Lecture 8A: Binary Dependent Variable

Introduction ot Econometrics,Fall 2018

Zhaopeng Qu

Nanjing University

11/1/2018

The Linear Probability Model(LPM)

Introduction

- So far the dependent variable (Y) has been continuous:
    - testscore
    - average hourly earnings
    - GDP growth rate
- What if Y is discrete?
    - Binary outcomes: LPM, logit and probit
    - 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

- If a variable is binary,then the expecation of it is

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

- Then we have the probability of Y conditional on X

$$E[Y|X_{1i}, ..., X_{ki}] = Pr(Y = 1|X_{1i}, ..., X_{ki})$$

The linear probability model

- The conditional expectation equals the probability that $Y_i = 1$ conditional on $X_{1i}, ..., X_{ki}$:

$$E[Y|X_{1i}, ..., X_{ki}] = Pr(Y = 1|X_{1i}, ..., X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

- The population coefficient β_j equals the change in the probability that $Y_i = 1$ associated with a unit change in X_j .

$$\frac{\partial Pr(Y_i = 1|X_{1i}, ..., X_{ki})}{\partial X_j} = \beta_j$$

The linear probability model

- Assumptions are the same as for general multiple regression model:
- Review: Multiple regression model with continuous dependent variable
- Advantages of the linear probability model:
  - easy to estimate
  - Coeffcient estimates are easy to interpret
- Disadvantages of the linear probability model
  - Predicted probability can be above 1 or below 0!(it doesn't make sense)
  - Error terms are heteroskedastic

An Example: Mortgage applications

- 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?
- Boston HMDA data: a data set on mortgage applications collected by the Federal Reserve Bank in Boston.

Variable	Description	Mean	SD
deny	$= 1$ if application is denied	0.120	0.325
pi_ratio	monthly loan payments / monthly income	0.331	0.107
black	$= 1$ if applicant is black	0.142	0.350

An Example: Mortgage applications

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

```
coeftest(myfit,vcov. = sandwich)
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.079910   0.031953 -2.5008  0.01246 *
## hmda_small$pi_rat    0.603535   0.098441  6.1309 1.02e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

An Example: Mortgage applications

- The estimated OLS coefficient on the payment to income ratio $\hat{\beta}_1 = 0.60$
- The estimated coefficient is significantly different from 0 at a 1% significance level.
- How should we interpret $\hat{\beta}_1$?
    - An original one: “payments/monthly income ratio increase *1*,then **probability being denied** will also increase *0.6*.”
    - More reasonable one: “payments/monthly income ratio increase *0.1(10%)*,then probability being denied will also increase *0.06(6%)*”.

An Example: Mortgage applications

- What is the effect of race on the probability of denial, holding constant the P/I ratio? To keep things simple, we focus on differences between black applicants and white applicants.

```
coeftest(myfit2,vcov. = sandwich)
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         -0.090514   0.028582 -3.1669  0.001561 **
## hmda_small$pi_rat    0.559195   0.088610  6.3107 3.303e-10 **
## hmda_small$black     0.177428   0.024931  7.1169 1.455e-12 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

An Example: Mortgage applications: Race

- The coefficient on black, *0.177*, indicates that an African American applicant has a *17.7%* higher probability of having a mortgage application denied than a white applicant, holding constant their payment-to-income ratio.
- This coefficient is significant at the 1% level (the t-statistic is 7.11).

LPM: shortcomings

- Always suffer heteroskedasticity.
    - Always use heteroskedasticity robust standard errors!
- In the linear probability model the predicted probability can be below 0 or above 1!

