optimal GMM estimator (which uses $\hat{\Phi}^{-1}$ as W) is called the "minimum chi-square estimator" since the above has a chi-square distribution

- GMM distance statistic (GMM criterion function statistic)

$H_0: \ f(\theta) = 0$ (linear or nonlinear)

We can have an unrestricted GMM estimator, $\hat{\theta}$. as well as a restricted GMM estimator (that imposes the null), $\tilde{\theta}$.

Using the same weighting matrix, $\hat{\Phi}^{-1}$ (from unrestricted estimator)

$$\frac{1}{n}\left(\sum g_i(\tilde{\theta})\right)' \hat{\Phi}^{-1} \left(\sum g_i(\tilde{\theta})\right) - \frac{1}{n}\left(\sum g_i(\hat{\theta})\right)' \hat{\Phi}^{-1} \left(\sum g_i(\hat{\theta})\right) \sim \chi^2_m$$

where m = # of restrictions.

(like LR type test) "quasi-LR" Wooldridge, p370

* advantage over Wald test (Hayashi, p.223)

this LR type test is invariant to different types of nonlinear restrictions: see Wald test in the (MLE) note.

Note Sargan's statistic (in 2SLS)

under homoskedasticity in a linear model

$var(g(\theta)) = \sigma^2 (Z'Z)$,

using $\Sigma^{-1} = [\sigma^2 (Z'Z)]^{-1}$ gives

$$\text{Sargan's stat} = n \frac{(Z'\hat{e})'(Z'Z)^{-1}(Z'\hat{e})}{\hat{\sigma}^2} \sim \chi^2_{L-K}$$

Also, LR test can be used to test $R\beta = r$

$$n\frac{(Z'\tilde{e})'(Z'Z)^{-1}(Z'\tilde{e})}{\hat{\sigma}^2} - \frac{n(Z'\hat{e})'(Z'Z)^{-1}(Z'\hat{e})}{\hat{\sigma}^2} \sim \chi^2_m$$

where the same error variances $\hat{\sigma}^2$ are used.

Other estimators

· LAD (least absolute deviation) estimator solves

Wooldridge, p348

$$\text{Min } \frac{1}{n} \Sigma | y_i - f(x_i, \theta) |$$

... Median estimator (robust to outliers, non-normal dist. not symmetric.)

... "quantile" regression Koenker & Bassett (1978)

"distribution of wages"

; factors that affect 75percentile, 25 percentile of the income group, for instance.

(more on this, later)

M-estimation

Huber (1967) : "M" for max or min

obj: Min $\frac{1}{N}\sum_{i=1}^{N} q(w_i, \theta)$ .. sample mean of $q(\cdot)$

eg1) $y_i = m(x_i, \theta) + e_i$

; nonlinear model

say, $m(x_i, \theta) = \exp(x_i\theta)$ or $\frac{\exp(x_i\theta)}{1+\exp(x_i\theta)}$

then, $q(w_i, \theta) = (y_i - m(x_i, \theta))^2$ with $w_i = (y_i, x_i)$

$\Rightarrow$ Nonlinear LS

eg2) MLE

$q(w_i, \theta) = -\log \mathcal{L}_i = -\log f(y_i)$

Under certain regularity conditions, (compact, continuous, identification)

$$\text{Max}\left| \frac{1}{N}\sum q(w_i, \theta) - E[q(w, \theta)] \right| \xrightarrow{p} 0$$

"uniform weak law of large numbers"

and

$\hat{\theta} \xrightarrow{p} \theta_0$; consistent

Some issues: identification & nuisance parameters

Score vector

FOC: $\partial q(w,\theta)/\partial\theta \overset{let}{=} s(w,\theta) \Rightarrow$ set 0

$$\sum_{i=1}^{N} s(W_i,\hat\theta) = 0 \quad \Leftrightarrow \quad E[s(w,\theta)] = 0$$

Note $s(w_i,\hat\theta) = \begin{pmatrix} \partial q/\partial\hat\theta_1 \\ \vdots \\ \partial q/\partial\hat\theta_p \end{pmatrix} = p\times1$ vector

Note why zero?

