Introduction to limited dependent variable models

Discrete/qualitative regression models

Where are we in this course?

1. General Introduction to econometrics

2. Introduction to Stata as an econometric software

3. Understanding data and data management before and during the analysis process

4. Introduction to Regression Analysis

5. Introduction to limited dependent variable models

6. Introduction to binary choice (probability) models

7. Introduction to Discrete (categorical) choice models (ordered and nominal dependent variables)

8. Count Data and other econometric models (theoretical)

What are discrete/qualitative regression models?

- The linear regression model assumes that the dependent variable *Y* is **quantitative and continuous** (e.g WTP in ugx)

- On the other hand, the **explanatory variables** are either quantitative, qualitative (or dummy), or a mixture thereof

- What if the dependent variable is **discrete** (qualitative) rather than **continuous?**

- We need discrete models where the dep variable takes **few** but **discrete values e.g** variable y is **binary** ({1,0} or any other 2 values)

- Binary choice models assume that individuals are faced with a choice between two alternatives and the choice depends on identifiable characteristics
  - Technology adoption – 1 if individual adopts a new technology (0 otherwise)
  - Member of farmer group – 1 if individual is in the group (0 otherwise)

What are discrete/qualitative regression models?

- Things are more complicated when the dep variable y can assume more than two values
- We can classify the cases into: (a) categorical variables or (b) non-categorical variables
- Categorical can further be classified as Unordered or ordered variable

Mode of produce transport to mkt
y=1; *if mode of transport is head*
y=2; *if mode of transport is motorbike*
y=3; *if mode of transport is truck*

Employer of household head
y=1*; if occupation is civil servant*
y=2; private employed
y=3; *private self employed*
y=4; *no employment*

- In these examples we can define dep variables in **<u>any</u>** order desired
- Ordered variable examples

Average monthly expenditure packaging materials (Ugx)
y=1 if spends less than 1,000
y=2 if spends more than 1,000 but < 2000
y=3 if spends more than 2,000 but < 4000
y=4 if spends more than 4,000

Preferences measured on a scale
y=1*; if intensely dislikes*
y=2; if moderately dislikes
y=3; *if neutral*
y=4; *if moderately like*
Y=5; *if intensely like*

What are discrete/qualitative regression models?

- In models where *Y* is qualitative, our objective is to find the **probability** of something happening, e.g owning a house, participating in labour force etc.

- Hence, qualitative response regression models are often known as **probability models**.

- We seek answers to the following questions:

  1. How do we estimate qualitative response regression models?
     - Can we simply estimate them with the usual OLS procedures?

  2. Are there special inference problems?
     - In other words, is the hypothesis testing procedure any different from the ones we have learned so far?

  3. If a dep variable is qualitative, how can we measure the goodness of fit of such models?
     - Is the conventionally computed R^2 of any value in such models?

What are discrete/qualitative regression models?

4. Once we go beyond the dichotomous dep variable case, how do we estimate and interpret the polychotomous regression models?

   - How do we handle models in which the dep variable**,** is, an **ordered categorical** variable (ordinal), or
   - How do we handle models in which the dep variable has no inherent ordering (farmer, teacher, lawyer, etc)?

5. How do we model **count data,** or **rare event** data phenomena, such as the number of extension visits in a year, the number of articles published by a college professor in a year, etc

- We want to provide answers to some of these questions at the elementary level

- Lets start with **binary response** regression model.
  - There are three approaches to developing a probability model for a binary response variable:
    1. The **linear probability model (LPM)**
    2. The **logit model**
    3. The **probit model**

Binary outcome models

LPM, Probit, Logit

The Linear Probability Model (LPM)

- Why LPM first
  - its simple and can be estimated by OLS

- To fix ideas, consider the following regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Where X_i = **family income** and

$$Y_i = \begin{cases} 1: & \text{if family owns a house} \\ 0: & \text{if family does not own a house} \end{cases}$$

