Intermediate Social Statistics Hilary 2009 Lecture : Binary Discrete Dependent Variable

Raymond Duch

Nuffield College

February 2, 2010

Continuous Dependent Variables

- Up until now, we have been treating all of our models as if Y were continuous.
- Today we'll consider the class of models where Y is non-continuous.
- Examples of continuous Y might include:
  - Presidential approval rates
  - Policy mood
  - Congressional polarization
  - Political tolerance
  - International trade
  - Globalization
  - Others?

Discrete Dependent Variables

- Lots of dependent variables cannot be characterized as continuous
- Those fall into several categories, such as (with examples):
  - Count (terrorist bombings)
  - Binary (votes)
  - Ordered (agree-to-disagree scales)
  - Multinomial (candidates in a primary; parties in multiparty election)
- And we'll treat these separately.

Functional Form of Discrete Models

All of our models will resemble probability models, of the sort:

Prob(event j occurs) = Prob(Y=j) = F[stochastic component, systematic component]

Illustrations of Bivariate Dependent Variable

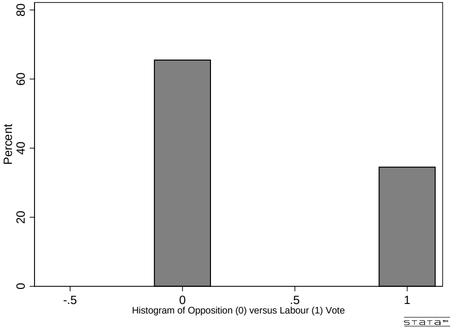
- This either occurs when the situation is genuinely binary—e.g., vote Labour or vote Opposition
- Or when the situation is continuous in the underlying (but unobserved) reality, but binary in observation—e.g., the decision to make, or not make, campaign contributions, which in a latent sense is a (continuous) probability model, but all we observe is [contribute, do not vote contribute].

An Example: Vote Preference of UK Citizens

- The example we will focus on in this lecture is from a 2004 survey of the voting preferences of U.K. citizens.
- The binary choice is vote Labour or vote for one of the opposition parties.

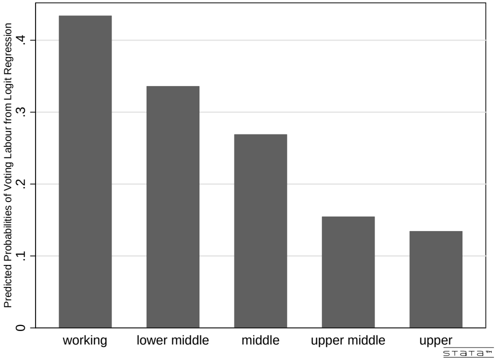
Illustrations of Bivariate Dependent Variable

Figure: Frequency of Labour versus Opposition Vote: UK 2004



Insights from Limited-Dependent Variable Models

Figure: Predicting Vote Choice Based on Class: UK 2004



What's wrong with the linear probability model?

- Why not just estimate:

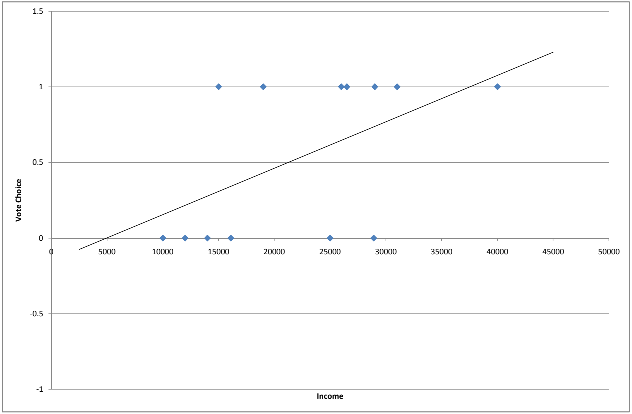
$$y = \mathbf{x}\beta + \varepsilon$$

- where y = 0 or y = 1?
- In terms of our example from the 2004 European Election study this would suggest,

$$\text{Labour Vote} = \beta_o + \beta_1 * \text{lrself} + \varepsilon$$

What's wrong with the linear probability model?

Figure: Estimating Hypothetical Vote Choice Model with OLS Regression



What's wrong with the linear probability model?

- First, you can see where ε will be heteroskedastic.
- The variance of it will be lowest around $p = 0.5$, and highest close to 0 and 1.
- But we can fix this with GLS. So this isn't too too too serious.

What's wrong with the linear probability model?

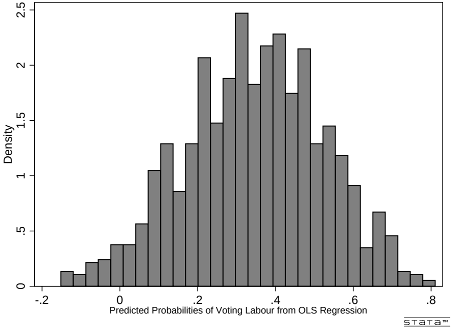
- Much more seriously, the model—you can see why—will make nonsense predictions, with $p < 0$ and $p > 1$.
- That will also produce negative variances. We can see this more clearly by estimating the following model using OLS:
- Labour vote = retnat + class + union + southwest + urban + lrself + own + ε

What's wrong with the linear probability model?

Figure: Stata OLS Estimation of Labour Vote Model

```
Tuesday February 2 13:49:00 2010    Page 1

                                              ___  ____  ____  ____  ____(R)
                                             /__    /   ____/   /   ____/
                                            ___/   /   /___/   /   /___/
                                              Statistics/Data Analysis

1 . regr incumvote retnat class union southwest urban lrself own

      Source |       SS       df       MS              Number of obs =     785
-------------+------------------------------           F(  7,   777) =   18.66
       Model |  25.6683298     7  3.66690426           Prob > F      =  0.0000
    Residual |  152.693454   777  .196516671           R-squared     =  0.1439
-------------+------------------------------           Adj R-squared =  0.1362
       Total |  178.361783   784  .227502275           Root MSE      =   .4433

------------------------------------------------------------------------------
   incumvote |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retnat |  -.1367916   .0194311    -7.04   0.000    -.1749353    -.098648
       class |  -.0715786   .0162049    -4.42   0.000    -.1033892   -.0397679
       union |   .0948849   .0384663     2.47   0.014     .0193747     .170395
   southwest |  -.1605439   .0587451    -2.73   0.006    -.2758619    -.045226
       urban |   .0599385   .0344262     1.74   0.082    -.0076409    .1275179
      lrself |  -.0257666     .00704    -3.66   0.000    -.0395864   -.0119468
         own |  -.1111831   .0380606    -2.92   0.004    -.1858969   -.0364694
       _cons |   .9525068   .0732016    13.01   0.000     .8088105    1.096203
------------------------------------------------------------------------------

2 .
3 . predict yhat
  (option xb assumed; fitted values)
  (339 missing values generated)
```

What's wrong with the linear probability model?



Generating Predicted Probabilities for OLS

Figure: Predicted Probability of Voting: The Data

```
Tuesday February 2 10:52:54 2010   Page 1

                                        (R)
                     Statistics/Data Analysis

1 . list yhat incumvote retnat class union southwest urban lrself own in 1454/1474
```

	yhat	incumv~e	retnat	class	union	southw~t	urban	lrself	own
1454.	-.0357318	0	worse	upper middle	0	0	0	7	1
1455.	.2446636	.	worse	lower middle	0	0	1	4	1
1456.	.27961	.	worse	lower middle	1	0	0	4	1
1457.	-.0414531	.	worse	middle	0	0	0	right	1
1458.	.0184854	0	worse	middle	0	0	1	right	1
1459.	.3673287	.	same	working	0	0	0	5	1
1460.	.06973	0	worse	middle	0	0	0	right	0
1461.	.	.	better	working	0	0	0	.	0
1462.	.5384504	1	same	working	0	0	1	5	0
1463.	.1299316	.	worse	working	0	1	1	5	1
1464.	.2422033	.	worse	middle	1	0	1	5	1
1465.	.4269786	.	same	working	0	0	0	7	0
1466.	.	.	worse	.	0	0	1	5	1
1467.	.	0	worse	middle	0	0	1	.	1
1468.	.6155921	1	better	working	0	0	1	3	1
1469.	.4788004	.	same	working	0	0	1	3	1
1470.	.3853604	0	worse	working	1	0	1	5	1
1471.	-.1563826	.	worse	upper	0	1	1	5	1
1472.	.	0	worse	working	0	0	1	.	0
1473.	.4785118	.	same	working	0	0	0	5	0
1474.	.	.	better	.	1	0	1	5	1

