Intermediate Social Statistics Hilary 2009 Lecture 5: Bi-variate Probit and Ordered Probit

Raymond Duch

Nuffield College

February 17, 2009

Predictions in Logit and Probit

```
     +-------------------------------------------+
     | yhat_probitt   yhat_logit  incumvote  predict_vote |
     |-------------------------------------------|
  1. | .3928562    .3869703          1           0 |
  2. |  .560674    .5643634          .           1 |
  3. | .3432808    .3406143          1           0 |
  4. | .5578907    .5597182          .           1 |
  5. |  .560674    .5643634          .           1 |
     |-------------------------------------------|
  6. | .4942218    .4918689          0           0 |
  7. | .5892738    .5929783          .           1 |
  8. | .7399474    .7458074          0           1 |
  9. | .4524617    .4533052          0           0 |
 10. | .3928562    .3869703          0           0 |
     |-------------------------------------------|
 11. | .3445356    .3425307          1           0 |
 12. | .2229143    .2179674          0           0 |
 13. | .7816748    .7864491          0           1 |
 14. | .4901605    .4919069          .           0 |
 15. | .3007564    .2955182          .           0 |
     |-------------------------------------------|
 16. |        .           .          1           1 |
 17. | .2117439    .2085973          .           0 |
 18. |        .           .          .           1 |
 19. | .2395462    .2338773          0           0 |
 20. | .4267031    .4243549          .           0 |
```

Predictions in Logit and Probit

```
. summarize correct_1 correct_0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   correct_1 |      1500        .114    .3179173          0          1
   correct_0 |      1500    .2993333    .4581188          0          1
```

PCP versus ePCP

$$ePCP = \frac{1}{N}(\sum_{y_i=1}^{N} p_i + \sum_{y_i=0}^{N}(1 - p_i)) \quad (1)$$

Confidence Intervals on Probit Predictions

```
Confidence intervals by delta method
                              95% Conf. Interval
  Pr(y=1|x):          0.3929   [ 0.2980,    0.4877]
  Pr(y=0|x):          0.6071   [ 0.5123,    0.7020]
       retnat       class      union  southwest       urban      lrself         own
x=          3           1          0          0           1           5           0

. prvalue in 35

probit: Predictions for incumvote

Confidence intervals by delta method
                              95% Conf. Interval
  Pr(y=1|x):          0.1443   [ 0.0945,    0.1941]
  Pr(y=0|x):          0.8557   [ 0.8059,    0.9055]

       retnat       class      union  southwest       urban      lrself         own
x=          3           3          0          0           1           5           1
. prvalue in 68

probit: Predictions for incumvote

Confidence intervals by delta method
                              95% Conf. Interval
  Pr(y=1|x):          0.2010   [ 0.1094,    0.2926]
  Pr(y=0|x):          0.7990   [ 0.7074,    0.8906]
       retnat       class      union  southwest       urban      lrself         own
x=          2           4          0          0           0           6           0
```

Deriving The Likelihood Function

- We are modeling actually a latent quantity that gives rise to the observed discrete outcomes.
- Think of this underlying latent quantity as a "probability" or a "random utility".
- We model the unobserved net utilities y^* of the choices y via the model:

$$y^* = \mathbf{x_i}\beta + \varepsilon_i \tag{2}$$

Deriving The Likelihood Function

- With either iid N(0,1) for probit.
- We don't observe the net utility of the choice , just whether it was made or not.
- So we observe
  - whether a person voted for Labour or for the Opposition;
  - whether an individual made a campaign contribution or did not;
  - whether a country went to war or did not;
  - whether someone died or did not.
- We posit that

$$
\begin{aligned}
y_i &= 0 \text{ if } y_i^* \leq 0 \\
&= 1 \text{ if } y_i^* > 0
\end{aligned}
$$

The Probit Estimator

Some simple algebra gives us our estimator.

$$
\begin{aligned}
Prob(y_i = 1|\mathbf{x_i}) &= Prob(y_i^* > 0) \\
&= Prob(\mathbf{x_i}\beta + \varepsilon_i) > 0 \\
&= Prob(\varepsilon_i > -\mathbf{x_i}\beta) \\
&= Prob(\frac{\varepsilon_i}{\sigma} > \frac{-\mathbf{x_i}\beta}{\sigma})
\end{aligned} \tag{3}
$$

