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

EC 471
Spring 2004

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 = get into college, or not
- Y = person smokes, or not
- Y = 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 the predicted value $\hat{Y}$ mean when Y is binary? For example, what does $\hat{Y} = 0.26$ mean?
- What does β_1 mean when Y is binary? Is $\beta_1 = \frac{\Delta Y}{\Delta X}$?

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***:
  - $E(Y|X=x) = \Pr(Y=1|X=x)$ = prob. that $Y = 1$ given x
  - $\hat{Y}$ = the ***predicted probability*** that $Y_i = 1$, given X
- β_1 = change in probability that $Y = 1$ for a given Δx:

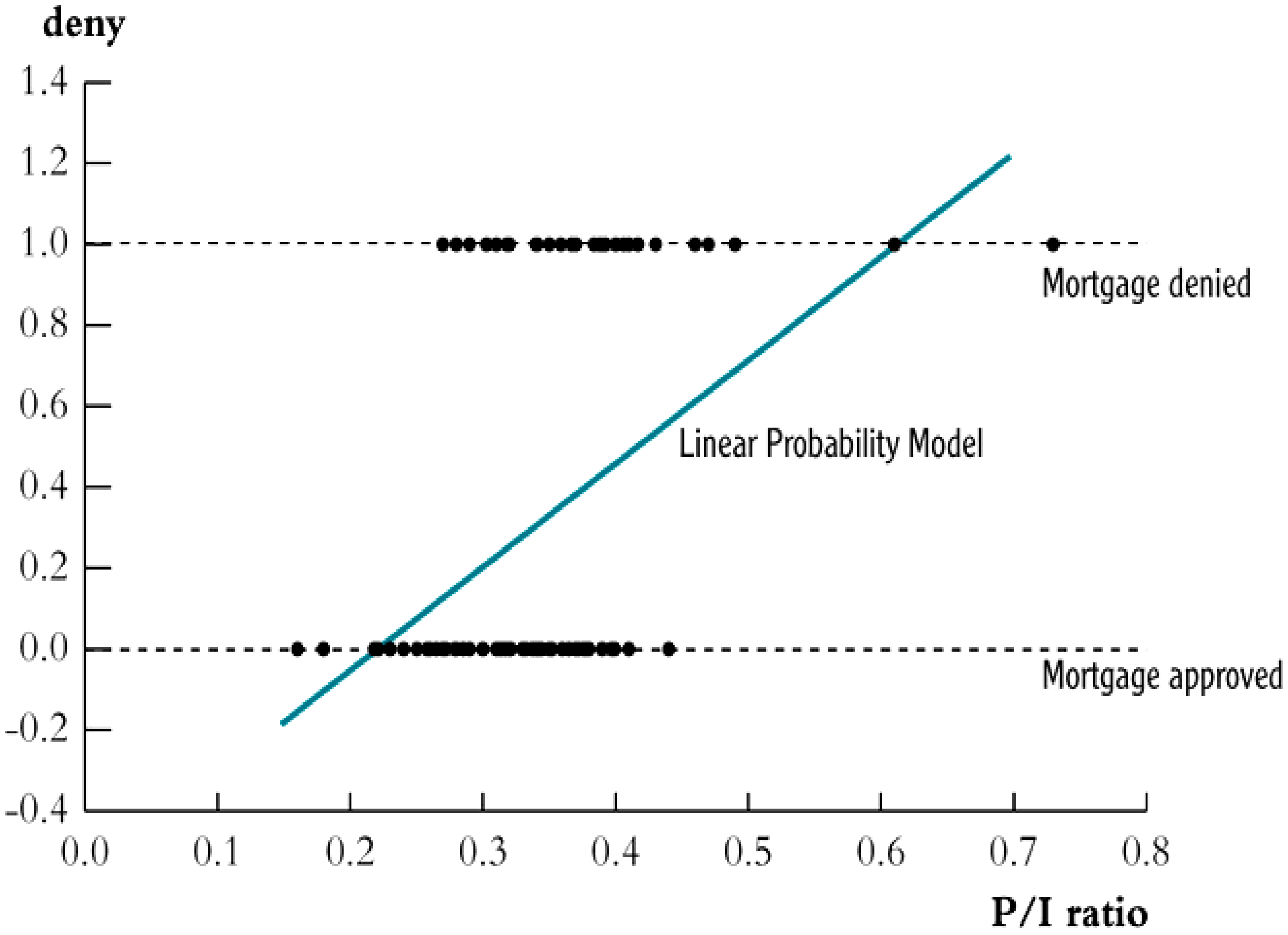
$$\beta_1 = \frac{\Pr(Y = 1 \mid X = x + \Delta x) - \Pr(Y = 1 \mid X = x)}{\Delta x}$$

