Lecture 16

Panel DataModels (II)

– Non-linear Panel Models –

Read
- Greene WP (survey paper)
- Panel RE Models Review in Stata
- Survey paper by Arellano & Bonhomme

Panel Modelling (as we know)

i) FE $\quad y_{it} = \beta' x_{it} + \alpha_i + \varepsilon_{it}$

ii) RE $\quad y_{it} = \beta' x_{it} + u_{it}, \quad u_{it} = \alpha_i + \varepsilon_{it}$

iii) Random parameters* $\quad y_{it} = \beta_i' x_{it} + \varepsilon_{it}, \quad \beta_i = \beta + v_i, \quad E(v_i) = 0$, $Var(v_i) = \Sigma$

$$\Rightarrow y_{it} = (\beta + v_i)' x_{it} + \varepsilon_{it} = \beta' x_{it} + u_{it}, \quad u_{it} = v_i' x_{it} + \varepsilon_{it}$$

.. groupwise heteroskedastic model; somewhat restrictive

eg) $\log L_i = \sum_t \log \Phi\left(q_i \dfrac{\beta' x_i}{1 + x_i \Sigma x_i} \right)$

(1) Nonlinear Models with Fixed Effects

$$\log \mathcal{L} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log g(y_{it}, \beta' x_{it} + \alpha_i, \theta)$$

$$\frac{\partial \ln L}{\partial \beta} = 0, \quad \frac{\partial \ln L}{\partial \alpha_i} = 0 \ (i = 1, \cdots, N), \quad \frac{\partial \ln L}{\partial \theta} = 0$$

(incidental parameter problem!)

Another key problem: $\hat{\beta}$ is a function of $\hat{\alpha}_i$, while $\hat{\alpha}_i$ is biased when T_i is small. Thus, $\hat{\beta}$ becomes biased!

$$Var(\hat{\alpha}_i) = O(T_i^{-1}) \Rightarrow Var(\hat{\beta}) = O(T_i^{-1})$$

Note "A sequence c_T is of order T^δ, denoted $O(T^\delta)$ if plim $\frac{1}{T^\delta} c_T$ is a finite non-zero constant."

eg) $Var(\hat{\beta}) \dfrac{1}{T_i^{-1}} = T_i \cdot Var(\hat{\beta}) \rightarrow$ something

Note "A sequence c_T is of order less than T^δ, denoted $o(T^\delta)$ if plim $\frac{1}{T^\delta} c_T$ equals 0."

* Conditioning-out Method for FE Models

there are a few cases where the conditional density provide consistent estimator of β.

Then, the conditional likelihood function is not a function of α_i.

Examples

i) FE logit model

" $\sum_{t}^{T_i} y_{it}$ is a sufficient statistic "

the conditional density, conditional on $\sum_{t=1}^{T_i} y_{it}$, is free of α_i.

ii) FE Poisson model

$$\log L = \sum_i \sum_t^{T_i} \left[-\exp(\alpha_i)\exp(\beta' X_{it}) + y_{it}(\beta' X_{it} + \alpha_i) - \log y_{it}! \right]$$

$- \log y_{it}!$ can be ignored

$$\frac{\partial \ln L}{\partial \alpha_i} = -\exp(\alpha_i)\sum_{t=1}^{T_i}\left[\exp(\beta' X_{it}) + \sum_{t=1}^{T_i} y_{it}\right] \Rightarrow 0$$

$$\Rightarrow \exp(\alpha_i) = \frac{\sum_{t=1}^{T_i} y_{it}}{\sum_{t=1}^{T_i} \exp(\beta' X_{it})}$$

plug this into $\log \mathcal{L}$ to obtain the concentrated likelihood function.

then, it is free of α_i.

this also works for neg. binomial models.

iii) Cox-proportional duration model
(not panel, but adopts a similar idea)

Note this method does not work for probit and Tobit models. It works for linear exponential conditional mean equations

Read Greene WP (p. 14)

Also, Maddala book, p. 245.

