Limited Dependent Variable Model in Panel Data

Contents

Panel Probit/Logit Model

We consider the following model

$$y^*_{it} = x_{it}\beta + u_i + \epsilon_{it}$$

In which u_i is time-invariant unobservable term and

$$y_{it} = 1(0) \quad if \quad y^*_{it} > 0(< 0)$$

of course we still have

$$pr(y_{it} = 1|x_{it}, u_i) = pr(\epsilon_{it} > -x_{it}\beta - u_i)$$

if ϵ_{it} follows normal distribution, then

$$pr(y_{it} = 1|x_{it}, u_i) = \Phi(-x_{it}\beta - u_i)$$

if ϵ_{it} follows logit distribution, then

$$pr(y_{it} = 1|x_{it}, u_i) = \frac{exp(x'_{it}\beta + u_i)}{1 + exp(x'_{it}\beta + u_i)}$$

Under either condition, we can write down the likelihood of a person i:

$$f(y_{i1},,,y_{iT}|u_i, x_{i1},,,x_{iT}) \equiv \prod_{t=1}^{T} [pr(y_{it} = 1|x_{it}, u_i)]^{y_{it}} [pr(y_{it} = 0|x_{it}, u_i)]^{1-y_{it}}$$

$$= \prod_{t=1}^{T} [\Lambda(x'_{it}\beta + u_i)]^{y_{it}} [1 - \Lambda(x'_{it}\beta + u_i)]^{1-y_{it}}$$

On the other hand we have

$$f_y(y_{i1},,,y_{iT}|x_{i1},,,x_{iT}) = \int f(y_{i1},,,y_{iT}, u_i|x_{i1},,,x_{iT})du_i$$

$$= \int f(y_{i1},,,y_{iT}|u_i, x_{i1},,,x_{iT})g(u_i|x_{i1},,,x_{iT})du_i \tag{1}$$

in which we use g to denote the distribution of u_i.We next discuss how to deal with u_i.

u_i Independent of $x_{i1},,x_{iT}$

This is random effect model .Under this assumption, it is quite easy to process. The likelihood function for person i is now

$$\int f(y_{i1},,,y_{iT}|u_i,x_{i1},,,x_{iT})g(u_i|x_{i1},,,x_{iT})du_i$$

$$= \int f(y_{i1},,,y_{iT}|u_i,x_{i1},,,x_{iT})g(u_i)du_i$$

$$= \int \prod_{t=1}^{T} [\Lambda(x'_{it}\beta + u_i)]^{y_{it}} [1 - \Lambda(x'_{it}\beta + u_i)]^{1-y_{it}} g(u_i)du_i$$

We can calculate this integral numerically if we know the distribution of $g(u_i)$.

u_i Not independent of $x_{i1},,x_{iT}$

This is fixed effect model. Under this assumption, it is often very hard to calculate $g(u_i|x_{i1},,,x_{iT})$! Then, following the idea in the linear panel model, we may want to get rid of u_i, i.e., let the u_i disappear in the likelihood function!

To do this, let's see the (1) again. The distribution of $y_{i1},,,y_{iT}$ depends on u_i. Recall the definition of'sufficient statistics': conditioning on the sufficient statistics, the distribution of sample has nothing to do with population parameter. Can we find a 'sufficient statistics' for $y_{i1},,,y_{iT}$ such that, conditioned on the this statistics, the joint distribution of $y_{i1},,,y_{iT}$ does not depend on u_i?

It turns out that if the ϵ_{it} follows logit distribution, and by conditioning on $\sum_{t=1}^{T} y_{it} = s$, the conditional likelihood function for $y_{i1},,,y_{iT}$ would no longer contain the μ_i

To make this idea concrete, we assume there is only two periods, 1, and 2. Therefore we have

- If $\sum_{t=1}^{T} y_{it} = 0$, Then it is certain that $y_{i1} = y_{i2} = 0$. Therefore we have

$$pr(y_{i1} = 0, y_{i2} = 0|u_i, x_{i1}, x_{i2}, \sum_{t=1}^{T} y_{it} = 0) = 1$$

- If $\sum_{t=1}^{T} y_{it} = 2$, Then it is certain that $y_{i1} = y_{i2} = 1$. Therefore we have

$$pr(y_{i1} = 1, y_{i2} = 1|u_i, x_{i1}, x_{i2}, \sum_{t=1}^{T} y_{it} = 2) = 1$$

- If $\sum_{t=1}^{T} y_{it} = 1$, Then there are two possible cases: $y_{i1} = 1, y_{i2} = 0$,or $y_{i1} = 0, y_{i2} = 1$ Therefore we have

$$pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2}, \sum_{t=1}^{T} y_{it} = 1) = \frac{pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2})}{pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2}) + pr(y_{i1} = 0, y_{i2} = 1|u_i, x_{i1}, x_{i2})}$$

$$= \frac{\frac{exp(x'_{i1}\beta+u_i)}{1+exp(x'_{i1}\beta+u_i)} \frac{1}{1+exp(x'_{i2}\beta+u_i)}}{\frac{exp(x'_{i1}\beta+u_i)}{1+exp(x'_{i1}\beta+u_i)} \frac{1}{1+exp(x'_{i2}\beta+u_i)} + \frac{exp(x'_{i2}\beta+u_i)}{1+exp(x'_{i2}\beta+u_i)} \frac{1}{1+exp(x'_{i1}\beta+u_i)}} = \frac{exp(x'_{i1}\beta + u_i)}{exp(x'_{i1}\beta + u_i) + exp(x'_{i2}\beta + u_i)}$$

$$= \frac{exp(x'_{i1}\beta)}{exp(x'_{i1}\beta) + exp(x'_{i2}\beta)}$$

similarly we have

$$pr(y_{i1} = 0, y_{i2} = 1|u_i, x_{i1}, x_{i2}, \sum_{t=1}^{T} y_{it} = 1) = \frac{exp(x'_{i2}\beta)}{exp(x'_{i1}\beta) + exp(x'_{i2}\beta)}$$

