Econometrics II: Econometric Modelling

Jürgen Meinecke

Research School of Economics, Australian National University

5 October, 2018

Assignment 2

You can already start working on Assignment 2

After today's lecture and tutorial you should be able to solve most of the assignment already

Deadline: 17 October at 12:00pm (noon, mid-day)

Reminder: my deadlines are *very* sharp!
(If you submit at 12:01pm you will receive a mark of zero!)

No extensions given under any circumstances!

Note: I do not offer any help on solving the assignment!

Roadmap

Introduction

Models with Binary Dependent Variable

- Linear Probability Model: OLS
- Nonlinear Probability Models: Probit Model
- Nonlinear Probability Models: Logit Model

So far we looked at the dependent variable Y_i as being continuous

- longevity after heart attack
- gdp growth
- income

What if Y_i is binary?

- Y_i = get into college, or not
- Y_i = person smokes, or not
- Y_i = mortgage application is accepted, or not

Mortgage Denial and Race The Boston Fed HMDA Dataset

- Individual applications for single-family mortgages made in 1990 in the greater Boston area
- 2380 observations, collected under Home Mortgage Disclosure Act (HMDA)

Variables

- Dependent variable:
  - Is the mortgage denied or accepted?
- Independent variables:
  - income, wealth, employment status
  - other loan, property characteristics
  - race of applicant

A natural starting point is the linear regression model with a single regressor:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

But:

- What does β_1 mean when Y is binary? Is $\beta_1 = \frac{\Delta Y}{\Delta X}$?
- What does the line $\beta_0 + \beta_1 X$ mean when Y is binary?
- What does the predicted value $\hat{Y}$ mean when Y is binary? For example, what does $\hat{Y} = 0.26$ mean?

In the linear probability model, β_1 is the change in that predicted probability for a unit change in X_i

Linear probability model: $Y_i = \beta_0 + \beta_1 X_i + u_i$

$$\text{E}[Y_i|X_i] = \text{E}[\beta_0 + \beta_1 X_i + u_i|X_i] = \beta_0 + \beta_1 X_i,$$

under OLS Assumption 1, $\text{E}[u_i|X_i] = 0$

At the same time, when Y_i is binary,

$$\begin{aligned} \text{E}[Y_i|X_i] &= 1 \cdot \Pr(Y = 1|X_i) + 0 \cdot \Pr(Y = 0|X_i) \\ &= \Pr(Y = 1|X_i) \end{aligned}$$

Connecting the dots:

$$\Pr(Y = 1|X) = \beta_0 + \beta_1 X_i$$

When Y is binary, the linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

is called the *linear probability model* because

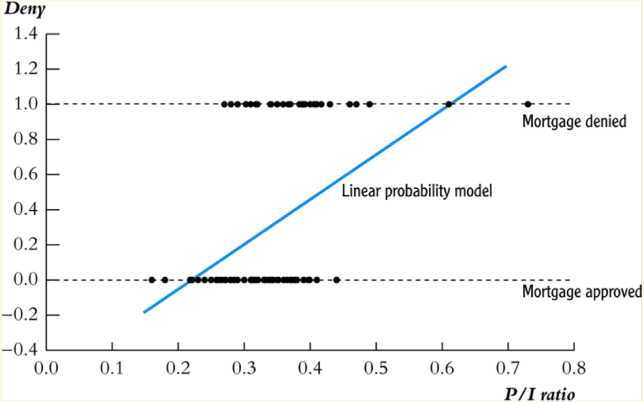
$$\Pr(Y_i = 1|X_i) = \beta_0 + \beta_1 X_i$$

The probability is modeled to be linear

The coefficient β_1 = change in probability that $Y_i = 1$ for a unit change in X_i:

$$\beta_1 = \Pr(Y_i = 1|X_i = x + 1) - \Pr(Y_i = 1|X_i = x)$$

Example: Mortgage denial versus ratio of debt payments to income (P/I ratio) in a subset of the HMDA data set (n = 127)



$$\widehat{deny} = \underset{(0.032)}{-0.080} + \underset{(0.098)}{0.604} \cdot P/Iratio \quad (n = 2,380)$$

Predicted probability of denial:

- for applicant with P/I ratio = .3:
  $\widehat{\Pr}(deny = 1|P/Iratio = .3) = -0.080 + 0.604 \cdot 0.3 = 0.101$
- for applicant with P/I ratio = .4:
  $\widehat{\Pr}(deny = 1|P/Iratio = .4) = -0.080 + 0.604 \cdot 0.4 = 0.162$
- for applicant with P/I ratio = .5:
  $\widehat{\Pr}(deny = 1|P/Iratio = .5) = -0.080 + 0.604 \cdot 0.5 = 0.222$
- for applicant with P/I ratio = .6:
  $\widehat{\Pr}(deny = 1|P/Iratio = .6) = -0.080 + 0.604 \cdot 0.6 = 0.282$
- for applicant with P/I ratio = .7:
  $\widehat{\Pr}(deny = 1|P/Iratio = .7) = -0.080 + 0.604 \cdot 0.7 = 0.343$