Mortgage applications: Predicted value

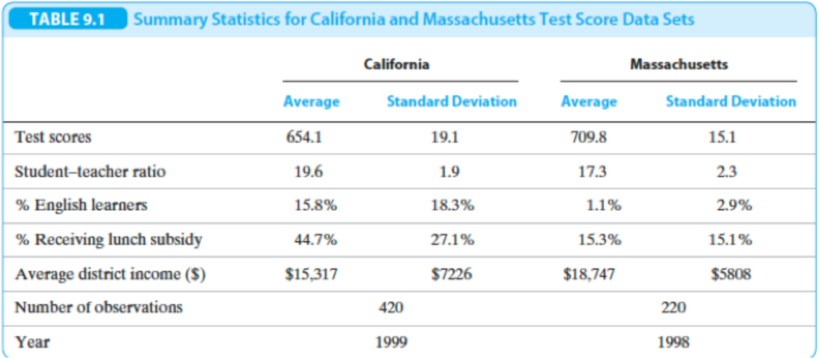


TABLE 9.1 Summary Statistics for California and Massachusetts Test Score Data Sets

	California		Massachusetts	
	Average	Standard Deviation	Average	Standard Deviation
Test scores	654.1	19.1	709.8	15.1
Student–teacher ratio	19.6	1.9	17.3	2.3
% English learners	15.8%	18.3%	1.1%	2.9%
% Receiving lunch subsidy	44.7%	27.1%	15.3%	15.1%
Average district income ($)	$15,317	$7226	$18,747	$5808
Number of observations	420		220	
Year	1999		1998	

Nonlinear probability model

Introduction

- Probabilities cannot be less than 0 or greater than 1
- To address this problem we will consider nonlinear probability models

$$\begin{aligned} Pr(Y_i = 1) &= G(Z) \\ &= G(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i}) \end{aligned}$$

- where $Z = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i}$ and $0 \leq g(Z) \leq 1$
- 2 types nonlinear functions

1. Probt

$$G(Z) = \phi(Z)$$

2. Logit

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

Probit Model with one regression

- Probit regression models the probability that Y = 1
  - Using the cumulative standard normal distribution function $\Phi(Z)$ and $0 \leq Phi(Z) \leq 1$
  - evaluated at $Z = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i}$
  - since $\Phi(z) = Pr(Z \leq z)$ we have that the predicted probabilities of the probit model are between 0 and 1.
- For example
  - Suppose we have only one regressor X and $Z = -2 + 3X_1$
  - like We want to know the probability that $Y = 1$ when $X_1 = 0.4$
  - Then $Z = -2 + 3 \times 0.4 = -0.8$
  - So the probability $Pr(Y = 1) = Pr(Z \leq -0.8) = \Phi(-0.8)$

Probit Model

- $Pr(Y = 1) = Pr(Z \leq -0.8) = \Phi(-.8) = 0.2119$

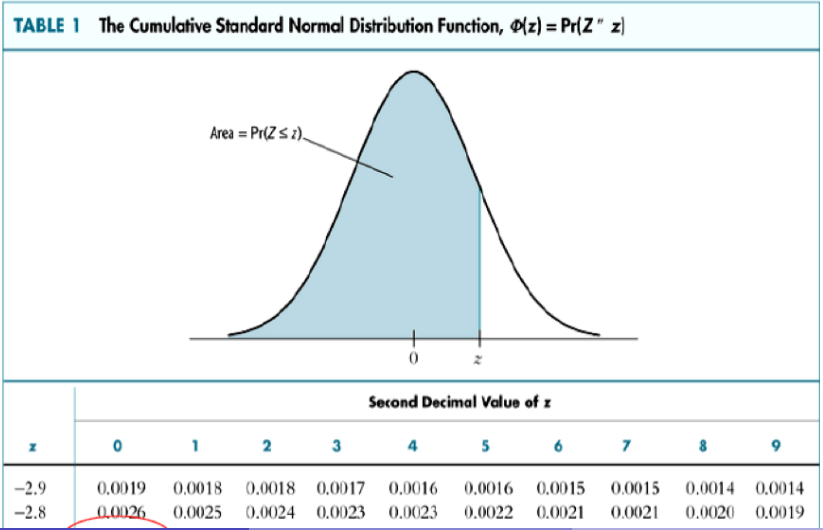


TABLE 1 The Cumulative Standard Normal Distribution Function, Φ(z) = Pr(Z " z)

Area = Pr(Z ≤ z)
0
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

Probit Model with multiple regressors

- The probit model with multiple regressor, the probit population regression model with two regressors, X1 and X2, is

