Note Title 5/5/2006

Pairwise-Difference Estimators

- See Honore + Powell (2001).

- Some of these estimators are inspired by the properties of nonlinear panel data models:

Partially Linear Logit Model

- Logit model with fixed effects:

$$y_{it} = 1\{\alpha_i + x_{it}\beta - \varepsilon_{it} \geq 0\}$$

- with $\varepsilon_{it} \sim$ iid logistic [conditional on α_i].

- Let $G(x) = \dfrac{\exp\{x\}}{1+\exp\{x\}}$

- Suppose we observe two time periods for i: $t=1,2$.

- Fix an individual i. Then

$$Pr(y_{i2}=1 \mid x_{i1}, x_{i2}, y_{i1}+y_{i2}=1)$$

$$= \frac{G(\alpha_i + x_{i2}\beta)[1-G(\alpha_i + x_{i1}\beta)]}{G(\alpha_i+x_{i2}\beta)[1-G(\alpha_i+x_{i1}\beta)] + (1-G(\alpha_i+x_{i2}\beta))G(\alpha_i+x_{i1}\beta)}$$

- Cancelling the terms in the denominator of $G(x)$ and $1-G(x)$, we get:

$$= \frac{\exp\{\alpha_i + x_{i2}\beta\}}{\exp\{\alpha_i + x_{i2}\beta\} + \exp\{\alpha_i + x_{i1}\beta\}}$$

- Multiplying and dividing by $\exp\{\alpha_i\}$, we get:

$$= \frac{\exp\{x_{i2}\beta\}}{\exp\{x_{i2}\beta\} + \exp\{x_{i1}\beta\}}$$

$$= \frac{\exp\{(x_{i2}-x_{i1})\beta\}}{\exp\{(x_{i2}-x_{i1})\beta\} + 1}$$

Immediately,

$$\Pr(y_{i2}=0 \mid x_{i1}, x_{i2}, y_{i1}+y_{i2}=1)$$

$$= 1 - \Pr(y_{i2}=1 \mid x_{i1}, x_{i2}, y_{i1}+y_{i2}=1)$$

$$= \frac{1}{\exp\{(x_{i2}-x_{i1})\beta\}+1}$$

$$= \frac{\exp\{(x_{i1}-x_{i2})\beta\}}{1+\exp\{(x_{i1}-x_{i2})\beta\}}$$

- So by considering only those individuals for which $y_{i1} \neq y_{i2}$, we can express their log-likelihood function as:

$$y_{i1} \log\left(\frac{\exp\{(x_{i1}-x_{i2})\beta\}}{1+\exp\{(x_{i1}-x_{i2})\beta\}}\right)$$
$$+ y_{i2} \log\left(\frac{\exp\{(x_{i2}-x_{i1})\beta\}}{1+\exp\{(x_{i2}-x_{i1})\beta\}}\right)$$

- So β can be estimated by

$$\max_{\beta} \sum_{i: y_{i1} \neq y_{i2}} \left[y_{i1} \log\left(\frac{\exp\{(x_{i1} - x_{i2})\beta\}}{1 + \exp\{(x_{i1} - x_{i2})\beta\}} \right) + y_{i2} \log\left(\frac{\exp\{(x_{i2} - x_{i1})\beta\}}{1 + \exp\{(x_{i2} - x_{i1})\beta\}} \right) \right]$$

$\Rightarrow$ Now, suppose that in the context of a cross-sectional data set, we have:

$$y_i = 1\!\!1\{x_i\beta + g(w_i'\gamma) - \varepsilon_i \geq 0\}$$

- where $g(\cdot)$ is an unknown, invertible transformation.