Generating Predicted Probabilities for OLS

Figure: Predicted Probability of Voting: The Data

Tuesday February 2 13:35:30 2010   Page 1

(R)
Statistics/Data Analysis

```
1 . list yhat incumvote retnat class union southwest urban lrself own in 1454/1474, nolabel
```

	yhat	incumv~e	retnat	class	union	southw~t	urban	lrself	own
1454.	-.0357318	0	3	4	0	0	0	7	1
1455.	.2446636	.	3	2	0	0	1	4	1
1456.	.27961	.	3	2	1	0	0	4	1
1457.	-.0414531	.	3	3	0	0	0	10	1
1458.	.0184854	0	3	3	0	0	1	10	1
1459.	.3673287	.	2	1	0	0	0	5	1
1460.	.06973	0	3	3	0	0	0	10	0
1461.	.	.	1	1	0	0	0	.	0
1462.	.5384504	1	2	1	0	0	1	5	0
1463.	.1299316	.	3	1	0	1	1	5	1
1464.	.2422033	.	3	3	1	0	1	5	1
1465.	.4269786	.	2	1	0	0	0	7	0
1466.	.	.	3	.	0	0	1	5	1
1467.	.	0	3	3	0	0	1	.	1
1468.	.6155921	1	1	1	0	0	1	3	1
1469.	.4788004	.	2	1	0	0	1	3	1
1470.	.3853604	0	3	1	1	0	1	5	1
1471.	-.1563826	.	3	5	0	1	1	5	1
1472.	.	0	3	1	0	0	1	.	0
1473.	.4785118	.	2	1	0	0	0	5	0
1474.	.	.	1	.	1	0	1	5	1

Generating Predicted Probabilities for OLS

Figure: Predicted Probability of Voting: The Data

Variable	Value	Beta	Result
incumb_vote			
retnat	3	-0.14	-0.41
class	4	-0.07	-0.29
union	0	0.09	0.00
southwest	0	-0.16	0.00
urban	0	0.06	0.00
lrself	7	-0.03	-0.18
own	1	-0.11	-0.11
constant	1	0.95	0.95
			-0.03

How to address limitations of OLS?

- Any continuous probability distribution defined over the real line would work.
- We use the normal because it's widely studied (which produces probit), and the logistic because it's mathematically convenient (logs—which produces logit).
- Ideologues might have reasons to prefer one to the other. If your results hinge on using one versus the other, you have problems.
- What we need is a probability model that looks like the following:

$$E[y|\mathbf{x}] = 0[1 - F(\mathbf{x}\beta)] + 1[F(\mathbf{x}\beta)] = \mathbf{F}(\mathbf{x}\beta)$$

The Logit Data Generating Process

The general problem with binary data is identifying a data generating process, a probability function, that maps our systematic component $E(y_i|X) = \mathbf{x_i}\beta$ into the unit interval, i.e., between 0 and 1.

$$Prob(y_i = 1|\mathbf{x_i}) = F(\mathbf{x_i}\beta)$$

The logistic distribution is like a normal with longer tails (i.e., more extreme values are likely).

$$Prob(y_i = 1|\mathbf{x_i}) = \frac{1}{1 + e^{-\mathbf{x_i}\beta}} = \Lambda(\mathbf{x_i}\beta)$$

where Λ is the logistic cumulative distribution function.

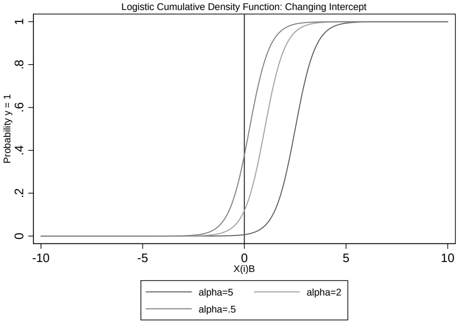
The Logit Data Generating Process

We can generate example of logistic cumulative density function using Stata:

Here we manipulate alpha:

- twoway function y=1/(1+exp(5+2*(-x))), range(-10 10) xline(0) scheme(s1mono)
- || function y=1/(1+exp(2+2*(-x))), range(-10 10)
- || function y=1/(1+exp(.5+2*(-x))), range(-10 10)

The Logit Data Generating Process: Alpha



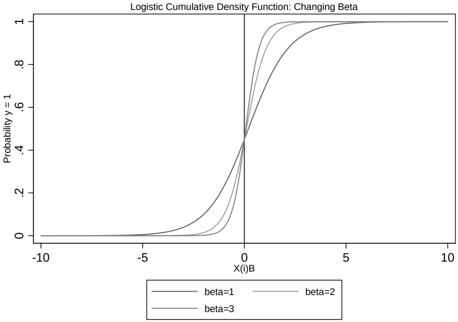
The Logit Data Generating Process

We can generate example of logistic cumulative density function using Stata:

Here we manipulate beta:

- twoway function y=1/(1+exp(5+2*(-x))), range(-10 10) xline(0) scheme(s1mono)
- || function y=1/(1+exp(2+2*(-x))), range(-10 10)
- || function y=1/(1+exp(.5+2*(-x))), range(-10 10)

The Logit Data Generating Process: Beta



The Likelihood Function

- maximum likelihood provides a convenient and powerful method for estimating the parameters of the logit model.
- a key assumption is that the data are identically and independently distributed
- which allows us to form a likelihood function for the whole data from the product of the likelihoods for each observation:

$$(y_1, y_2, ..., y_n) = P(y_1)P(y_2)...P(y_n)$$

$$= \prod_{y_{i=1}} F(\mathbf{x_i}\beta) \prod_{y_{i=0}} [1 - F(\mathbf{x_i}\beta)]$$

The Likelihood Function

In Likelihood notation:

$$L = \prod_{y_{i=1}}^{N} F(\mathbf{x_i}\beta)^{y_i} \prod_{y_{i=0}}^{N} [1 - F(\mathbf{x_i}\beta)]^{1-y_i}$$

- Each observation thus contributes something to the likelihood,
- either in the first part when y_i = 1,
- or in the second part when y_i = 0 (so 1 - y_i = 1).

The Log Likelihood Function

As is typical with MLE, it is easier to work with the log-likelihood:

$$lnL = \sum_{y_{i=1}}^{N} y_i lnF(\mathbf{x_i}\beta) + \sum_{y_{i=0}}^{N} (1 - y_i) ln[1 - F(\mathbf{x_i}\beta)]$$

- The only unknowns here are is the vector of β
- But there is no simple analytic solution so this is typically accomplished iteratively.
- An example using the Labour incumbent voting data. Lets do a couple of iterations by hand.

The Likelihood Function

- In this example is Labour incumbent vote and takes on a value of 1 or 0.
- The independent variable, is income category that ranges in value from 10 (10,000) to 100 (100,000 or greater).

$$lnL = \sum_{y_{i=1}}^{N=10} y_i lnF(\alpha + \beta_1 Income) + \sum_{y_{i=0}}^{N=10} (1 - y_i) ln[1 - F(\alpha + \beta_1 Income)]$$

The Likelihood Function

- This can simply be calculated by hand
- Remember that the CDF for the logit is

$$F(\alpha + \beta_1 Income) = \frac{1}{1 + e^{-(\alpha + \beta_1 Income)}}$$

which is equivalent to

$$F(\alpha + \beta_1 Income) = \frac{e^{(\alpha + \beta_1 Income)}}{1 + e^{(\alpha + \beta_1 Income)}}$$