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

- For example
  - Suppose we have only one regressor X and $Z = -1.6 + 2X_1 + 0.5X_2$
  - $X_1 = 0.4$ and $X_2 = 1$, then $Z = -1.6 + 2 \times 0.4 + 0.5 \times 1 = -0.3$
  - So the probability $Pr(Y = 1) = Pr(Z \leq -0.3) = \Phi(-0.3)$

Example: Mortgage Applications

- we fit a probit model: mortgage denial (deny) and the payment-toincome ratio (P/I ratio)

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

- *What is the change in the predicted probability* that an application will be deniedif P/I ratio incrreases from *0.3 to 0.4*?
- The probability of denial when $P/I\ ratio = 0.3$

$$\Phi(-2.19 + 2.97 \times 0.3) = \Phi(-1.3) = 0.097$$

- The probability of denial when $P/I\ ratio = 0.4$

$$\Phi(-2.19 + 2.97 \times 0.4) = \Phi(-1.0) = 0.159$$

- The estimated change in the probability of denial is $0.159 - 0.097 = 0.062$

Effect of a change in X: Marginal Effects

- For nonlinear models, the ME varies with the point of evaluation
  - *Marginal Effect at a Representative Value* (MER):ME at $X = X^*$ (at representative values of the regerssors)
  - *Marginal Effect at Mean* (MEM): ME at $X = \bar{X}$(at the sample mean of the regressors)
  - *Average Marginal Effect* (AME): average of ME at each $X = X_i$ (at sample values and then average)

Example: Mortgage applications: marginal effect

- Because the probit regression function is nonlinear, the effect of a change in X depends on the starting value of X.

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

- *Marginal Effect at Mean* (MEM):(at the sample mean of the regressors)

$$\frac{\partial Pr(deny = 1|P/I\ ratio)}{\partial P/I\ ratio}_{at\ mean} = \Phi(-2.19 + 2.97 \times 0.331) \times 2.97$$

Case II: The explanatory variable is discrete.

- If xj is a discrete variable then we should not rely on calculus in evaluating the effect on the response probability.
- Assume X_2 is a dummy variable, then partial effect of X_2 changing from 0 to 1:

$$G(\beta_0+\beta_1X_{1,i}+\beta_2\times1+...+\beta_kX_{k,i})-G(\beta_0+\beta_1X_{1,i}+\beta_2\times0+...+\beta_kX_{k,i})$$

