Recent Advances in the Field of Trade Theory and Policy Analysis Using Micro-Level Data

July 2012

Bangkok, Thailand

Cosimo Beverelli

(World Trade Organization)

Content

a) Binary dependent variable models in cross-section

b) Binary dependent variable models with panel data

c) Examples of firm-level analysis

b) Binary dependent variable models in cross-section

- Binary outcome
- Latent variable
- Linear probability model (LMP)
- Probit model
- Logit model
- Marginal effects
- Odds ratio in logit model
- Maximum likelihood (ML) estimation
- Rules of thumb

Binary outcome

- In many applications the dependent variable is not continuous but qualitative, discrete or mixed:
  - Qualitative: car ownership (Y/N)
  - Discrete: education degree (Ph.D., University degree,…, no education)
  - Mixed: hours worked per day

- Here we focus on the case of a binary dependent variable
  - Example with firm-level data: exporter status (Y/N)

Binary outcome (ct'd)

- Let y be a binary dependent variable:

$$y = \begin{cases} 1 & with\, probability\, p \\ 0 & with\, probability\, 1-p \end{cases}$$

- A regression model is formed by parametrizing the probability p to depend on a vector of explanatory variables $\boldsymbol{x}$ and a $K \times 1$ parameter vector β
- Commonly, we estimate a conditional probability:

$$p_i = \Pr[y_i = 1|\boldsymbol{x}] = F(\boldsymbol{x}_i{}'\beta) \qquad (1)$$

where $F(\cdot)$ is a specified function

Intuition for $F(\cdot)$: latent variable

- Imagine we wanted to estimate the effect of $\boldsymbol{x}$ on a continuous variable y^*
- The "index function" model we would like to estimate is:

$$y_i{}^* = \boldsymbol{x}_i{}'\beta - \varepsilon_i$$

- However, we do not observe y^* but only the binary variable y

$$y = \begin{cases} 1 & if\ y^* > 0 \\ 0 & otherwise \end{cases}$$

Intuition for $F(\cdot)$: latent variable (ct'd)

- There are two ways of interpreting ${y_i}^*$:

1. Utility interpretation: ${y_i}^*$ is the additional utility that individual i would get by choosing $y_i = 1$ rather than $y_i = 0$

2. Threshold interpretation: ε_i is a threshold such that if $\boldsymbol{x}_i{}'\beta > \varepsilon_i$, then $y_i = 1$

- The parametrization of p_i is:

$$
\begin{aligned}
p_i = \Pr[y = 1|\boldsymbol{x}] = \Pr[y^* > 0|x] = \Pr[\,\boldsymbol{x}'\beta - \varepsilon > 0|x] \\
= \Pr[\varepsilon < \boldsymbol{x}'\beta] = F[\boldsymbol{x}'\beta]
\end{aligned}
$$

where $F(\cdot)$ is the CDF of ε

Linear probability model (LMP)