The effect on the probability of denial of an increase in P/I ratio by 0.10 (successively) is to increase the probability by 0.0604, that is, by 6.04 percentage points

That, of course, is exactly equal to one tenth of $\hat{\beta}_1$

Probability changes linearly

Next include *black* as a regressor:

$$\widehat{deny} = \underset{(0.032)}{-0.091} + \underset{(0.098)}{0.559}P/Iratio + \underset{(0.025)}{0.177}black$$

Predicted probability of denial:

- for black applicant with P/I ratio = .3:
  $\widehat{\Pr}(deny = 1) = -0.091 + 0.559 \cdot 0.3 + 0.177 \cdot 1 = 0.254$
- for white applicant, P/I ratio = .3:
  $\widehat{\Pr}(deny = 1) = -0.091 + 0.559 \cdot 0.3 + 0.177 \cdot 0 = 0.077$

Difference = 0.177 = 17.7 percentage points

Coefficient on black is significant at the 5% level

Still plenty of room for omitted variable bias…

The linear probability model models $\Pr(Y_i = 1|X_i)$ as a linear function of X_i

- Advantages
  - simple to estimate and to interpret
  - inference is the same as for multiple regression (use heteroskedasticity-robust standard errors)
- Disadvantages
  - A LPM says that the change in the predicted probability for a given change in X_i is the same for all values of X_i, but that doesn't always seem sensible
  - Also, LPM predicted probabilities can be < 0 or > 1!
- These disadvantages can be solved by using a *nonlinear* probability model: probit and logit regression

Roadmap

The problem with the linear probability model is that it models the probability of Y=1 as being linear:

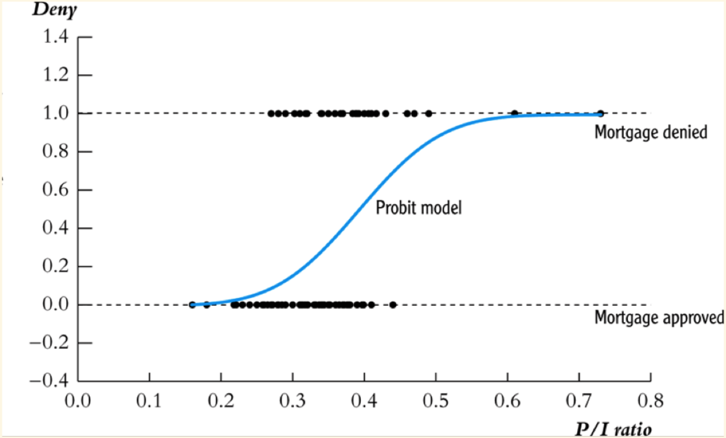
$$\Pr(Y_i = 1|X_i) = \beta_0 + \beta_1 X_i$$

Instead, we want:

1. $\Pr(Y_i = 1|X_i)$ to be have different slopes as X_i changes
2. $0 \leq \Pr(Y_i = 1|X_i) \leq 1$ for all X_i

This requires using a *nonlinear* functional form for the probability

How about an "S-curve"...



The probit model satisfies these conditions:

1. $\Pr(Y_i = 1|X_i)$ is now nonlinear in X_i for $\beta_1 > 0$
2. $0 \leq \Pr(Y_i = 1|X_i) \leq 1$ for all X_i

Probit regression models the probability that $Y_i = 1$ using the cumulative standard normal distribution function, $\Phi(z)$, evaluated at $z = \beta_0 + \beta_1 X$

The probit regression model is,

$$\Pr(Y_i = 1|X_i) = \Phi(\beta_0 + \beta_1 X_i)$$

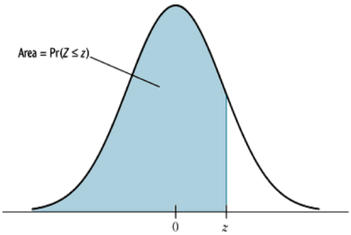
where Φ is the cumulative normal distribution function and $z = \beta_0 + \beta_1 X_i$ is the "z-value" or "z-index" of the probit model

Example: Suppose $\beta_0 = -2, \beta_1 = 3, X = 0.4$, so

$$\Pr(Y_i = 1|X_i = 0.4) = \Phi(-2 + 3 \times 0.4) = \Phi(-0.8)$$

STAT1008 refresher: $\Pr(Y_i = 1|X_i = 0.4)$ = area under the standard normal density to left of $z = -0.8$, which is …

TABLE 1 **The Cumulative Standard Normal Distribution Function, $\Phi(z)$ = Pr(Z " z)**



z	Second Decimal Value of z									
	0	1	2	3	4	5	6	7	8	9
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121

$\Pr(z \leq -0.8) = 0.2119$

(The good old days when people still used standard normal tables. . .)