The Probit Estimator

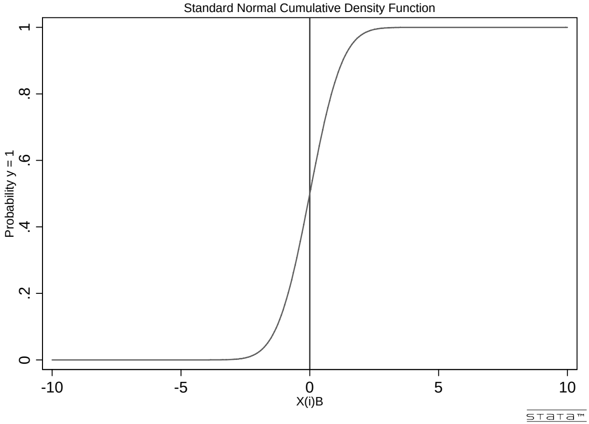
Because the error has a normal distribution this becomes

$$Prob(y_i = 1|\mathbf{x_i}) = 1 - \Phi(\frac{-\mathbf{x_i}\beta}{\sigma}) = \Phi(\frac{\mathbf{x_i}\beta}{\sigma}) \tag{4}$$

Similarly,

$$Prob(y_i = 0|\mathbf{x_i}) = 1 - \Phi(\frac{\mathbf{x_i}\beta}{\sigma}) \tag{5}$$

This is simply the standard normal cumulative density function and as you can see it closely resembles the logistic CDF that we examined last week.


Standard Normal Cumulative Density Function
Probability y = 1
0
.2
.4
.6
.8
1
-10
-5
0
5
10
X(i)B
stata™

Probit Estimation: More MLE

Maximum likelihood provides a convenient and powerful method for estimating the parameters of this probit 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:

$$
\begin{aligned}
(y_1, y_2, ..., y_n) &= P(y_1)P(y_2)...P(y_n) \\
&= \prod_{y_{i=1}} \Phi(\mathbf{x_i}\beta) \prod_{y_{i=0}} [1 - \Phi(\mathbf{x_i}\beta)]
\end{aligned} \tag{6}
$$

In Log Likelihood notation:

$$
LnL = \sum_{y_{i=1}}^{N} \{ y_i ln\Phi(\mathbf{x_i}\beta) + (1 - y_i) ln[1 - \Phi(\mathbf{x_i}\beta)] \} \tag{7}
$$

Probit Estimation: More MLE

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 - $y_i = 1$).

$$lnL = \sum_{y_{i=1}}^{N} y_i ln\Phi(\mathbf{x_i}\beta) + \sum_{y_{i=0}}^{N}(1-y_i)ln[1-\Phi(\mathbf{x_i}\beta)] \qquad (8)$$

- The only unknowns here are is the vector of β
- Can use calculus to optimize to solve for the unknown vector of
- But there is no simple analytic solution so this is typically accomplished iteratively.
- Lets do a couple of iterations by hand

Probit Estimation: More MLE

- Lets return to our example from the UK 2004 Election study.
- In this example the dependent variable is Labour incumbent vote and it 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).

$$lnL = \sum_{y_{i=1}}^{N} y_i ln\Phi(\alpha+\beta_1*(Income)) + \sum_{y_{i=0}}^{N}(1-y_i)ln[1-\Phi(\alpha+\beta_1*(Income))] \quad (9)$$

Probit Estimation: More MLE

- The two unknowns in this likelihood function are α and β_1.
- For this example, lets pick as our starting values α=1.2 and β_1=-.02.
- And lets calculate the likelihood for a respondent with an income of 10 (10,000 GBP) and who expressed a Labour preference (vote=1).

$$lnL = 1ln\Phi(1.2 - .02 * (Income))$$

$$= -.173$$

Probit Estimation: More MLE

- And for the person who expressed a preference for an opposition candidate and who had an income of 50 (50,000 GBP)...