Example: linear probability model, HMDA data

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

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

$$\widehat{deny} = -.080 + \underset{(.032)}{} .604 \; P/I\ ratio \qquad (n = 2380)$$
$$(.032)\ (.098)$$

- What is the predicted value for *P/I ratio* = .3?

$$\widehat{\Pr(deny = 1 \mid P/Iratio = .3)} = -.080 + .604 \times .3 = .151$$

- Calculating "effects:" increase *P/I ratio* from .3 to .4:

$$\widehat{\Pr(deny = 1 \mid P/Iratio = .4)} = -.080 + .604 \times .4 = .212$$

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* (*what?*).

Next include *black* as a regressor:

$$\widehat{deny} = -.091 + .559 P/I\ ratio + .177 black$$
$$\quad (.032) \quad (.098) \qquad\qquad (.025)$$

Predicted probability of denial:

- for black applicant with *P/I ratio* = .3:

$$\widehat{\Pr(deny = 1)} = -.091 + .559 \times .3 + .177 \times 1 = .254$$

- for white applicant, *P/I ratio* = .3:

$$\widehat{\Pr(deny = 1)} = -.091 + .559 \times .3 + .177 \times 0 = .077$$

- **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:

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

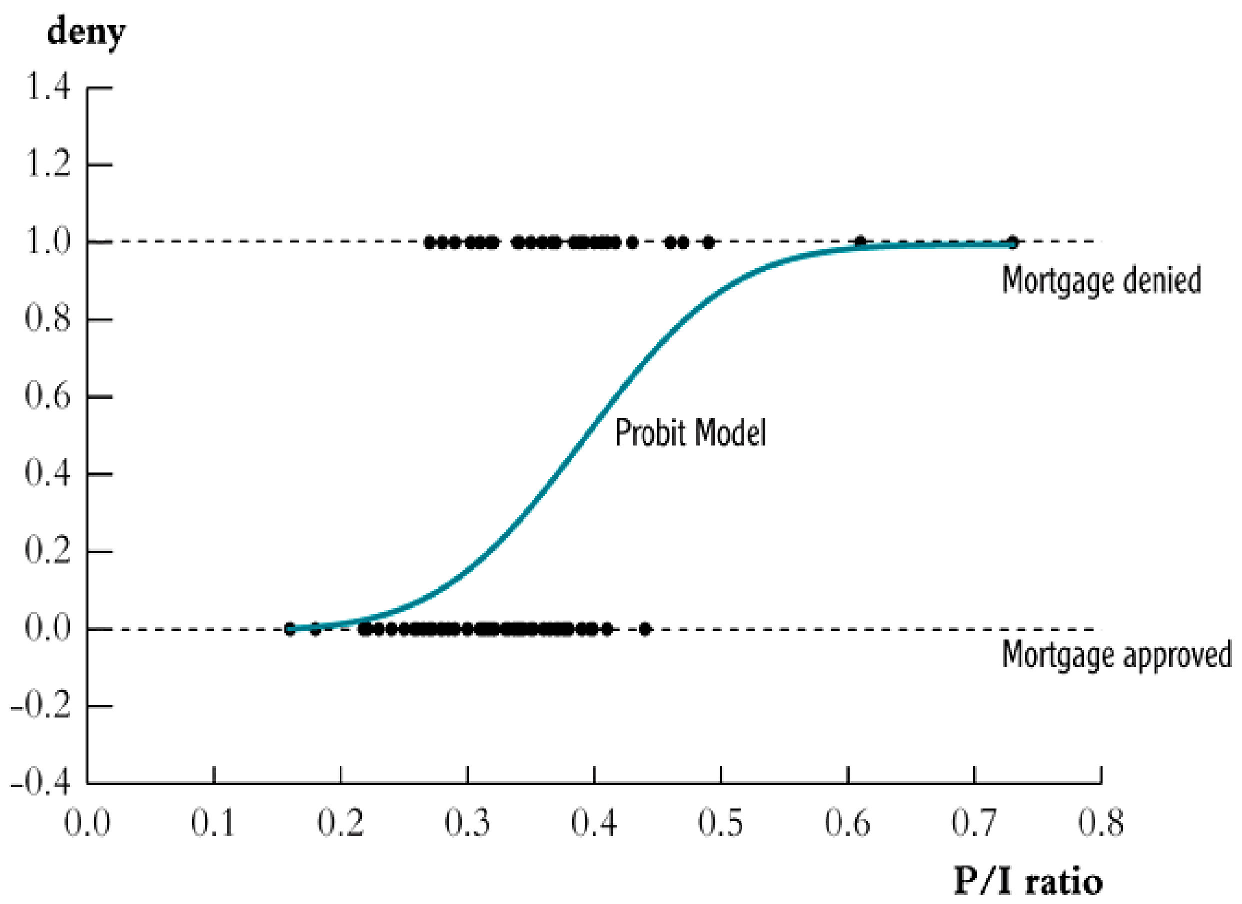
Instead, we want:

