ECON4150 - Introductory Econometrics

Lecture 15: Binary dependent variables

Monique de Haan
(moniqued@econ.uio.no)

Stock and Watson Chapter 11
Lecture Outline

• The linear probability model

• Nonlinear probability models
  • Probit
  • Logit

• Brief introduction of maximum likelihood estimation

• Interpretation of coefficients in logit and probit models
Introduction

- So far the dependent variable ($Y$) has been continuous:
  - average hourly earnings
  - traffic fatality rate

- What if $Y$ is binary?
  - $Y$ = get into college, or not; $X$ = parental income.
  - $Y$ = person smokes, or not; $X$ = cigarette tax rate, income.
  - $Y$ = mortgage application is accepted, or not; $X$ = race, income, house characteristics, marital status ...
The linear probability model

- Multiple regression model with continuous dependent variable

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \]

- The coefficient \( \beta_j \) can be interpreted as the change in \( Y \) associated with a unit change in \( X_j \)

- We will now discuss the case with a binary dependent variable

- We know that the expected value of a binary variable \( Y \) is

\[ E[Y] = 1 \cdot Pr(Y = 1) + 0 \cdot Pr(Y = 0) = Pr(Y = 1) \]

- In the multiple regression model with a binary dependent variable we have

\[ E[Y_i|X_{1i}, \cdots, X_{ki}] = Pr(Y_i = 1|X_{1i}, \cdots, X_{ki}) \]

- It is therefore called the **linear probability model**.
Mortgage applications

Example:

- Most individuals who want to buy a house apply for a mortgage at a bank.
- Not all mortgage applications are approved.
- What determines whether or not a mortgage application is approved or denied?
- During this lecture we use a subset of the Boston HMDA data ($N = 2380$)
  - a data set on mortgage applications collected by the Federal Reserve Bank in Boston

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>deny</td>
<td>1 if mortgage application is denied</td>
<td>0.120</td>
<td>0.325</td>
</tr>
<tr>
<td>pi_ratio</td>
<td>anticipated monthly loan payments / monthly income</td>
<td>0.331</td>
<td>0.107</td>
</tr>
<tr>
<td>black</td>
<td>1 if applicant is black, = 0 if applicant is white</td>
<td>0.142</td>
<td>0.350</td>
</tr>
</tbody>
</table>
Mortgage applications

- Does the payment to income ratio affect whether or not a mortgage application is denied?

\[
\text{. regress deny pi\_ratio, robust}
\]

| deny   | Robust Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------|--------------|-----------|------|-----|----------------------|
| pi\_ratio | .6035349     | .0984826  | 6.13 | 0.000| .4104144  .7966555 |
| _cons  | -.0799096    | .0319666  | -2.50| 0.012| -.1425949 -.0172243|

- The estimated OLS coefficient on the payment to income ratio equals \( \hat{\beta}_1 = 0.60 \).

- The estimated coefficient is significantly different from 0 at a 1% significance level.

- How should we interpret \( \hat{\beta}_1 \)?
The linear probability model

- The conditional expectation equals the probability that \( Y_i = 1 \) conditional on \( X_{1i}, \ldots, X_{ki} \):

\[
E [Y_i|X_{1i}, \ldots, X_{ki}] = Pr(Y_i = 1|X_{1i}, \ldots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}
\]

- The population coefficient \( \beta_j \) equals the change in the probability that \( Y_i = 1 \) associated with a unit change in \( X_j \).

\[
\frac{\partial Pr(Y_i = 1|X_{1i}, \ldots, X_{ki})}{\partial X_j} = \beta_j
\]

In the mortgage application example:

- \( \hat{\beta}_1 = 0.60 \)

- A change in the payment to income ratio by 1 is estimated to increase the probability that the mortgage application is denied by 0.60.

- A change in the payment to income ratio by 0.10 is estimated to increase the probability that the application is denied by 6% (0.10*0.60*100).
The linear probability model

Assumptions are the same as for general multiple regression model:

1. \( E(u_i|X_{1i}, X_{2i}, \ldots, X_{ki}) = 0 \)
2. \((X_{1i}, \ldots, X_{ki}, Y_i)\) are i.i.d.
3. Big outliers are unlikely
4. No perfect multicollinearity.

