4. *Binary Response Models – Static Models*

4.1 Introduction

Examples: (i) y=1 individual is unemployed; y=0 is employed

(ii) y=1 union member; (iii) y=1 firm is still operating; etc.

Interested in the **response probability**:

$P(y=1|\mathbf{x}) = P(y=1|x_1,\ldots, x_k)$

Want the partial effect of x_j on the response probability $\frac{\partial P(y=1)}{\partial x_j}$. **IMP**

If x_j is binary, then we would wish to calculate the difference in the response probability between x=1 and x=0 cases.

NOTE: For binary variables, $E(y|x)=P(y=1|x)=p$ say; then $Var(y|x)=p(1-p)$.

MODELS

Model $P(y=1|x) = E(y|x) = G(\mathbf{x}\boldsymbol{\beta})$

- $0<G(.)<1$; $\mathbf{x}\boldsymbol{\beta}$ is called the index.
- Better to take G to be a cdf.

1. Linear Prob Model (LPM): P(y=1|x)=**xβ;**

   Model specified as: y = **xβ** + error

   Advantage – the techniques for linear panel models can be applied.

   However, there are some problems with this model (heterosk, forecasting, effects of a change in x being fixed…...

2. <u>Probit & Logit</u>;

   Std. Normal G – Probit; Logistic G – Logit [G(z)=$\frac{exp(z)}{1+exp(z)}$]

4.2 Latent variable model interpretation

Consider $y^* = \mathbf{x}\boldsymbol{\beta} + e$; $y=1[y^*>0]$

- e is indep over i and cont distributed and indep of **x.**
- Assume the dist of e is symmetric about 0
- This implies….. [1-G(-z))=G(z) for all z].
- Thus, $P(y=1|\mathbf{x}) = P(y^*>0|\mathbf{x}) = P(e>-\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) = 1-G(-\mathbf{x}\boldsymbol{\beta}) = G(\mathbf{x}\boldsymbol{\beta})$
- G(.) is the distribution of e!

- Threshold value does not matter
- Scale normalisation required: $P(y=1|\mathbf{x}) = P(y^{*}>0|\mathbf{x}) = P(y^{*}/\sigma>0)$
- So set probit: variance=1; But logit: var = $\pi^2/3$.
- Sometimes written as: $y = 1\{\mathbf{x}\boldsymbol{\beta} + e > 0\}$; $1\{.\}$ is called the indicator function.

4.3 <u>Coefficient interpretation</u>

$$\frac{\partial P(y=1)}{\partial x_j} = g(\mathbf{x}\boldsymbol{\beta})\,\beta_j \qquad \text{where } g(z) = \frac{dG}{dz}(z)$$

Partial effects depend on **x**. g(z) is positive. So sign of the partial effect is the same as the sign of the coefficient.

Can evaluate the partial effects at different values of x.

4.4 Estimation

LPM can be estimated by OLS or WLS.

Probit/Logit requires MLE.

For each individual, density of y given **x**

$$= f(y|\mathbf{x};\boldsymbol{\beta}) = [G(\mathbf{x}_i\boldsymbol{\beta})]^{y}\,[1-G(\mathbf{x}_i\boldsymbol{\beta})]^{1-y} \qquad y=0,1$$

Likelihood function is globally concave and maximisation of this will give you the max-lik-est.

Can use std tests such as Wald, LR or LM to test most of the hypotheses of interest.

4.5 For Panel Data

$y_{it}^{*} = \mathbf{x_{it}}\beta + c_i + u_{it}$ $\qquad$ $y_{it} = 1[y_{it}^{*} > 0]$

i.e. $y_{it} = 1\{ \mathbf{x_{it}}\beta + c_i + u_{it} > 0\}$ $\quad$ (t=1,…,T; i=1,…, N)

$\text{Prob}(y_{i1}=1|\mathbf{x_{it},}c_i) = P(y_{it}^{*} > 0|\mathbf{x_{it},}c_i) = P(u_{it} > -\mathbf{x_{it}}\beta - c_i|\mathbf{x_{it},}c_i)$

$= 1 - G(-\mathbf{x_{it}}\beta - c_i) = G(\mathbf{x_{it}}\beta + c_i)$ $\quad$ if the dist is sym

4.5.1 Probit

$P(y_{it}|\mathbf{x_{i1},\ldots,x_{iT}}, c_i) = P(y_{it}|\mathbf{x}_{it},c_i)=\Phi(\mathbf{x_{it}}\boldsymbol{\beta}+c_i)$ $\quad t=1,..,T$ (1)

<u>Assume</u> (i) st. exog $\mathbf{x}_{it.}$ (ii) $y_{i1},\ldots, y_{iT}$ are indep conditional on $\mathbf{x}_i$ and c_i.

