ECON-C4200 - Econometrics II

Lecture 5: Limited dependent variable models

Otto Toivanen

Learning outcomes

- At the end of lectures 5 & 6, you

1. understand what a **L**imited **D**ependent **V**ariable (LDV) is
2. what a **Discrete choice model** is
3. what one can and cannot identify with a discrete choice model
4. what a **L**inear **P**robability **M**odel (LPM) is
5. how to estimate a LPM
6. how to interpret the parameters of a LPM
7. how to make a discrete choice model consistent with probability theory
8. what a likelihood function is
9. what **Probit** and **Logit** models are and how to estimate them
10. what **marginal effects** are and how to calculate them

What is a LDV?

- It is a variable that can take on only **restricted** values.
    1. The share of income spent on item j is between zero and one
    2. The number of products one buys is a non-negative integer.
    3. A firm either invests in R&D or it does not

Discrete choice

- Choice is discrete if the set of alternatives is limited (= you can count the alternative choices).
- Discrete choice is an example of LDV models; the class of LDV models is much wider, but we concentrate on discrete choice.
- How do we model decisions in economics?
- Utility maximization.
- How to do this when choices discrete (cannot differentiate...)?

Discrete choice

- Example: buying a product.
- Denote utility from buying U.
- Allow U to vary with characteristics of the individual: $U = U(X)$.
- Price same for everybody: p.
- Note: we are discussing the simplest discrete choice models where there are two options. The models generalize to (much) more complicated settings.

Discrete choice

- Utility from buying (assuming ”quasilinear” preferences - preferences are linear in the **numeraire** good):

$$U(X) - p \tag{1}$$

- What is the utility from not buying?
- Hard to know, may very across individuals.
- Let’s **normalize** to 0 $\rightarrow$ we identify differences in utility (we have an example of this later).

Discrete choice

- How does utility change if individual j buys the product?

$$[U(X_j) - p] - 0 = U(X_j) - p \tag{2}$$

- When does j buy? If and only if

$$U(X_j) - p > 0 \tag{3}$$

Discrete choice

- Denote "buy" $\rightarrow Y = 1$
- Denote "don't buy" $\rightarrow Y = 0$

$$Y = 1 \Leftrightarrow U(X_j) - p > 0 \tag{4}$$

Discrete choice

- How to relate this to an econometric model?
- Let's introduce an error term.

$$U(X_j) = \beta_0 + \beta_1 X_j + \epsilon_j \tag{5}$$

$$Y = 1 \Leftrightarrow \beta_0 + \beta_1 X_j + \epsilon_j - p > 0 \tag{6}$$

Notice that in our model where p same for everybody, it "goes into" the constant term.

Discrete choice

- Interpretation:

  $\mathbb{E}[Y_j|X_j] = \beta_0 + \beta_1 X_j =$ expected utility for consumer j from buying the good.

  $\mathbb{E}[Y_j|X_j] = \beta_0 + \beta_1 X_j =$ probability of individual j buying the good.

Choice of vertical integration

- Yes / no $\rightarrow$ 0 / 1 (or the other way round).
- Example Gil, R. (2015). Does vertical integration decrease prices? evidence from the paramount antitrust case of 1948. *American Economic Journal: Economic Policy*, *7*(2), 162–91
- Question: Should movie studios own cinemas?
- Variable: `VI_Ever = 1` in Gil's paper if cinema *i* vertically integrated, 0 otherwise.
- Let's take a cross-section of the 1^{st} year of each theatre.
- We concentrate on the 1^{st} year as then the courts did not (yet) restrict VI.

How to choose VI?

- Let's assume

$$\pi_i^{VI} = \alpha_0 + \alpha_1 size_i + \epsilon_i^{VI} \quad (7)$$

$$\pi_i^{noVI} = \gamma_0 + \gamma_1 size_i + \epsilon_i^{noVI} \quad (8)$$

- A theatre is VI iff it is profitable, i.e.,

$$\pi_i^{VI} - \pi_i^{noVI} \geq 0$$

.

How to choose VI?

$$\alpha_0 + \alpha_1 size_i + \epsilon_i^{VI} - (\gamma_0 + \gamma_1 size_i + \epsilon_i^{noVI}) \geq 0 \tag{9}$$

$$(\alpha_0 - \gamma_0) + (\alpha_1 - \gamma_1) \times size_i + (\epsilon_i^{VI} - \epsilon_i^{noVI}) \geq 0 \tag{10}$$

$$\beta_0 + \beta_1 size_i + \epsilon_i \geq 0 \tag{11}$$

- NOTE: we can measure the **difference** in profits (utility), not the level.