Recall : Mean value theorem

$$f(x) = f(a) + f'(b)(x-a) \qquad \text{where } a<b<x$$

using this,

$$\sum_{i=1}^{N} s(w_i,\hat\theta) = \sum_{i=1}^{N} s(w_i,\theta) + \sum \underbrace{H(w_i,\theta)}_{\text{Hessian}} (\hat\theta-\theta_0)$$

$$= 0 \qquad\qquad \text{Hessian} = \frac{\partial^2 q(w,\theta)}{\partial\theta\,\partial\theta'} = p\times p$$

or

$$\frac{1}{\sqrt N}\sum_{i=1}^{N} s(w_i,\theta) + \frac{1}{N}\sum_{i=1}^{N} H(w_i,\theta)\sqrt N(\hat\theta-\theta_0)$$

$$\Rightarrow \sqrt N(\hat\theta-\theta_0) = \left[\frac{1}{N}\sum_{i=1}^{N} H(w_i,\theta_0)\right]^{-1}\left[-\frac{1}{\sqrt N}\sum_i S_i(\theta_0)\right]$$

$$\overset{let}{=} \frac{1}{\sqrt N}\sum_{i=1}^{N} r_i(\theta_0) + o_p(1) \qquad \text{where } \frac{1}{N}\sum H(w_i,\hat\theta) \xrightarrow{p} A_0$$

where $\boxed{r_i(\theta_0) \equiv -A_0^{-1} s_i(\theta_0)}$

ex) MLE $\quad \hat\theta = \theta_0 - \left(\frac{\partial^2 \ln L_i}{\partial\theta\,\partial\theta'}\right)^{-1}\left(\frac{\partial \ln L_i}{\partial\theta}\right) \Rightarrow \hat\theta-\theta_0 = -A_0^{-1}s_i(\theta_0)$

Thus

$$Var(\sqrt{N}\,\hat\theta) = Var\left(\frac{1}{\sqrt{N}} \Sigma r_i(\theta_0)\right)$$

$$= Var\left[A_0^{-1} \left(-\frac{1}{\sqrt{N}} \Sigma s_i(\theta_0)\right) \right]$$

$$= A_0^{-1} \underbrace{Var\left(\frac{1}{\sqrt{N}} \Sigma s_i(\theta_0)\right)}_{Var(s(w,\theta_0)) \overset{let}{=} B_0} A_0^{-1}$$

$$= A_0^{-1} B_0 A_0^{-1}$$

where $A_0 = E[H(w, \theta_0)]$

$B_0 = E[s(w,\theta_0) \cdot s(w,\theta_0)'] = Var(s(w,\theta_0))$

That is,

$$\sqrt{N}(\hat\theta - \theta_0) \longrightarrow N[0,\ A_0^{-1} B_0 A_0^{-1}]$$

.. asymptotic <u>normality</u> by CLT

.. sandwich estimate of the variance matrix

<u>Example</u> $y = \exp(X\beta) + u$

where $m(x,\beta) = \exp(X\beta)$

Min $q(x,\beta) = [y - \exp(X\beta)]^2/2$

score $s(x,\beta) = \dfrac{\partial q(x,\beta)}{\partial \beta} = -(\underbrace{e^{X\beta}x}_{\frac{\partial m(x,\beta)}{\partial\beta}})'[\underbrace{y - \exp(X\beta)}_{u = y - \exp(X\beta)}] \Rightarrow 0$

$B_0 = Var(\text{Score}) = E\left(\underbrace{u^2}_{\text{as in robust variance}} \left(\frac{\partial m(x,\beta)}{\partial\beta}\right)' \left(\frac{\partial m(x,\beta)}{\partial\beta}\right)\right] = *$

Hessian

$$H(x,\beta) = \frac{\partial m(x,\beta)}{\partial \beta}' \frac{\partial m(x,\beta)}{\partial \beta} - \frac{\partial^2 m(x,\beta)}{\partial \beta \partial \beta'} \cdot u$$

$$= e^{2x\beta} x'x - e^{x\beta} x'x \cdot u$$

$$A_0 = E[H(x,\beta)] = e^{x\beta} x'x$$

$$Var(\hat{\beta}) = \frac{1}{N} A_0^{-1} B_0 A_0^{-1}$$

which can be estimated by

$$\left(\sum_i^N \hat{A}_i\right)^{-1} \left(\sum_i^N \hat{s}_i \hat{s}_i'\right) \left(\sum_i^N \hat{A}_i\right)^{-1}$$

or

$$\left(\sum \hat{H}_i\right)^{-1} \left(\sum \hat{s}_i \hat{s}_i'\right) \left(\sum \hat{H}_i\right)^{-1}$$

using $\hat{s}_i = -(e^{x_i\hat{\beta}} x_i)'(y_i - e^{x_i\hat{\beta}})$

$\hat{H}_i$ and $\hat{A}_i$ can be defined accordingly.

