Probit and Logit Models

Ani Katchova

Outline

- Linear probability model
- Probit and logit models
- Maximum likelihood estimation
- Coefficients
- Predicted probabilities
- Marginal effects
    - Marginal effect at the means
    - Average marginal effect
- Goodness of fit measures
    - Pseudo R-squared
    - Percent correctly predicted

Binary dependent variable

- A binary dependent variable has two outcomes: 0 or 1.
- Examples: working or not working, has insurance or does not have insurance, etc.
- The outcome of interest is denoted as 1.
    - $y = 1$ if working, $y = 0$ if not working.
- If the outcome of not working is of interest, then it would be denoted as 1.
    - $y = 1$ if not working, $y = 0$ if working.
- There are typically fewer outcomes of interest, i.e. fewer 1s in the data.

Linear probability model (LPM)

- A linear probability model is a linear regression model where the dependent variable is a binary variable.
- Linear probability model with binary dependent variable $y = 0$ or 1.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_k x_k + u = x\beta + u$
  - where $x\beta$ is expressed in a matrix form.
- Expected value of y is $E(y) = x\beta$.
- Because the binary variable y has two outcomes 0 or 1, the expected value for y is the probability of y being 1, $P(y = 1)$.
- $E(y) = 1 * P(y = 1) + 0 * P(y = 0) = P(y = 1)$
- Example: if 30% of y are 1 and the rest are zero, then $E(y) = P(y = 1) = 0.3$
- The <u>linear probability model</u> for the probability of the outcome $y = 1$ is

$$P(y = 1) = x\beta$$

Advantages and disadvantages of LPM

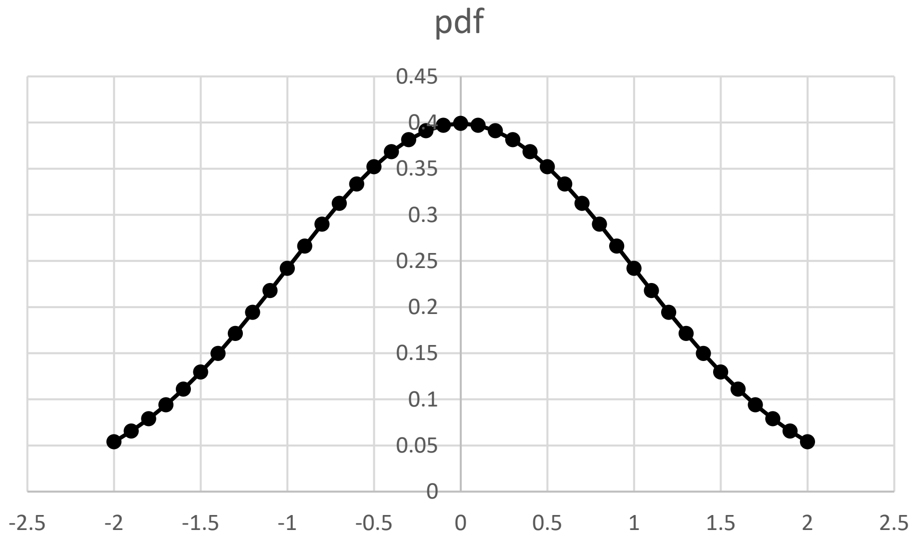
- Advantages of LPM
  - Easy to estimate and interpret (coefficients are marginal effects)
  - The coefficients and predictions are reasonably good
- Disadvantages of LPM
  - Not the best model for binary dependent variable (probit or logit models are better)
  - Predicted probabilities can be less than 0 or greater than 1
  - Marginal effects are the coefficients, which are constant/do not vary with x
  - Heteroscedasticity because the variance is not constant
  - $var(y) = P(y = 1) * [1 - P(y = 1)]$