- $0 \leq \Pr(Y = 1|X) \leq 1$ 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”…

FIGURE 9.2 **Probit Model of the Probability of Denial, Given the P/I Ratio**

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 $z = \beta_0 + \beta_1 X$:

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

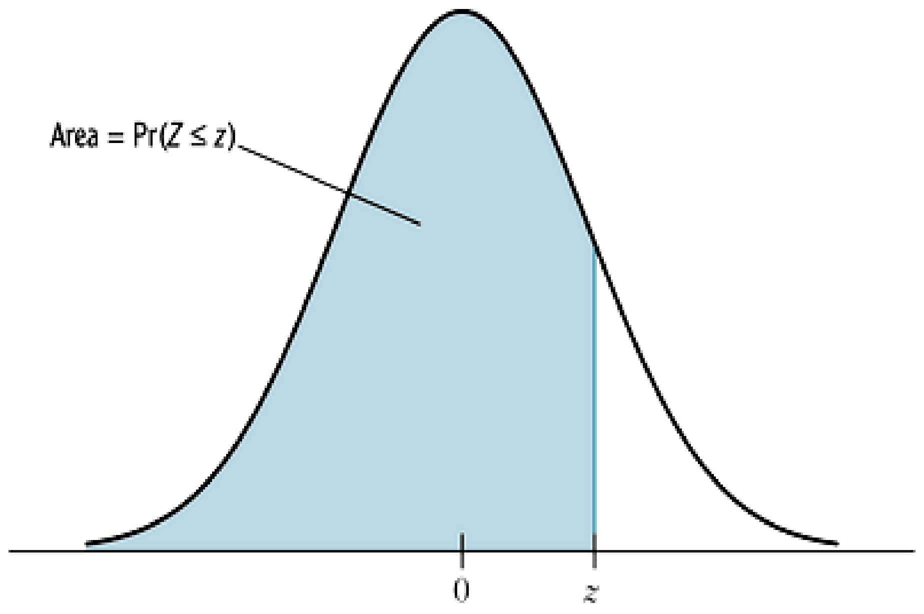
- Φ is the **cumulative normal distribution function**.
- $z = \beta_0 + \beta_1 X$ is the "z-value" or "z-index" of the probit model.

Example: Suppose $\beta_0 = -2$, $\beta_1 = 3$, $X = .4$, so

$$\Pr(Y = 1|X=.4) = \Phi(-2 + 3\times.4) = \Phi(-0.8)$$

$\Pr(Y = 1|X=.4)$ = area under the standard normal density to left of $z = -.8$, which is…

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



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

$$\Pr(Z \leq -0.8) = .2119$$

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:
  - z-value $= \beta_0 + \beta_1 X$
  - $\hat{\beta}_0 + \hat{\beta}_1 X$ is the predicted z-value, given X
  - β_1 is the change in the z-value for a unit change in X

STATA Example: HMDA data

```
. probit deny p_irat, r;

Iteration 0:   log likelihood =  -872.0853      We'll discuss this later
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
------------------------------------------------------------------------------
```

$$\widehat{\Pr(deny=1 \mid P/Iratio)} = \Phi(-2.19 + 2.97 \times P/I\ ratio)$$

$$(.16) \quad (.47)$$

STATA Example: HMDA data, ctd.

