0.1 Practical Methods for Micro-Panel Data Analysis

Myoung-jae Lee
Singapore Management University

(i) Often $T = 2$ or 3 used; N large T small; i for individuals often omitted.

(ii) Mostly 'fixed effect' ('related effect') models; rarely 'random effect'.

(iii) Based mostly on Lee (2002) and the literature since 2002; other overviews in Arellano and Honoré (2001), Arellano (2003), and Hsiao (2003),

1. Linear Models

1.1 Getting $T \times 1$ vector $y_i = q_i' \eta + v_i$

1.2 Moments, IVE and GMM

1.3 Handling Individual Effect δ_i

1.3.1 Error Differencing

1.3.2 Regressor Differencing

1.3.3 δ-Splitting and MDE

2. Limited Dependent Variables

2.1 Conditional Logit and Panel Probit

2.2 Conditional Poisson

2.3 Censored Model

2.4 Dynamic Models

2.4.1 Dynamic Panel Probit

2.4.2 Dynamic Count Response

2.4.3 Dynamic Censored Response

APPENDIX: How to Get SD or CI

0.1.1 1. Linear Models

1.1 Getting $T\times 1$ Vector $y_i = q_i'\eta + v_i$ Suppose, for $i = 1, ..., N$ and $t = 1,\ 2,\ 3\ (= T)$,

$$y_{it} = 1\cdot \underset{1\times 1}{\tau_t} + \underset{1\times k_{\tilde{c}}}{\tilde{c}_i'}\tilde{\alpha} + \underset{1\times k_x}{x_{it}'}\beta + \delta_i + u_{it} \qquad ((1.1))$$

where τ_t, $\tilde{\alpha}$ and β are parameters, $\tilde{c}_i$ is time-constant regressors, x_{it} is time-variant regressors, and $\delta_i + u_{it}$ is an error term.

An example is

y_{it} : ln(wage)

τ_t : effect of economy on y_{it} common to all i

$\tilde{c}_i$: race, schooling years

x_{it} : work hours, local unemployment rate

δ_i : ability, IQ, or productivity

u_{it} : unobserved residential information

Define

$$\tilde{k} \equiv k_{\tilde{c}} + k_x, \quad \tilde{\gamma} \equiv \begin{bmatrix} \tilde{\alpha} \\ \beta \end{bmatrix}, \quad \tilde{w}_{it} \equiv \begin{bmatrix} \tilde{c}_i \\ x_{it} \end{bmatrix}, \quad v_{it} \equiv \delta_i + u_{it},$$

to rewrite the model as

$$y_{it} = 1\cdot \underset{1\times 1}{\tau_t} + \underset{1\times \tilde{k}}{\tilde{w}_{it}'}\tilde{\gamma} + v_{it}. \qquad ((1.2))$$

Define

$$c_i \equiv (1, \tilde{c}_i')',\ \alpha \equiv (\tau_1, \tilde{\alpha}),\ w_{it} \equiv (c_i, x_{it}')',\ \gamma \equiv (\alpha', \beta')',\ k \equiv k_c + k_x.$$

If $\tau_1 = \tau_2 = \tau_3$, then

$$y_{it} = \underset{1\times k}{w_{it}'}\gamma + v_{it}, \qquad ((1.3))$$

to be used sometimes to simplify exposition.

Assume only *iid of* (w'_{it}, v_{it}) *across* i *while allowing for arbitrary dependence and heterogeneity across* t *within a given* i. (1.2) allows endogenous regressors and lagged dependent variables as regressors.

Define

$$y_i \equiv \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{iT} \end{bmatrix}, \quad x'_i \equiv \begin{bmatrix} x'_{i1} \\ x'_{i2} \\ x'_{i3} \end{bmatrix}, \quad u_i \equiv \begin{bmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \end{bmatrix}.$$

Write the stacked time-effects as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \equiv I_3 \tau = \tau.$$

Define

$$m_3 \equiv \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Delta\tau \equiv \begin{bmatrix} \tau_2 - \tau_1 \\ \tau_3 - \tau_1 \end{bmatrix}, \quad \tau^* \equiv \begin{bmatrix} \Delta\tau \\ \tau_1 \end{bmatrix}$$

to get

$$I_3 \tau = (m_3, 1_3) \cdot \tau^* :$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \tau_2 - \tau_1 \\ \tau_3 - \tau_1 \\ \tau_1 \end{bmatrix}$$

The 1_3 column is the analog for the usual intercept in cross-section models; it becomes a time-constant regressor.

Observe

$$1'_3 \otimes c_i = [1, 1, 1] \otimes c_i = [c_i, ..., c_i]$$

and finally write (1.2) as

$$\underset{3\times 1}{y_i} = \underset{3\times 2}{m_3}\Delta\tau + \underset{3\times k_c}{(1_3'\otimes c_i)'}\alpha + \underset{3\times k_x}{x_i'}\beta + \underset{3\times 1}{1_3}\delta_i + \underset{3\times 1}{u_i}$$

$$= \underset{3\times 2}{m_3}\Delta\tau + \underset{3\times k}{w_i'}\gamma + v_i = \underset{3\times(k+2)}{q_i'}\eta + v_i \qquad ((1.4))$$

where

$$w_i' \equiv (w_{i1}, w_{i2}, w_{i3})' = ((1_3'\otimes c_i)', x_i'),$$
$$q_i' \equiv (m_3, w_i'), \quad \eta \equiv (\Delta\tau', \gamma')', \quad v_i \equiv 1_3\delta_i + u_i.$$