Why use the cumulative normal probability distribution?

- The "S-shape" gives us what we want:
  - $\Pr(Y_i = 1|X_i)$ to be increasing in X_i for $\beta_1 > 0$, and
  - $0 \leq \Pr(Y_i = 1|X_i) \leq 1$ for all X_i
- Easy to use:
  the probabilities are computed rapidly by Stata
- Relatively straightforward interpretation:
  - $\beta_0 + \beta_1 X_i =$ z-value
  - $\hat{\beta}_0 + \hat{\beta}_1 X_i$ is the predicted z-value, given X_i
  - β_1 is the change in the z-value for a unit change in X_i

STATA Example: HMDA data

```
probit deny p_irat, robust

Iteration 0:   log likelihood =  -872.0853
Iteration 1:   log likelihood =  -835.6633
Iteration 2:   log likelihood = -831.80534
Iteration 3:   log likelihood = -831.79234
Probit estimates                                  Number of obs   =       2380
                                                  Wald chi2(1)    =      40.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -831.79234                       Pseudo R2       =     0.0462
------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      p_irat |   2.967908   .4653114     6.38   0.000     2.055914    3.879901
       _cons |  -2.194159   .1649721   -13.30   0.000    -2.517499    -1.87082
------------------------------------------------------------------------------
```

$$\widehat{\Pr}(deny = 1 | P/I\ ratio) = \Phi(\underset{(0.16)}{-2.19} + \underset{(0.47)}{2.97} \cdot P/I\ ratio)$$

$$\widehat{\Pr}(deny = 1 | P/I\ ratio) = \Phi(\underset{(0.16)}{-2.19} + \underset{(0.47)}{2.97} \cdot P/I\ ratio)$$

- Positive coefficient: does this make sense?

$$\begin{aligned}
\widehat{\Pr}(deny = 1 | P/I\ ratio = 0.3) &= \Phi(-2.19 + 2.97 \cdot 0.3) \\
&= \Phi(-1.30) = .097 \\
\widehat{\Pr}(deny = 1 | P/I\ ratio = 0.4) &= \Phi(-2.19 + 2.97 \cdot 0.4) \\
&= \Phi(-1.00) = 0.158 \\
\widehat{\Pr}(deny = 1 | P/I\ ratio = 0.5) &= \Phi(-2.19 + 2.97 \cdot 0.5) \\
&= \Phi(-0.71) = 0.240 \\
\widehat{\Pr}(deny = 1 | P/I\ ratio = 0.6) &= \Phi(-2.19 + 2.97 \cdot 0.6) \\
&= \Phi(-0.41) = 0.342 \\
\widehat{\Pr}(deny = 1 | P/I\ ratio = 0.7) &= \Phi(-2.19 + 2.97 \cdot 0.7) \\
&= \Phi(-0.11) = 0.456
\end{aligned}$$

Comparison of LPM and probit predicted probabilities

P/I ratio	LPM	Probit
0	-8.0%	1.4%
0.1	-2.0%	2.9%
0.2	4.1%	5.5%
0.3	10.1%	9.7%
0.4	16.2%	15.5%
0.5	22.2%	24.0%
0.6	28.2%	34.2%
0.7	34.3%	45.6%
0.8	40.3%	57.4%
0.9	46.4%	68.5%
1	52.4%	78.2%

Probit has marginal probabilities that are increasing at an increasing rate at first, then at decreasing rate

Adding explanatory variables is straightforward

$$\Pr(Y_i = 1 | X_{1i}, X_{2i}) = \Phi(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})$$

- Φ is still the cumulative normal distribution function
- $z = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$ is the "z-value" or "z-index" of the profit model
- β_1 is the effect on the z-score of a unit change in X_{1i}, holding constant X_{2i}