- $\varepsilon_i | x_i, w_i \sim$ iid logistic

- Suppose we have two observations, i, j for which $w_i'\gamma = w_j'\gamma$ and therefore $g(w_i'\gamma) = g(w_j'\gamma)$. Then we can proceed analogously with the partially linear logit model:

- For any pair of individuals (i, j) such that $y_i \neq y_j$, the conditional log-likelihood can be expressed as

$$y_i \log\left(\frac{\exp\{(x_i - x_j)\beta\}}{1+\exp\{(x_i - x_j)\beta\}}\right) + y_j \log\left(\frac{\exp\{(x_j - x_i)\beta\}}{1+\exp\{(x_j - x_i)\beta\}}\right)$$

- If $w_i'\gamma \sim$ continuous, then $\Pr(w_i'\gamma = w_j'\gamma) = 0$. In this case we need to use a kernel-weighed objective function. We would estimate $\hat{\beta}$ by maximizing:

$$\boxed{\frac{1}{h}\binom{N}{2}^{-1}\sum_{i<j} K\left(\frac{(w_i - w_j)'\hat{\gamma}}{h}\right) 1\!\!1\{y_i \neq y_j\} \times \left[y_i \log\left(\frac{\exp\{(x_i - x_j)\beta\}}{1+\exp\{(x_i - x_j)\beta\}}\right) + y_j \log\left(\frac{\exp\{(x_j - x_i)\beta\}}{1+\exp\{(x_j - x_i)\beta\}}\right)\right]}$$

- where we are already incorporating the fact that γ must be estimated (estimator denoted by $\hat{\gamma}$).

Partially Linear Tobit Models

- Consider a censored-regression model with fixed-effects:

$$y_{it} = \max\{0, \alpha_i + x_{it}\beta + \varepsilon_{it}\}$$

- Our very own Bo Honore (1992, Econometrica) showed that β can be estimated by:

$$\underset{\beta}{\text{minimize}} \sum_{i=1}^{N} \ell(y_{i1}, y_{i2}, \Delta x_i \beta) \equiv S_n(\beta)$$

where $\Delta x_i = x_{i1} - x_{i2}$

and

$$\ell(z_1, z_2, d) = \begin{cases} 0 & \text{if } z_1 \leq \max\{0, d\}, \\ & \& \; z_2 \leq \max\{0, -d\} \\ |z_1 - z_2 - d| & \text{otherwise} \end{cases}$$

[Note: $\varphi(z_1, z_2, \cdot)$ is continuous, with left and right derivatives everywhere]

- He also showed we could use an alternative objective function:

$$\underset{\beta}{\text{minimize}} \sum_{i=1}^{N} \lambda(y_{i1}, y_{i2}, \Delta x_i \beta) \equiv T_n(\beta)$$

where

$$\lambda(z_1, z_2, \delta) = \begin{cases} z_1^2 - 2z_1(z_2 + \delta) & \text{if } \delta \leq -z_2 \\ (z_1 - z_2 - \delta)^2 & \text{if } -z_2 < \delta < z_1 \\ z_2^2 - 2z_2(z_1 - \delta) & \text{if } z_1 \leq \delta \end{cases}$$

- Adapting this to the context of a partially linear Tobit model:

$$y_i = \max\{0, g(w_i'\gamma) + x_i\beta + \varepsilon_i\}$$

We would then estimate β by:

$$\min_{\beta} \frac{1}{n}\binom{N}{2}^{-1}\sum_{i<j} k\left(\frac{(w_i - w_j)'\hat{\gamma}}{h}\right) \ell(y_i, y_j, (x_i - x_j)\beta)$$

or

$$\min_{\beta} \frac{1}{n}\binom{N}{2}^{-1}\sum_{i<j} k\left(\frac{(w_i - w_j)'\hat{\gamma}}{h}\right) \lambda(y_i, y_j, (x_i - x_j)\beta)$$

