Introductory Econometrics I

Limited Dependent Variable Models*

Yingjie Feng

School of Economics and Management

Tsinghua University

June 1, 2024

Review: Assumptions for OLS

- Recall the classical linear model (CLM) assumptions for OLS regression:
  - MLR.1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$
  - MLR.2: random sampling from the population
  - MLR.3: no perfect collinearity in the sample
  - MLR.4: $\mathbb{E}[u|x_1, ..., x_k] = \mathbb{E}[u] = 0$ (exogenous explanatory variables)
  - MLR.5: $\mathbb{V}[u|x_1, ..., x_k] = \mathbb{V}[u] = \sigma^2$ (homoskedasticity)
  - MLR.6: $u|x_1, \cdots, x_k \sim \mathsf{Normal}(0, \sigma^2)$
- We have endeavored to relax MLR.2, MLR.4, MLR.5, MLR.6
- MLR.3 is mild (just used to guarantee the existence of OLS estimators)
- Now, we focus on MLR.1 (linear in parameters), which may be inappropriate in some scenarios.
  - Note: Linearity in x_j is not a big issue, since x_j can be nonlinear transformation of underlying variables (log, quadratics, etc.)

Outline

Limited Dependent Variables

- In many scenarios, the dependent variable takes values in a *restricted* range
  - Binary response: labor force participation
  - Multiple choice: choice of transportation mode
  - Positive integers: number of times arrested during a year
- We have discussed binary response before.
  - Linear probability model (LPM) works (some issues need to be taken care of, e.g., heteroskedasticity)
  - But LPM does have several disadvantages
    - The predicted probability is not guaranteed to be between 0 and 1
    - The partial effect of x on the response probability is a constant, which is inappropriate in this case
- Today, we will focus on binary responses and discuss two nonlinear models, **logit** and **probit**, that help address these issues.

Binary Choice

- Suppose $y \in \{0, 1\}$ is binary. Our goal is to characterize response probability

$$\mathbb{P}(y = 1|\boldsymbol{x}) = \mathbb{P}(y = 1|x_1, \cdots, x_k) = \mathbb{E}[y|x_1, \cdots, x_k]$$

  - Example: y is an employment indicator, and $\boldsymbol{x}$ can be education, age, etc.

- Recall that in linear probability model, we set

$$\mathbb{P}(y = 1|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

- By contrast, we can specify a nonlinear model, for example,

$$\mathbb{P}(y = 1|\boldsymbol{x}) = G\Big(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\Big)$$

  G is some known nonlinear function such that $0 \leq G(z) \leq 1$ for all $z \in \mathbb{R}$

Outline

1. Limited Dependent Variables
2. Logit and Probit Models
3. Maximum Likelihood Estimation of Logit/Probit Models

Logit and Probit Models

- In **logit** model, we set

$$G(z) = \frac{\exp(z)}{1 + \exp(z)} = \Lambda(z)$$

  - $\Lambda(z)$ is the cumulative distribution of standard logistic distribution

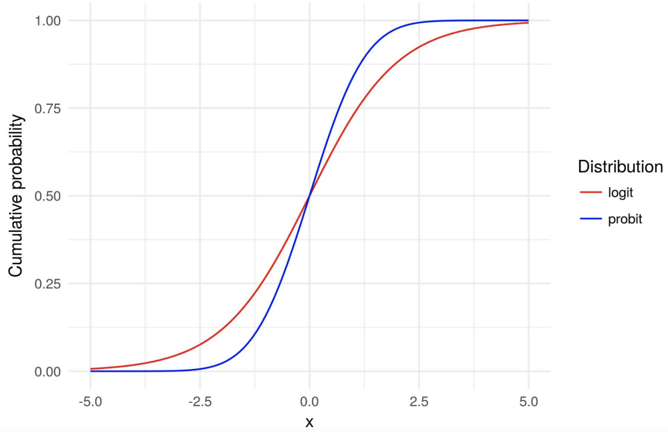
- In **probit** model, we set

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv$$

  - $\Phi(z)$ is the cumulative distribution of standard normal distribution

- Recall a cumulative distribution function take values between zero and one
  - $G(z) \to 0$ as $z \to -\infty$
  - $G(z) \to 1$ as $z \to +\infty$

Logit and Probit Models



Justification: Latent Variable Models

- We can justify logit/probit by latent variable models

$$y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + e, \quad y^* \text{ not observed}$$

$$y = \mathbb{1}(y^* > 0) = \begin{cases} 0 & \text{if } y^* \leq 0 \\ 1 & \text{if } y^* > 0 \end{cases}$$