Note If the model is correctly specified, under some regularity conditions, we have

$$E[s(w,\theta_0)\, s(w,\theta_0)'] = \sigma_0^2\, E[H(w,\theta_0)]$$

i.e. $B_0 = A_0$.

then

$$var(\hat{\beta}) = \frac{1}{N} A_0^{-1}$$

eg) $Var(\hat{\beta}) = \hat{\sigma}^2 \left[\sum \exp(2x_i\hat{\beta}) x_i' x_i\right]^{-1}$ in the previous example.

Two-step M-estimators

1st stage $\quad y_2 = h(z, \gamma) + \varepsilon \quad \Rightarrow \hat{\gamma}$

2nd stage $\quad y = m(W, \theta; \hat{\gamma}) + u \quad$; given $\hat{\gamma}$

$$\text{Min} \sum_{i=1}^{N} q(w_i, \theta; \hat{\gamma})$$

Examples ① Hausman test for endogeneity (also in nonlinear models)

$$y_1 = x_1\beta_1 + y_2\beta_2 + c\hat{e} + u$$

where $\hat{y}_2 = x_1\hat{c}_1 + z_2\hat{c}_2 + \hat{e}$ $\qquad \hat{\gamma} = (\hat{c}_1, \hat{c}_2)$

point u & e are correlated.

Thus $var(\hat{\beta})$ should be adjusted if $c \neq 0$.

② Selection Bias Model

$$y = x\beta + c\hat{\lambda} + u$$

$$\hat{\lambda} = \frac{\phi(z\hat{\gamma})}{\Phi(z\hat{\gamma})} \quad \text{from the probit model}$$

③ probit model with generated regressors

$$P = \Phi(X\beta + c\hat{z}) \quad \text{... example in lecture 3}$$

④* Weighted NLS

$$\text{Min} \sum (y_i - m(l_i, \theta))^2 / h(z_i, \hat{\gamma})$$

where $h(z_i, \hat{\gamma}) = z_i^2 \hat{\gamma}$

(ie if heteroskedasticity depends on z_i^2)

Properties of 2-step M-estimators

Consistent : Uniform WLLN
Asy. normal : CLT

Issues

i) $\hat{\theta}_{2\text{-step}}$ can be consistent in some cases even if $h(z, \gamma)$ is not correctly specified (ie $\hat{\gamma}$ is not consistent). In other cases, $\hat{\theta}$ is inconsistent

... weighted NLS is an example.

Note QMLE (Probit, Poisson, Fractional Logit models) using the sandwich estimator is good enough even if the model is not correctly specified (mean equation : its functional form)

ii) The standard error (ie. $\mathrm{var}(\hat{\theta})$) needs to be adjusted by the 1st stage estimation in most cases. But in some cases, the adjustment is not necessary (weighted NLS, for example)

Questions : when do we need to correct for it
How ?

Point : This is a generalization of the issue of generated regressors.

1st stage $\Rightarrow \hat{\gamma}$ $\quad y_2 = h(z, \gamma)$
2nd stage $\Rightarrow \hat{\theta}$ $\quad y = m(w, \theta; \hat{\gamma})$

From earlier note, we have

$$\sqrt{N}(\hat{\theta}-\theta_0) = A_0^{-1}\Big(-\frac{1}{\sqrt{N}}\sum s_i(\theta_0,\hat{\gamma})\Big) + o_p(1) \quad \cdots (1)$$

Here, the question is whether the following holds.

$$\frac{1}{\sqrt{N}}\sum s_i(\theta;\hat{\gamma}) = \frac{1}{N}\sum S_i(\theta_0,\gamma_0^*) + o_p(1) \quad \cdots (*)$$

where γ_0^* is the true value of γ

If this holds, $\hat{\theta}$ would be consistent.

Consider a Taylor expansion,

$$\frac{1}{\sqrt{N}}\sum s_i(\theta_0,\hat{\gamma}) = \frac{1}{N}\sum S_i(\theta_0,\gamma^*) + F_0\cdot\sqrt{N}(\hat{\gamma}-\gamma^*) + o_p(1)$$

where $F_0 = E\left[\frac{\partial S(w,\theta_0,\hat{\gamma})}{\partial\gamma}\right] \quad \cdots (**)$

point If either $\hat{\gamma}=\gamma^*$ or $F_0=0$, then $(*)$ holds

Note In the weighted LS, $F_0=0$ holds regardless of whether $h(z,\gamma)$ is correctly specified.