This gives the density of $(y_{i1},\ldots., y_{iT})$ conditional on $\mathbf{x}_i$ and c_i

$$= f(y_{i1},\ldots., y_{iT}| \mathbf{x}_i,c_i; \beta) = \prod_{t=1}^{T} f(y_{it} \mid \mathbf{x_{it}}, c_i; \boldsymbol{\beta}) \quad (2)$$

where $\prod_{t=1}^{T} f(y_t \mid \mathbf{x_t}, c; \boldsymbol{\beta}) = \left[\Phi\left(\mathbf{x}_t\boldsymbol{\beta}+c\right)\right]^{y_t}\left[1-\Phi\left(\mathbf{x_t}\boldsymbol{\beta}+c\right)\right]^{1-y_t}$ (3)

NOTES:

1. Called **random effects probit.**

2. Can't estimate the c_i as parameters since as N gets larger, the number of c_i gets larger – **incidental parameter problem.**

3. Can't do a transformation to eliminate the c_i prior to estimation (like the WG).

4. So assume c is random and integrate it out of the likelihood to get the unconditional likelihood:

$$f(y_{i1},\ldots, y_{iT}|\mathbf{x}_i; \boldsymbol{\beta}) = \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T} f\,(y_{it} \mid \mathbf{x_{it}}, c_i; \boldsymbol{\beta}) \right] g(c_i)\, dc_i \qquad (4)$$

where it is assumed that $c|\mathbf{x} \sim f(c_i)$ with some parameters. Also note, I have assumed that c and **x** are independent (distributionally).

Zero correlation assumption is not enough.

5. The above integral may not have a closed-form. If so, need to use numerical approximation to the integral: Gaussian-Hermite quadrature or discrete approximation or use simulation methods…..

6. Some popular choices for g(c): Normal or discrete approximation.

7. Under Normality assumption for g(c) $\quad c_i|\mathbf{x_i} \sim N(0,\sigma^2_c)$ (5)

$$f(y_{i1},\ldots, y_{iT}|\mathbf{x}_i; \boldsymbol{\beta}, \boldsymbol{\sigma}^2_{\mathbf{c}}) = \int_{-\infty}^{+\infty}\left[\prod_{t=1}^{T} f\,(y_{it} \mid \mathbf{x_{it}}, c_i; \boldsymbol{\beta})\right] \frac{1}{\sigma_c}\,\varphi\left(\frac{c}{\sigma_c}\right) dc \qquad (6)$$

8. (5) implies that c and **x** are indep and normal.

9. Can measure $\rho = \frac{\sigma_c^2}{1+\sigma_c^2}$ = correlation of (c_i+u_{it}) over time – **fixed.** Can generalise this by allowing different correlations between different periods. But will need to estimate using multivariate probit routines…..

10. OR allow for correlation by assuming a special AR(1) type distribution for u_{it}.

11. Can estimate a pooled probit. Because of the distributional assumptions, c_i+u_{it} will be normally distributed. Will get consistent parameter est of

$[\beta/(1+\sigma_c^2)^{1/2}]$ (Robinson 1982). But std. Errors will be wrong since the serial correlation will not be accounted for. Use outer-product of the score.

11. If not happy with indep assumption, use Mundlak's or Chamberlain's formulation to account for correlation: $c_i|\mathbf{x}_i \sim N(\psi+\bar{\mathbf{x}}_i\boldsymbol{\xi},\sigma_a^2)$

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i \qquad (7)$$

Note the problem with time-invariant variables!