Advantages of the linear probability model:

- Easy to estimate
- Coefficient estimates are easy to interpret

Disadvantages of the linear probability model

- Predicted probability can be above 1 or below 0!
- Error terms are heteroskedastic
The linear probability model: heteroskedasticity

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \]

- The variance of a Bernoulli random variable (CH 2 S&W):
  \[ \text{Var}(Y) = \text{Pr}(Y = 1) \times (1 - \text{Pr}(Y = 1)) \]

- We can use this to find the conditional variance of the error term

\[
\begin{align*}
\text{Var}(u_i | X_{1i}, \cdots, X_{ki}) &= \text{Var}(Y_i - (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) | X_{1i}, \cdots, X_{ki}) \\
&= \text{Var}(Y_i | X_{1i}, \cdots, X_{ki}) \\
&= \text{Pr}(Y_i = 1 | X_{1i}, \cdots, X_{ki}) \times (1 - \text{Pr}(Y_i = 1 | X_{1i}, \cdots, X_{ki})) \\
&= (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \times (1 - \beta_0 - \beta_1 X_{1i} - \cdots - \beta_k X_{ki}) \\
&\neq \sigma_u^2
\end{align*}
\]

- Solution: Always use heteroskedasticity robust standard errors when estimating a linear probability model!
The linear probability model: shortcomings

In the linear probability model the predicted probability can be below 0 or above 1!

**Example**: linear probability model, HMDA data
Mortgage denial v. ratio of debt payments to income (P/I ratio) in a subset of the HMDA data set ($n = 127$)
Nonlinear probability models

• Probabilities cannot be less than 0 or greater than 1

• To address this problem we will consider nonlinear probability models

\[ Pr(Y_i = 1) = G(Z) \]

\[ \text{with} \quad Z = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} \]

\[ \text{and} \quad 0 \leq G(Z) \leq 1 \]

• We will consider 2 nonlinear functions

1. Probit

\[ G(Z) = \Phi(Z) \]

2. Logit

\[ G(Z) = \frac{1}{1 + e^{-Z}} \]
Probit regression models the probability that $Y = 1$

- Using the cumulative standard normal distribution function $\Phi(Z)$
- evaluated at $Z = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$
- since $\Phi(z) = Pr(Z \leq z)$ we have that the predicted probabilities of the probit model are between 0 and 1

Example

- Suppose we have only 1 regressor and $Z = -2 + 3X_1$
- We want to know the probability that $Y = 1$ when $X_1 = 0.4$
- $z = -2 + 3 \cdot 0.4 = -0.8$
- $Pr(Y = 1) = Pr(Z \leq -0.8) = \Phi(-0.8)$
Pr(bit)
Logit regression models the probability that $Y = 1$

- Using the cumulative standard logistic distribution function

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

- evaluated at $Z = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$

- since $F(z) = Pr(Z \leq z)$ we have that the predicted probabilities of the probit model are between 0 and 1

Example

- Suppose we have only 1 regressor and $Z = -2 + 3X_1$
- We want to know the probability that $Y = 1$ when $X_1 = 0.4$
- $z = -2 + 3 \cdot 0.4 = -0.8$
- $Pr(Y = 1) = Pr(Z \leq -0.8) = F(-0.8)$
\( \Pr(Y = 1) = \Pr(Z \leq -0.8) = \frac{1}{1 + e^{0.8}} = 0.31 \)
Logit & probit

Standard Logistic CDF and Standard Normal CDF

- logistic
- normal
How to estimate logit and probit models

- In lecture 11 we discussed regression models that are nonlinear in the independent variables
  - these models can be estimated by OLS

- Logit and Probit models are nonlinear in the coefficients $\beta_0, \beta_1, \ldots, \beta_k$
  - these models can’t be estimated by OLS

- The method used to estimate logit and probit models is Maximum Likelihood Estimation (MLE).

- The MLE are the values of $(\beta_0, \beta_1, \ldots, \beta_k)$ that best describe the full distribution of the data.
The **likelihood function** is the joint probability distribution of the data, treated as a function of the unknown coefficients.

The **maximum likelihood estimator (MLE)** are the values of the coefficients that maximize the likelihood function.

MLE’s are the parameter values “most likely” to have produced the data.

Let’s start with a special case: The MLE with no $X$

- We have $n$ i.i.d. observations $Y_1, \ldots, Y_n$ on a binary dependent variable
- $Y$ is a Bernoulli random variable
- There is only 1 unknown parameter to estimate:
  - The probability $p$ that $Y = 1$,
  - which is also the mean of $Y$
Step 1: write down the likelihood function, the joint probability distribution of the data

- $Y_i$ is a Bernoulli random variable we therefore have

  $$Pr(Y_i = y) = Pr(Y_i = 1)^y \cdot (1 - Pr(Y_i = 1))^{1-y} = p^y (1 - p)^{1-y}$$