$$\widehat{\Pr(deny=1\,|\,P/I\,ratio)} = \Phi(-2.19 + 2.97 \times P/I\ ratio)$$

$$(.16) \quad (.47)$$

- Positive coefficient: *does this make sense*?
- Standard errors have usual interpretation
- Predicted probabilities:

$$\widehat{\Pr(deny=1\,|\,P/I\,ratio=.3)} = \Phi(-2.19+2.97\times .3)$$

$$= \Phi(-1.30) = .097$$

- Effect of change in *P/I ratio* from .3 to .4:

$$\widehat{\Pr(deny=1\,|\,P/I\,ratio=.4)} = \Phi(-2.19+2.97\times .4) = .159$$

Predicted probability of denial rises from .097 to .159

Probit regression with multiple regressors

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

- Φ 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.
- β_1 is the effect on the z-score of a unit change in X_1, holding constant X_2

STATA Example: HMDA data

```
. probit deny p_irat black, r;

Iteration 0:   log likelihood =  -872.0853
Iteration 1:   log likelihood = -800.88504
Iteration 2:   log likelihood =  -797.1478
Iteration 3:   log likelihood = -797.13604

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
------------------------------------------------------------------------------
```

We'll go through the estimation details later…

STATA Example: predicted probit probabilities

```
. probit deny p_irat black, r;

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
------------------------------------------------------------------------------

.  sca 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
```

NOTE

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

STATA Example: HMDA data, ctd.

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

$$= \Phi(\underset{(.16)}{-2.26} + \underset{(.44)}{2.74}\times P/I\ ratio + \underset{(.08)}{.71}\times black)$$

- Is the coefficient on *black* statistically significant?
- Estimated effect of race for *P/I ratio* = .3:

$$\widehat{\Pr(deny=1\,|\,.3,1)} = \Phi(-2.26+2.74\times.3+.71\times1) = .233$$

$$\widehat{\Pr(deny=1\,|\,.3,0)} = \Phi(-2.26+2.74\times.3+.71\times0) = .075$$

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

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

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

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

Why bother with logit if we have probit?

- Historically, numerically convenient
- In practice, very similar to probit

STATA Example: HMDA data

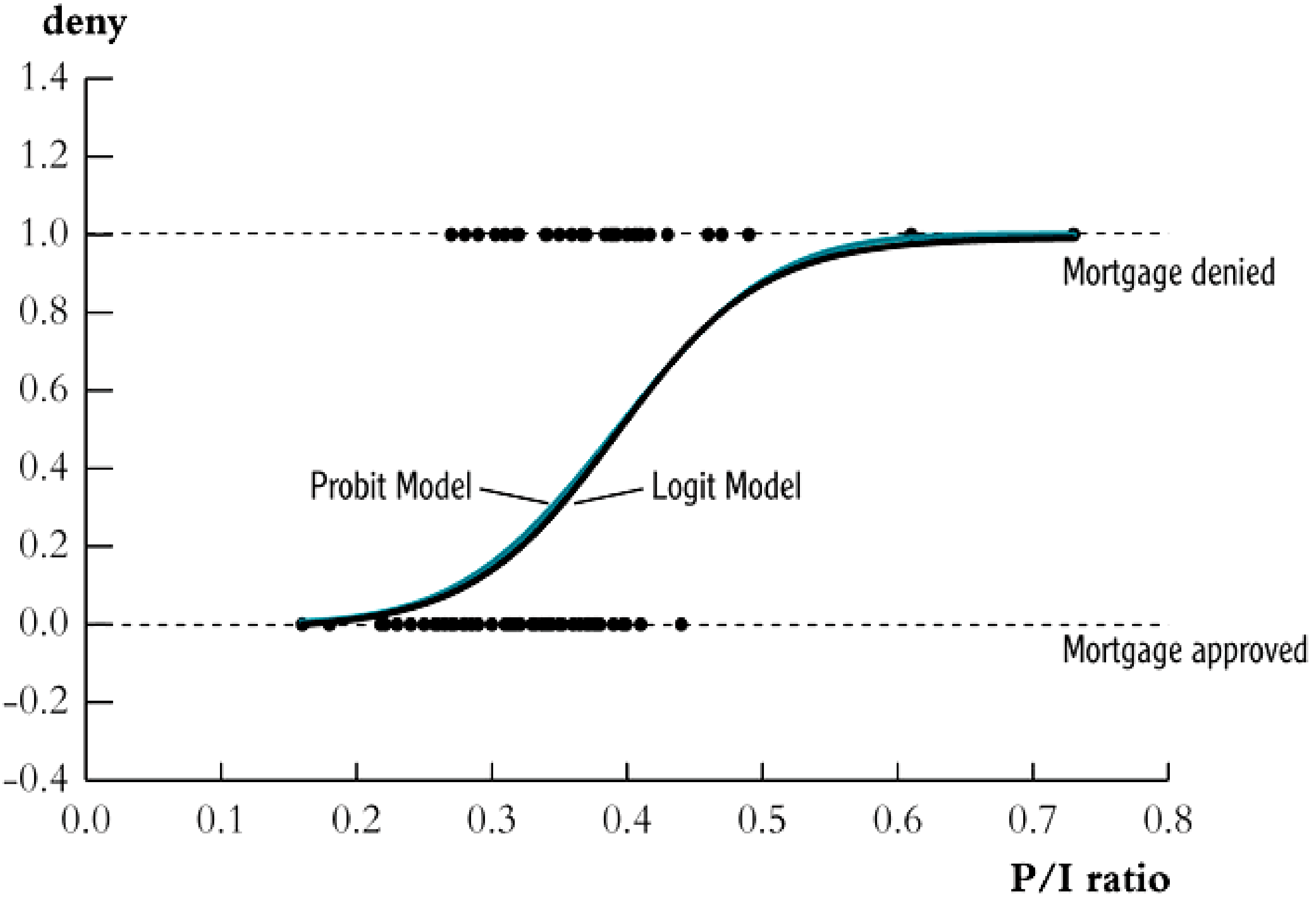
```
. logit deny p_irat black, r;

Iteration 0:   log likelihood =  -872.0853      Later…
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
------------------------------------------------------------------------------

.  dis "Pred prob, p_irat=.3, white: "
>     1/(1+exp(-(_b[_cons]+_b[p_irat]*.3+_b[black]*0)));

Pred prob, p_irat=.3, white: .07485143
```