- Assume e is independent of $\boldsymbol{x}$ and follows the distribution given by $G(\cdot)$.

$$\begin{aligned} \mathbb{P}(y = 1|\boldsymbol{x}) = \mathbb{P}(y^* > 0|\boldsymbol{x}) &= \mathbb{P}(e > -\beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k|\boldsymbol{x}) \\ &= 1 - G(-\beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k) \end{aligned}$$

- For logit/probit, G is symmetric w.r.t. zero, i.e., $G(-z) = 1 - G(z)$. So

$$\mathbb{P}(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

Partial Effects

- Partial effect (PE) of some x_j on the response probability
  - In LPM:
  $$\frac{\partial}{\partial x_j}\mathbb{P}(y=1|\boldsymbol{x}) = \beta_j$$
  - In nonlinear models:
  $$\frac{\partial}{\partial x_j}\mathbb{P}(y=1|\boldsymbol{x}) = \beta_j G'(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k), \quad G'(z)=\frac{d}{dz}G(z)$$
- For probit/logit, G is strictly increasing $(G'(\cdot)>0)$. So the partial effect of x_j has the same sign as β_j
- But PE is no longer a constant; it depends on particular values of the explanatory variables. Some parameters of interest are

$$\begin{aligned}
&\text{PE of } x_j \text{ at average of each } x_l\text{:} && \beta_j G'\Big(\beta_0+\beta_1\mathbb{E}[x_1]+\cdots+\beta_k\mathbb{E}[x_k]\Big)\\
&\text{Average PE of } x_j\text{:} && \mathbb{E}\Big[\beta_j G'(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)\Big]
\end{aligned}$$

Partial Effects

- Partial effect (PE) of some x_j on the response probability
  - In LPM:
  $$\frac{\partial}{\partial x_j}\mathbb{P}(y=1|\boldsymbol{x}) = \beta_j$$
  - In nonlinear models:
  $$\frac{\partial}{\partial x_j}\mathbb{P}(y=1|\boldsymbol{x}) = \beta_j G'(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k), \quad G'(z)=\frac{d}{dz}G(z)$$
- Sometimes you want to see the effect of a dummy variable, say x_1. Then, the goal is to estimate
  $$G(\beta_0+\beta_1+\beta_2 x_2+\cdots+\beta_k x_k) - G(\beta_0+\beta_2 x_2+\cdots+\beta_k x_k)$$
  - Again, you want to choose particular values $x_2,\cdots,x_k$ (or take average)

Partial Effects

- Once we have parameter estimates $\hat{\beta}_j$'s, the estimates of these effects are available.
  - The estimated partial effect of x_j evaluated at the average of each x_l (PEA_j)

$$\widehat{PEA}_j = \hat{\beta}_j G'(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \cdots + \hat{\beta}_k \bar{x}_k)$$

  - The estimated average partial effect of x_j (APE_j)

$$\widehat{APE}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_j G'(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik})$$

- In Stata,
  - use `probit/logit` to estimate the parameters in the model (via maximum likelihood method)
  - use `margins` to obtain the partial effects

Example

- Women's labor force participation (MROZ.DTA)
- Linear regression (LPM)

```
. reg inlf nwifeinc educ exper age
```

Source	SS	df	MS			
Model	34.1681574	4	8.54203935			
Residual	150.559598	748	.201282885			
Total	184.727756	752	.245648611			

Number of obs	=	753
F(4, 748)	=	42.44
Prob > F	=	0.0000
R-squared	=	0.1850
Adj R-squared	=	0.1806
Root MSE	=	.44865

inlf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nwifeinc	-.0037402	.0015182	-2.46	0.014	-.0067206	-.0007599
educ	.0358477	.0076757	4.67	0.000	.0207792	.0509162
exper	.0231594	.0022352	10.36	0.000	.0187714	.0275475
age	-.0111411	.0022111	-5.04	0.000	-.0154819	-.0068004
_cons	.430941	.1342506	3.21	0.001	.1673883	.6944938

