ECON 613: Applied Econometrics

Methods for Cross-sectional Data

January 29, 2019

Introduction

Binary response models are models where the variable to be explained y is a random variable taking on the values zero and one which indicate whether or not a certain event has occured.

- $y = 1$ if a person is employed
- $y = 1$ if a family contributes to a charity during a particular year
- $y = 1$ if a firm has a particular type of pension plan
- $y = 1$ if a worker goes to college
- Regardless of what y stands for, we refer to $y = 1$ as a success and $y = 0$ as a failure.

An OLS regression of y on dependent variables denoted x ignores the discreteness of the dependent variable and does not constrain predicted probabilities to be between zero and one.

Linear Probability Model (1)

$$P(y = 1) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. \tag{1}$$

- If x_1 is continuous, β_1 is the change in the probability of success given one unit increase in x_1
- If x_1 is discrete, β_1 is the difference in the probability of success when $x_1 = 1$ and $x_1 = 0$, holding other x_j fixed.

Linear Probability Model (2)

Given that y is a random variable (Bernouilli), we also have

$$E(y \mid x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. \quad (2)$$
$$Var(y \mid x) = x\beta(1 - x\beta) \quad (3)$$

Implications:

- OLS regression of y on $x_1, x_2, \ldots, x_k$ produces consistent and unbiaised estimators of the β_j.
- Heteroskedacticity, which can be dealt with using standard heteroskedasticity-robust standard errors.
- Problem: OLS fitted values may not be between zero and one.

Maximum Likelihood Estimation: Logit and Probit Models

For binary outcome data, the dependent variable y takes one of two values. We let

$$y = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases} \tag{4}$$

Parametrize conditional probabilities:

$$p_i = F_\epsilon(X\beta) \tag{5}$$

And, Marginal Effects

$$\frac{\partial Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F'_\epsilon(X\beta)\beta_j \tag{6}$$

Probit Model (1)

The probit model corresponds to the case where $F(x)$ is the cumulative standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(\tfrac{1}{2}X^2)dX \tag{7}$$

Where $F(X\beta) = \Phi(X\beta)$.

Probit Model (2)

- Consider the latent approach

$$y^\star = X\beta + \epsilon \tag{8}$$

where $\epsilon \sim N(0,1)$. Think of $y^\star$ as the net utility associated with some action. If the action yields positive net utility, it is undertaken otherwise it is not.

- We would care only about the sign of $y^\star$

$$y = \begin{cases} 1, & \text{if } p \\ 0, & \text{with probability } 1-p \end{cases} \tag{9}$$

- Probabilities

$$\begin{aligned} Pr(y=1) &= Pr(y^\star > 0) = Pr(X\beta + \epsilon \geq 0) \\ &= Pr(\epsilon \geq -X\beta) = Pr(\epsilon \leq X\beta) = \Phi(X\beta) \end{aligned}$$

Logit Model

The logit model specifies the cdf function $F(x)$ is now the logistic function

$$\Lambda(x) = \frac{1}{1+\exp(-x)} = \frac{exp(x)}{1+\exp(x)} \tag{10}$$

The logit model is most easily derived by assuming that

$$\log\left(\frac{P}{1-P}\right) = X\beta \tag{11}$$

The logarithm of the odds (ratio of two probabilities) is equal to $X\beta$

Maximum Likelihood Estimation

- Likelihood can not be defined as a joint density function.
- Outcome of a Bernouilli trial

$$f(y_i \mid x_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \tag{12}$$

- Given the independence of individuals, the likelihood can be written

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}(1 - F(x_i\beta))^{1-y_i} \tag{13}$$

Log Likelihood

- The log likelihood

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} y_i lnF(x_i\beta) + (1 - y_i) ln(1 - F(x_i\beta)) \tag{14}$$

- First order conditions

$$\sum_{i=1}^{n} \frac{y_i - F(x_i\beta)}{F(x_i\beta)(1 - F(x_i\beta))} F^{'}(x_i\beta)x_i = 0 \tag{15}$$

Empirical considerations

- Probit and logit yield same outcomes. Only difference is how parameters are scaled.
- The natural metric to compare models is the fitted log-likelihood provided that the models have the same number of parameters.
- Although estimated parameters are different, marginal effects are quite similar.

Pseudo R2

$$R^2_{\text{Binary}} = 1 - \frac{\mathcal{L}(\hat{\beta})}{N[\bar{y} ln \bar{y} + (1 - \bar{y}) ln(1 - \bar{y})]} \tag{16}$$

Predicted Outcomes

- The criterion $\sum_i (y_i - \hat{y}_i)^2$ gives the number of wrong predictions.
  - average rule: let $\hat{y} = 1$ when $\hat{p} = F(X\beta) > 0.5$
  - Receiver Operating Characteristics (ROC) curve plots the fractions of $y = 1$ correctly classified against the fractions of $y = 0$ incorrectly specified as the cutoffs $\hat{p} = F(X\beta) > c$ varies.

Example: Affairs

Infidelity data, known as Fair's Affairs. Cross-section data from a survey conducted by Psychology Today in 1969.

Table: Descriptive Characteristics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
affairs	601	1.5	3.3	0	0	0	12
age	601	32.5	9.3	18	27	37	57
yearsmarried	601	8.2	5.6	0	4	15	15
religiousness	601	3.1	1.2	1	2	4	5
education	601	16.2	2.4	9	14	18	20
occupation	601	4.2	1.8	1	3	6	7
rating	601	3.9	1.1	1	3	5	5

Understanding the determinants of having an affair

Table:

	Dependent variable:					
	Affairs					
	(1)	(2)	(3)	(4)	(5)	(6)
I(gender)male	0.140	0.120	0.210^{*}	0.200^{*}	0.200^{*}	0.210
	(0.110)	(0.110)	(0.120)	(0.120)	(0.120)	(0.130)
I(age)		0.007	-0.024^{**}	-0.023^{**}	-0.023^{**}	-0.023^{**}
		(0.006)	(0.010)	(0.010)	(0.010)	(0.010)
I(yearsmarried)			0.066^{***}	0.055^{***}	0.064^{***}	0.064^{***}
			(0.017)	(0.018)	(0.019)	(0.019)
I(children)yes				0.250	0.270^{*}	0.270^{*}
				(0.160)	(0.160)	(0.160)
I(religiousness)					-0.200^{***}	-0.200^{***}
					(0.051)	(0.051)
I(education)						-0.005
						(0.026)
Constant	-0.740^{***}	-0.970^{***}	-0.560^{**}	-0.670^{***}	-0.150	-0.067
	(0.078)	(0.200)	(0.240)	(0.250)	(0.280)	(0.480)
Log Likelihood	-337.000	-336.000	-328.000	-327.000	-319.000	-318.000
Note:						*p$<$0.1; **p$<$0.05; ***p$<$0.01

Identification considerations

- β is identified up to a scale.
- We observe only whether $X\beta + \epsilon > 0$.
- Implication for the interpretation of the coefficients.