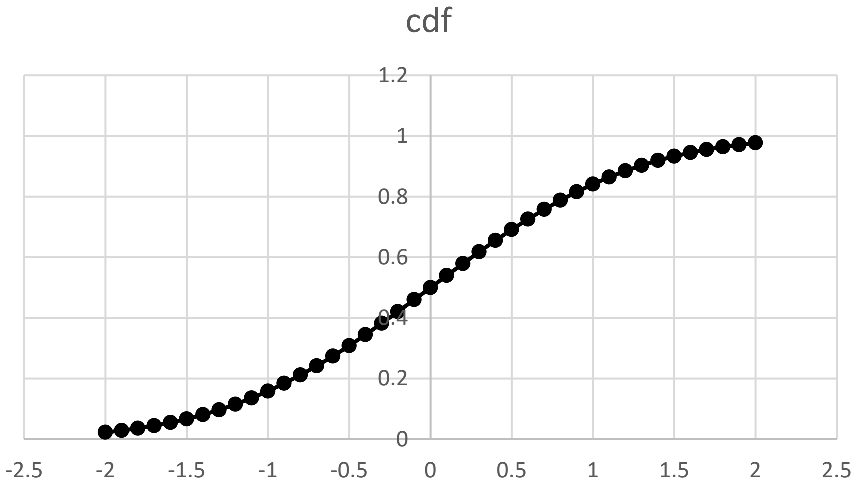
Linear versus non-linear probability models

- The linear probability model estimate the probability of $y = 1$ as a linear function of the independent variables.
  - $P(y = 1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_k x_k = x\beta$
- The probit and logit models estimate the probability of $y = 1$ as a non-linear function G of the independent variables.
  - $P(y = 1) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_k x_k) = G(x\beta)$
  - G is a non-linear function that transforms $x\beta$ to be between 0 and 1 because $P(y = 1)$ is a probability.

Normal distribution – pdf and cdf

- The probability density function (pdf) of the normal distribution ϕ shows the probability that y is between two numbers.
- The cumulative density function (cdf) of the normal distribution Φ shows the probability that y is less than a given number.





Probit model

- The probit model uses the cumulative density function (cdf) of the normal distribution Φ.
- $P(y=1) = \Phi(x\beta) = \int_{-\infty}^{x\beta} \phi(z)dz$
- $P(y=1)$ will be a number between 0 and 1 because the cdf of the normal distribution is a number between 0 and 1.

Logit model

- The logit model uses the logistic function:
- $P(y=1) = G(x\beta) = \frac{\exp(x\beta)}{1+\exp(x\beta)} = \frac{e^{x\beta}}{1+e^{x\beta}}$
- $P(y=1)$ will be a number between 0 and 1 because $\exp(x\beta)$ is positive.
- The probability of $y=0$ is:
- $P(y=0) = 1 - P(y=1) = 1 - \frac{\exp(x\beta)}{1+\exp(x\beta)} = \frac{1}{1+\exp(x\beta)}$

Likelihood function

- The likelihood is the probability that the outcome for observation i is y_i.
  - The likelihood of $y_i = 1$ is $P(y_i = 1)$.
  - The likelihood of $y_i = 0$ is $P(y_i = 0)$.
- The likelihood function is defined as: $P(y_i = 1)^{y_i}\, P(y_i = 0)^{1-y_i}$
  - The likelihood of $y_i = 1$ is $P(y_i = 1)^1\, P(y_i = 0)^{1-1} = P(y_i = 1)$
  - The likelihood of $y_i = 0$ is $P(y_i = 1)^0\, P(y_i = 0)^{1-0} = P(y_i = 0)$

Maximum likelihood estimation

- The likelihood function is: $P(y_i = 1)^{y_i}\, P(y_i = 0)^{1-y_i}$
- Taking logs and summing up over all observations i.
- The log likelihood function is:
- $\sum_{i=1}^{n}(y_i * logP(y_i = 1) + (1 - y_i) * logP(y_i = 0))$
- Substituting $P(y = 1) = G(x\beta)$ into the log likelihood function.
- $\sum_{i=1}^{n}(y_i * \log(G(x\beta)) + (1 - y_i) * \log(1 - G(x\beta))$
- The β coefficients are obtained by maximizing the log likelihood function.