1.2 Moments, IVE and GMM Four types of orthogonalities between regressors and error terms:

$$\begin{array}{ll} \text{summing (SUM):} & E(\sum_t x_t v_t) = 0 \\ \text{contemp. (CON):} & E(x_t v_t) = 0 \quad \forall t \\ \text{predeter. (PRE):} & E(x_s v_t) = 0 \quad \forall s \le t \\ \text{strictly exo. (EXO):} & E(x_s v_t) = 0 \quad \forall s, t \end{array}$$

The following holds:

$$\text{SUM} \Leftarrow \text{CON} \Leftarrow \text{PRE} \Leftarrow \text{EXO}.$$

SUM is the moments for the LSE treating the panel as NT cross-section observations, for the LSE moment condition is

$$N^{-1}\Sigma_i\Sigma_t x_{it}v_{it} = 0 \implies E(\Sigma_t x_t v_t) = 0.$$

This LSE is similar to the '*between group estimator (BET)*' which is the LSE applied to $\bar{y}_i \equiv T^{-1}\sum_t y_{it}$ and $\bar{x}_i \equiv T^{-1}\sum_t x_{it}$.

In CON, only contemporaneous correlations are zero. In PRE, x_t can be correlated with v_s if $t > s$; e.g., rational expectation models with

$$E(v_t|x_1, ..., x_t) = 0, \text{ not with } E(v_t|x_1, ..., x_T) = 0.$$

In EXO, x_s and v_t are uncorrelated $\forall s, t$.

Moment conditions other than the above may be used as well. For example,

$$E(x_s v_t) = 0 \quad \forall s < t \quad (\text{not } s \le t \text{ as in PRE})$$

allowing a contemp. relation for x_t and v_t.

For IVE under PRE, observe

$$
\begin{aligned}
t &= 1: E(v_1 w_1) = 0, \\
t &= 2: E(v_2 w_1) = 0,\ E(v_2 w_2) = 0, \\
t &= 3: E(v_3 w_1) = 0,\ E(v_3 w_2) = 0,\ E(v_3 w_3) = 0.
\end{aligned}
$$

Remove redundant moments due to c_i appearing in all w_{it}'s to get

$$
\begin{aligned}
t &= 1: E(v_1 w_1) = 0, \\
t &= 2: E(v_2 x_1) = 0,\ E(v_2 w_2) = 0, \\
t &= 3: E(v_3 x_1) = 0,\ E(v_3 x_2) = 0,\ E(v_3 w_3) = 0.
\end{aligned}
$$

For IVE, set up the instrument matrix

$$z = \text{diag}\{w_1, (x_1', w_2')', (x_1', x_2', w_3')'\}$$

to observe

$$
z \cdot v = \underset{[\{k_x \cdot (1+2+3)\} + k_c \cdot 3] \times 3}{\begin{bmatrix} w_1 & 0 & 0 \\ 0 & x_1 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & x_1 \\ 0 & 0 & x_2 \\ 0 & 0 & w_3 \end{bmatrix}} \underset{3\times 1}{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}} = \begin{bmatrix} w_1 v_1 \\ \hline x_1 v_2 \\ w_2 v_2 \\ \hline x_1 v_3 \\ x_2 v_3 \\ w_3 v_3 \end{bmatrix}.
$$

With z_i, the IVE for η is

$$
\begin{aligned}
h_{ive} = \; & \{\sum_i q_i z_i' \, (\sum_i z_i z_i')^{-1} \, \sum_i z_i q_i'\}^{-1} \\
& \cdot \sum_i q_i z_i' \, (\sum_i z_i z_i')^{-1} \, \sum_i z_i y_i.
\end{aligned}
$$

The GMM more efficient than the IVE is

$$
\begin{aligned}
h_{gmm} &= (\sum_i q_i z_i' \ C_N^{-1} \ \sum_i z_i q_i')^{-1} \\
&\cdot \sum_i q_i z_i' \ C_N^{-1} \ \sum_i z_i y_i;
\end{aligned}
$$

$C_N \equiv (1/N)\sum_i z_i \hat{v}_i \hat{v}_i' z_i'$, $\hat{v}_i \equiv y_i - q_i' h_{ive}$, and

$$
\sqrt{N}(h_{gmm} - \eta) \sim N(0, \ \{(\sum_i q_i z_i'/N) \ C_N^{-1} \ (\sum_i q_i z_i'/N)\}^{-1}).
$$

Also, with $\tilde{v}_i \equiv y_i - q_i' h_{gmm}$,

$$
\{(1/\sqrt{N})\sum_i z_i \tilde{v}_i\}'\{(1/N)\sum_i z_i \tilde{v}_i \tilde{v}_i' z_i'\}^{-1}(1/\sqrt{N})\sum_i z_i \tilde{v}_i
$$

is the GMM over-identification test statistic for $H_o : E(zv) = 0$.

1.3 Handling Individual Effect δ_i When δ is related to some components w_t^1 in $w_t = (w_t^{o\prime}, w_t^{1\prime})'$, one solution is using w_s^o as instruments for w_t^1. Three other solutions are:

1. *Error Differencing*: use w_t as instruments for $v_t - v_{t-1}$ free of δ.