$$lnL = (1-0)ln[1-\Phi(1.2-.02*(Income))]$$
$$= -.866$$

Predictions in Logit and Probit

	vote	income	alpha	beta	logindi~1	loglike	alpha2	beta2	logindi~2	loglike2
1.	1	10	1.2	-.02	-.1727538	-4.882811	2.38	-.052	-.0319477	-3.484201
2.	1	20	1.2	-.02	-.2380737	-4.882811	2.38	-.052	-.0944455	-3.484201
3.	1	30	1.2	-.02	-.3205539	-4.882811	2.38	-.052	-.2308079	-3.484201
4.	0	40	1.2	-.02	-1.065434	-4.882811	2.38	-.052	-.9621029	-3.484201
5.	0	50	1.2	-.02	-.8657396	-4.882811	2.38	-.052	-.5326208	-3.484201
6.	1	60	1.2	-.02	-.6931472	-4.882811	2.38	-.052	-1.471199	-3.484201
7.	0	70	1.2	-.02	-.5460044	-4.882811	2.38	-.052	-.1096304	-3.484201
8.	0	80	1.2	-.02	-.4224764	-4.882811	2.38	-.052	-.0382607	-3.484201
9.	0	90	1.2	-.02	-.320554	-4.882811	2.38	-.052	-.010782	-3.484201
10.	0	100	1.2	-.02	-.2380737	-4.882811	2.38	-.052	-.0024041	-3.484201

Predictions in Logit and Probit

As it turns out α=2.38 and β_1=-.052 maximize the log likelihood.

```
. probit vote income

Iteration 0:   log likelihood = -6.7301167
Iteration 1:   log likelihood =  -3.846122
Iteration 2:   log likelihood = -3.5128548
Iteration 3:   log likelihood = -3.4841193
Iteration 4:   log likelihood = -3.4837079
Iteration 5:   log likelihood = -3.4837078

Probit regression                                 Number of obs   =         10
                                                  LR chi2(1)      =       6.49
                                                  Prob > chi2     =     0.0108
Log likelihood = -3.4837078                       Pseudo R2       =     0.4824

------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |  -.0524063   .0292291    -1.79   0.073    -.1096943    .0048817
       _cons |   2.384189   1.496149     1.59   0.111    -.5482103    5.316588
------------------------------------------------------------------------------
```

Probit Regression Results

Here are the results of a probit model with Labour as the dependent variable and the same set of independent variables we have used in previous examples...

```
. probit  incumvote retnat class union southwest urban lrself own

Iteration 0:   log likelihood = -507.77976
Iteration 1:   log likelihood = -444.89944
Iteration 2:   log likelihood = -443.61601
Iteration 3:   log likelihood = -443.61387

Probit regression                                 Number of obs   =        785
                                                  LR chi2(7)      =     128.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -443.61387                       Pseudo R2       =     0.1264

------------------------------------------------------------------------------
   incumvote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retnat |  -.4202345   .0610461    -6.88   0.000    -.5398826   -.3005864
       class |   -.232117   .0510478    -4.55   0.000    -.3321689   -.1320651
       union |   .2976013   .1156959     2.57   0.010     .0708414    .5243611
   southwest |   -.510433   .1970207    -2.59   0.010    -.8965865   -.1242795
       urban |    .189083   .0632399     2.99   0.003      .065135     .313031
      lrself |  -.0795816   .0216375    -3.68   0.000    -.1219903   -.0371729
         own |  -.3172383   .1135865    -2.79   0.005    -.5398637   -.0946129
       _cons |    1.16181   .2511572     4.63   0.000     .6695505    1.654069
------------------------------------------------------------------------------
```

Interpreting the Probit Coefficients

So if we're going to find the marginal effects of a change in some x on the probability of y = 1 (for example, the marginal effect of changing left-right self identification on Labour vote probabilities), we have to note that the parameters of the model, β, are not the marginal effect of x on y. In general:

$$\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x_k}} = \frac{d\Phi(\mathbf{x}\beta)}{d(\mathbf{x_k})} = \phi(\mathbf{x}\beta)\beta_k \tag{10}$$

Probit Regression Results

where ϕ denotes the standard Normal Probability Density Function (PDF). And it's important to note that the values of the marginal effects (which is what the partials of y wrt x are) will change with x. That's what makes them nonlinear. (Think of a line. The marginal effect of x on y does not change anywhere over the course of that line. It is always β. Not here.)

Probit Regression Results

One of the most meaningful strategies for calculating the impact of x on y is to evaluate the impact of incrementing x_k by δ is to increment x_k by δ for each observation in the data and then calculate the average change in y.

$$\frac{\Delta Pr(y_i = 1|\mathbf{x_i})}{\Delta x_{i,k}} \tag{11}$$
$$= Pr(y_i \quad = \quad 1|\mathbf{x_i}, x_{i,k} + \delta) - Pr(y_i = 1|\mathbf{x_i}, x_{i,k})$$

Predictions in Logit and Probit

For each individual, i, a change in the variable x_k to $x_k + \delta$, the predicted probability of an event changes by $\frac{\Delta Pr(y_i=1|\mathbf{x_i})}{\Delta x_{i,k}}$, holding all other variables at their actual values for that individual i.