STATA Example, ctd.: Predicted probit probabilities

```
probit deny p_irat black, robust

Probit estimates                                  Number of obs   =       2380
                                                  Wald chi2(2)    =     118.18
                                                  Prob > chi2     =     0.0000
Log likelihood = -797.13604                       Pseudo R2       =     0.0859
------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      p_irat |   2.741637   .4441633     6.17   0.000     1.871092    3.612181
       black |   .7081579   .0831877     8.51   0.000      .545113    .8712028
       _cons |  -2.258738   .1588168   -14.22   0.000    -2.570013   -1.947463
------------------------------------------------------------------------------

Now we are computing predicted probability
scalar z1 = _b[_cons]+_b[p_irat]*.3+_b[black]*0;
display "Pred prob, p_irat=.3, white: " normprob(z1);
Pred prob, p_irat=.3, white: .07546603

'b[_cons]' is the estimated intercept (-2.258738)
'b[p_irat]' is the coefficient on p_irat (2.741637)
'scalar' creates a new scalar which is the result of a calculation
'display' prints the indicated information to the screen
```

$$\widehat{\Pr}(deny = 1|P/I, black) =$$

$$\Phi(\underset{(0.16)}{-2.26} + \underset{(0.44)}{2.74} \times P/I\ ratio + \underset{(0.08)}{0.71} \times black)$$

- Is the coefficient on black statistically significant?
- Estimated effect of race for P/I ratio = 0.3:

$$\widehat{Pr}(deny = 1|0.3, 1) = \Phi(-2.26 + 2.74 \cdot 0.3 + 0.71 \cdot 1) = 0.233$$

$$\widehat{Pr}(deny = 1|0.3, 0) = \Phi(-2.26 + 2.74 \cdot 0.3 + 0.71 \cdot 0) = 0.075$$

- Difference in rejection probabilities = 0.158 (15.8 percentage points)
- Still plenty of room for omitted variable bias!

Easier way to calculate predicted probabilities in Stata

```
margins, at(pi_rat=0.3 black=(0 1))

Adjusted predictions                              Number of obs   =       2380
Model VCE    : OIM

Expression   : Pr(deny), predict()

1._at        : pi_rat          =          .3
               black           =           0

2._at        : pi_rat          =          .3
               black           =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |    .075466   .0060601    12.45   0.000     .0635884    .0873436
          2  |   .2332795   .0233158    10.01   0.000     .1875815    .2789776
------------------------------------------------------------------------------
```

Roadmap

Logit regression models the probability of $Y_i = 1$, given X_i, as the cumulative standard logistic distribution function, evaluated at $z = \beta_0 + \beta_1 X_i$:

$$\Pr(Y_i = 1|X_i) = F(\beta_0 + \beta_1 X_i)$$

where F is the cumulative logistic distribution function:

$$F(z) = \frac{1}{1 + e^{-z}}$$

Because logit and probit use different probability functions, the coefficients (β's) are different in logit and probit

$$\Pr(Y_i = 1|X_i) = F(\beta_0 + \beta_1 X_i)$$

where $F(\beta_0 + \beta_1 X_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}}$

Example: $\beta_0 = -2, \beta_1 = 3, X = 0.4$, as before

$$\beta_0 + \beta_1 X = -2 + 3 \times 0.4 = -0.8$$

$$\Pr(Y = 1|X = 0.4) = \frac{1}{1 + e^{-(-0.8)}} = 0.31$$

Compare that to 0.21 for Φ

Illustrates: logit has fatter tails

(but in the center they are quite similar)

Why bother with logit if we have probit?

- The main reason is historical: logit is computationally faster & easier, but that doesn't matter nowadays
- In practice, logit and probit are very similar - since empirical results typically don't hinge on the logit/probit choice, both tend to be used in practice

STATA Example: HMDA data

```
logit deny p_irat black, robust

Iteration 0:   log likelihood =  -872.0853
Iteration 1:   log likelihood =  -806.3571
Iteration 2:   log likelihood = -795.74477
Iteration 3:   log likelihood = -795.69521
Iteration 4:   log likelihood = -795.69521
Logit estimates                                   Number of obs   =       2380
                                                  Wald chi2(2)    =     117.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -795.69521                       Pseudo R2       =     0.0876
------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      p_irat |   5.370362   .9633435     5.57   0.000     3.482244    7.258481
       black |   1.272782   .1460986     8.71   0.000     .9864339     1.55913
       _cons |  -4.125558    .345825   -11.93   0.000    -4.803362   -3.447753
------------------------------------------------------------------------------
```

Predicted probabilities in Stata

```
margins, at(pi_rat=0.3 black=(0 1))

Adjusted predictions                              Number of obs   =       2380
Model VCE    : OIM

Expression   : Pr(deny), predict()

1._at        : pi_rat          =          .3
               black           =           0

2._at        : pi_rat          =          .3
               black           =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0748514   .0063373    11.81   0.000     .0624305    .0872724
          2  |   .2241459   .0239438     9.36   0.000     .1772169    .2710749
------------------------------------------------------------------------------
```

The predicted probabilities from the probit and logit models are very close in these HMDA regressions: