Logit/Probit using Panel Data

The model: $p_{it} = \Pr[y_{it} = 1] = E(y_{it} | x_{it}) = F(x_{it}'\beta)$

and

$$y_{it} = 1 \ \ if \ \ y_{it}^* > 0$$
$$= 0 \ \ if \ \ y_{it}^* \leq 0$$

where: $y_{it}^* = x_{it}'\beta + u_{it}$

Then,

$$\Pr[y_{it} = 1] = \Pr[y_{it}^* > 0] = \Pr[u_{it} > -x_{it}'\beta] = F(x_{it}'\beta)$$

Fixed-Effects Logit Model

For panel data, it is more likely to have individual fixed effect. Then, the model can be

$$y_{it}^{*} = x_{it}'\beta + \alpha_i + \varepsilon_{it}$$

with

$$\Pr[y_{it} = 1] = \Pr[y_{it}^{*} > 0] = \Pr[\varepsilon_{it} > -x_{it}'\beta - \alpha_i] = F(x_{it}'\beta + \alpha_i)$$

Equality of the last term can be held if F is symmetric around zero.

However, with the fixed effect α_i and a fixed T, when $n \to \infty$, number of α_i will increase, then, α_i cannot be consistently estimated for fixed T.

This is called *incidental parameters problem.*

Max. Conditional Likelihood Function

To solve the problem, maximize conditional likelihood function:

$$L_c = \prod_{i=1}^{N} \Pr\left(y_{i1}, \ldots, y_{iT} \middle| \sum_{t=1}^{T} y_{it} \right)$$

<u>Proof</u>: Assume 2 periods

<u>Unconditional Function</u>: $L_c = \prod_{i=1}^{N} \Pr(y_{i1}) \Pr(y_{i2})$

The condition $\sum_{t=1}^{2} y_{it} = y_{i1} + y_{i2}$ can be *0*, *1*, or *2*

1. $y_{i1} + y_{i2} = 0$ then $\Pr[y_{i1} = 0, y_{i2} = 0 | y_{i1} + y_{i2} = 0] = 1$
2. $y_{i1} + y_{i2} = 2$ then $\Pr[y_{i1} = 1, y_{i2} = 1 | y_{i1} + y_{i2} = 2] = 1$

For the first two cases $\log 1 = 0$, log-likelihood will not change, thus, they are not matter.

Max. Conditional Likelihood Function

<u>Proof</u>: Assume 2 periods (Cont.)

For Logit model: $\Pr[y_{it}=1]=\dfrac{e^{\alpha_i+x_{it}'\beta}}{1+e^{\alpha_i+x_{it}'\beta}}$

3.1 $y_{i1}+y_{i2}=1$ then

$$\Pr[y_{i1}=1,y_{i2}=0\,|\,y_{i1}+y_{i2}=1]=\frac{\Pr[y_{i1}=1,y_{i2}=0]}{\Pr[y_{i1}+y_{i2}=1]}$$

$$=\frac{e^{\alpha_i+x_{i1}'\beta}}{e^{\alpha_i+x_{i1}'\beta}+e^{\alpha_i+x_{i2}'\beta}}=\left(\frac{e^{\alpha_i+x_{i1}'\beta}}{e^{\alpha_i+x_{i1}'\beta}}\right)\frac{e^{\alpha_i+x_{i1}'\beta}}{e^{\alpha_i+x_{i1}'\beta}+e^{\alpha_i+x_{i2}'\beta}}=\frac{1}{1+e^{(x_{i2}-x_{i1})'\beta}}$$

3.2 $y_{i1}+y_{i2}=1$ then

$$\Pr[y_{i1}=0,y_{i2}=1\,|\,y_{i1}+y_{i2}=1]=\frac{\Pr[y_{i1}=0,y_{i2}=1]}{\Pr[y_{i1}+y_{i2}=1]}$$

$$=\frac{e^{(x_{i2}-x_{i1})'\beta}}{1+e^{(x_{i2}-x_{i1})'\beta}}$$

Fixed-Effects Probit Model

By using maximize conditional log-likelihood function, fixed-effect logit can be consistently estimated.

Due to different functional form, maximize conditional log-likelihood function will not lead to consistent estimators of the fixed-effect probit model.

Test for Fixed-Effects

Hausman-type test based on the difference between conditional MLE and logit MLE can be used and is χ^2 with K degree of freedom.

Random-Effects Model

Random effects MLE assumes $\alpha_i \sim N[0, \sigma_\alpha^2]$

The model can be estimated by maximizing the following log-likelihood function:

$$\sum_{t=1}^{n_i} \ln f(\mathrm{y}_{it} | \mathrm{x}_{it}, \beta, \sigma_\alpha^2)$$

where:

$$f(\mathrm{y}_i | \mathrm{X}_i, \beta, \sigma_\alpha^2) = \int_{-\infty}^{\infty} f(\mathrm{y}_i | \mathrm{X}_i, \beta, \sigma_\alpha^2) \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left(\frac{-\alpha_i}{2\sigma_\alpha^2}\right)^2 d\alpha_i$$

Since there is no closed-form solution for this log-likelihood function, the estimation is usually employed numerically using *Guass-Hermite quadrature methods*.

Test Pooled vs Panel Logit Estimators

To test whether the pooled estimator or the panel estimator should be employed, the test whether $\rho = 0$ should be performed.

where: $$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$$

If $\rho = 0$, it means that panel-level variance component σ_α^2 is unimportant.

Then, panel logit estimator is no different from the pooled logit estimator.

Test Pooled vs Panel Probit Estimators

To test whether the pooled probit estimator or the panel probit estimator should be employed, the test whether $\rho = 0$ should be performed.

where: $$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + 1}$$

If $\rho = 0$, it means that panel-level variance component σ_α^2 is unimportant.

Then, panel probit estimator is no different from the pooled probit estimator.