- Next we examine the asymptotics of these types of estimators

Note Title 5/5/2006

Asymptotics of Kernel-Weighted Pairwise-Differenced Estimators

- We will focus on cases in which the "control function" or pairwise-differencing criterion is an invertible transformation of the L-dimensional

$$w_i \gamma_0 \in \mathbb{R}^L$$

- with γ_0 replaced (possibly) by a first-step estimator $\hat{\gamma}$. The estimator $\hat{\beta}$ minimizes an objective function of the form:

$$Q_n(\hat{\gamma}, b) = \binom{N}{2}^{-1} \frac{1}{h^L} \sum_{i<j} K\left(\frac{(w_i - w_j)\hat{\gamma}}{h}\right) s(v_i, v_j; b)$$

where – without loss of generality – $s(v_i, v_j; b)$ is symmetric in its first two arguments. Minimization is over $b \in \mathcal{B} \subset B$ is the parameter space).

- Assuming that $s(v_i, v_j, \cdot)$ has left derivatives, $\hat{\beta}$ can be assumed to satisfy a vector of "first-order conditions" (recall censored LAD)

$$\boxed{\underbrace{\binom{N}{2}^{-1} \frac{1}{h^L} \sum_{i<j} K\left(\frac{(W_i - W_j)\hat{\gamma}}{h}\right) t(v_i, v_j; \hat{\beta})}_{\text{some combination of left-derivatives of } s(v_i, v_j, \cdot) \text{ w.r.t } \beta \text{ is } O_p(N^{-1/2})} = O_p(N^{-1/2})}$$

- To illustrate asymptotics, let us assume $W_i'\hat{\gamma} \in \mathbb{R}$ (real-valued index). so $L = 1$...

- We assume that our first-step estimator $\hat{\gamma}$ satisfies:

$$\hat{\gamma} = \gamma_0 + \frac{1}{N} \sum_{i=1}^{N} \psi_i + O_p(N^{-1/2})$$

where $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_i \xrightarrow{d} N(0, E[\psi\psi'])$ and $\therefore$

$N(\hat{\gamma} - \gamma_0)^2 = O_p(1)$.

- A simple Taylor approximation yields:

$$K\left(\frac{(W_i - W_j)\hat{\gamma}}{h}\right) = K\left(\frac{(W_i - W_j)\gamma_0}{h}\right)$$

$$+ K^{(1)}\left(\frac{(W_i - W_j)\gamma_0}{h}\right) \frac{(W_i - W_j)}{h} (\hat{\gamma} - \gamma_0)$$

$$+ O_p(1/N)$$

$K^{(1)}(\psi)$ denotes first derivative of $K(\psi)$

- therefore:

$$o_p(N^{-1/2}) = \binom{N}{2}^{-1} \frac{1}{h} \sum_{i<j} K\left(\frac{(W_i - W_j)\gamma_0}{h}\right) t(V_i, V_j; \hat{\beta})$$

$$+ \binom{N}{2}^{-1} \frac{1}{h^2} \sum_{i<j} K^{(1)}\left(\frac{(W_i - W_j)\gamma_0}{h}\right) (W_i - W_j)(\hat{\gamma} - \gamma_0)\, t(V_i, V_j; \hat{\beta})$$

$$+ o_p(N^{-1/2})$$

- where the last $o_p(N^{-1/2})$ follows from $N(\hat{\gamma} - \gamma_0)^2 = O_p(1)$ and the appropriate assumptions about h (for example, that $N^{1/2} h \to \infty$), as well as the assumptions: $K^{(2)}(\psi)$ uniformly bounded and $E[W^2 t(V_i, V_j; \beta)] < \infty$ $\forall \beta \in B$.

Using $\hat{\gamma} - \gamma_0 = \frac{1}{N} \sum_{k=1}^{N} \psi_k + O_p(N^{-1/2})$, we get

$$O_p(N^{-1/2}) = \binom{N}{2}^{-1} \frac{1}{h} \sum_{i<j} K\left(\frac{(W_i - W_j)\gamma_0}{h}\right) t(V_i, V_j; \hat{\beta})$$

$$+ \binom{N}{2}^{-1} \frac{1}{h^2} \sum_{i<j} K^{(1)}\left(\frac{(W_i - W_j)\gamma_0}{h}\right)(W_i - W_j)\left[\frac{1}{N} \sum_{k=1}^{\tilde{N}} \psi_k + O_p(N^{-1/2})\right]$$

$$\times t(V_i, V_j; \hat{\beta})$$

- We have U-statistics that are not in closed-form (they are functions of $\hat{\beta}$). We must make assumptions about how these objects behave uniformly in the parameter space B.