2. *Regressor Differencing*: use $w_t^1 - w_{t-1}^1$ as instruments for v_t if $w_t^1 = \delta + \omega_t$ with ω_t unrelated to δ.

3. *δ-Splitting*: absorb the part of δ related to w_t^1 into $w_t'\gamma$, leaving in v_t only the part of δ unrelated to w_t^1.

1.3.1 Error Differencing/Transforming When $T = 2$,

$$y_t - y_{t-1} = \tau_t - \tau_{t-1} + (x_t - x_{t-1})'\beta + u_t - u_{t-1}$$

is free of δ; LSE or IVE can be applied. For a generic T, apply mean differencing:

$$v_{it} - \frac{\sum_{t=1}^T v_{it}}{T} = u_{it} - \frac{\sum_{t=1}^T u_{it}}{T}.$$

Multiply the 'mean-differencing matrix'

$$Q_T \equiv I_T - \frac{1_T 1_T'}{T}$$

to the y_i equation to get

$$\begin{aligned} Q_T y_i &= Q_T m_T \Delta\tau + Q_T x_i'\beta + Q_T u_i \equiv Q_T w_i^{*\prime}\gamma + Q_T u_i, \\ \underset{T\times(T-1+k_x)}{w_i^{*\prime}} &\equiv (m_T, x_i'), \quad \gamma^* \equiv (\Delta\tau', \beta')'; \end{aligned}$$

e.g.,

$$Q_T y_i = (y_{i1} - \bar{y}_i, ..., y_{iT} - \bar{y}_i)'.$$

The '*within group estimator (WIT)*' is

$$g_{wit} = (\sum_i w_i^* Q_T w_i^{*\prime})^{-1} \cdot \sum_i w_i^* Q_T y_i.$$

As a digression, suppose

$$\begin{aligned} y_{it} &= \underbrace{\bar{w}_i'\gamma_1}_{\text{permanent}} + \underbrace{(w_{it}-\bar{w}_i)'\gamma_2}_{\text{transitory part}} + \delta_i + u_{it} \\ &\Longrightarrow \bar{y}_i = \bar{w}_i'\gamma_1 + \delta_i + \bar{u}_i \quad \text{(for BET)} \\ &\Longrightarrow y_{it} - \bar{y}_i = (w_{it} - \bar{w}_{it})'\gamma_2 + u_{it} - \bar{u}_i. \quad \text{(for WIT)} \end{aligned}$$

If $\gamma_1 = \gamma_2 = \gamma$, we get $y_{it} = w_{it}'\gamma + \delta_i + u_{it}$.

Another digression is that, sometimes when N is small relative to T, dummy variables are used for all i:

$$\begin{aligned} y_j &= \tilde{x}_j'\tilde{\beta} + \Sigma_{i=1}^N \delta_i d_{ij} + u_j, \quad j = 1, ..., NT \\ d_{ij} &= 1 \text{ if datum } j \text{ is for person } i \text{ and } 0 \text{ otherwise.} \end{aligned}$$

Let $\hat{y}_j$ ($\hat{x}_j$) denote the LSE residual of y_j ($\tilde{x}_j$) on $d_{1j}, ..., d_{Nj}$. $\tilde{\beta}$ can be estimated by the LSE of $\hat{y}_j$ on $\hat{x}_j$, which is WIT. This can be seen in the LSE of Y $(NT \times 1)$ on $I_N \otimes 1_T$:

$$\begin{aligned} &\{(I_N \otimes 1_T)'(I_N \otimes 1_T)\}^{-1}(I_N \otimes 1_T)'Y \\ = \ &(I_N \otimes 1_T'1_T)^{-1}(I_N \otimes 1_T')Y \\ = \ &(I_N \otimes T^{-1})(\Sigma_{t=1}^T y_{1t}, ..., \Sigma_{t=1}^T y_{Nt})' = (\bar{y}_1, ...\bar{y}_N)'. \end{aligned}$$

Ln(wage) Equation				
	BET	WIT	GMM-PRE	GMM-EXO
$\tau_2 - \tau_1$			-.004(.15)	.00(.47)
$\tau_3 - \tau_1$		.002(.04)	.001(.04)	.004(.17)
τ_1	.413(.58)		1.377(1.2)	.412(.45)
age	.043(1.3)	.079(1.4)	.062(1.1)	.076(1.5)
$\frac{age^2}{100}$	-.033(-.9)	-.081(-1.3)	-.053(-.7)	-.076(-1.1)
edu	.064(5.2)		.010(.24)	.064(2.0)
#kids	.003(.11)	.044(.76)	.027(.16)	-.064(-.41)
ln(hour)	.013(.57)	-.105(-3.2)	.030(.36)	-.009(-.15)
married	.114(.88)	.008(.06)	-.634(-.91)	-.069(-.15)
salaried	.259(3.7)	.091(2.0)	.271(2.9)	.125(2.1)
self-emp.	-.454(-4.2)	-.200(-1.8)	-.291(-2.6)	-.278(-3.0)
unem.rate	-.012(-.6)	-.030(-1.5)	-.039(-2.1)	-.031(-2.4)
p-value for GMM over-ID test:			.288	.035

1.3.2 Regressor Differencing/Transforming Error-differencing removes all time-constants (e.g., education) along with δ. In regressor differencing, the error term—i.e., the model equation—stays intact and all time-constants are kept. The IVE and GMM are of this type.