Descriptive statistics

Stata code

```
tabstat vi_ever capacity_1000, stat(mean sd min max n)
scatter vi_ever capacity_1000 if capacity_1000 < 6, ///
  xtitle("Capacity, 000 seats") ///
  ytitle("Vertical integration = 1") ///
  note("x-axis censored at 6 000 seats") ///
  graphregion(color(white)) bgcolor(white)

  graphexport "scatter_gil.pdf", replace
```

Descriptive statistics

```
. tabstat vi_ever capacity_1000, stat(mean sd p10 p25 p50 p75 p90 min max n)
```

stats	vi_ever	cap~1000
mean	.4351145	1.735972
sd	.496404	1.638655
p10	0	.485
p25	0	.8
p50	0	1.4
p75	1	2.2
p90	1	3.172
min	0	.115
max	1	23
N	393	393

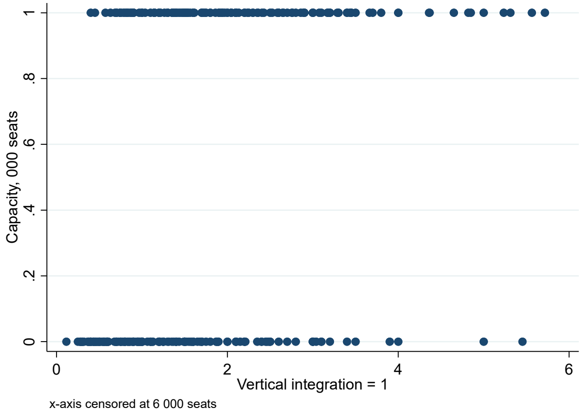
1

Descriptive statistics

```
. pwcorr vi_ever capacity_1000, sig

             |  vi_ever cap~1000
-------------+------------------
     vi_ever |   1.0000
             |
             |
capacit~1000 |   0.4206   1.0000
             |   0.0000
```

2

Descriptive statistics



3

How to estimate an LDV model I

- Linear regression $\rightarrow$ **L**inear **P**robability **M**odel (LPM).
- Works...
- What is the interpretation of the regression function?

How to estimate an LDV model I

```
. regr vi_ever capacity_1000, robust

Linear regression                               Number of obs     =        393
                                                F(1, 391)         =       3.27
                                                Prob > F          =     0.0715
                                                R-squared         =     0.0701
                                                Root MSE          =      .4793
```

vi_ever	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
capacity_1000	.0802159	.0443814	1.81	0.071	-.0070401	.1674718
_cons	.295862	.075513	3.92	0.000	.1473997	.4443243

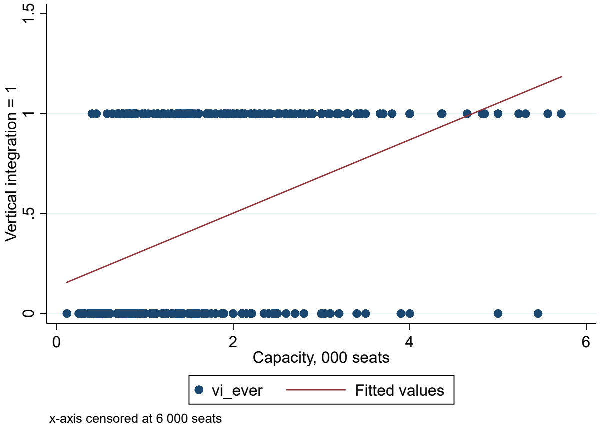
4

Descriptive statistics

Stata code

```
twoway scatter vi_ever capacity_1000 if capacity_1000 < 6 || ///
  lfit vi_ever capacity_1000 if capacity_1000 < 6, ///
  xtitle("Capacity, 000 seats") ///
  ytitle("Vertical integration = 1") ///
  note("x-axis censored at 6 000 seats") ///
  graphregion(color(white)) bgcolor(white)

  graphexport "scatter_gil_2.pdf", replace
```

How to estimate an LDV model I



5

What features does LPM have?

1. Good: Coefficients are **marginal effects** $= \partial Y / \partial X$ (derivatives).
2. "Bad: Predicted probabilities" may be < 0 and/or > 1.
3. To take into account: Error terms are heteroscedastic by design ($\rightarrow$ use robust se).

- Does all this matter? Depends what you want to do.
- In (very) large data sets, LPM is just fine if your interest is in the marginal effects only.