4.5.2 Logit

$$P(y_{it}|\mathbf{x}_{it},c_i)=\frac{exp(\mathbf{x}_{it}+c_i)}{1+exp(\mathbf{x}_{it}+c_i)}=\Lambda(\mathbf{x}_{it}+c_i) \qquad (8)$$

As before assume $y_{i1},\ldots,y_{iT}$ are indep conditional on $\mathbf{x}_i$ and c_i.

1. Can do what we did before with the probit…assume a dist for c and integrate it out of the likelihood function – **see (4)**. **Assumptions important!**

2. OR (unlike in the probit model) we can use Conditional Max Lik to estimate the $\boldsymbol{\beta}$ without specifying the distribution of c – **FE logit!** (Chamberlain (1980)).

The density of $(y_{i1},\ldots, y_{iT})$ conditional on $\mathbf{x}_i$ and c_i

$$= f(y_{i1},\ldots, y_{iT} | \mathbf{x}_i, c_i; \boldsymbol{\beta}) = \prod_{t=1}^{T} f(y_{it} | \mathbf{x}_{it}, c_i; \boldsymbol{\beta}) \qquad (2)$$

where $\prod_{t=1}^{T} f(y_t | \mathbf{x}_t, c; \boldsymbol{\beta}) = \left[\Lambda\left(\mathbf{x}_t\boldsymbol{\beta} + c\right)\right]^{y_t} \left[1 - \Lambda\left(\mathbf{x}_t\boldsymbol{\beta} + c\right)\right]^{1-y_t} \qquad (9)$

Find a <u>minimal set of sufficient statistics</u> t_i such that conditioning on t_i

eliminates the c,

$f(y_{i1},\ldots, y_{iT}|t_i, \mathbf{x}_i, c_i; \boldsymbol{\beta}) = f(y_{i1},\ldots, y_{iT}| t_i, \mathbf{x}_i; \boldsymbol{\beta})$

[example: WG estimation: $t = \bar{y}$]

The contribution to the log-lik of the i-th individual [from (8)] is

$$= c_i \sum_{t=1}^{T} y_{it} + \left[\sum_{t=1}^{T} (\mathbf{x_{it}} . y_{it}) \right] \boldsymbol{\beta} - \sum_{t=1}^{T} log \left[1 + exp(\mathbf{x_{it}} \boldsymbol{\beta} + c_i) \right] \qquad (10)$$

The minimal sufficient statistic is $\sum_{t=1}^{T} y_{it}$. Conditioning on this will get-rid of the c_i.

3. Conditional MLE **Example**: T=2

$\sum_{t=1}^{T} y_{it} = y_{i1} + y_{i2}$ which is 0, 1 or 2.

<u>First consider</u> $\sum_{t=1}^{T} y_{it} = 1$ **:**

$P[y_{i1}+y_{i2}=1] = P[y_{i1}=1 \ \& \ y_{i2}=0] + P[y_{i1}=0 \ \& \ y_{i2}=1] = P(1,0)+P(0,1)$

Conditional on c_i, the y_{i1} and y_{i2} are independent.

Hence,

$$P[y_{i1}+y_{i2}=1] = \frac{exp(\mathbf{x_{i1}\beta} + c_i) + exp(\mathbf{x_{i2}\beta} + c_i)}{[1+exp(\mathbf{x_{i1}\beta} + c_i)][1+exp(\mathbf{x_{i2}\beta} + c_i)]}$$

and

$P[(1,0)|y_{i1}+y_{i2}=1]$

$$= \frac{P[(1,0)]}{P[y_{i1}+y_{i2}=1]} = \frac{exp(c_i).exp(\mathbf{x_{i1}\beta})}{exp(c_i)\left[exp(\mathbf{x_{i1}\beta}) + exp(\mathbf{x_{i2}\beta})\right]}$$

$$= \frac{exp(\mathbf{x_{i1}\beta} - \mathbf{x_{i2}\beta})}{[exp(\mathbf{x_{i1}\beta} - \mathbf{x_{i2}\beta}) + 1]} = \frac{exp[(\mathbf{x_{i1}} - \mathbf{x_{i2}})\boldsymbol{\beta}]}{[exp\{(\mathbf{x_{i1}} - \mathbf{x_{i2}})\boldsymbol{\beta}\} + 1]} \qquad (11)$$

Hence, $P[(0,1)|y_{i1}+y_{i2}=1]$ does not contain c_i.