The *Linear projection* of λ on z is

$$E(\lambda z')\ E^{-1}(zz')\ z = B'z, \text{ where } B \equiv E^{-1}(zz')\ E(z\lambda');$$

B is the 'linear projection coefficient', and either $E(z) = 0$ or z should have 1 as its component. Split λ into two parts $B'z$ and $\varepsilon \equiv \lambda - B'z$; $COR(\varepsilon, z) = 0$ by construction.

Suppose, for a regressor m_t,

$$E(m_t) \text{ and } E(\delta m_t) \text{ are not functions of } t,$$

which is a moment-stationarity assumption. Linearly project m_{it} on $(1, \delta_i)$ to get

$$m_{it} = \phi_i + (m_{it} - \phi_i) \equiv \phi_i + \lambda_{it}.$$

λ_{it} is uncorrelated with ϕ_i by construction;

$$m_{it} - m_{i,t-1} = \lambda_{it} - \lambda_{i,t-1}$$

can be used as an instrument.

1.3.3 δ-Splitting and MDE

Chamberlain (1982) rewrites δ_i as

$$\delta_i = 1 \cdot \zeta_o + \tilde{c}_i' \zeta_{\tilde{c}} + \Sigma_\tau x_{i\tau}' \zeta_\tau + \nu_i, \;\; \nu_i \equiv \delta_i - \zeta_o - \tilde{c}_i' \zeta_{\tilde{c}} - \Sigma_\tau x_{i\tau}' \zeta_\tau;$$

δ_i is (linearly) projected on $(1, \tilde{c}_i', x_{i1}', ..., x_{iT}')$, and $(\zeta_o, \zeta_c', \zeta_1', ..., \zeta_T')$ are the (linear) projection coefficients.

Substitute δ_i into $y_{it} = \tau_t + \tilde{c}_i' \tilde{\alpha} + x_{it}' \beta + \delta_i + u_{it}$ to get

$$y_{it} = \tau_t + \zeta_o + \tilde{c}_i'(\tilde{\alpha} + \zeta_{\tilde{c}}) + x_{it}'(\beta + \zeta_t) + \Sigma_{\tau \neq t} x_{i\tau}' \zeta_\tau + v_{it}, \quad v_{it} \equiv \nu_i + u_{it};$$

With δ gone, each wave can be estimated by LSE or IVE.

When each wave is estimated separately, there occurs problems: e.g., T-many estimates for $\tilde{\alpha} + \zeta_{\tilde{c}}$. How do we combine these? This is done by a *minimum distance estimator (MDE)*, which is a (weighted) *average* of T-many estimates.

The advantage of δ-splitting is that projection does not require any restriction (no assumption that $E(\delta | c_i, x_{i1}, ..., x_{iT})$ is a linear function of $c_i, x_{i1}, ..., x_{iT}$). A disadvantage is that $\tilde{\alpha}$ is not identified due to $\zeta_{\tilde{c}}$. This can be avoided omitting $\tilde{c}_i$ from the variables on which δ_i is projected. In this case, the resulting error term may be correlated with $\tilde{c}_i$ (use IVE then).

Holtz-Eakin et al. (1988,89) project y_t on

$$1, \;\; y_{t-1}, ..., y_{t-J}, \;\; x_{t-1}, ..., x_{t-J}, \;\; \delta_i \quad \text{to get}$$

$$y_{it} = \alpha_{0t} + \Sigma_{j=1}^{J} \alpha_{jt} y_{i,t-j} + \Sigma_{j=1}^{J} \beta_{jt} x_{i,t-j} + \Phi_t \delta_i + u_{it}.$$

The projection yields PRE type conditions:

$$E(u_{it}) = 0,\ E(y_{is} u_{it}) = 0,\ E(x_{is} u_{it}) = 0, \quad t - J \leq s \leq t - 1.$$

Removing δ_i with a 'quasi-differencing' $y_{it} - (\Phi_t / \Phi_{t-1}) y_{i,t-1}$, they estimate the model with IVE. 'Granger non-causality' of x_t on y_t ($\beta_{jt} = 0\ \forall j, t$) can be tested.

0.1.2 2. Limited Dependent Variables

2.1 Conditional Logit and Panel Probit Suppose

$$\begin{aligned} y^*_{it} &= \tau_t + \tilde{c}'_i\tilde{\alpha} + x'_{it}\beta + \delta_i + u_{it} = \tau_t + \tilde{w}'_{it}\tilde{\gamma} + v_{it}, \\ y_{it} &= 1[y^*_{it} > 0], \quad \text{where} \end{aligned}$$

τ_t is the time-effect common to all i,

$\tilde{c}_i$ is time-constant regressors,

x_{it} is time-variant regressors,

$\tilde{w}_{it}$ is $(\tilde{c}'_i, x'_{it})'$, the parameter $\tilde{\gamma}$ is $(\tilde{\alpha}', \beta')'$,

$v_{it} \equiv \delta_i + u_{it}$ is a composite error.