Example: Mortgage applications: Race

```
##
## Call:
## glm(formula = hmda_small$deny ~ hmda_small$pi_rat + hmda_s
##     family = binomial(link = "probit"), data = hmda_small)
##
## Deviance Residuals:
##     Min        1Q    Median        3Q       Max
## -2.1208  -0.4762  -0.4251  -0.3550   2.8799
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -2.25879    0.13669 -16.525  < 2e-16 ***
## hmda_small$pi_rat   2.74178    0.38047   7.206 5.75e-13 ***
## hmda_small$black    0.70816    0.08335   8.496  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Example: Mortgage applications: Race

- we fit a probit model: mortgage denial (deny) and the payment-toincome ratio (P/I ratio) and race

$$Pr(deny \widehat{= 1|}P/I\ ratio) = \Phi(-2.26 + 2.74 P/I\ ratio + 0.71 black)$$

- The probability of denial when $black = 0$,thus whites(non-blacks) is

$$\Phi(-2.26 + 2.97 \times 0.3 + 0.71 \times 0) = \Phi(-1.3) = 0.097$$

- The probability of denial when $black = 1$,thus blacks is

$$\Phi(-2.26 + 2.97 \times 0.3 + 0.71 \times 1) = \Phi(-0.59) = 0.2776$$

- so the difference between whites and blacks at $P/Iratio = 0.3$ is $0.2776 - 0.097 = 0.19$, which means probability of denial for blacks is 19% higher than that for whites.

Logit Model

- Logit regression models the probability that Y $= 1$
- Using the cumulative standard logistic distribution function

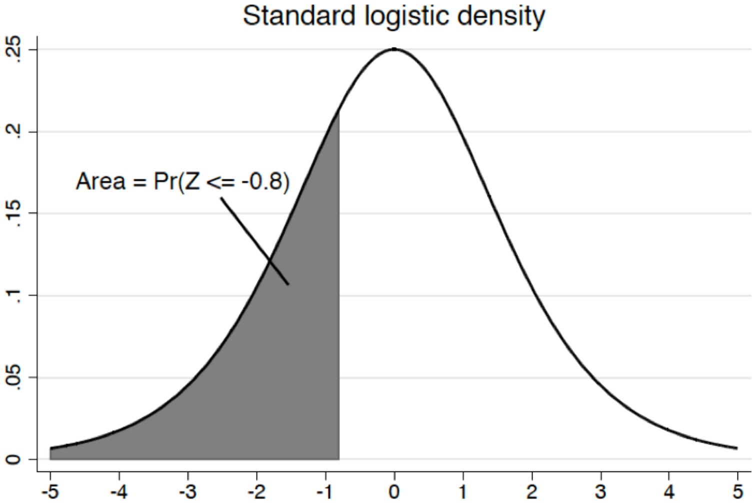
$$F(Z) = \frac{1}{1+e^{-z}}$$

  - evaluated at $Z = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i}$
  - since $F(z) = Pr(Z \leq z)$ we have that the predicted probabilities of the probit model are between 0 and 1.

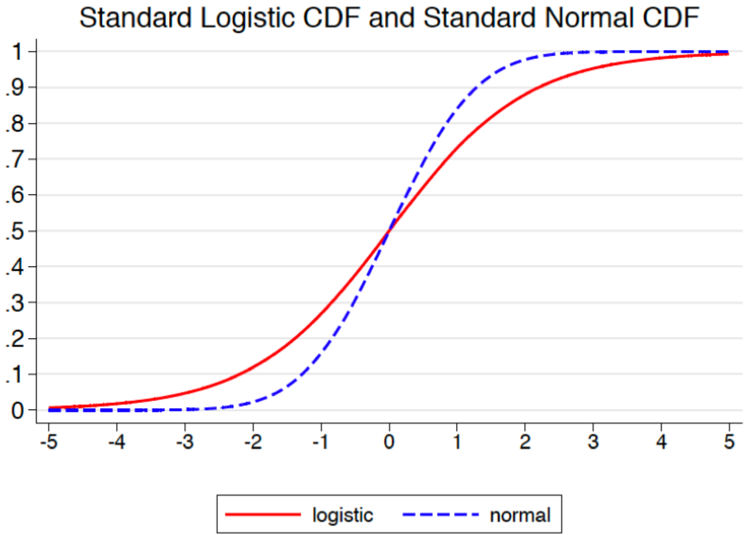
- For example
  - Suppose we have only one regressor X and $Z = -2 + 3X_1$
  - We want to know the probability that $Y = 1$ when $X_1 = 0.4$
  - Then $Z = -2 + 3 \times 0.4 = -0.8$
  - So the probability $Pr(Y = 1) = Pr(Z \leq -0.8) = F(-0.8)$

Logit Model

- $Pr(Y=1) = Pr(Z \leq -0.8) = \frac{1}{1+e^{0.8}} = 0.31$



Logit v.s. Probit



How to estimate Logit and Probit models

- 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 $\beta_0, \beta_1, ..., \beta_k$
  - these models can NOT be estimated by OLS
- The method used to estimate logit and probit models is **Maximum Likelihood Estimation** (MLE).

Maximum Likelihood Estimation

Introdution to MLE

- 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.

Introdution to MLE

- $X_1, X_2, X_3, ...X_n$ have joint density denoted

$$f_\theta(x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n|\theta)$$

- Given observed values $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$,the likelihood of θ is the function

$$lik(\theta) = f(x_1, x_2, ..., x_n|\theta)$$

considered as a function of θ.

Maximum likelihood estimation

- Lets start with a special case: The MLE with only Y
- Suppose that we are flipping a coin that may be biased, so that the probability of a heads may not be 0.5. Maybe we???re interested in estimating the probability of a heads.
- Let $Y = 1(heads)$ be a binary variable that indicates whether or not a heads is observed. The outcome of a toss is a Bernoulli random variable: $Pr(Y=1)=p$