- We must analyze the $\boxed{\text{U-processes}}$

- Let $z = (W, V)$

$$\mathcal{F}_1 = \left\{ f : f(z_1, z_2) = \frac{1}{h} K\left(\frac{(W_1 - W_2)\gamma_0}{h}\right) t(V_1, V_2; \beta) \right\}$$

$\longrightarrow$ Focus on U-process $\boxed{U_{n,2} f : f \in \mathcal{F}_1}$

– Let

$p_N^A(z_i;\beta)$ denote the projection of

$$\binom{N}{2}^{-1} \frac{1}{h} \sum_{i<j} K\left(\frac{(w_i - w_j)\gamma_0}{h}\right) t(v_i, v_j; \beta)$$

– Suppose we show that

$$\boxed{\begin{aligned} &\binom{N}{2}^{-1} \frac{1}{h} \sum_{i<j} K\left(\frac{(w_i - w_j)\gamma_0}{h}\right) t(v_i, v_j; \beta) \\ &\quad = \frac{1}{N} \sum_{i=1}^{N} p_N^A(z_i;\beta) + \varepsilon_N(\beta) \\ &\text{where} \quad \sup_{\beta \in B} |\varepsilon_N(\beta)| = O_p(N^{-1/2}) \end{aligned}}$$

– this could be achieved for example if we show that $U_{n,2} f$ is Euclidean and invoke one of Sherman's results for U-processes.

– There's still a second U-statistic [of order 3] floating around

$$\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\left[\frac{1}{N}\sum_{k=1}^{N}\psi_k+o_p(N^{-1/2})\right]$$
$$\times t(V_i,V_j;\hat{\beta})$$

$$=\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\frac{1}{N}\sum_{k=1}^{N}\psi_k\, t(V_i,V_j;\hat{\beta})$$

$$\underbrace{+\,o_p(N^{-1/2})}$$

Assumes

$$\sum_{k=1}^{N}\psi_k$$

$$\sup_{\beta\in B}\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\,t(V_i,V_j;\beta)$$
$$=O_p(1)$$

- It would be enough to assume that $t(V_i,V_j,\beta)$ is dominated by some $\phi(Z_i,Z_j)$ w.p. 1 $\forall\,\beta\in B$.

- So we have to focus on:

$$\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\frac{1}{N}\sum_{k=1}^{N}\psi_k\, t(V_i,V_j;\hat{\beta})$$

- This can be split into two components:

  * The sum over terms in which $k=i$ or $k=j$
  * The sum over terms in which $k\neq i$ and $k\neq j$

$$\underbrace{\frac{1}{N}\underbrace{\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\psi_i\, t(V_i,V_j;\beta)}_{O_p(1)\text{ uniformly in } B}}$$

therefore, multiplied by $\frac{1}{N}$ is easily $o_p(N^{-1/2})$

- Same for the case $k=j$:

$$\underbrace{\frac{1}{N}\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\psi_j\, t(V_i,V_j;\beta)}_{o_p(N^{-1/2})}$$