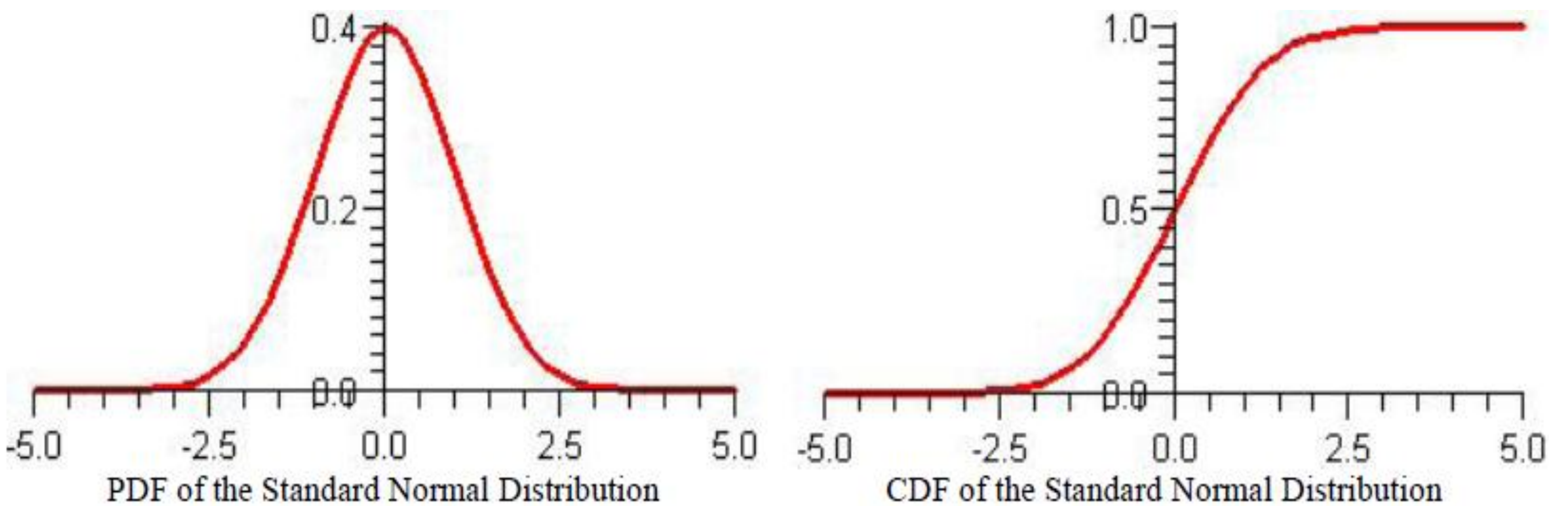
- The LPM does not use a CDF, but rather a linear function for $F(\cdot)$
- Therefore, equation (1) becomes:

$$p_i = \Pr[y_i = 1|\boldsymbol{x}] = \boldsymbol{x}_i{}'\beta$$

- The model is estimated by OLS with error term ε_i
- From basic probability theory, it should be the case that $0 \leq p_i \leq 1$
- This is not necessarily the case in the LPM, because $F(\cdot)$ in not a CDF (which is bounded between 0 and 1)
  - Therefore, one could estimate predicted probabilities $\hat{p}_i = \boldsymbol{x}_i{}'\widehat{\beta}$ that are negative or exceed 1
- Moreover, $V(\varepsilon_i) = \boldsymbol{x}_i{}'\beta(1 - \boldsymbol{x}_i{}'\beta)$ depends on $\boldsymbol{x}_i$
  - Therefore, there is heteroskedasticity (standard errors need to be robust)
- However, LPM provides a good guide to which variables are statistically significant

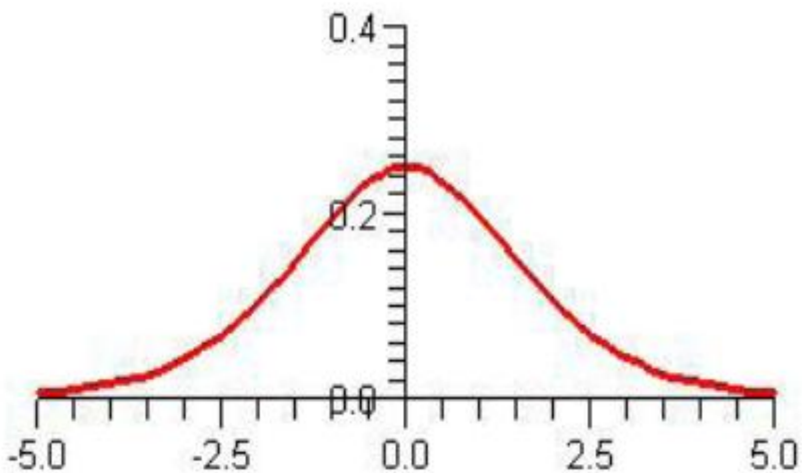
Probit model

- The probit model arises if $F(\cdot)$ is the CDF of the normal distribution, $\Phi(\cdot)$
- So $\Phi(\boldsymbol{x}'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz$, where $\phi(\cdot) \equiv \Phi'(\cdot)$ is the normal pdf