- Yi is a Bernoulli random variable, therefore the density for a single observation is

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

Maximum likelihood estimation

- **Step 1**: write down the likelihood function, the joint probability distribution of the data
- $Y_1, ..., Y_n$ are i.i.d, the joint probability distribution is therefore the product of the individual distributions

$$
\begin{aligned}
Pr(Y_1 = y_1, ..., Y_n = y_n) &= Pr(Y_1 = y_1) \times ... \times Pr(Y_n = y_n) \\
&= p^{y_1}(1-p)^{1-y_1} \times ... \times p^{y_n}(1-p)^{1-y_n} \\
&= p^{(y_1+y_2+...+y_n)}(1-p)^{n-(y_1+y_2+...+y_n)}
\end{aligned}
$$

Maximum likelihood estimation

- We have the likelihood function:

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

- **Step 2**: Maximize the likelihood function
- Easier to maximize the logarithm of the likelihood function

$$ln(f_{bernouilli}(p; Y_1 = y_1, ..., Y_n = y_n)) = \left(\sum y_i\right) ln(p) + \left(n - \sum y_i\right) ln(1-$$

- 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 and setting it to zero.

$$\frac{d}{dp} ln(f_{bernouilli}(p; Y_1 = y_1, ..., Y_n = y_n)) = \frac{\sum y_i}{p} - \frac{n - \sum y_i}{1-p} = 0$$

- Solving for p yields the MLE; that is, $\hat{p}_{MLE}$ satisfies

$$\hat{p}_{MLE} = \frac{1}{n} \sum y_i = \overline{Y}$$

MLE of the probit model

- Step 1: write down the likelihood function

$$\begin{aligned} Pr(Y_1 = y_1, ..., Y_n = y_n) &= Pr(Y_1 = y_1) \times ... \times Pr(Y_n = y_n) \\ &= p^{y_1}(1-p)^{1-y_1} \times ... \times p^{y_n}(1-p)^{1-y_n} \end{aligned}$$

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

$$p_i = \phi(X_{1i}, ..., X_{ki}) = \phi(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})$$

- substituting for pi gives the likelihood function:

$$\Bigg[\phi(\beta_0 + \beta_1 X_{11} + ... + \beta_k X_{k1})^{y_1}(1 - \phi(\beta_0 + \beta_1 X_{11} + ... + \beta_k X_{k1}))$$

$$... \times \Bigg[\phi(\beta_0 + \beta_1 X_{1n} + ... + \beta_k X_{kn})^{y_n}(1 - \phi(\beta_0 + \beta_1 X_{1n} + ... + \beta_k X_{kn}$$

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

$$
\begin{aligned}
ln(f_{probit}(\beta_0, ..., \beta_k; Y_1, ..., Y_n | X_{1i}, ..., X_{ki}, i = 1, ..., n)) = \\
\sum Y_i \times ln[\phi(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})] \\
+ \sum (1 - Y_i) \times ln[1 - \phi(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})]
\end{aligned}
$$

MLE of the logit model

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

$$Pr(Y_1 = y_1, ..., Y_n = y_n) = p^{y_1}(1-p)^{1-y_1} \times ... \times p^{y_n}(1-p)^{1-y_n}$$

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

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})}}$$

MLE of the logit model

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

$$
\begin{aligned}
& ln(f_{logit}(\beta_0, ..., \beta_k; Y_1, ..., Y_n | X_{1i}, ..., X_{ki}, i = 1, ..., n)) \\
& \quad = \sum Y_i \times ln\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})}}\right) \\
& \quad + \sum (1 - Y_i) \times ln\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})}}\right)
\end{aligned}
$$

MLE estimator in practice

- There is no simple formula for the probit and logit MLE, the maximization must be done using **numerical algorithm** on a computer.