- u_i is independently distributed random variable with mean 0

- This model looks like a typical linear regression model
  - *but because the regressand is binary, or dichotomous, it is called a* ***linear probability model (LPM)***

The Linear Probability Model (LPM)

- We write the LPM as follows to interpret Y as a probability

$$P_i = \begin{cases} 1 & \text{when } \beta_0 + \beta_1 X_i \geq 1 \\ \beta_0 + \beta_1 X_i & \text{when } 0 < \beta_0 + \beta_1 X_i < 1 \\ 0 & \text{when } \beta_0 + \beta_1 X_i \leq 0 \end{cases}$$

- We can interpret $E(Y_i|X_i)$ as the *conditional probability* that the event will occur given X_i i.e. $\Pr(Y_i = 1|X_i)$

- Thus, in our example, $E(Y_i|X_i)$ gives the probability of a family owning a house and whose income is the given amount X_i

- Assuming $E(u_i = 0)$ (to obtain unbiased estimators), we obtain

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

- If P_i = probability that Y_i = 1 (that is, the event occurs) then the variable Y_i has the following (probability) distribution:

Y_i	Probability
0	$1 - P_i$
1	P_i
Total	1

The Linear Probability Model (LPM)

- That is, Y_i follows the **Bernoulli probability distribution.**

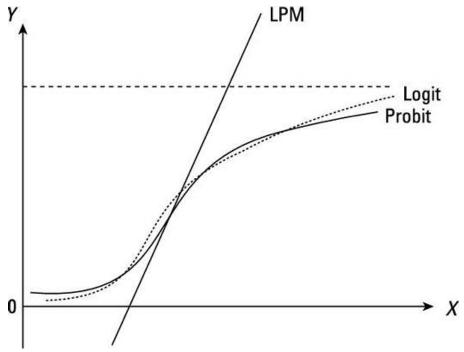
- Running a regression on a binary outcome variable

1. Stochastic component

$$Y_i \sim Bernouli = {P_i}^{y_i}(1-P_i)^{1-y_i} = \begin{cases} P_i & for\ \ y=1 \\ 1-P_i & for\ y=0 \end{cases}$$

2. Systematic component $\Pr(Y_i = 1|\beta) \equiv E(Y_i) \equiv P_i \equiv X_i\beta$

3. Y_i and Y_j are independent given $\forall\ i \neq j$, conditional on X



- Quiz: what's good? What's bad?
- For some x, $\Pr(Y) \notin [0,1]$
- But models are approximations . May be ok for middling P?
- Unlikely to get the uncertainties right

The Linear Probability Model (LPM)

- Assume we would like to find the determinants of willingness to pay for tree seedlings
  - the dependent variable is 1 if the person is willing to pay
  - and 0 if the person is not willing to pay

- As long as explanatory variables are not correlated with the error term, this simple method (LPM) is perfectly fine.

- However, a caution is necessary. When the dependent variable is a dummy variable, the variance depends X values

$$var(u|X) = X\beta[1 - X\beta]$$

- As we have seen, the LPM is plagued by several problems, such as:
  - non-normality of u_i, *(solved by increasing sample size)*
  - heteroscedasticity of u_i , *(solved by using robust standard errors)*
  - possibility of $\hat{Y}_i$ lying outside the 0–1 range, and
  - the generally lower R^2 values.

Alternatives to LPM--The logistic regression (Logit) model

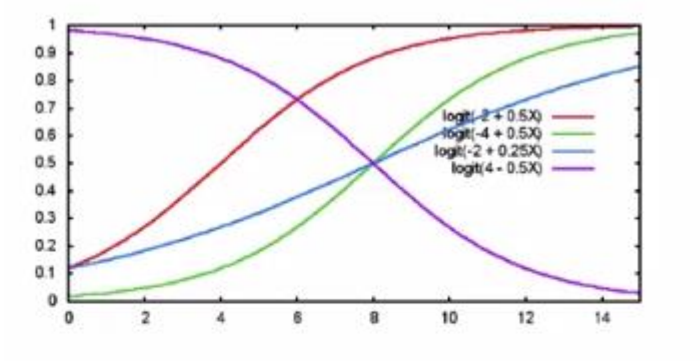
The model

1. Stochastic component

$$Y_i \sim Bernouli = \pi_i{}^{yi}(1-\pi_i)^{1-yi} = \begin{cases} \pi_i & for\ \ y = 1 \\ 1-\pi_i & for\ y = 0 \end{cases}$$