(2) Nonlinear Models with Random Effects

$$\mathcal{L}_i = f(y_{i1}, \cdots, y_{iT_i}) = \int f(y_{i1}, \cdots, y_{iT_i} \mid u_i) f(u_i)\, du_i \quad \text{or } f(y_{it}, \cdots, y_{iT_i}, u_i)$$

$$= \int \prod_{t=1}^{T_i} f(y_{it} \mid u_i) f(u_i)\, du_i$$

- Three cases

i) Exact integration & closed form

Poisson regression with a gamma density

$$\underset{f(y_{it} \mid u_i)}{\uparrow} \qquad \underset{f(u_i)}{\uparrow}$$

"this works for a cross-sectional count model as well."

ii) Approximation by Hermite quadrature

Butler & Moffitt's (1982) method, using a normal density $f(u_i)$: earlier note

$$\int \prod_{t=1}^{T_i} f(y_{it} \mid u_i) f(u_i)\, du_i$$

$$= \frac{1}{\sigma_u \sqrt{2\pi}} \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} g(y_{it}, \beta' X_{it} + u_i, \theta) \exp\left(\frac{-u_i^2}{2\sigma^2}\right) du_i$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} g(y_{it}, \beta' X_{it} + \sigma v_i, \theta) \cdot \frac{1}{\sqrt{\pi}} \exp(-v_i^2)\, dv_i$$

$$= \sum_{h=1}^{H} w_h \prod_{t=1}^{T_i} g(y_{it}, \beta' X_{it} + \sigma z_h, \theta)$$

where z_h & w_h are nodes and weights for the Hermite quadrature of degree H.

Applications : RE probit (logit), RE poisson (also neg bin, intreg)
RE Tobit, RE Selection, ..
(almost all RE models & PA models) "population average"

iii) Simulated Maximum Likelihood

$$\int \underbrace{\prod_{t=1}^{T_i} f(y_{it}|u_i)\cdot f(u_i)\,du_i}$$

$$= E\left[\prod_{t=1}^{T_i} f(y_{it}|u_i)\right] \overset{let}{=} E[F(u_i|\theta)]$$

If $(u_{i1}, \dots, u_{iR})$ is a random sample of iid draws from multi-variate distributions, then

$$\text{plim}\ \frac{1}{R}\sum_{r=1}^{R} F(u_{ir}|\theta) = E[F(u_i|\theta)]$$

the simulated integral can be inserted into the likelihood function. Then do MLE.

Note · How many R should be used is a question.
· Does it produce the same results as in the exactly integrated counterpart.
see references (wp, p. 18)

(3) Generalized Equation Estimation (GEE) : STATA covers this!
Also called Generalized Linear Models (GLM)
by McCullagh & Nelder (1983) -- popular in other applied sciences

GEE (GLM) fits generalized linear models of y

$$g\{E(y)\} = X\beta, \quad y \sim F$$

i) $g(\ )$ is the canonical link
ii) F ~ distribution

Eg 1) $E(y) = X\beta$, $y \sim$ Normal
link = identity, $F \sim$ normal

eg2) $\text{logit}\{E(y)\} = X\beta$, $y \sim$ Bernoulli

$\log\left[\frac{E(y)}{1-E(y)}\right]$

link : logit function

F : Bernoulli

eg3) $\log[E(y)] = X\beta$, $y \sim$ Poisson

link : log function $\leftarrow E(y) = \exp(X\beta)$

F : Poisson

eg4) $\Phi^{-1}(E(y)] = X\beta$, $y \sim$ Bernoulli

link : inverse Gaussian cumulative.

Estimation : Iterative weighted LS using the generalized estimating equation.

For details, see McCullagh & Nelder (1983)

Note i) The method is extended from the quasi-score function by Wedderburn (1974).

ii) there may be many possible links, and there are no theoretical justifications.

(see Greene, WP. p25)

Only a few links make sense.

(e.g. count data model

: only log-transformation appears proper

probit or logit link's do not make sense.]

"advantage" $\Rightarrow$ iii) In panel models, group correlations can be modeled. ("xtgee" in Stata)

; $\text{Corr}(\varepsilon_{it}, \varepsilon_{is}) = \rho_{ts}$

(4) Latent Class Models (LIMDEP version 8.0 covers this!)

Express

$$f(y_{it}) = \sum_{j=1}^{M} P_j \, g(y_{it}, \beta' x_{it} + \alpha_j, \theta)$$: M latent classes!