The impact of this change evaluated over the entire sample of respondents is simply the mean $\Delta Pr(y_i = 1|\mathbf{x_i})$

Predictions in Logit and Probit

Lets work through an example from the probit model of Labour vote preference. Here we want to determine the impact on the probability of voting Labour of a unit change in the economic evaluation variable (retnat), i.e., where $x_k = retnat$

$$\begin{aligned}\frac{\Delta Pr(y_i = 1|\mathbf{x_i})}{\Delta x_{i,k}} &= \qquad (12) \\ \Pr(Labour_i &= 1|\mathbf{x_i}, retnat_i + 1) - Pr(Labour_i = 1|\mathbf{x_i}, retnat_i)\end{aligned}$$

Generated Estimated Effects

For each individual in data set the estimate of the $Pr(Labour_i = 1|\mathbf{x_i}, retnat_i)$ is simply

$$\begin{aligned} Pr(Labour_i &= 1|\mathbf{x_i}, retnat_i) \\ &= 1 - \Phi(\alpha + \beta_1 * (retnat) + \beta_2 * (class) + \beta_3 * (union) \\ &+ \beta_4 * (southwest) + \beta_5 * (urban) + \beta_6 * (lrself) + \beta_7 * (own)) \\ &= 1 - \Phi(1.16 - 0.42 * (retnat) - 0.23 * (class) + 0.297 * (union) \\ &- 0.51 * (southwest) + 0.189 * (urban) - 0.079 * (lrself) - 0.317 * (own)) \end{aligned} \quad (13)$$

Generated Estimated Effects

And for each individual the estimate of the $Pr(Labour_i = 1|\mathbf{x_i}, retnat_i + 1)$ is simply

$$\begin{aligned} Pr(Labour_i &= 1|\mathbf{x_i}, retnat_i + 1) \\ &= 1 - \Phi(\alpha + \beta_1 * (retnat) + \beta_2 * (class) + \beta_3 * (union)\beta_4 * (southwest) \\ &+ \beta_5 * (urban) + \beta_6 * (lrself) + \beta_7 * (own)) \\ &= 1 - \Phi(1.16 - 0.42 * (retnat) - 0.23 * (class) + 0.297 * (union) - 0.51 * (southwest) \\ &+ 0.189 * (urban) - 0.079 * (lrself) - 0.317 * (own)) \end{aligned} \tag{14}$$

The Delta Method

	retnat	class	union	southw~t	urban	lrself	own	incumvote	XB_1	yhat_1	XB_2	yhat_2	delta
40.	worse	middle	non-union	0	rural	5	own	Opposition	-1.321308	.0931993	-1.321308	.0931993	0
41.	worse	middle	union	0	large town	4	rent	Labour	-.2487211	.4017883	-.2487211	.4017883	0
42.	worse	lower middle	non-union	0	small-medium town	4	own	Labour	-.8205267	.205958	-.8205267	.205958	0
43.	worse	lower middle	union	1	large town	5	own	Opposition	-.923857	.1777804	-.923857	.1777804	0
44.	better	lower middle	non-union	1	rural	left	own	Opposition	-.4408289	.3296684	-.8610634	.1946016	-.1350669
45.	same	working	non-union	0	large town	5	own	Opposition	-.0586738	.476606	-.4789082	.316002	-.160604
46.	same	lower middle	non-union	0	small-medium town	5	own	Opposition	-.4798737	.3156586	-.9001082	.1840313	-.1316273
47.	same	middle	non-union	0	large town	7	own	Opposition	-.6820709	.2475971	-1.102305	.1351645	-.1124326
48.	same	upper middle	non-union	0	rural	6	own	Opposition	-1.212772	.1126084	-1.633007	.0512338	-.0613747
49.	worse	middle	non-union	0	small-medium town	7	own	Opposition	-1.291388	.0982845	-1.291388	.0982845	0
50.	same	working	union	0	large town	6	own	.	.1593459	.5633019	-.2608886	.3970892	-.1662126
51.	worse	middle	union	0	small-medium town	6	own	Opposition	-.9142056	.1803044	-.9142056	.1803044	0
52.	worse	working	non-union	0	small-medium town	3	own	Labour	-.508828	.3054364	-.508828	.3054364	0
53.	same	.	non-union	0	small-medium town	5	own	.	.	.	.	.	.
54.	better	middle	union	0	rural	4	own	Labour	-.1036564	.458721	-.5238909	.3001772	-.1585438
55.	better	middle	non-union	0	small-medium town	5	own	Opposition	-.2917562	.3852365	-.7119907	.2382353	-.1470013
56.	worse	working	non-union	1	small-medium town	4	own	.	-1.098843	.1359183	-1.098843	.1359183	0
57.	same	middle	union	0	large town	5	own	Opposition	-.2253065	.4108704	-.645541	.2592883	-.1515821
58.	better	middle	non-union	0	large town	3	own	Labour	.0564899	.5225242	-.3637446	.3580244	-.1644999
59.	better	middle	non-union	0	small-medium town	left	own	Opposition	.0265701	.5105987	-.3936644	.3469144	-.1636842
60.	better	middle	non-union	0	small-medium town	6	own	Opposition	-.3713378	.355193	-.7915723	.214305	-.1408879