  - $Pr(Y_i = 1) = p^1 (1 - p)^0 = p$
  - $Pr(Y_i = 0) = p^0 (1 - p)^1 = 1 - p$

- $Y_1, \ldots, Y_n$ are i.i.d, the joint probability distribution is therefore the product of the individual distributions

  $$Pr(Y_1 = y_1, \ldots, Y_n = y_n) = Pr(Y_1 = y_1) \times \ldots \times Pr(Y_n = y_n)$$

  $$= [p^{y_1} (1 - p)^{1-y_1}] \times \ldots \times [p^{y_n} (1 - p)^{1-y_n}]$$

  $$= p^{y_1+y_2+\ldots+y_n} (1 - p)^{n-(y_1+y_2+\ldots+y_n)}$$
We have the likelihood function:

$$f_{\text{Bernoulli}}(p; Y_1 = y_1, \ldots, Y_n = y_n) = p^{\sum y_i} (1 - p)^{n - \sum y_i}$$

**Step 2:** Maximize the likelihood function w.r.t $p$

- Easier to maximize the logarithm of the likelihood function

$$\ln(f_{\text{Bernoulli}}(p; Y_1 = y_1, \ldots, Y_n = y_n)) = \left( \sum_{i=1}^{n} y_i \right) \cdot \ln(p) + \left( n - \sum_{i=1}^{n} y_i \right) \ln(1 - p)$$

- Since the logarithm is a strictly increasing function, maximizing the likelihood or the log likelihood will give the same estimator.
Maximum likelihood estimation

- Taking the derivative w.r.t $p$ gives

$$\frac{d}{dp} \ln(f_{\text{Bernoulli}}(p; Y_1 = y_1, \ldots, Y_n = y_n)) = \frac{\sum_{i=1}^{n} y_i}{p} - \frac{n - \sum_{i=1}^{n} y_i}{1 - p}$$

- Setting to zero and rearranging gives

$$(1 - p) \times \sum_{i=1}^{n} y_i = p \times (n - \sum_{i=1}^{n} y_i)$$

$$\sum_{i=1}^{n} y_i - p \sum_{i=1}^{n} y_i = n \cdot p - p \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} y_i = n \cdot p$$

- Solving for $p$ gives the MLE

$$\hat{p}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{Y}$$
MLE of the probit model

Step 1: write down the likelihood function

\[ Pr(Y_1 = y_1, \ldots, Y_n = y_n) = Pr(Y_1 = y_1) \times \ldots \times Pr(Y_n = y_n) \]
\[ = [p_1^{y_1}(1 - p_1)^{1-y_1}] \times \ldots \times [p_n^{y_n}(1 - p_n)^{1-y_n}] \]

- so far it is very similar as the case without explanatory variables except that \( p_i \) depends on \( X_{1i}, \ldots, X_{ki} \)

\[ p_i = \Phi (X_{1i}, \ldots, X_{ki}) = \Phi (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \]

- substituting for \( p_i \) gives the likelihood function:

\[ \left[ \Phi (\beta_0 + \beta_1 X_{11} + \cdots + \beta_k X_{k1})^{y_1} (1 - \Phi (\beta_0 + \beta_1 X_{11} + \cdots + \beta_k X_{k1}))^{1-y_1} \right] \times \ldots \]
\[ \times \left[ \Phi (\beta_0 + \beta_1 X_{1n} + \cdots + \beta_k X_{kn})^{y_n} (1 - \Phi (\beta_0 + \beta_1 X_{1n} + \cdots + \beta_k X_{kn}))^{1-y_n} \right] \]
MLE of the probit model

Also with obtaining the MLE of the probit model it is easier to take the logarithm of the likelihood function

**Step 2:** Maximize the log likelihood function

\[
\ln \left[ f_{\text{probit}} (\beta_0, \ldots, \beta_k; \ Y_1, \ldots, Y_n | X_{1i}, \ldots, X_{ki}, i = 1, \ldots, n) \right]
\]

\[
= \sum_{i=1}^{n} Y_i \ln [\Phi (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})]
\]

\[
+ \sum_{i=1}^{n} (1 - Y_i) \ln [1 - \Phi (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})]
\]

w.r.t \( \beta_0, \ldots, \beta_1 \)

- There is no simple formula for the probit MLE, the maximization must be done using numerical algorithm on a computer.
MLE of the logit model

**Step 1:** write down the likelihood function

\[ Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i} \]

- very similar to the Probit model but with a different function for \( p_i \)

\[ p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})}} \]

**Step 2:** Maximize the log likelihood function w.r.t \( \beta_0, \ldots, \beta_1 \)

\[ \ln \left[ f_{\text{logit}} (\beta_0, \ldots, \beta_k; Y_1, \ldots, Y_n| X_{1i}, \ldots, X_{ki}, i = 1, \ldots, n) \right] \]

\[ = \sum_{i=1}^{n} Y_i \ln \left( 1 / \left[ 1 + e^{-(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})} \right] \right) \]

\[ + \sum_{i=1}^{n} (1 - Y_i) \ln \left( 1 - \left( 1 / \left[ 1 + e^{-(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})} \right] \right) \right) \]

- There is no simple formula for the logit MLE, the maximization must be done using numerical algorithm on a computer.
The estimated MLE coefficient on the payment to income ratio equals $\hat{\beta}_1 = 2.97$.