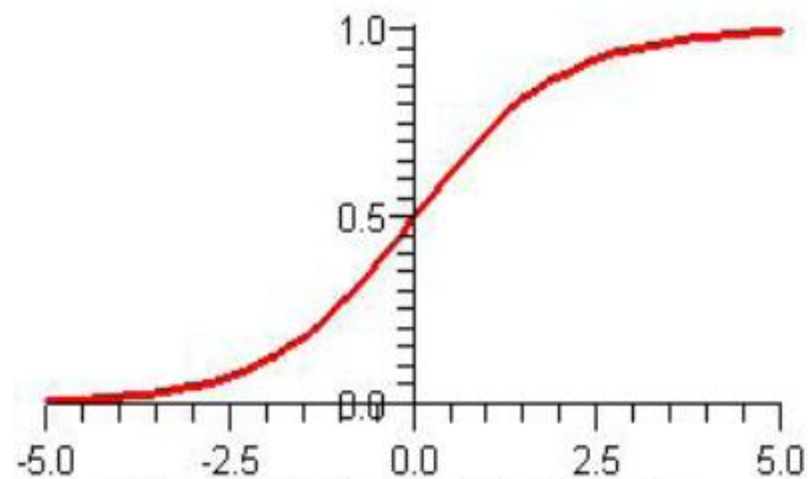


Logit model

- The logit model arises if $F(\cdot)$ is the CDF of the logistic distribution, $\Lambda(\cdot)$
- So $\Lambda(\boldsymbol{x}'\beta) = \frac{e^{x'\beta}}{1-e^{x'\beta}}$



PDF of the Standard Logistic Distribution



CDF of the Standard Logistic Distribution

Marginal effects

- For the model $p_i = \Pr[y_i = 1|\boldsymbol{x}] = F(\boldsymbol{x}_i{}'\beta) - \varepsilon_i$, the interest lies in estimating the marginal effect of the j'th regressor on p_i:

$$\frac{\partial p_i}{\partial x_{ij}} = F'(\boldsymbol{x}_i{}'\beta)\beta_j$$

- In the LPM model, $\frac{\partial p_i}{\partial x_{ij}} = \beta_j$

- In the probit model, $\frac{\partial p_i}{\partial x_{ij}} = \phi(\boldsymbol{x}_i{}'\beta)\beta_j$

- In the logit model, $\frac{\partial p_i}{\partial x_{ij}} = \Lambda(\boldsymbol{x}'\beta)[1 - \Lambda(\boldsymbol{x}_i{}'\beta)]\beta_j$

Odds ratio in logit model

- The odds ratio $\mathrm{OR} \equiv p/(1-p)$ is the probability that $y = 1$ relative to the probability that $y = 0$
- An odds ratio of 2 indicates, for instance that the probability that $y = 1$ is twice the probability that $y = 0$
- For the logit model:

$$
\begin{gathered}
p = e^{\boldsymbol{x}'\beta}/(1 + e^{\boldsymbol{x}'\beta}) \\
\mathrm{OR} = p/(1-p) = e^{\boldsymbol{x}'\beta} \\
\ln(OR) = \boldsymbol{x}'\beta
\end{gathered}
$$

(the log-odds ratio is linear in the regressors)

- β_j is a semi-elasticity
- If $\beta_j = 0.1$, a one unit increase in regressor j increases the odds ratio by a multiple 0.1
- See also [here]

Maximum likelihood (ML) estimation

- Since y_i is Bernoulli distributed ($y_i = 0, 1$), the density (pmf) is:

$$f(y_i|x_i) = {p_i}^{y_i}(1 - p_i)^{1-y_i}$$

Where $p_i = F(\boldsymbol{x}_i{}'\beta)$

- Given independence over i's, the log-likelihood is:

$$\mathcal{L}_N(\beta) = \sum_{i=1}^{N} \{y_i \ln F(\boldsymbol{x}_i{}'\beta) + (1 - y_i) \ln(1 - F(\boldsymbol{x}_i{}'\beta))\}$$

- There is no explicit solution for $\hat{\beta}_{MLE}$, but if the log-likelihood is concave (as in probit and logit) the iterative procedure usually converges quickly
- There is no advantage in using the robust sandwich form of the VCV matrix unless $F(\cdot)$ is mis-specified
- If there is cluster sampling, standard errors should be clustered

Rules of thumb

- The different models yield different estimates $\hat{\beta}$
- This is just an artifact of using different formulas for the probabilities
- It is meaningful to compare the marginal effects, not the coefficients
- At any event, the following rules of thumb apply:

$$\hat{\beta}_{Logit} \cong 4\ \hat{\beta}_{LPM}$$

$$\hat{\beta}_{Probit} \cong 2.5\ \hat{\beta}_{LPM}$$

$$\hat{\beta}_{Logit} \cong 1.6\ \hat{\beta}_{Probit}$$

(or $\hat{\beta}_{Logit} \cong (\frac{\pi}{\sqrt{3}})\ \hat{\beta}_{Probit}$)