Ordered Probit

- Consider the 4-point democratic satisfaction measure from the 2004 UK Election study.
- If we use this as a dependent variable and estimate using OLS, all differences will be treated as equal.
- OLS will also generate predicted y that will be "below" the "not at all satisfied" option, and "above" the "very satisfied" option.

Ordered Probit

We assume that individual opinion is continuous, but unobserved. So just as we did with probit, we treat this is a latent regression model, where:

$$y^* = \mathbf{x_i}\beta + \varepsilon_i \tag{15}$$

Ordered Probit

- Again the y^* is unobserved, but we do observe discrete outcomes.
- Here I work out the example of the democratic satisfaction dependent variable
- It has four discrete ordered outcomes: 1, 2, 3 and 4.

Ordered Probit

$$
\begin{aligned}
y_i &= 1 \text{ if } y_i^* < \gamma_1 \\
&= 2 \text{ if } \gamma_1 \geq y_i^* < \gamma_2 \\
&= 3 \text{ if } \gamma_2 \geq y_i^* < \gamma_3 \\
&= 4 \text{ if } \gamma_3 \leq y_i^*
\end{aligned} \tag{16}
$$

Ordered Probit

So, thinking of this as a probit, we have:

$$\begin{aligned} Prob(y_i &= 1|\mathbf{x_i}) = Prob(y_i^* < \gamma_1) \\ &= Prob(\mathbf{x_i}\beta + \varepsilon_i < \gamma_1) \\ &= Prob(\varepsilon_i < \gamma_1 - \mathbf{x_i}\beta) \\ &= \Phi(\gamma_1 - \mathbf{x_i}\beta) \end{aligned} \tag{17}$$

Ordered Probit

$$\begin{aligned} Prob(y_i &= 2|\mathbf{x_i}) = Prob(\gamma_1 \leq y_i^* < \gamma_2) \\ &= Prob(\gamma_1 \leq \mathbf{x_i}\beta + \varepsilon_i < \gamma_2) \\ &= Prob(\varepsilon_i < \gamma_2 - \mathbf{x_i}\beta) - Prob(\varepsilon_i < \gamma_1 - \mathbf{x_i}\beta) \\ &= \Phi(\gamma_2 - \mathbf{x_i}\beta) - \Phi(\gamma_1 - \mathbf{x_i}\beta) \end{aligned} \tag{18}$$

Ordered Probit

$$\begin{aligned} Prob(y_i &= 3|\mathbf{x_i}) = Prob(\gamma_2 \leq y_i^* < \gamma_3) \\ &= Prob(\gamma_2 \leq \mathbf{x_i}\beta + \varepsilon_i < \gamma_3) \\ &= Prob(\varepsilon_i < \gamma_3 - \mathbf{x_i}\beta) - Prob(\varepsilon_i < \gamma_2 - \mathbf{x_i}\beta) \\ &= \Phi(\gamma_3 - \mathbf{x_i}\beta) - \Phi(\gamma_2 - \mathbf{x_i}\beta) \end{aligned} \tag{19}$$

Ordered Probit

$$
\begin{aligned}
Prob(y_i &= 4|\mathbf{x_i}) = Prob(y_i^* \geq \gamma_3) \\
&= Prob(\mathbf{x_i}\beta + \varepsilon_i \geq \gamma_3) \\
&= Prob(\varepsilon_i \geq \gamma_3 - \mathbf{x_i}\beta) \\
&= 1 - \Phi(\gamma_3 - \mathbf{x_i}\beta)
\end{aligned} \tag{20}
$$