Not α_i.

where $0 \le P_j \le 1$, $\sum_j P_j = 1$

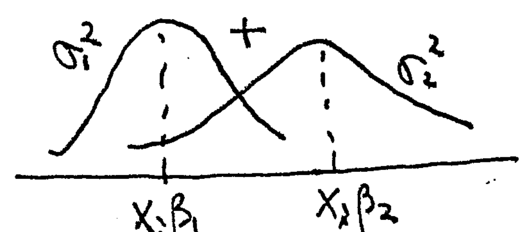
point: If there are heterogeneity, heterogeneous groups can be divided by M groups. Then the distribution is given as a MIXTURE of distributions with different parameters (α_j).

Note A regime switching model is a special case of such treatment. ($M=2$)

Quandt & Ramsey (1978) : regime switching
Maddala & Nelson (1975) : disequilibrium model

$$y_i = \begin{cases} X_i \beta_1 + e_{i1} & \text{with prob } \lambda \\ X_i \beta_2 + e_{i2} & \text{with prob } 1-\lambda \end{cases}$$

$$f(y_i) = \lambda \cdot \frac{1}{\sigma_1} \phi\left(\frac{y_i - X_i\beta_1}{\sigma_1}\right) + (1-\lambda) \frac{1}{\sigma_2} \phi\left(\frac{y_i - X_i\beta_2}{\sigma_2}\right)$$



heterogeneous parameters in this example $= (\beta_1, \beta_2)$ & (σ_1^2, σ_2^2) ⇒

: weighted average of two mixture of distributions. (weight $= \lambda$)

<u>Note</u> In a cross-sectional model, the unobserved heterogeneity can be modeled as θ_j

... Firstly suggested by Heckman & Singer (1984)

: duration models

Example) Probit Model : Keane & Geweke (1999)

$$y_i = 1(\beta' x_i + \varepsilon_i > 0)$$

$$f(\varepsilon_i) = \sum_{j=1}^{M} p_j \frac{1}{\sigma_j} \phi\left(\frac{\varepsilon_i - \mu_j}{\sigma_j}\right)$$

(parameters : p_j, μ_j, σ_j)

$$P(y_i = 1) = \int_{-\beta' x_i}^{\infty} \left[\sum_{j=1}^{M} p_j \left(\frac{1}{\sigma_j}\right) \Phi\left(\frac{\varepsilon_i - \mu_j}{\sigma_j}\right)\right] d\varepsilon_i$$

<u>Note</u> $\sum_{j}^{M} p_j = 1$ need to be satisfied.

We can consider

$$p_j = \frac{e^{\theta_j}}{\sum_{j=1}^{M} e^{\theta_j}}, \quad j = 1, \cdots, M, \quad \theta_M = 0$$

(normalization)

- we may write $\theta_j = \gamma_j' z_i$ (dependent on z_i)

- In panel

$$p_{ij} = \frac{e^{\theta_{ij}}}{\sum_{j=1}^{M} e^{\theta_{ij}}}, \quad j = 1, \cdots, M, \quad \theta_M = 0$$

<u>Estimation</u> MLE, E-M algorithm

$\frac{\partial \ln L}{\partial p} = 0 \Rightarrow \hat{p}_j$ is a function of $\hat{\beta}$] iterates!

$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \hat{\beta}$ " $\hat{p}_j$

Summary of Nonlinear panel data Models

"INT" "Integrating out" method (Gauss-Hermite quadrature)

"CONT" "Conditioning out" method (Conditional likelihood)

"SML" simulated ML.

	FE	RE	SML	Latent class (LIMDEP)	GEE (STATA)
(1) Probit	dummy (inconsistent)	INT		✓	✓
(2) Logit	CONT	INT		✓	✓
(3) Poisson	CONT	INT		✓	✓
(4) Neg Bin	CONT	INT		✓	✓
(5) TOBIT (Truncation)	dummy (inconsistent) *	INT	✓ (truncation)	✓	
(6) SELECTION	dummy (inconsistent) *	—	✓	✓	
(7) Interval Regression	—	INT		✓	
(8) Multiple Responses	—	INT (ordered logit/probit)	✓ (multinomial probit)		
(9) Frontier	dummy (inconsistent)	INT		✓	

Note Some recent developments are discussed in the survey paper. they include semi-parametric methods.