NOTE: the probit predicted probability is .07546603

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.



Estimation and Inference in Probit (and Logit) Models (SW Section 9.3)

Probit model:

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

- Estimation and inference
  - How to estimate β_0 and β_1?
  - What is the sampling distribution of the estimators?
  - Why can we use the usual methods of inference?
- First discuss *nonlinear least squares* (easier to explain)
- Then discuss *maximum likelihood* estimation (what is actually done in practice)

Probit estimation by nonlinear least squares

Recall OLS:

$$\min_{b_0, b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$$

- The result is the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

In probit, we have a different regression function – the nonlinear probit model. So, we could estimate β_0 and β_1 by ***nonlinear least squares***:

$$\min_{b_0, b_1} \sum_{i=1}^{n} [Y_i - \Phi(b_0 + b_1 X_i)]^2$$

Solving this yields the ***nonlinear least squares*** estimator of the probit coefficients.

Nonlinear least squares, ctd.

$$\min_{b_0, b_1} \sum_{i=1}^{n} [Y_i - \Phi(b_0 + b_1 X_i)]^2$$

How to solve this minimization problem?

- Calculus doesn’t give and explicit solution.
- Must be solved *numerically* using the computer, e.g. by “trial and error” method of trying one set of values for (b_0,b_1), then trying another, and another,…
- Better idea: use specialized minimization algorithms

In practice, nonlinear least squares isn’t used because it isn’t efficient – an estimator with a smaller variance is…

Probit estimation by maximum likelihood

The ***likelihood function*** is the conditional density of $Y_1,\ldots,Y_n$ given $X_1,\ldots,X_n$, treated as a function of the unknown parameters β_0 and β_1.

- The maximum likelihood estimator (MLE) is the value of (β_0, β_1) that maximize the likelihood function.
- The MLE is the value of (β_0, β_1) that best describe the full distribution of the data.
- In large samples, the MLE is:
  - consistent
  - normally distributed
  - efficient (has the smallest variance of all estimators)

Special case: the probit MLE with no *X*

$$Y = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } 1-p \end{cases} \quad \text{(Bernoulli distribution)}$$

Data: $Y_1,\ldots,Y_n$, i.i.d.