- Therefore, uniformly in B:

$$\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{i<j}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\frac{1}{N}\sum_{k=1}^{N}\Psi_k\, t(V_i,V_j;\beta)$$

$$=\frac{1}{N}\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{\substack{i<j\\ k\neq i,\\ k\neq j}}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\Psi_k\, t(V_i,V_j;\beta)$$

$$+\,O_p(N^{-1/2})$$

- Before proceeding, we should tidy-up the last expression: Let

$$C=\{(i,j,k),(i,k,j),(j,k,i)\}$$

Denote $U=(W,V,\Psi)$ and define

$$H_N(U_i,U_j,U_k;\beta)=$$

$$\sum_{C}\frac{1}{h^2}K^{(1)}\left(\frac{(W_i-W_j)\gamma_0}{h}\right)(W_i-W_j)\Psi_k\, t(V_i,V_j;\beta)$$

Then:

$$\frac{1}{N}\binom{N}{2}^{-1}\frac{1}{h^2}\sum_{\substack{i<j\\ k\neq i\\ k\neq j}} K^{(1)}\left(\frac{(W_i-W_j)'\delta_0}{h}\right)(W_i-W_j)\,\psi_k\, t(U_i,U_j;\beta)$$

$$=\underbrace{\frac{1}{3}\frac{(N-1)}{N}}_{\longrightarrow\,\frac{1}{3}}\binom{N}{3}^{-1}\sum_{i<j<k} H_N(U_i,U_j,U_k;\beta)$$

- Let $P_N^{\beta}(U_i;\beta)$ denote the projection of $\underbrace{\left(\frac{1}{3}\right)\binom{N}{3}^{-1}\sum_{i<j<k} H_N(U_i,U_j,U_k;\beta)}_{U\text{-process}}$

- Suppose we can show that

$$\binom{N}{3}^{-1}\sum_{i<j<k} H_N(U_i,U_j,U_k;\beta)$$

$$=\frac{1}{N}\sum_{i=1}^{N} P_N^{\beta}(U_i;\beta)+v_N(\beta)$$

where $\sup_{\beta\in B}|v_N(\beta)| = o_p(N^{-1/2})$

- This could be accomplished for example - as above - by showing that the U-process

$$\binom{N}{3}^{-1} \sum_{i<j<k} H_N(U_i, U_j, U_k; \beta) : \beta \in B$$

is Euclidean...

- Thus, if all of the above holds, we'd have:

$$\binom{N}{2}^{-1} \frac{1}{h} \sum_{i<j} K\left(\frac{(W_i - W_j)\gamma_0}{h}\right) t(V_i, V_j; \beta)$$

$$+ \binom{N}{2}^{-1} \frac{1}{h^2} \sum_{i<j} K^{(1)}\left(\frac{(W_i - W_j)\gamma_0}{h}\right) (W_i - W_j)(\hat{\gamma} - \gamma_0)\, t(V_i, V_j; \beta)$$

$$= \underbrace{\frac{1}{N} \sum_{i=1}^{N} P_N^A(Z_i; \beta) + \frac{1}{N} \sum_{i=1}^{N} P_N^B(U_i; \beta) + o_p(N^{-1/2})}_{}$$

Uniformly over the parameter space B.

- Therefore, since $\hat{\beta} \in B$ and given the "first-order conditions" satisfied by $\hat{\beta}$:

$$\frac{1}{N}\sum_{i=1}^{N} \rho_N^A(z_i, \hat{\beta}) + \frac{1}{N}\sum_{i=1}^{N} \rho_N^B(v_i, \hat{\beta}) = O_p(N^{-1/2})$$

Consistency of $\hat{\beta}$:

- Suppose that uniformly in B:

$$\frac{1}{N}\sum_{i=1}^{N} \rho_N^A(z_i; \beta) \xrightarrow{p} E[\rho_N^A(z_i; \beta)]$$

$$\frac{1}{N}\sum_{i=1}^{N} \rho_N^B(z_i; \beta) \xrightarrow{p} E[\rho_N^B(z_i; \beta)]$$

- The usual "dominance" assumption would suffice for this. Then, given all our previous work, we'd have $\hat{\beta} \xrightarrow{p} \beta_0$ if

$$\left. E[\rho_N^A(z_i; \beta)] + E[\rho_N^B(z_i; \beta)] = 0 \right\} \text{ whenever } \beta = \beta_0$$

- Alternatively we could've worked directly with the objective function (not the "F.O.C") and shown that it converges uniformly in B to the population conditional expectation that identifies β_0

- **Asymptotic Normality:**

- Suppose we've shown that $\hat{\beta} \xrightarrow{p} \beta_0$.