Maximum likelihood estimation

- The probit and logit model coefficients are obtained by maximizing the log likelihood function.
- $\max \sum_{i=1}^{n}(y_i * logP(y_i = 1) + (1 - y_i) * logP(y_i = 0))$
  - If the outcome $y_i = 1$, the predicted probability $P(y_i = 1)$ is maximized (e.g. 0.8 or 0.9).
  - If the outcome $y_i = 0$, $P(y_i = 0)$ is maximized or equivalently the predicted probability $P(y_i = 1)$ is minimized (e.g. 0.1 or 0.2).
- The maximum likelihood estimators are consistent, asymptotically normal, and asymptotically efficient if the assumptions hold.

Maximum likelihood estimation versus OLS estimation

- The probit and logit model coefficients are obtained by maximizing the log likelihood function (if the outcome $y = 1$, the predicted probability $P(y = 1)$ is maximized)
- $\max \sum_{i=1}^{n} (y_i * logP(y_i = 1) + (1 - y_i) * logP(y_i = 0))$

- The OLS coefficients are obtained by minimizing the sum of squared residuals (difference between actual value y and predicted values $\hat{y}$)
- $\min \sum_{i=1}^{n} \hat{u}^2 = \sum_{i=1}^{n} (y - \hat{y})^2 = \sum_{i=1}^{n} (y - x\hat{\beta})^2$

Example

- Model to explain if women are in the labor force or not.
- $P(inlf = 1) = G(\beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 age + \beta_5 kidslt6)$
  - $inlf$ is a binary 0 or 1 variable for whether women are in labor force or not.
  - $nwifeinc$ is non-wife income.
  - $kidslt6$ is number of kids under 6 years old.
- 57% of the women are in the labor force and the rest are not. The unconditional probability of being in the labor force is 0.57. $P(y = 1) = 0.57$

Variable	Mean	Std. Dev.	Min	Max
inlf	0.57	0.50	0	1
nwifeinc	20.13	11.63	-0.03	96
educ	12.29	2.28	5	17
exper	10.63	8.07	0	45
age	42.54	8.07	30	60
kidslt6	0.24	0.52	0	3

LPM, probit, and logit model – coefficients

	LPM	Probit	Logit
VARIABLES	inlf	inlf	inlf
nwifeinc	-0.003**	-0.011**	-0.020**
	(0.001)	(0.005)	(0.008)
educ	0.039***	0.132***	0.223***
	(0.007)	(0.025)	(0.043)
exper	0.022***	0.069***	0.118***
	(0.002)	(0.007)	(0.013)
age	-0.019***	-0.058***	-0.095***
	(0.002)	(0.008)	(0.013)
kidslt6	-0.275***	-0.886***	-1.464***
	(0.033)	(0.117)	(0.200)
Constant	0.770***	0.765*	1.153
	(0.135)	(0.440)	(0.742)

- The coefficients are different for the probit and logit models. The logit coefficients are about 1.6 times the probit coefficients.
- Interpretation of the coefficient on education: women with higher education are more likely to be in the labor force.
- Interpretation of the coefficient on age: women who are older are less likely to be in the labor force.
- The magnitudes of the coefficients are not interpreted.

Predicted probabilities

- After estimating the models and obtaining the coefficients $\hat{\beta}$, the predicted probabilities can be calculated as:
- $P(y_i = 1) = G\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + .. + \hat{\beta}_k x_{ik}\right) = G(x_i \hat{\beta})$
  - If the actual value $y_i = 1$ and the predicted probability $P(y_i = 1)$ is above 0.5, it is a correct prediction.
  - If the actual value $y_i = 0$ and the predicted probability $P(y_i = 1)$ is below 0.5, it is also a correct prediction.
  - Otherwise, it will be incorrect prediction.
- The average of the predicted probabilities will be the unconditional probability, which is the sample average $\bar{y}$.

Actual values and predicted probabilities

	Actual value	LPM predicted probability	Probit predicted probability	Logit predicted probability
obs i	inlf y	Inlfhat_lpm	Inlfhat_probit	Inlfhat_logit
1	1	0.65	0.67	0.68
2	1	0.73	0.77	0.77
3	1	0.61	0.63	0.64
4	1	0.72	0.76	0.76
601	0	0.31	0.27	0.26
602	0	0.35	0.31	0.29
603	0	0.39	0.35	0.34
604	0	0.67	0.69	0.70
605	0	-0.19	0.01	0.03

For the LPM, the predicted probabilities can be a negative number (e.g. -0.19 for observation 605) or a number above 1, which is not reasonable.
For the probit and logit models, the predicted probabilities for the first four observations with inlf=1 are all above 0.5 (correct prediction).
The predicted probabilities for observation 604 with inlf=0 are above 0.5 which in an incorrect prediction.
The predicted probabilities for the other four observations with inlf=0 are below 0.5 (correct prediction).
The average of inlf is 0.57, which is also the average of the predicted probabilities.