Example

- Women's labor force participation (MROZ.DTA)
- Logit estimation

```
. logit inlf nwifeinc educ exper age

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -438.66346
Iteration 2:   log likelihood = -437.87224
Iteration 3:   log likelihood = -437.86993
Iteration 4:   log likelihood = -437.86993

Logistic regression                               Number of obs   =        753
                                                  LR chi2(4)      =     154.01
                                                  Prob > chi2     =     0.0000
Log likelihood = -437.86993                       Pseudo R2       =     0.1496
```

inlf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0188992	.0076755	-2.46	0.014	-.0339428	-.0038555
educ	.1852325	.0407518	4.55	0.000	.1053604	.2651046
exper	.1165652	.0129169	9.02	0.000	.0912485	.141882
age	-.0506876	.0112326	-4.51	0.000	-.072703	-.0286722
_cons	-.6320442	.6734115	-0.94	0.348	-1.951906	.6878181

Example

- Women's labor force participation (MROZ.DTA)
- Probit estimation, with **robust** standard error

```
. probit inlf nwifeinc educ exper age, robust

Iteration 0:   log pseudolikelihood =  -514.8732
Iteration 1:   log pseudolikelihood = -438.85547
Iteration 2:   log pseudolikelihood = -438.71697
Iteration 3:   log pseudolikelihood = -438.71695

Probit regression                                 Number of obs   =        753
                                                  Wald chi2(4)    =     114.67
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -438.71695                 Pseudo R2       =     0.1479
```

inlf	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0110665	.0046413	-2.38	0.017	-.0201632	-.0019698
educ	.1089496	.0245473	4.44	0.000	.0608378	.1570614
exper	.0685406	.0078943	8.68	0.000	.0530681	.0840131
age	-.0314559	.0066097	-4.76	0.000	-.0444107	-.018501
_cons	-.303706	.4047309	-0.75	0.453	-1.096964	.489552

Example

- Women's labor force participation (MROZ.DTA)
- After probit/logit, you can use **margins** to get partial/marginal effects

```
. margins, dydx(*)

Average marginal effects                          Number of obs     =        753
Model VCE    : Robust

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper age
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0036778	.0015222	-2.42	0.016	-.0066613	-.0006943
educ	.0362077	.0078099	4.64	0.000	.0209006	.0515149
exper	.0227784	.0021427	10.63	0.000	.0185787	.0269781
age	-.0104539	.0021128	-4.95	0.000	-.0145949	-.0063129

This table reports the average partial effects (APE)

Final Remarks

- Coefficient from LPM, logit and probit cannot be directly compared, but you can compare the partial effects

- The distribution function $G(\cdot)$ of standard logistic and that of standard normal are similar. In practice, the logit and probit estimates usually do not differ much. If they do, maybe you want to re-specify your model.

- Various robust standard errors are still available in probit/logit estimation.

Outline

Review: Maximum Likelihood Estimation

- Recall the idea of MLE is to maximize the likelihood of the observed sample by choosing the parameters
- For example, assume $(y_i, \boldsymbol{x}_i)$ are i.i.d. and the conditional density function of y given $\boldsymbol{x}$ is $f(y|\boldsymbol{x};\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a vector of parameters of interest

$$\text{individual likelihood :} \quad L_i(\boldsymbol{\beta}) := f(y_i|\boldsymbol{x}_i;\boldsymbol{\beta})$$

$$\text{joint likelihood :} \quad L(\boldsymbol{\beta}) := \prod_{i=1}^{n} L_i(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{x}_i;\boldsymbol{\beta})$$

$$\text{log likelihood :} \quad \ell(\boldsymbol{\beta}) := \sum_{i=1}^{n} \log f(y_i|\boldsymbol{x}_i;\boldsymbol{\beta})$$

- Maximum likelihood estimator:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta})$$

- Technical notes: as before, our anlaysis is conditional on all explanatory variables, so we only characterize conditional distribution of y given $\boldsymbol{x}$

MLE of Logit/Probit Models

- Conditional on $\boldsymbol{x} = (1, x_1, \cdots, x_k)$, y can only take two values, 0 and 1.
- In statistics, such binary variables are said to follow *Bernoulli* distribution.

$$y_i = \begin{cases} 1 & \text{with prob. } G(\boldsymbol{x}_i\boldsymbol{\beta}) \\ 0 & \text{with prob. } 1 - G(\boldsymbol{x}_i\boldsymbol{\beta}) \end{cases}$$

- The conditional density of y_i given $\boldsymbol{x}_i$ can be written as

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

- MLE:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \Big(y_i \log[G(\boldsymbol{x}_i\boldsymbol{\beta})] + (1 - y_i) \log[1 - G(\boldsymbol{x}_i\boldsymbol{\beta})] \Big)$$