Derivation of the likelihood starts with the density of Y_1:

$$\Pr(Y_1 = 1) = p \text{ and } \Pr(Y_1 = 0) = 1-p$$

so

$$\Pr(Y_1 = y_1) = p^{y_1}(1-p)^{1-y_1}$$ (*verify this for y_1=0, 1!*)

Joint density of (Y_1,Y_2):

Because Y_1 and Y_2 are independent,

$$\Pr(Y_1 = y_1, Y_2 = y_2) = \Pr(Y_1 = y_1) \times \Pr(Y_2 = y_2)$$

$$= [p^{y_1}(1-p)^{1-y_1}] \times [p^{y_2}(1-p)^{1-y_2}]$$

Joint density of $(Y_1,..,Y_n)$:

$$\Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

$$= [p^{y_1}(1-p)^{1-y_1}] \times [p^{y_2}(1-p)^{1-y_2}] \times \ldots \times [p^{y_n}(1-p)^{1-y_n}]$$

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

The likelihood is the joint density, treated as a function of the unknown parameters, which here is p:

$$f(p;Y_1,\ldots,Y_n) = p^{\sum_{i=1}^n Y_i}(1-p)^{\left(n-\sum_{i=1}^n Y_i\right)}$$

The MLE maximizes the likelihood. Its standard to work with **the log likelihood**, $\ln[f(p;Y_1,\ldots,Y_n)]$:

$$\mathbf{ln[\boldsymbol{f}(\boldsymbol{p};\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{\boldsymbol{n}})] =} \left(\sum_{i=1}^n Y_i\right)\ln(p) + \left(n - \sum_{i=1}^n Y_i\right)\ln(1-p)$$

The probit likelihood with one X

The derivation starts with the density of Y_1, given X_1:

$$\Pr(Y_1 = 1|X_1) = \Phi(\beta_0 + \beta_1 X_1)$$

$$\Pr(Y_1 = 0|X_1) = 1 - \Phi(\beta_0 + \beta_1 X_1)$$

so

$$\Pr(Y_1 = y_1|X_1) = \Phi(\beta_0 + \beta_1 X_1)^{y_1}[1 - \Phi(\beta_0 + \beta_1 X_1)]^{1-y_1}$$

The probit likelihood function is the joint density of $Y_1,\ldots,Y_n$ given $X_1,\ldots,X_n$, treated as a function of β_0, β_1:

$$\mathrm{f}(\beta_0,\beta_1;\, Y_1,\ldots,Y_n|X_1,\ldots,X_n)$$

$$= \{\Phi(\beta_0 + \beta_1 X_1)^{Y_1}[1 - \Phi(\beta_0 + \beta_1 X_1)]^{1-Y_1}\} \times$$

$$\ldots \times \{\Phi(\beta_0 + \beta_1 X_n)^{Y_n}[1 - \Phi(\beta_0 + \beta_1 X_n)]^{1-Y_n}\}$$

The probit likelihood function:

f(β_0,β_1; Y_1,…,Y_n|X_1,…,X_n)

$$= \{\Phi(\beta_0 + \beta_1 X_1)^{Y_1}[1-\Phi(\beta_0 + \beta_1 X_1)]^{1-Y_1}\} \times$$

$$\ldots \times \{\Phi(\beta_0 + \beta_1 X_n)^{Y_n}[1-\Phi(\beta_0 + \beta_1 X_n)]^{1-Y_n}\}$$

- Can't solve for the maximum explicitly
- Must maximize using numerical methods
- As in the case of no *X*, in large samples:
  - $\hat{\beta}_0^{MLE}$, $\hat{\beta}_1^{MLE}$ are consistent
  - $\hat{\beta}_0^{MLE}$, $\hat{\beta}_1^{MLE}$ are normally distributed (more later…)
  - Their standard errors can be computed
  - Testing, confidence intervals proceeds as usual
- For multiple *X*'s, see SW App. 9.2

Measures of fit

The R^2 and $\bar{R}^2$ don't make sense here (*why?*). So, two other specialized measures are used:

1. The ***fraction correctly predicted*** = fraction of Y's for which predicted probability is >50% (if Y_i=1) or is <50% (if Y_i=0).

2. The ***pseudo-R^2*** measure the fit using the likelihood function: measures the improvement in the value of the log likelihood, relative to having no X's (see SW App. 9.2). This simplifies to the R^2 in the linear model with normally distributed errors.

Note (Lee): eople report these values, but practically, they carry no important meaning, I presume.

Testing Hypothesis

(a) One restriction on each coefficient

Use usual t-stat.

(b) More than one restriction

Use Likelihood Ratio (LR) Test

.. Chi-square (χ) test..

$$\mathbf{LR = 2\,(\log L_u - \log L_R)} \sim \chi \text{ with df} = g$$

$\log L_u$ = Unrestricted log-lik

$\log L_R$ = Restricted log-lik

Example: Extra-marital affairs

Y = 1 (affairs), or 0 (no)

```
. log using "C:\EC471\fair.log"

. insheet using "C:\EC471\fair.txt"
(13 vars, 601 obs)

. probit y  age edu kids occupation rating_m religion sex yrs_marr

Iteration 0:   log likelihood = -337.68849
Iteration 1:   log likelihood = -305.53338
Iteration 2:   log likelihood = -305.19816
Iteration 3:   log likelihood = -305.19796

Probit estimates                                  Number of obs   =        601
                                                  LR chi2(8)      =      64.98
                                                  Prob > chi2     =     0.0000
Log likelihood = -305.19796                       Pseudo R2       =     0.0962

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0245844   .0104178    -2.36   0.018    -.0450028   -.0041659
         edu |   .0112622   .0295165     0.38   0.703    -.0465891    .0691135
```

```
        kids |   .2166441   .1651681     1.31   0.190    -.1070795    .5403677
  occupation |   .0136686   .0414037     0.33   0.741    -.0674813    .0948184
    rating_m |  -.2717912   .0534747    -5.08   0.000    -.3765998   -.1669826
    religion |  -.1854684   .0516258    -3.59   0.000    -.2866532   -.0842837
         sex |   .1734569   .1379911     1.26   0.209    -.0970007    .4439144
    yrs_marr |   .0543435   .0188087     2.89   0.004     .0174792    .0912078
       _cons |   .7794021   .5125492     1.52   0.128    -.2251759     1.78398
------------------------------------------------------------------------------
```

Now, we wish to test if the coefficients of AGE and EDU are jointly insignificant.

H_o: $\beta_1 = \beta_2 = 0$ H_a: H_o is not true

Unrestricted Log likelihood = -305.19796

We run a restricted probit model (without age and edu).

```
. probit y   kids occupation rating_m religion sex yrs_marr

Iteration 0:   log likelihood = -337.68849
Iteration 1:   log likelihood = -308.34709
Iteration 2:   log likelihood = -308.08918
Iteration 3:   log likelihood = -308.08906

Probit estimates                                  Number of obs   =        601
```

```
                                                  LR chi2(6)      =      59.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -308.08906                       Pseudo R2       =     0.0877

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |   .2367733   .1639708     1.44   0.149    -.0846035    .5581501
  occupation |   .0134218   .0369199     0.36   0.716    -.0589399    .0857835
    rating_m |  -.2642769   .0522959    -5.05   0.000     -.366775   -.1617789
    religion |   -.188488   .0514507    -3.66   0.000    -.2893295   -.0876464
         sex |   .1174134   .1325305     0.89   0.376    -.1423415    .3771684
    yrs_marr |   .0223682   .0130136     1.72   0.086     -.003138    .0478744
       _cons |   .4233281   .3149532     1.34   0.179    -.1939689    1.040625
------------------------------------------------------------------------------
```

Restricted Log likelihood = −308.08906

LR = 2(logL$_U$ – logL$_R$)= 2 * (-305.19796 – (−308.08906))
= 5.65

5% Critical value of the chi-square test with df 2 = 5.99

Thus, we do not reject the null. The two coefficients are jointly insignificant.

Simply, in STATA

```
. test age edu

 ( 1)  age = 0
 ( 2)  edu = 0

           chi2(  2) =    5.65
         Prob > chi2 =    0.0594
```

Note: overall insignificance (all slope coefficients = 0)

```
LR chi2(8)      =      64.98    Prob > chi2     =     0.0000
```

Summary

(SW Section 9.5)

- If Y_i is binary, then $E(Y| X) = \Pr(Y=1|X)$
- Three models:
  - linear probability model (linear multiple regression)
  - probit (cumulative standard normal distribution)
  - logit (cumulative standard logistic distribution)
- LPM, probit, logit all produce predicted probabilities
- Effect of ΔX is 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