IVE (for 'regressor differencing') is not applicable, for the y_{it} eq. is not solvable for v_{it}; only 'error-differencing' and 'δ-splitting' can remove the relation between δ_i and $\tilde{w}_{it}$. In error-differencing, absorb $\tilde{c}_i$ into δ_i, for both will be removed by the differencing.

Conditional logit (CLOG) with $T = 2$ assumes

u_{it} is logistic independently of $(\delta_i, x_{i1}, x_{i2})$,

and *iid* across i *and* t,

and maximizes, for b $(\Delta x_i \equiv (1, x'_{i2} - x'_{i1})')$,

$$\sum_i d_i \left[y_{i1} \ln \frac{1}{1 + \exp(\Delta x'_i b)} + y_{i2} \ln \frac{\exp(\Delta x'_i b)}{1 + \exp(\Delta x'_i b)} \right]$$

where $d_i = 1$ if $y_{i1} \neq y_{i2}$ and 0 otherwise; the intercept in b is for $\tau_2 - \tau_1$. CLOG is error-differencing.

For $T \geq 3$, apply CLOG to each possible pair, to combine the estimates with MDE. For *ordered discrete responses (ODR),* collapse ODR into binary; apply CLOG to each possible binary version; use MDE. Also multinomial CLOG is available.

The CLOG dynamics is limited. First, $u_{i1}, ..., u_{iT}$ are iid: $v_{it} = \delta_i + u_{it}$, $t = 1, ..., T$, are related only through δ_i; auto-correlation of v_{it} is constant over t. Second, u_{it} is independent of $(\delta_i, x_{i1}, ..., x_{iT})$, not just of (δ_i, x_{it}), or of $(\delta_i, x_{i1}, ..., x_{it})$; these three are of type EXO, CON, and PRE.

One disadvantage of the EXO in CLOG is that, constraining u_{it} to be independent of the future regressors, the future x_{it} cannot be adjusted depending on the past u_{is}. Another disadvantage is that $y_{i,t-1}$ is not allowed in x_{it}: if $y_{i,t-1}$ is in x_{it}, then u_{it} becomes dependent on $x_{i,t+1}$.

Panel probit assumes

$$\delta_i = \zeta_o + \tilde{c}_i' \zeta_{\tilde{c}} + x_{i1}' \zeta_1 +, \cdots, + x_{iT}' \zeta_T + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2).$$

Differently from the linear model, this is an assumption, not projection. Plug this into the y_{it}^* eq. to get

$$y_{it}^* = \tau_t + \zeta_o + \tilde{c}_i'(\alpha + \zeta_{\tilde{c}}) + x_{it}'(\beta + \zeta_t) + \Sigma_{j \neq t} x_{ij}' \zeta_j + \varepsilon_i + u_{it}.$$

Divide both sides by $\sigma_t \equiv SD(\varepsilon_i + u_{it})$ and apply the usual probit to each wave. The remaining steps are similar to those for the linear model projection approach with MDE; an extra complication is σ_t varying over t.

2.2 Conditional Poisson Defining

$$\Delta x_{it1} \equiv (1, x_{it}' - x_{i1}')',$$

Conditional poisson (CPOI) of Hausman et al. (1984) for count responses maximizes for b

$$\sum_{t=1}^{T} y_{it} \cdot [\ \Delta x_{it1}' b - \ln \{ \sum_{s=1}^{T} \exp(\Delta x_{is1}' b) \}\ ].$$

Wooldridge (1999) shows that CPOI needs only

$$E(y_{it}|\delta_i, x_{i1}, ..., x_{iT}) = E(y_{it}|\delta_i, x_{it}) = \exp(\tau_t + x_{it}'\beta + \delta_i).$$

The second equality specifies a regression function. The first equality is an EXO, because given x_{it} and δ_i, the dist. of y_{it} is that of an 'error term' in y_{it}, and knowing $x_{i\tau}$, $\tau \neq t$, is not informative for the error term. These assumptions are much weaker than the original ones for CPOI.

2.3 Censored Model When $T = 2$, for the censored model

$$y_{it} = \max(y^*_{it}, 0),$$

Honoré's (1992) assumes that u_{i1} and u_{i2} are *exchangeable* given $(\delta_i, x_{i1}, x_{i2})$—an EXO—to propose an estimator minimizing

$$\Sigma_{i=1}^{N}[\ \{\max(y_{i2}, \Delta x_i' b) - \max(y_{i1}, -\Delta x_i' b) - \Delta x_i' b\}^2$$
$$-2 \cdot 1[y_{i2} < \Delta x_i' b]\ (y_{i2} - \Delta x_i' b) y_{i1}\ -2 \cdot 1[y_{i1} < -\Delta x_i' b]\ (y_{i1} + \Delta x_i' b) y_{i2}\ ].$$

Private Transfer (Dae-Woo panel; Kang & Lee (2003))		
	96/97, related	96/97, unrelated
Public transfers	-0.998 (-3.33)	-0.742 (-4.27)
Pre-transfer income	-0.002 (-0.11)	-0.013 (-1.67)
# elderly above 60	-0.529 (-0.04)	35.66 (4.89)
Household size	-66.52 (-1.80)	-26.61 (-4.51)
agriculture/fishery/part-time	91.18 (1.25)	116.58 (6.14)
unemployed/non-paid	101.06 (1.38)	185.04 (8.89)