Marginal effects in the linear probability model

- The linear probability model:
- $P(y = 1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_k x_k = x\beta$
- The coefficient on x_j is β_j.
- The marginal effect of x_j on the probability of $y = 1$ is the coefficient β_j.

$$\frac{\Delta P(y = 1)}{\Delta x_j} = \beta_j$$

- In the LPM, the marginal effects are the coefficients.
- The marginal effect explains the effect of the independent variable on the probability that $y = 1$.

Marginal effects in the probit and logit model

- The probit and logit model:
- $P(y = 1) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_k x_k) = G(x\beta)$
- The coefficient on x_j is β_j.
- The marginal effect of x_j on the probability of $y = 1$ is

$$\frac{\Delta P(y = 1)}{\Delta x_j} = G'(x\beta) * \beta_j$$

- In the probit and logit model, the marginal effects are the coefficients multiplied by a scale factor $G'(x\beta)$, which is the derivative of the G function.
- The marginal effect explains the effect of the independent variable on the probability that $y = 1$ (by how much the probability of $y = 1$ increases when x_j increases by 1 unit).

Marginal effects in the probit and logit model

- The marginal effect of x_j on the probability of $y = 1$ is
- $\frac{\Delta P(y=1)}{\Delta x_j} = G'(x\beta) * \beta_j$
- The probit model: $P(y = 1) = \Phi(x\beta)$
- The marginal effect in the probit model: $\frac{\Delta P(y=1)}{\Delta x_j} = \phi(x\beta) * \beta_j$
  - Φ is the cdf and ϕ is the pdf of the normal distribution.
- The logit model: $P(y = 1) = \frac{\exp(x\beta)}{1+\exp(x\beta)}$
- The marginal effect in the logit model: $\frac{\Delta P(y=1)}{\Delta x_j} = \frac{\exp(x\beta)}{[1+\exp(x\beta)]^2} * \beta_j$

Marginal effect at the mean and average marginal effect

- The marginal effect depends on x. $\frac{\Delta P(y=1)}{\Delta x_j} = G'(x\beta) * \beta_j$
- The marginal effect at the mean is calculated at the mean value of x, which is $\bar{x}$.
- $\frac{\Delta P(y=1)}{\Delta x_j} = G'(\bar{x}\beta) * \beta_j$
- The average marginal effect is calculated for each observation, and then averaged across all observations.
- $\frac{\Delta P(y=1)}{\Delta x_j} = \overline{G'(x_i\beta)} * \beta_j = \frac{1}{n}\sum_{i=1}^{n} G'(x_i\beta) * \beta_j$
- The marginal effect at the mean uses the means of the variables, but there may not be such “average” individual (e.g. mean for variable female is 0.3). The average marginal effect makes more sense. In practice, the marginal effects will be similar.

Marginal effect for an indicator variable

- If the model is: $P(y = 1) = G(\beta_0 + \beta_1 * d_1 + \beta_2 x_2)$
- with an independent variable d_1 which an indicator variable taking values of 0 or 1, the marginal effect is calculated as:
- $G(\beta_0 + \beta_1 * 1 + \beta_2 x_2) - G(\beta_0 + \beta_1 * 0 + \beta_2 x_2)$
- The marginal effect of d_1 being 1 instead of 0 is the difference in $P(y = 1)$ if $d_1 = 1$ and the probability of $P(y = 1)$ if $d_1 = 0$.