2. Systematic component $\Pr(Y_i = 1|\beta) = \frac{1}{1+e^{-x_i\beta}} = \Lambda(x_i\beta)$

3. Y_i and Y_j are independent given $\forall\ i \neq j$, conditional on X



Quiz: what's good? What's bad?
$\Pr(Y) \in [0,1]$ for any y
One change for probit $\pi_i = \Phi(x_i\beta)$
Could be more flexible, for now

The logit log-likelihood

Probability density of all the data

$$P(y|\pi) = \prod_{i=1}^{n} \pi_i^{yi}(1-\pi_i)^{1-y_i}, \pi_i = \frac{1}{1+e^{-x_i\beta}}$$

Log-likelihood

$$lnL(\beta|y) = \sum_{i=1}^{n}\{y_i ln\pi_i + (1-y_i)\ln(1-\pi_i)\}$$

$$\sum_{i=1}^{n}\{-y_i ln\frac{1}{1+e^{-x_i\beta}} + (1-y_i)\ln(1-\frac{1}{1+e^{-x_i\beta}})\}$$

$$\sum_{i=1}^{n} \ln(1+e^{1-2y_i)x_i\beta})$$

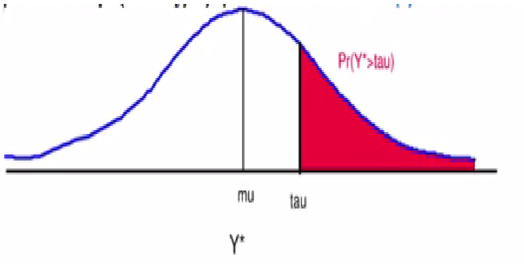
What does this give us?

The logit log-likelihood

- It gives a function into which;
  - we can guess values of β
  - or query this function with chosen values of β that we put into the function
  - ask this function how likely is any particular value of β to be observed in the population
- We find the value of β with the highest likelihood
  - these are known as the maximum likelihood estimates
- How do we interpret β?
  - For a logit, they are interpreted as the log odds of a particular event of interest happening given the X

Latent variable modelling –logit

- Continuous unobserved variable: Y^*, animal health, voting propensity



- A Model $P(y_i^*|\mu_i), \mu_{i=x_i\beta,\ \ Y_i \perp Y_j|X}$
- What model has Y^* observed and $P(.)$ normal?
- With observation mechanism $y_i = \begin{cases} 1,\ \ y_i^* \leq if\ i\ is\ alive \\ 0,\ \ y_i^* > 0\ if\ i\ is\ dead \end{cases}$
- If only y_i is observed, and Y^* is standardized logistic,
- $P(y_i^*|\mu_i) = STL(y^*|\mu_i) = \frac{\exp(y_i^*-\mu_i)}{[1+\exp(y_i^*-\mu_i)]^2}$ $\rightsquigarrow$ logit model
- Proof: $\Pr(Y_i = 1|\mu_i) = \Pr(y_i^* \leq 0) = \int_{-\infty}^{0} STL(y_i^*|\mu_i)dy_i^*$

  $= F_{stl}(0|\mu_i = [1 + \exp(-X_i\beta]^{-1}$ $\rightsquigarrow$ logit functional form