2.4 Dynamic Models

2.4.1 Dynamic Panel Probit

Lee and Tae (2005) assume

$$
\begin{aligned}
y_{i1} &= 1[w_{i1}'\alpha + \alpha_\delta \delta_i + u_{i1} > 0], \quad COR(u_1, u_t) = 0 \ \forall t = 2, ..., T \\
y_{it} &= 1[\beta_y y_{i,t-1} + \beta_{yz} y_{i,t-1} z_{it} + w_{it}'\beta + \delta_i + u_{it} > 0], \quad t = 2, ..., T \\
\delta_i &= \bar{x}_i' \mu + \eta_i.
\end{aligned}
$$

The δ equation comes from

$$
\begin{aligned}
\delta_i &= x_{i1}'\mu_1 +, \ ..., \ + x_{iT}'\mu_T + \eta_i \\
&= (\Sigma_t x_{it}')\mu_0 + \eta_i \quad \text{under } \mu_1 =, ..., = \mu_T \equiv \mu_0 \\
&= \bar{x}_i'(\mu_0 T) \ + \ \eta_i = \bar{x}_i'\mu \ + \ \eta_i, \quad \text{where } \mu \equiv \mu_0 T
\end{aligned}
$$

Substitute the δ equation to get

$$
\begin{aligned}
y_{i1} &= 1[w_{i1}'\alpha + \bar{x}_i'\alpha_\delta \mu + \alpha_\delta \eta_i + u_{i1} > 0], \\
y_{it} &= 1[\beta_y y_{i,t-1} + \beta_{yz} y_{i,t-1} z_{it} + w_{it}'\beta + \bar{x}_i'\mu + \eta_i + u_{it} > 0].
\end{aligned}
$$

Modelling δ_i appears in Chamberlain (1984), and modelling y_{i0} in Heckman (1981).

Assume

$$
\begin{aligned}
u_{i1} \sim \ & N(0, \sigma_1^2), \ u_{i2}, ..., u_{iT} \text{ are iid } N(0, \sigma_u^2), \quad \eta_i \sim N(0, \sigma_\eta^2), \\
& u_{i1}, \ u_{i2}, ..., u_{iT}, \ \eta_i \text{ are independent of one another, and} \\
& \text{independent of } w_{i1}, ..., w_{iT}.
\end{aligned}
$$

With $\Phi(a)^y\{1-\Phi(a)\}^{1-y} = \Phi\{a \ (2y-1)\}$, the log-likelihood function is $(\zeta \equiv \eta/\sigma_\eta)$

$$\sum_i \ln [\int \Phi\{ (w_1' \frac{\alpha}{\sigma_1} + \bar{x}_i' \frac{\mu\alpha_\delta}{\sigma_1} + \zeta \frac{\alpha_\delta \sigma_\eta}{\sigma_1}) (2y_{i1} - 1) \}$$
$$\cdot \prod_{t=2}^{T} \Phi\{ (y_{i,t-1} \frac{\beta_y}{\sigma_u} + y_{i,t-1} z_{it}' \frac{\beta_{yz}}{\sigma_u} + w_{it}' \frac{\beta}{\sigma_u}$$
$$+ \bar{x}_i' \frac{\mu}{\sigma_u} + \zeta \frac{\sigma_\eta}{\sigma_u}) (2y_{it} - 1) \} \phi(\zeta) d\zeta].$$

The identified parameters for $y_{i2}, ..., y_{iT}$:

$$\frac{\beta_y}{\sigma_u}, \frac{\beta_{yz}}{\sigma_u}, \frac{\beta}{\sigma_u}, \frac{\mu}{\sigma_u}, \frac{\sigma_\eta}{\sigma_u};$$

σ_η / σ_u shows the importance of η_i. It should have been $\delta_i = \bar{w}_i' \mu + \eta_i$: the coefficients of c_i in w_{it} includes those from $\bar{w}_i$.

Female Work or Not (KLIPS panel; Lee and Tae (2005))			
y_{t-1}	0.571 (2.57)	ed3	-0.201 (-2.42)
y_{t-1}*married	0.664 (4.18)	ed4	-0.624 (-1.60)
y_{t-1}*ed4	0.327 (1.67)	ed5	-0.809 (-3.62)
y_{t-1}*age20	-0.538 (-2.31)	ed6	1.541 (3.25)
y_{t-1}*age30	-0.547 (-2.09)		
age	0.289 (11.1)		
age2	-0.335 (-11.5)		
ch1	0.030 (0.30)	$\overline{ch1}$	-1.289 (-6.54)
ch2	-0.002 (-0.03)	$\overline{ch2}$	-0.286 (-1.68)
ch3	-0.135 (-1.67)	$\overline{ch3}$	0.318 (2.76)
age20*ed3	0.356 (3.49)	age30*ed3	0.187 (1.92)
age20*ed4	1.988 (4.88)	age30*ed4	0.342 (0.84)
age20*ed5	2.411 (8.80)	age30*ed5	1.425 (5.81)
income	-0.025 (-3.75)	$\overline{income}$	-0.204 (-7.37)
job training	0.298 (2.40)	$\overline{job\ training}$	2.030 (6.16)
married	-1.485 (-4.46)	$\overline{married}$	0.656 (2.07)
σ_η / σ_u	1.395 (15.4)		

2.4.1 Dynamic Count Response 'Integer-valued AR(1) process' is

$$y_t = \rho \circ y_{t-1} + v_t, \quad 0 < \rho < 1,$$

where $\rho \circ y_{t-1}$ is $B(y_{t-1}, \rho)$—binomial with #trials y_{t-1} and success probability ρ, and v_t is Poisson with parameter λ independent of $\rho \circ y_{t-1}$; v_t, $t = 1, ..., T$, are independent.