Marginal effects

VARIABLES	Probit marginal effect at mean inlf	Probit average marginal effects inlf	Logit marginal effects at mean inlf	Logit average marginal effects inlf
nwifeinc	-0.004**	-0.003**	-0.005**	-0.004**
	(0.002)	(0.001)	(0.002)	(0.001)
educ	0.051***	0.040***	0.054***	0.040***
	(0.010)	(0.007)	(0.010)	(0.007)
exper	0.027***	0.021***	0.029***	0.021***
	(0.003)	(0.002)	(0.003)	(0.002)
age	-0.023***	-0.018***	-0.023***	-0.017***
	(0.003)	(0.002)	(0.003)	(0.002)
kidslt6	-0.346***	-0.271***	-0.355***	-0.266***
	(0.046)	(0.031)	(0.049)	(0.031)

Unlike the coefficients, the marginal effects in the probit and logit model are similar. The marginal effects at the mean and the average marginal effects are similar. The magnitude of the marginal effects can be interpreted.
For each additional year of education, women are 5.1% <u>more likely</u> to be in the labor force.
For each additional child less than 6 years old, women are 34.6% <u>less likely</u> to be in the labor force.

Pseudo R-squared

- Pseudo R-squared, aka McFadden R-squared, measures the goodness of fit for a probit or logit model. It compares the log-likelihood of a model with that of a model with only a constant.
- $Pseudo\ R^2 = 1 - \frac{LL_{ur}}{LL_0}$
    - LL_{ur} is the log likelihood for the unrestricted model with all independent variables.
    - LL_0 is the log likelihood for the restricted model with only a constant.
    - If the independent variables do not explain the dependent variable then the log likelihoods for the restricted and unrestricted models (LL_0 and LL_{ur}) will be the similar and the pseudo R-squared will be 0.
    - If the independent variables explain the dependent variable very well, then because the log likelihood for the unrestricted model LL_{ur} will be maximized (which is negative number that will approach 0) and the pseudo R-squared will approach 1.
- The pseudo R-squared indicates how well the model predicts the outcome and how well the model improves on a null model with only an intercept, but the magnitude is not interpreted.
- A higher pseudo R-squared with the same dependent variable but different independent variables would indicate that the model predicts the outcome better.

Pseudo R-squared example

- $Pseudo\ R^2 = 1 - \frac{LL_{ur}}{LL_0} = 1 - \frac{-406.30}{-514.87} = 0.21$
  - LL_{ur} is the log likelihood for the unrestricted model with all independent variables
  - LL_0 is the log likelihood for the restricted model with only a constant.
- The pseudo R-squared shows how well the model predicts the outcome, with a higher pseudo R-squared being preferred.

Percent correctly predicted

- Percent correctly predicted is a goodness of fit measure, which shows the percent of correct predictions to total predictions for the binary outcome.
- The actual outcome y is 0 or 1.
- The predicted probability $P(y = 1)$ is a number between 0 and 1. If the predicted probability is greater than 0.5, then $\hat{y} = 1$, otherwise $\hat{y} = 0$.
- There are four cases of making correct or incorrect predictions.

	Actual $y = 1$	Actual $y = 0$
Predicted $\hat{y} = 1$	Correct prediction	Incorrect prediction
Predicted $\hat{y} = 0$	Incorrect prediction	Correct prediction

- Percent correctly predicted is the proportion of correct predictions to total predictions.
- Percent correctly predicted should have high values (perhaps above 70%).

Percent correctly predicted example

- After estimating the probit model and getting the predicted values $\hat{y}$, the table below shows the number of correct and incorrect predictions. The total number of predictions is 753.

	Actual $y = 1$	Actual $y = 0$
Predicted $\hat{y} = 1$	Correct prediction 346	Incorrect prediction 115
Predicted $\hat{y} = 0$	Incorrect prediction 82	Correct prediction 210

- Percent correctly predicted = (346+210)/753=73.84%.
- For the probit model, 73.84% of the outcomes are correctly predicted.
- For the logit model, percent correctly predicted = (345+211)/753=73.84%.

Review questions

- Describe and give examples of a binary dependent variable.
- Explain the functional form for the linear probability model, the probit model, and the logit model.
- How are the coefficients interpreted?
- Describe the marginal effect at the mean and the average marginal effect.
- How are the marginal effects interpreted?
- Describe the goodness of fit measures for the probit and logit model.