Regression with a Binary Dependent Variable
(SW Ch. 9)

So far the dependent variable ($Y$) has been continuous:

- district-wide average test score
- traffic fatality rate

But we might want to understand the effect of $X$ on a binary variable:

- $Y = \text{get into college, or not}$
- $Y = \text{person smokes, or not}$
- $Y = \text{mortgage application is accepted, or not}$
Example: Mortgage denial and race, The Boston Fed HMDA data set

• 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
The Linear Probability Model  
(SW Section 9.1)  
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 \( \beta_1 \) 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?
The linear probability model, ctd.

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

Recall assumption #1: \( E(u_i|X_i) = 0 \), so

\[ E(Y_i|X_i) = E(\beta_0 + \beta_1 X_i + u_i|X_i) \]
\[ = \beta_0 + \beta_1 X_i \]

When \( Y \) is binary,

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

so

\[ E(Y|X) = \Pr(Y=1|X) \]
The linear probability model, ctd.
When $Y$ is binary, the linear regression model

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

is called the **linear probability model**.

- The predicted value is a *probability*:

  $$E(Y|X=x) = \Pr(Y=1|X=x)$$
  
  $$= \text{prob. that } Y = 1 \text{ given } x$$

  $\hat{Y}$ = the *predicted probability* that $Y_i = 1$, given $X$

- $\beta_1$ = change in probability that $Y = 1$ for a given $\Delta x$:

  $$\beta_1 = \frac{\Pr(Y = 1 | X = x + \Delta x) - \Pr(Y = 1 | X = x)}{\Delta x}$$
Example: linear probability model, HMDA data

Mortgage denial v. ratio of debt payments to income (P/I ratio) in the HMDA data set

**FIGURE 9.1** Scatterplot of Mortgage Application Denial and the Payment-to-Income Ratio

Mortgage applicants with a high ratio of debt payments to income (P/I ratio) are more likely to have their application denied (deny = 1 if denied; deny = 0 if approved). The linear probability model uses a straight line to model the probability of denial, conditional on the P/I ratio.
Linear probability model: HMDA data

\[ den = -0.080 + 0.604P/I \text{ ratio} \]
\[ \quad \text{(0.032) (0.098)} \]
\[ (n = 2380) \]

- What is the predicted value for \( P/I \) ratio = 0.3?
  \[ \Pr(deny=1 \mid P/I \text{ ratio} = 0.3) \]
  \[ = -0.080 + 0.604 \times 0.3 = 0.151 \]

- Calculating “effects:” increase \( P/I \) ratio from 0.3 to 0.4:
  \[ \Pr(deny=1 \mid P/I \text{ ratio} = 0.4) \]
  \[ = -0.080 + 0.604 \times 0.4 = 0.212 \]
The effect on the probability of denial of an increase in $P/I$ ratio from .3 to .4 is to increase the probability by .061, that is, by 6.1 percentage points.
Next include race as a regressor

deny =
\[-0.091 + 0.559P/I \text{ ratio} + 0.177\text{black}\]

$\begin{pmatrix}
0.032 \\
0.098 \\
0.025
\end{pmatrix}$

Predicted probability of denial:
• for black applicant with $P/I \text{ ratio} = .3$:

$$Pr \ (deny = 1) = -0.091 + 0.559 \times 0.3 + 0.177 \times 1$$

$$= 0.254$$

• for white applicant, $P/I \text{ ratio} = 0.3$:

$$Pr \ (deny = 1) = -0.091 + 0.559 \times 0.3 + 0.177 \times 0$$

$$= 0.077$$
• difference = .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: Summary

- Models probability as a linear function of $X$

- Advantages:
  - Simple to estimate and to interpret
  - Inference is the same as for multiple regression (need heteroskedasticity-robust standard errors)

- Disadvantages:
  - Does it make sense that the probability should be linear in $X$?
• Predicted probabilities can be <0 or >1!