Motivated by

$$E(y_t|y_{t-1}) = \rho \cdot y_{t-1} + \lambda,$$

Blundell et al. (2002) specify λ as a function of regressors:

$$E(y_t|\delta, Y_{t-1}, W_t) = E(y_t|\delta, y_{t-1}, w_t) = \rho y_{t-1} + \exp(\tau_t + \delta + x_t'\beta)$$
$$\text{where } W_t \equiv (w_1, ..., w_t)', \quad Y_t \equiv (y_1, ..., y_t)'.$$

The first equality is a PRE, and the second is specifying a regression function.

To derive moment conditions, define

$$\lambda_t \equiv \exp(\tau_t + \delta + x_t'\beta), \quad e_t \equiv y_t - \rho y_{t-1} - \lambda_t$$
$$\Longrightarrow y_t = \rho y_{t-1} + \lambda_t + e_t, \quad E(e_t|\delta, Y_{t-1}, W_t) = 0.$$

Also define

$$\begin{aligned} s_t(\rho, \gamma) &\equiv (y_t - \rho y_{t-1})\frac{\lambda_{t-1}}{\lambda_t} - (y_{t-1} - \rho y_{t-2}) \\ &= (\lambda_t + e_t)\frac{\lambda_{t-1}}{\lambda_t} - (\lambda_{t-1} + e_{t-1}) = e_t\frac{\lambda_{t-1}}{\lambda_t} - e_{t-1} \end{aligned}$$

to get the moment condition:

$$\begin{aligned} & E\{s_t(\rho, \gamma)|\delta, Y_{t-2}, W_{t-1}\} \\ = \; & E\{\, E(e_t\frac{\lambda_{t-1}}{\lambda_t}|\delta, Y_{t-1}, W_t)\; |\delta, Y_{t-2}, W_{t-1} \,\} \\ = \; & E\{\, \frac{\lambda_{t-1}}{\lambda_t} \cdot E(e_t|\delta, Y_{t-1}, W_t)\; |\delta, Y_{t-2}, W_{t-1} \,\} = 0. \end{aligned}$$

Apply nonlinear GMM. The case with $\rho = 0$ was proposed by Chamberlain (1992) and Wooldridge (1997).

2.4.3 Dynamic Censored Response (not practical yet with convergence problem)

For a dynamic censored response

$$y_{it} = \max(0,\ \gamma y_{i,t-1} + x_{it}'\beta\ +\ \delta_i + u_{it}),$$

suppose $\gamma \geq 0$ and

$$u_t, u_s \text{ are identically distributed given } (x_t, x_s, \delta).$$

Honoré (1993) defines 'pseudo-residual'

$$\begin{aligned} e_{ts}(\gamma,\beta) &\equiv \max(0,\ (x_t - x_s)'\beta,\ y_t - \gamma y_{t-1}) - x_t'\beta \\ &= \max(-x_t'\beta,\ -x_s'\beta,\ y_t - \gamma y_{t-1} - x_t'\beta) \\ &= \max(-x_t'\beta,\ -x_s'\beta,\ \delta + u_t). \end{aligned}$$

Since e_{ts} and e_{st} are identically distributed,

$$\begin{aligned} 0 \quad &= \quad E(e_{ts} - e_{st}|\delta, x_t, x_s) \\ \Longrightarrow \quad & 0 = E[\ \{\max(0,\ \Delta x_{ts}'\beta,\ y_t - \gamma y_{t-1}) - \Delta x_{ts}'\beta \\ & - \max(0,\ -\Delta x_{ts}'\beta,\ y_s - \gamma y_{s-1})\} \cdot (\text{functions of } x_t, x_s)\ ]. \end{aligned}$$

If

$$u_t, u_s \text{ are exchangeable given } (x_t, x_s, \delta),$$

then $(e_{ts} - e_{st})|(x_t, x_s, \delta)$ is symmetric around 0, which implies

$$E\{h(e_{ts} - e_{st})|\delta, x_t, x_s\} = 0 \quad \text{for any } h \text{ with } h(a) = h(-a).$$

Honoré and Hu (2004) strengthen the assumption to

$$u_1, ..., u_T \text{ are iid given } (x_1, ..., x_T, \delta, y_0) \qquad \text{((IID))}$$

to propose a version identified globally.

Hu (2002) considers

$$y_{it} = \max(0,\ \gamma y^*_{i,t-1} + x'_{it}\beta + \delta_i + u_{it});$$

the latent, not observed, lagged response appears. This is relevant if the censoring is only a data problem while the economic agent experiences the latent variable; e.g., top-coded income or censored duration. In this case, use ($t - s > 1$ under $T \geq 4$)

$$E\{\ 1[y_{s-1}> 0, y_s> 0, y_{t-1}> 0, y_t> 0]\ h(e_{ts} - e_{st})\ |\delta, x_t, x_s\ \} = 0.$$

The main proposal of Hu (2002) is in fact a version requiring only $T = 3$ under (IID).