The estimated coefficient is positive and significantly different from 0 at a 1% significance level.

How should we interpret $\hat{\beta}_1$?
The estimate of $\beta_1$ in the probit model CANNOT be interpreted as the change in the probability that $Y_i = 1$ associated with a unit change in $X_1$!!

- In general the effect on $Y$ of a change in $X$ is the expected change in $Y$ resulting from the change in $X$
- Since $Y$ is binary the expected change in $Y$ is the change in the probability that $Y = 1$

In the probit model the predicted change the probability that the mortgage application is denied when the payment to income ratio increases from

0.10 to 0.20:

$$\Delta \Pr(Y_i = 1) = \Phi(-2.19 + 2.97 \cdot 0.20) - \Phi(-2.19 + 2.97 \cdot 0.10) = 0.0495$$

0.30 to 0.40:

$$\Delta \Pr(Y_i = 1) = \Phi(-2.19 + 2.97 \cdot 0.40) - \Phi(-2.19 + 2.97 \cdot 0.30) = 0.0619$$
The probit model satisfies these conditions:

I. \( \Pr(Y = 1|X) \) to be increasing in \( X \) for \( \beta_1 > 0 \), and

II. \( 0 \leq \Pr(Y = 1|X) \leq 1 \) for all \( X \)

- All predicted probabilities are between 0 and 1!
Logit: mortgage applications

. logit deny pi_ratio

Iteration 0:  log likelihood =  -872.0853
Iteration 1:  log likelihood =  -830.96071
Iteration 2:  log likelihood =  -830.09497
Iteration 3:  log likelihood =  -830.09403
Iteration 4:  log likelihood =  -830.09403

Logistic regression                               Number of obs   =       2380
LR chi2(   1)      =     83.98
Prob > chi2     =     0.0000
Log likelihood =  -830.09403

                  Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------- -------- -------- -------- -------- ------------------------
deny             
   pi_ratio      5.884498   .7336006    8.02    0.000      4.446667    7.322328
   _cons       -4.028432   .2685763   -15.00    0.000     -4.554832   -3.502032

• The estimated MLE coefficient on the payment to income ratio equals \( \hat{\beta}_1 = 5.88 \).

• The estimated coefficient is positive and significantly different from 0 at a 1% significance level.

• How should we interpret \( \hat{\beta}_1 \)?
Also in the Logit model:

The estimate of $\beta_1$ CANNOT be interpreted as the change in the probability that $Y_i = 1$ associated with a unit change in $X_1$!!

In the logit model the predicted change in the probability that the mortgage application is denied when the payment to income ratio increases from

**0.10 to 0.20:**

$$\triangle Pr(Y_i = 1) = \left( \frac{1}{1 + e^{-(-4.03+5.88\cdot0.20)}} \right) - \left( \frac{1}{1 + e^{-(-4.03+5.88\cdot0.10)}} \right) = 0.023$$

**0.30 to 0.40:**

$$\triangle Pr(Y_i = 1) = \left( \frac{1}{1 + e^{-(-4.03+5.88\cdot0.40)}} \right) - \left( \frac{1}{1 + e^{-(-4.03+5.88\cdot0.30)}} \right) = 0.063$$
The predicted probabilities from the probit and logit models are very close in these HMDA regressions:
We can easily extend the Logit and Probit regression models, by including additional regressors.

Suppose we want to know whether white and black applications are treated differentially.

Is there a significant difference in the probability of denial between black and white applicants conditional on the payment to income ratio?