- The differences between probit and logit are negligible if the interest lies in the marginal effects averaged over the sample

b) Binary dependent variable models with panel data

- Individual-specific effects binary models
- Fixed effects logit

Individual-specific effects binary models

- With panel data (each individual i is observed t times), the natural extension of the cross-section binary models is:

$$p_{it} = \Pr[y_{it} = 1 | x_{it}, \beta, \alpha_i] = \begin{cases} F(\alpha_i + \boldsymbol{x}'_{it}\beta) & in\ general \\ \Lambda(\alpha_i + \boldsymbol{x}'_{it}\beta) & for\ Logit\ model \\ \Phi(\alpha_i + \boldsymbol{x}'_{it}\beta) & for\ Probit\ model \end{cases}$$

- Random effects estimation assumes that $\alpha_i \sim N(0, \sigma^2{}_\alpha)$

Individual-specific effects binary models (ct'd)

- Fixed effect estimation is not possible for the probit model because there is an incidental parameters problem
  - Estimating α_i (N of them) along with β leads to inconsistent estimators of the coefficient itself if T is finite and $N \rightarrow \infty$ (this problem disappears as $N \rightarrow \infty$)
  - Unconditional fixed-effects probit models may be fit with the "probit" command with indicator variables for the panels. However, unconditional fixed-effects estimates are biased

- However, fixed effects estimation is possible with logit, using a conditional MLE that uses a conditional density (which describes a subset of the sample, namely individuals that "change state")

Fixed effects logit

- A conditional ML can be constructed conditioning on $\sum_t y_{it} = c$, where $0 < c < T$
- The functional form of $\Lambda(\cdot)$ allows to eliminate the individual effects and to obtain consistent estimates of β

- Notice that it is not possible to condition on $\sum_t y_{it} = 0$ or on $\sum_t y_{it} = T$
- Observations for which $\sum_t y_{it} = 0$ or $\sum_t y_{it} = T$ are dropped from the likelihood function
- That is, only the individuals that “change state” at least once are included in the likelihood function

Example

- T = 3
- We can condition on $\sum_t y_{it} = 1$ (possible sequences $\{0,0,1\}$, $\{0,1,0\}$ and $\{1,0,0\}$ or on $\sum_t y_{it} = 2$ (possible sequences $\{0,1,1\}$, $\{1,0,1\}$ and $\{1,1,0\}$)
- All individuals with sequences $\{0,0,0\}$ and $\{1,1,1\}$ are not considered

c) Examples of firm-level analysis

- Wakelin (1998)
- Aitken et al. (1997)
- Tomiura (2007)

Wakelin (1998)

- She uses a probit model to estimate the effects of size, average capital intensity, average wages, unit labour costs and innovation variables (exogenous variables) on the probability of exporting (dependent variable) of 320 UK manufacturing firms between 1988 and 1992
- Innovation variables include innovating-firms dummy, number of firm's innovations in the past and number of innovations used in the sector

- Non-innovative firms are found to be more likely to export than innovative firms of the same size…
- …However, the number of past innovations has a positive impact on the probability of an innovative firm exporting

Aitken et al. (1997)

- From a simple model of export behavior, they derive a probit specification for the probability that a firm exports
- The paper focuses on 2104 Mexican manufacturing firms between 1986 and 1990
- They find that locating near MNEs increases the probability of exporting
- Proximity to MNE increase the export probability of domestic firms regardless of whether MNEs serve local or export markets
- Region-specific factors, such as access to skilled labour, technology, and capital inputs, may also affect the probability of exporting
  - The export probability is positively correlated with the capital-labor ratio in the region

Tomiura (2007)

- How are internal R&D intensity and external networking related with the firm's export decision?
- Data from 118,300 Japanese manufacturing firms in 1998
- Logit model for the probability of direct export
- Export decision is defined as a function of R&D intensity and networking characteristics, while also controlling for capital intensity, firm size, subcontracting status, and industrial dummies
- 4 measures of networking status: computer networking, subsidiary networking, joint business operation, and participating in a business association