Latent variable modelling –probit

- Same setup as for logit, with one change

- Stochastic component $Y^* \sim P(y_i^*|\mu_i) = N(y_i^*|\mu_i, 1)$

- Systematic component becomes

- $\Pr(Y_i = 1|\mu_i) = \int_{-\infty}^{0} N(y_i^*|\mu_i, 1) dy_i^* = \Phi(X_i\beta)$

- Interpretation:

  - One unit of Y^*: one standard deviation
  - Interpret β: regression of Y^* on X
  - Interpret $\widehat{\beta}_i$: what happens to Y^* on average (or μ_i exactly) when X_j goes up by one unit, holding constant the other covariates

- Because the interpretation is disconnected with reality (observed) probit model first estimates are not useful so we opt for marginal effects

Utility approach-logit and probit

Definitions

- Utility from adoption: U_i^D
- Utility from non-adoption: U_i^R
- Utility difference, propensity to adopt: $Y^* \equiv U_i^D - U_i^R$
- Same Observation mechanism: $y_i = \begin{cases} 1, & y_i^* \leq 0 \; if \; i \; adopts \\ 0, & y_i^* > 0 \; if \; i \; does \; not \; adopt \end{cases}$

Assumptions

- $U_i^D \perp U_i^R | X$
- $U_i^k \sim P\left(U_i^k \middle| \eta_i^k\right) \; for \; k = \{D, R\}$

- If P(.) is normal: ↝ probit model
- If P(.) is a generalized extreme value: ↝ logit model

1. Of the three generalized justifications for the same binary model, which one do you prefer?
2. When would you choose LPM or logit or probit?

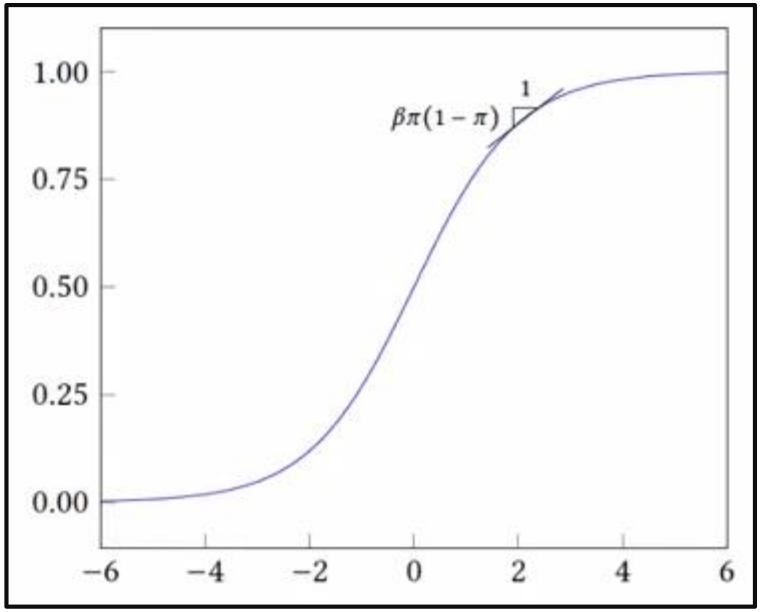
LPM, Logit and Probit Marginal effects

Derivative rule: $\frac{\partial \theta}{\partial X_j} = \frac{\partial g(X,\beta)}{\partial X_j}$

Linear: $\frac{\partial \mu}{\partial X_j} = \widehat{\beta}_j$

Logit: $\frac{\partial \pi}{\partial X_j} = \frac{\partial \frac{1}{1+e^{-X\beta}}}{\partial X_i} = \widehat{\beta}_j \hat{\pi}(1 - \hat{\pi})$

Probit: : $\frac{\partial \pi}{\partial X_j} = = f\ (\beta_0 + \beta_1 X_i)\widehat{\beta}_j$



Interpretation:

- Hold some variables constant at their means (or other values), move one particular X and observe what happens
  - percentage points change in Y given a very small change in X from its mean
  - Or difference in percentage point log odds between a category of interest compared to a base category