APPENDIX: How to Get SD or CI For an estimator b_N (for β) maximizing

$$Q_N(b) = \sum_i q(z_i, b),$$

the asymptotic variance for b_N can be estimated by (omit z_i)

$$\{\sum_i q_{ibb'}(b_N)\}^{-1} \cdot \sum_i q_{ib} q_{ib'}(b_N) \cdot \{\sum_i q_{ibb'}(b_N)\}^{-1}$$

where q_{ib} and $q_{ibb'}$ are the first and second derivatives of $q(z_i, b) \equiv q_i(b)$.

If q_{ib} is $k \times 1$, then its gth component at b_N can be obtained numerically with

$$\frac{q_i(b_N + \varepsilon \cdot c_g) - q_i(b_N - \varepsilon \cdot c_g)}{2\varepsilon}$$

where ε is a small constant, say 0.00001, and c_g is the gth column in I_k. $q_{ibb'}$ can be obtained applying this process to q_{ib}.

Alternatively, use *bootstrap percentile method* to get confidence intervals (CI). Draw N pseudo observations randomly with replacement from the original sample to get a pseudo sample and the pseudo estimate $b_N^{(1)}$. Repeat this, say, 500 times to get $b_N^{(1)}, ..., b_N^{(500)}$. Obtain the lower and upper 2.5% quantiles which yield a 95% CI.

Define $z_{\alpha/2}$ as the $(\alpha/2)$th quantile of $N(0,1)$; i.e., $\alpha/2 = \Phi(z_{\alpha/2})$. Denoting the empirical dist. function of the pseudo estimates as K and the $N(0,1)$ dist. function as Φ, the CI is

$$[K^{-1}(\alpha/2), \ \ K^{-1}(1-\frac{\alpha}{2})] = [K^{-1}\{\Phi(z_{\alpha/2})\}, \ \ K^{-1}\{\Phi(z_{1-\alpha/2})\}].$$

A '*biased-corrected (BC)*' CI is

$$K^{-1}\{\Phi\{z_{\alpha/2} + 2\Phi^{-1}(K(b_N))\}\}, \ \ K^{-1}\{\Phi\{z_{1-\alpha/2} + 2\Phi^{-1}(K(b_N))\}\}.$$

If b_N is the median in the pseudo estimates, then $K(b_N) = 0.5$ and $\Phi^{-1}(K(b_N)) = 0$; no BC. If $b_N <$ *median*, then $K(b_N) < 0.5$, and

$$\Phi^{-1}(K(b_N)) < 0 \implies \text{BC CI shifts left.}$$

REFERENCES

Arellano, M., 2003, Panel data econometrics, Oxford U. Press.

Arellano, M. and B. Honoré, 2001, Panel data models, in Handbook of Econometrics V, ed., J.J. Heckman and E.E. Leamer, Amsterdam, North-Holland.

Blundell, R., R. Griffith, and F. Windmeijer, 2002, Individual effects and dynamics in count data models, J. Econometrics 108, 113-131.

Chamberlain, G., 1982, Multivariate regression models for panel data, J. Econometrics 18, 5-46.

Chamberlain, G., 1984, Panel data, in Handbook of Econometrics II, ed., Z. Griliches and M. Intrilligator, North Holland.

Chamberlain, G., 1992, Comment: Sequential moment restrictions in panel data, J. Bus. and Econ. Stat. 10, 20-26.

Hausman, J.A., B.H. Hall, and Z. Griliches, 1984, Econometric models for count data with an application to the patents-R&D relationship, Econometrica 52, 909-938.

Heckman, J.J., 1981, The incidental parameters problem and the problem of initial conditions in estimating a discrete time discrete data stochastic process, in C.F. Manski and D. McFadden (ed.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press.

Holtz-Eakin, D., W. Newey, and H.S. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

Holtz-Eakin, D., W. Newey, and H.S. Rosen, 1989, The Revenue-expenditure nexus: Evidence from local government data, Intl. Econ. Rev. 30, 415-429.

Honoré, B.E., 1992, Trimmed LAD and LSE of truncated and censored regression models with fixed effects, Econometrica 60, 533-565.

Honoré, B.E., 1993, Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables, J. Econometics 59, 35-61.

Honoré, B.E. and L. Hu, 2004, Estimation of cross sectional and panel data censored regression models with endogeneity, J. Econometrics 122, 293-316.

Hsiao, C., 2003, Analysis of panel data, 2nd ed., Cambridge U. Press.

Hu, L., 2002, Estimation of a censored dynamic panel data model, Econometrica 70, 2499-2517.

Kang, S.J. and M.J. Lee, 2003, Analysis of private transfers with panel fixed effect censored model estimator, Economics Letters 80, 233-237

Lee, M.J., 2002, Panel data econometrics: methods-of-moments and limited dependent variables, Academic Press.

Lee, M.J. and Y.H. Tae, 2005, Analysis of Labor-Participation Behavior of Korean Women with Dynamic Probit and Conditional Logit, Oxford Bulletin of Econ. and Stat. 67, 71-91.

Wooldridge, J.M., 1997, Multiplicative panel data models without the strict exogeneity assumption, Econometric Theory 13, 667-678.

Wooldridge, J.M., 1999, Distribution-free estimation of some nonlinear panel data models, J. Econometrics 90, 77-97.