Taking the probability seriously

- If the dependent variable is $0/1$, then the model produces a probability.
- Probabilities are by definition in the support $[0, 1]$.
- What functional form would yield a mapping ("match") from X to Y that could be interpreted as a probability?
- Answer: any function that yields a prediction between 0 and 1.

Taking the probability seriously

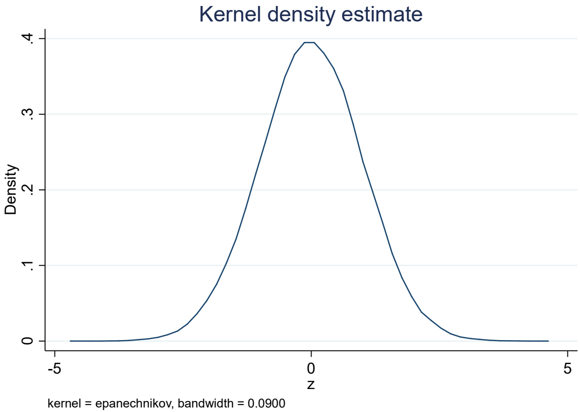
- What would be such a function? Answer: Cumulative density functions.
- Think of the normal distribution, denoted $\phi(z)$.
- But keep in mind that **any** cdf would work.
- We are going to discuss some popular choices in the next lecture.

Stata simulation data

Stata code

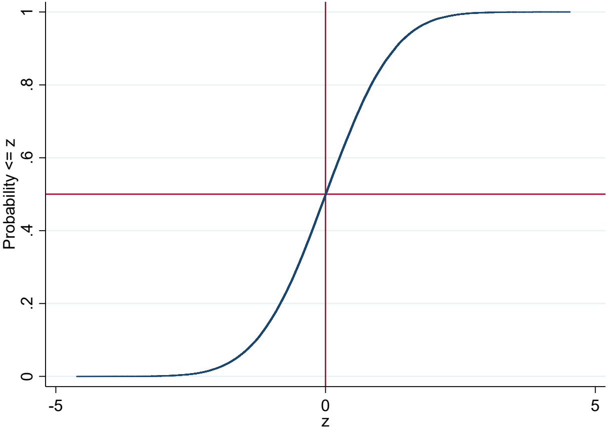
```
set obs 100000
gen z = invnorm(runiform())
/* NOTE: alternative
gen z = rnormal() */
kdensity z, ///
  graphregion(color(white)) bgcolor(white)

distplot z, yline(0.5) xline(0) /// // distplot from ssc
  graphregion(color(white)) bgcolor(white)
```

Stata simulation data



6

Stata simulation data



7

Taking the probability seriously

- If Y was continuous and between zero and one, we would write

  $Pr(Y = y|X = x) = \phi(\beta_0 + \beta_1 x).$

- But Y is discrete and also,
- where's the error term?
- Write $Y = 1 \Leftrightarrow \beta_0 + \beta_1 X + \epsilon \geq 0.$
- It then follows that $Y = 0 \Leftrightarrow \beta_0 + \beta_1 X + \epsilon < 0.$
- Notice how we have now divided all possible RHS values into those that deliver $Y = 0$ and those that deliver $Y = 1$.

Taking the probability seriously

- Find the lowest value of $\epsilon_i = \bar{\epsilon}_i$ for which $\beta_0 + \beta_1 X_i + \epsilon_i \geq 0$ holds.

  $\bar{\epsilon}_i = -(\beta_0 + \beta_1 X_i)$.

  $Pr(Y = 1|X = x) = \int_{-(\beta_0+\beta_1 x)}^{\infty} \phi(\epsilon) d\epsilon$.

  $= 1 - \Phi(-\beta_0 - \beta_1 X_i) = \Phi(\beta_0 + \beta_1 X_i)$.

- The last equality follows from the fact that the normal distribution is symmetric (= what it looks like for $\epsilon < 0$ is the mirror image of what it looks like for $\epsilon \geq 0$).
- Notice that we have now produced a probability that varies from theatre/individual to theatre/individual, depending on the value of X_i.
- Also notice that nothing in our derivation rested on us assuming ϵ is normally distributed.

Back to VI...

- We observe $1, ..., N$ theaters that either are or are not VI.
- Being VI means $Y = 1$ and $\beta_0 + \beta_1 x + \epsilon > 0$.
- $Pr(Y = 1|X = x) = \Phi(\beta_0 + \beta_1 x)$.
- $Pr(Y = 0|X = x) = 1 - \Phi(\beta_0 + \beta_1 x)$.