To answer this question we need to include two regressors:

- P/I ratio
- Black
Probit regression

Number of obs = 2380  
LR chi2(2) = 149.90  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.0859

Log likelihood = -797.13604

|     | Coef.   | Std. Err. | z      | P>|z|  | [95% Conf. Interval] |
|-----|---------|-----------|--------|------|---------------------|
| deny|         |           |        |      |                     |
| black| 0.7081579| 0.0834327 | 8.49   | 0.000| 0.5446328           |
| pi_ratio| 2.741637 | 0.3595888 | 7.62   | 0.000| 2.036856           |
| _cons| -2.258738| 0.129882  | -17.39 | 0.000| -2.513302         |

To say something about the size of the impact of race we need to specify a value for the payment to income ratio

Predicted denial probability for a white application with a P/I-ratio of 0.3 is

\[ \Phi(-2.26 + 0.71 \cdot 0 + 2.74 \cdot 0.3) = 0.0749 \]

Predicted denial probability for a black application with a P/I-ratio of 0.3 is

\[ \Phi(-2.26 + 0.71 \cdot 1 + 2.74 \cdot 0.3) = 0.2327 \]

Difference is 15.8%
Logit with multiple regressors

Logistic regression

Number of obs = 2380
LR chi2(2) = 152.78
Prob > chi2 = 0.0000
Pseudo R2 = 0.0876

Log likelihood = -795.6952

|        | Coef.   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|---------|-----------|-------|------|----------------------|
| deny   |         |           |       |      |                      |
| black  | 1.272782| 0.1461983 | 8.71  | 0.000| 0.9862385 1.559325   |
| pi_ratio| 5.370362| 0.7283192 | 7.37  | 0.000| 3.942883 6.797841    |
| _cons  | -4.125558| 0.2684161 | -15.37| 0.000| -4.651644 -3.599472  |

To say something about the size of the impact of race we need to specify a value for the payment to income ratio

Predicted denial probability for a white application with a P/I-ratio of 0.3 is

\[ \frac{1}{1 + e^{-(-4.13+5.37\cdot 0.30)}} = 0.075 \]

Predicted denial probability for a black application with a P/I-ratio of 0.3 is

\[ \frac{1}{1 + e^{-(-4.13+5.37\cdot 0.30+1.27)}} = 0.224 \]

Difference is 14.8%
Table 1: Mortgage denial regression using the Boston HMDA Data

<table>
<thead>
<tr>
<th>Dependent variable: $deny = 1$ if mortgage application is denied, $= 0$ if accepted</th>
<th>LPM</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>black</td>
<td>0.177***</td>
<td>0.71***</td>
<td>1.27***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.083)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>$P/I$ ratio</td>
<td>0.559***</td>
<td>2.74***</td>
<td>5.37***</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.44)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>constant</td>
<td>-0.091***</td>
<td>-2.26***</td>
<td>-4.13***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.16)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>difference $Pr(\text{deny}=1)$ between black and white applicant when $P/I$ ratio=0.3</td>
<td>17.7%</td>
<td>15.8%</td>
<td>14.8%</td>
</tr>
</tbody>
</table>
Both for the Linear Probability as for the Probit & Logit models we have to consider threats to

1 Internal validity

- Is there omitted variable bias?
- Is the functional form correct?
  - Probit model: is assumption of a Normal distribution correct?
  - Logit model: is assumption of a Logistic distribution correct?
- Is there measurement error?
- Is there sample selection bias?
- is there a problem of simultaneous causality?

2 External validity

- These data are from Boston in 1990-91.
- Do you think the results also apply today, where you live?
Distance to college & probability of obtaining a college degree

Linear regression

| college | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|---------|-------|-----------|------|------|----------------------|
| dist    | -0.012471 | 0.0031403  | -3.97 | 0.000 | -0.0186278 to -0.0063142 |
| _cons   | 0.2910057   | 0.0093045   | 31.28 | 0.000 | 0.2727633 to 0.3092481 |

Probit regression

| college | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|------|------|----------------------|
| dist    | -0.0407873 | 0.0109263  | -3.73 | 0.000 | -0.0622025 to -0.0193721 |
| _cons   | -0.5464198  | 0.028192    | -19.38 | 0.000 | -0.6016752 to -0.4911645 |

Logistic regression

| college | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|------|------|----------------------|
| dist    | -0.0709896 | 0.0193593  | -3.67 | 0.000 | -0.1089332 to -0.033046 |
| _cons   | -0.8801555  | 0.0476434   | -18.47 | 0.000 | -0.9735349 to -0.786776 |
The 3 different models produce very similar results.
• If $Y_i$ is binary, then $E(Y_i|X_i) = Pr(Y_i = 1|X_i)$

• Three models:

  1. linear probability model (linear multiple regression)
  2. probit (cumulative standard normal distribution)
  3. logit (cumulative standard logistic distribution)

• LPM, probit, logit all produce predicted probabilities

• Effect of $\Delta X$ is a change in conditional probability that $Y = 1$

• For logit and probit, this depends on the initial $X$

• Probit and logit are estimated via maximum likelihood
  • Coefficients are normally distributed for large $n$
  • Large-$n$ hypothesis testing, conf. intervals is as usual