We may consider (as in $(\hat{\theta}-\theta_0)$ in the above

$$\sqrt{N}(\hat{\gamma}-\gamma^*) = A_{0\gamma}^{-1}\Big(-\frac{1}{\sqrt{N}}\sum S_i(\gamma^*)\Big) + o_p(1)$$

$$\overset{\text{let}}{=} \frac{1}{\sqrt{N}}\sum r_i(\gamma^*) + o_p(1) \quad \cdots (***)$$

where $r_i(\gamma^*) = A_{0\gamma}^{-1}(-S_i(\gamma^*))$

Hessian in the 1st stage

score in the 1st stage

plug this into $(**)$, and (1).

$$\sqrt{N}(\hat{\theta}-\theta_0) = A_0^{-1}\left(-\frac{1}{\sqrt{N}}\sum g_i(\theta_0,\gamma^*)\right) + o_p(1) \quad \cdots (2)$$

where

$$g_i(\theta_0,\gamma^*) = s_i(\theta_0,\gamma^*) + \underbrace{F_0 \cdot \underbrace{r_i(\gamma^*)}_{***}}_{**}$$

thus, comparing (1) & (2), we note that $s_i(\theta_0,\gamma^*)$ is replaced with $g_i(\theta_0,\gamma^*)$.

Then

$$\text{var}(\hat{\theta}) = \frac{1}{N} A_0^{-1} \cdot \underbrace{\text{var}(g(\theta_0,\gamma^*)}_{\text{let this} = D_0} \cdot A_0^{-1}$$

$$\text{var}(\hat{\theta}) = \frac{1}{N} A_0^{-1} D_0 A_0^{-1}$$

this can be estimated by

$$\hat{F} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial s_i(\hat{\theta},\hat{\gamma})}{\partial \hat{\gamma}}\right)$$

$$\hat{g}_i(\hat{\theta},\hat{\gamma}) = \hat{s}_i + \hat{F}\hat{r}_i \quad \text{with } \hat{r}_i = A_{0\gamma}^{-1} \cdot s_i(\hat{\gamma})$$

from the 1st stage

$$\hat{D} = \frac{1}{N}\sum \hat{g}_i \hat{g}_i'$$

Note
If the coeff. of the generated regressor is insignificant, $\hat{g}_i = \hat{s}_i$. Then, no adj. is needed; Hausman test

Hypotheses Tests

Wald, LM (score) or QLR tests.

$$QLR = 2\left[\sum q(w_i,\tilde{\theta}) - \sum q(w_i,\hat{\theta})\right]$$

$$= \left(\frac{1}{\sqrt{N}}\sum \tilde{s}_i\right)' A_0^{-1} \left(\frac{1}{\sqrt{N}}\sum \tilde{s}_i\right) + o_p(1)$$

$\cdots$ same $\tilde{s}_i$ restricted scores.

Median & Quantile Regression

CCT 4.6) Koenker & Bassett (1978, *Econometrica*)

- Minimizes the sum of ABSOLUTE errors.

$$\text{Min} \sum_{i=1}^{N} | y_i - x_i'\beta | \quad \ldots \text{ Median estimator}$$

* Quantile estimator (qth quantile)

$$\text{Min} \sum_{y_i \geq x_i\beta} q \, | y_i - x_i'\beta_q | + \sum_{y_i < x_i\beta} (1-q) \, | y_i - x_i'\beta_q |$$

.. For every value of q, we obtain different β_q.

... least absolute deviation estimator

.. Better than OLS in the presence of outliers (robust estimator)

- Motivation

ex) Engel curve

Median engel elasticity		= 0.906	(OLS = 0.909)
10th quantile	"	= 0.879	
90th "	"	= 0.946	

ex) Effect of reducing class size
; positive peer effects in the lower tail of the achievement dist.

ex) Effect of job training program
: different effects on low income vs high income workers

- why absolute values?

Suppose that there are 99 obs. (ordered..)

50th obs = 10

51st obs = 12

If we choose $\beta = 12$, rather than 10,

$\Sigma | y_i - \beta |$ will { increase by 2 for 50 obs / decrease by 2 for 49 obs

The net effect $= 50 \times 2 - 49 \times 2 = 2$

Thus, choosing $\beta = 12$ leads to a higher sum.

we need to choose the 50th obs (median)

$\beta = 10$ to minimize $\Sigma | y_i - \beta |$

- Estimation

FOC is not feasible. why?

Thus, all numerical optimizations are not feasible.

Instead, we use linear programming methods.

- Distribution

$$\sqrt{N}(\hat{\beta}_q - \beta) \to N[0, A^{-1} B A^{-1}]$$

where $A = \text{plim} \frac{1}{N} \Sigma f_u(0|x_i) x_i x_i'$ ← **

$B = \text{plim} \frac{1}{N} \Sigma q(1-q) x_i x_i$

Note $f_u(0|x_i)$ is the conditional density of the error term $u_q = y - x\beta_q$ evaluated at $u_q = 0$.

This is not directly calculated.

(Thus, we use the bootstrap procedure)