• These disadvantages can be solved by using a nonlinear probability model: probit and logit regression
Probit and Logit Regression
(SW Section 9.2)

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

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

Instead, we want:

- $0 \leq \Pr(Y = 1|X) \leq 1$ for all $X$
- $\Pr(Y = 1|X)$ to be increasing in $X$ (for $\beta_1 > 0$)

This requires a *nonlinear* functional form for the probability. How about an “S-curve”…
The probit model uses the cumulative normal distribution function to model the probability of denial given the payment-to-income ratio or, more generally, to model $\Pr(Y = 1 \mid X)$. Unlike the linear probability model, the probit conditional probabilities are always between zero and one.
The probit model satisfies these conditions:

- $0 \leq \Pr(Y = 1|X) \leq 1$ for all $X$
- $\Pr(Y = 1|X)$ to be increasing in $X$
  (for $\beta_1 > 0$)
Probit regression models the probability that $Y=1$ using the cumulative standard normal distribution function, evaluated at $z = \beta_0 + \beta_1 X$:

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

- $\Phi$ is the cumulative normal distribution function.
- $z = \beta_0 + \beta_1 X$ is the “z-value” or “z-index” of the probit model.
Example:
Suppose $\beta_0 = -2$, $\beta_1 = 3$, $X = .4$,
so

$$\Pr(Y = 1 \mid X = .4) =$$

$$\Phi(-2 + 3x.4) = \Phi(-0.8)$$

$$\Pr(Y = 1 \mid X = .4) = \text{area under the standard normal density to left of } z = -.8, \text{ which is...}$$
\[ \text{Pr}(Z \leq -0.8) = .2119 \]
Probit regression, ctd.

Why use the cumulative normal probability distribution?

- The “S-shape” gives us what we want:
  - $0 \leq \Pr(Y = 1|X) \leq 1$ for all $X$
  - $\Pr(Y = 1|X)$ to be increasing in $X$ (for $\beta_1 > 0$)
- Easy to use – the probabilities are tabulated in the cumulative normal tables
- Relatively straightforward interpretation:
\( \hat{\beta}_0 + \hat{\beta}_1 X \) is the predicted \( z \)-value, given \( X \)

\( \hat{\beta}_1 \) is the change in the \( z \)-value for a unit change in \( X \)
Example: HMDA data, ctd.

\[
\Pr(\text{deny}=1 \mid \text{P/I ratio}) = \Phi(-2.19 + 2.97 \times P/I \text{ ratio})
\]

(16) (47)

• Positive coefficient: *does this make sense?*

• Standard errors have usual interpretation

• Predicted probabilities:

\[
\Pr(\text{deny}=1 \mid \text{P/I ratio} = 0.3) = \Phi(-2.19+2.97 \times 0.3)
\]

\[
= \Phi(-1.30) = .097
\]
• Effect of change in $P/I$ ratio from .3 to .4:

$$\Pr(\text{deny}=1 \mid P/I \text{ ratio} = 0.4) = \Phi(-2.19+2.97 \times 0.4) = .159$$

The predicted probability of denial rises from .097 to .159
Probit regression with multiple regressors

\[ \Pr(Y = 1|X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2) \]

- \( \Phi \) is the cumulative normal distribution function.
- \( z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \) is the “\( z \)-value” or “\( z \)-index” of the probit model.
- \( \beta_1 \) is the effect on the \( z \)-score of a unit change in \( X_1 \), holding constant \( X_2 \)

We’ll go through the estimation details later...
Example: HMDA data, ctd.

\[
\begin{align*}
\text{Pr}(\text{deny}=1 \mid \text{P/I, black}) & = \Phi(-2.26 + 2.74x \text{P/I ratio} + .71x \text{black}) \\
& = \Phi(-2.26 + 2.74 \times .3 + .71 \times 1) = .233 \\
\text{Pr}(\text{deny}=1 \mid .3,1) & = \Phi(-2.26 + 2.74 \times .3 + .71 \times 1) = .233 \\
\text{Pr}(\text{deny}=1 \mid .3,0) & = \Phi(-2.26 + 2.74 \times .3 + .71 \times 0) = .075
\end{align*}
\]
• Difference in rejection probabilities = .158 (15.8 percentage points)
• *Still plenty of room still for omitted variable bias*...
Logit regression

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

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

$F$ is the cumulative logistic distribution function:

$$F(\beta_0 + \beta_1 X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$
Logistic regression, ctd.

\[ \Pr(Y = 1 | X) = F(\beta_0 + \beta_1 X) \]

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

Example: \( \beta_0 = -3, \beta_1 = 2, X = .4, \)

so \( \beta_0 + \beta_1 X = -3 + 2 \times .4 = -2.2 \)

so

\[ \Pr(Y = 1 | X = .4) = 1/(1 + e^{-(-2.2)}) = .0998 \]

Why bother with logit if we have probit?

- Historically, numerically convenient
- In practice, very similar to probit
Predicted probabilities from estimated probit and logit models usually are very close.

**FIGURE 9.3** Probit and Logit Models of the Probability of Denial, Given the P/I Ratio

These logit and probit models produce nearly identical estimates of the probability that a mortgage application will be denied, given the payment-to-income ratio.