We can see that, due to the logit distribution, the u_i directly disappears. In summary, the above analysis implies that, for each individual, conditioning on his $\sum_{t=1}^{T} y_{it}$, we can easily find his 'conditional likelihood function'

$$f(y_{i1},,,y_{iT}|u_i, \sum_{t=1}^{T} y_{it}) = f(y_{i1},,,y_{iT}|\sum_{t=1}^{T} y_{it})$$

which is free of u_i. It is easy to see:

- Advantage: Since we only focus on the conditional likelihood function, then u_i is thrown away, and we do not need to do the integral. very convenient!
- Disadvantage: We lose the information for those individuals with $\sum_{t=1}^{T} y_{it} = T$ or 0.

In general, check the likelihood function condition on some statistics may be a good way to kill the undesirable parameters (in the cost of losing information). Similar techniques are used in the Cox model.

Poisson Regression in Panel Data

We next consider a model in which y_{it} is non-negative integer. Specifically,

$$pr(y_{it} = y|u_i, x_{it}) = \frac{e^{-\lambda_i(x_{it}\beta+\mu_i)}[\lambda_i(x_{it}\beta + \mu_i)]^y}{y!}$$

in which

$$\lambda_i(x_{it}\beta + \mu_i) = e^{x_{it}\beta+\mu_i}$$

Therefore we can write down the likelihood function for i:

$$f(y_{i1},,,y_{iT}|u_i, x_{i1},..,x_{iT}) = \prod_{t=1}^{T} \frac{e^{-\lambda_i(x_{it}\beta+\mu_i)}[\lambda_i(x_{it}\beta + \mu_i)]^{y_{it}}}{y_{it}!}$$

Of course, since u_i is unobservable, we still want to do take integral over it, as is shown in (1) ## u_i Independent of $x_{i1},,x_{iT}$ This is random effect model .Under this assumption, it is quite easy to process. The likelihood function for person i is now

$$\int f(y_{i1},,,y_{iT}|u_i, x_{i1},,,x_{iT})g(u_i|x_{i1},,,x_{iT})du_i$$

$$= \int f(y_{i1},,,y_{iT}|u_i, x_{i1},,,x_{iT})g(u_i)du_i$$

$$= \int \prod_{t=1}^{T} \frac{e^{-\lambda_i(x_{it}\beta+\mu_i)}[\lambda_i(x_{it}\beta + \mu_i)]^{y_{it}}}{y_{it}!} g(u_i)du_i$$

We can calculate this integral numerically if we know the distribution of $g(u_i)$. ## u_i Not independent of $x_{i1},,x_{iT}$ In this situation we have fixed effect model. As is in the panel logit model, we want to write down the 'conditional likelihood function'. We still take $T = 2$ as example. Suppose individual i has $y_{i1} + y_{i2} = 1$, and we want to see $f(y_{i1}, y_{i2}|u_i, x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1)$.There are two possible cases: $y_{i1} = 1, y_{i2} = 0$,or $y_{i1} = 0, y_{i2} = 1$.

$$pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2}, \sum_{t=1}^{T} y_{it} = 1) = \frac{pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2})}{pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2}) + pr(y_{i1} = 0, y_{i2} = 1|u_i, x_{i1}, x_{i2})}$$

Since

$$pr(y_{i1} = 1, y_{i2} = 0|u_i, x_{i1}, x_{i2}) = \frac{e^{-\lambda_i(x_{i1}\beta+\mu_i)}[\lambda_i(x_{i1}\beta + \mu_i)]^1}{1!} \frac{e^{-\lambda_i(x_{i2}\beta+\mu_i)}[\lambda_i(x_{i2}\beta + \mu_i)]^0}{0!}$$

$$pr(y_{i1}=0, y_{i2}=1|u_i, x_{i1}, x_{i2}) = \frac{e^{-\lambda_i(x_{i1}\beta+\mu_i)}[\lambda_i(x_{i1}\beta+\mu_i)]^0}{0!}\frac{e^{-\lambda_i(x_{i2}\beta+\mu_i)}[\lambda_i(x_{i2}\beta+\mu_i)]^1}{1!}$$

therefore we have

$$\frac{pr(y_{i1}=1, y_{i2}=0|u_i, x_{i1}, x_{i2})}{pr(y_{i1}=1, y_{i2}=0|u_i, x_{i1}, x_{i2}) + pr(y_{i1}=0, y_{i2}=1|u_i, x_{i1}, x_{i2})} = \frac{\lambda(x_{i1}\beta+u_i)}{\lambda(x_{i1}\beta+u_i)+\lambda(x_{i2}\beta+u_i)} = \frac{e^{x'_{i1}\beta}}{e^{x'_{i1}\beta}+e^{x'_{i2}\beta}}$$

similarly

$$pr(y_{i1}=0, y_{i2}=1|u_i, x_{i1}, x_{i2}, \sum_{t=1}^{T} y_{it}=1) = \frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta}+e^{x'_{i2}\beta}}$$

Thanks to the poisson distribution and the specification of λ, we get rid of the u_i when conditioning on $\sum_{t=1}^{T} y_{it}$