This gives us the conditional log likelihood when T=2 as

$$\sum_i log\left[\frac{y_{i1}\ exp(\mathbf{x_{i1}\beta}) + y_{i2} exp(\mathbf{x_{i2}\beta})}{exp(\mathbf{x_{i1}\beta}) + exp(\mathbf{x_{i2}\beta})}\right] \qquad (12)$$

- When $y_{i1}+y_{i2}=0$ or 2, the conditional likelihood contribution is 1. i.e. NO contributions from these individuals. Hence, might have problems with rare incidents.

- For general T, we have to consider $\sum y_{it}$ =1,2,..,(T-1).
- CMLE are consistent and the usual asy covar matrix can be used here.
- Conditional MLE can be extended to multinomial logit too (Chamberlain, 1980).

- Need variation in **x**.
- Time invariant regressors dropout.
- Only works in the case of a static model with logit assumption.

- Cannot do any predictions of prob(y=1) because of missing c.

4.6 Coefficient Interpretation (Wooldridge)

In cross-sectional models:

Partial effects/Marginal effects =

$$\frac{\partial P(y=1)}{\partial \mathrm{x}_j} = g(\mathbf{x}\boldsymbol{\beta})\,\beta_j \qquad \text{where } g(z) = \frac{dG}{dz}(z)$$

Now, have unobserved heterogeneity.

So calculate **average partial effects** APE – the expected value of PE over the dist of c

$$= \quad E_c[PE] = \quad E_c\left[\frac{\partial P(y=1|\mathbf{x},c)}{\partial x_j}\right] = E_c\left[\frac{\partial E(y|\mathbf{x},c)}{\partial x_j}\right]$$

Remember $P(y=1|\mathbf{x},c) = E(y|\,\mathbf{x},c)$.

Random Effects Probit with normally dist'd c and indep of **x**:

$$y_{it}^{*} = \mathbf{x_{it}}\boldsymbol{\beta} + c_i + u_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + v_{it}$$

Wooldridge (2002) section 2.2.5 shows how to calculate APE in general.

In this model, the APE can be obtained by working out the APE in the model with v_{it}.

<u>Calculation:</u>

$v_{it} | \mathbf{x_{it}} \sim \text{Normal}(0,\sigma^2_v)$; $\sigma^2_v = 1 + \sigma^2_c$ $(\sigma^2_u = 1)$

$P(y_{it}=1|\mathbf{x_{it}},c_i) = P(\mathbf{x_{it}}\boldsymbol{\beta} + c_i + u_{it} > 0) = P(u_{it} > -\mathbf{x_{it}}\boldsymbol{\beta} - c_i) = E(y_{it}=1|\mathbf{x_{it}},c_i)$

But $P(y_{it}=1|\mathbf{x_{it}}) = P(\mathbf{x_{it}}\boldsymbol{\beta} + v_{it} > 0) = P(v_{it} > -\mathbf{x_{it}}\boldsymbol{\beta}) = E(y_{it}=1|\mathbf{x_{it}})$

So, $E(y_{it}=1|\mathbf{x_{it}}) = \Phi(\mathbf{x_{it}}\boldsymbol{\beta}/\sigma_v) = E_c[E(y_{it}=1|\mathbf{x_{it}},c_i)] = E_c[\Phi(\mathbf{x_{it}}\boldsymbol{\beta}+c_i)]$

Hence APE evaluated at $\mathbf{x_0} = \partial\{E_c[\Phi(\mathbf{x_{it}}\boldsymbol{\beta}+c_i)]\}/\partial x_j$

$$= \partial\Phi(\mathbf{x_{it}}\boldsymbol{\beta}/\sigma_v)/\ \partial x_j = (\beta_j/\sigma_v)\ \phi(\mathbf{x_0}\boldsymbol{\beta}/\sigma_v)$$

- Calculation easy because of assumptions (heterogeneity is normally distributed independently of the **x).**
- More difficult with different assumptions – will need to do simulations.
- Can also use the effect on the log odds ratio in logit models.