- Assuming that the projections $P_N^A(Z_i, \cdot)$ and $P_N^B(U_i, \cdot)$ are smooth functions of β, then we'd have:

$$\frac{1}{N}\sum_{i=1}^{N} P_N^A(Z_i, \hat{\beta}) + \frac{1}{N}\sum_{i=1}^{N} P_N^B(U_i, \hat{\beta}) = O_p(N^{-1/2})$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big\{ P_N^A(Z_i, \beta_0) + \nabla_\beta P_N^A(Z_i, \bar{\beta})(\hat{\beta} - \beta_0)$$
$$+ P_N^B(U_i, \beta_0) + \nabla_\beta P_N^B(U_i, \bar{\beta})(\hat{\beta} - \beta_0)\Big\}$$
$$= O_p(N^{-1/2})$$

where $\bar{\beta}$ is between $\hat{\beta}$ and β_0

$\Rightarrow$

$$\hat{\beta} - \beta_0 = -\left(\frac{1}{N}\sum_{i=1}^{N}\left\{\nabla_\beta P_N^A(Z_i, \bar{\beta}) + \nabla_\beta P_N^B(U_i, \bar{\beta})\right\}\right)^{-1}$$
$$\times \frac{1}{N}\sum_{i=1}^{N}\left[P_N^A(Z_i, \beta_0) + P_N^B(U_i, \beta_0)\right]$$
$$+ O_p(N^{-1/2})$$

- Once again, $\hat{\beta} \xrightarrow{p} \bar{\beta}$ and some dominance assumption would yield:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\nabla_\beta \rho_N^A(Z_i,\bar{\beta}) + \nabla_\beta \rho_N^B(U_i,\bar{\beta})\right\}$$

$$\xrightarrow{p} E\left[\nabla_\beta \rho_N^A(Z_i,\beta_0) + \nabla_\beta \rho_N^B(U_i,\beta_0)\right]$$

- Assuming this matrix is nonsingular, then

$$\boxed{\begin{aligned} &N^{1/2}(\hat{\beta}-\beta_0) = \\ &E\left[\nabla_\beta \rho_N^A(Z_i,\beta_0) + \nabla_\beta \rho_N^B(U_i,\beta_0)\right]^{-1} \\ &\quad \times \frac{1}{N}\sum_{i=1}^{N}\left[\rho_N^A(Z_i,\beta_0) + \rho_N^B(U_i,\beta_0)\right] \\ &\quad + o_p(1) \end{aligned}}$$

- This characterizes the asymptotic distribution of $\sqrt{N}(\hat{\beta}-\beta_0)$

$\square$

Please
Note

* If the left-derivatives $t(v_{\bar{0}}, v_{\bar{j}}, \beta)$ are smooth functions of β, then we could do a Taylor approximation directly on them and then find the projections, proof would be simpler.

* We need $K(\cdot)$ and $K^{(1)}(\cdot)$ to have special properties (bias-reducing) in order for the projections $\rho_N^A(\cdot)$, $\rho_N^B(\cdot)$ to be tractable up to a term of order $o_p(N^{-1/2})$.

* Note; if $K(\cdot)$ is symmetric, then

$$\int K^{(1)}(\psi)\, d\psi = 0$$

that is why $\frac{1}{h^2}\int K^{(1)}\left(\frac{x-v}{h}\right) f(v)\, dv$

$$= \frac{1}{h}\int K^{(1)}(\psi)\, f(h\psi + x)\, d\psi = \underline{f'(x)} + h \cdot O(1)$$

assuming smoothness.