- the MLE is consistent and normally distributed in large samples.
- Because regression software commonly computes the MLE of the estimate coefficients, this estimator is easy to use in practice.

Statistical inference based on the MLE

- Because the MLE is normally distributed in large samples, statistical inference about the probit and logit coefficients based on the MLE proceeds in the same way as inference about the linear regression function coefficients based on the OLS estimator.
- That is, hypothesis tests are performed using the **t-statistic** and **95% confidence intervals** are formed as 1.96 standard errors. Tests of joint hypotheses on multiple coefficients use the **F-statistic** in a way similar to that discussed for the linear regression model.
- F-statistic and Chi-squared stistic

$$F_{stat} \longrightarrow \frac{\chi_q^2}{q}$$

where q is the number of restrictions being tested.

Measures of Fit

- R2 is a poor measure of fit for the linear probability model. This is also true for probit and logit regression.
- Two measures of fit for models with binary dependent variables

1. *fraction correctly predicted*
   - If Yi $=1$ and the predicted probability exceeds 50% or if Yi $=0$ and the predicted probability is less than 50%, then Yi is said to be correctly predicted.
2. **The pseudo-R2**
   - The pseudo-R2 compares the value of the likelihood of the estimated model to the value of the likelihood when none of the X???s are included as regressors.

$$pseudo-R2=1-\frac{ln(f^{max}_{probit})}{ln(f^{max}_{bernoulli})}$$

Comparing the LPM,Probit and Logit

- All three models???linear probability, probit, and logit???are just approximations to the unknown population regression function $E(Y|X) = Pr(Y = 1|X)$.
- LPM is easiest to use and to interpret, but it cannot capture the nonlinear nature of the true population regression function.
- Probit and logit regressions model this nonlinearity in the probabilities, but their regression coefficients are more difficult to interpret.
- So which should you use in practice?
  - *There is no one right answer*, and different researchers use different models.
  - *Probit and logit regressions frequently produce similar results*.

Comparing the LPM,Probit and Logit

- The marginal effects and predicted probabilities are much more similar across models.
- Coefficients can be compared across models, using the following rough conversion factors (Amemiya 1981)

$$\begin{aligned}\widehat{\beta}_{logit} &\simeq 4\widehat{\beta}_{ols}\\ \widehat{\beta}_{probit} &\simeq 2.5\widehat{\beta}_{ols}\\ \widehat{\beta}_{logit} &\simeq 1.6\widehat{\beta}_{probit}\end{aligned}$$

Example: Mortgage Applications(short regression)

Dependent variable: *deny* = 1 if mortgage application is denied, = 0 if accepted			
regression model	LPM	Probit	Logit
black	0.177*** (0.025)	0.71*** (0.083)	1.27*** (0.15)
P/I ratio	0.559*** (0.089)	2.74*** (0.44)	5.37*** (0.96)
constant	-0.091*** (0.029)	-2.26*** (0.16)	-4.13*** (0.35)
difference Pr(*deny*=1) between black and white applicant when *P/I ratio*=0.3	17.7%	15.8%	14.8%

Example: Mortgage Applications(long regression)

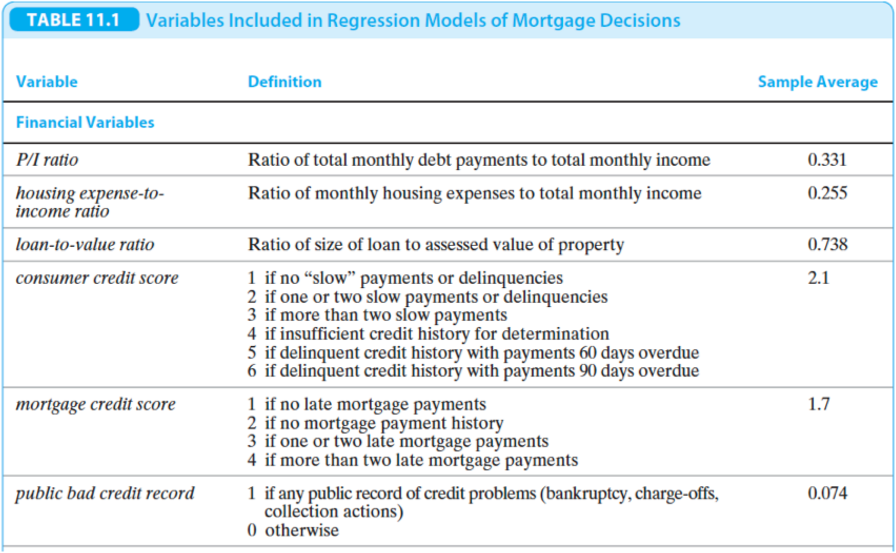
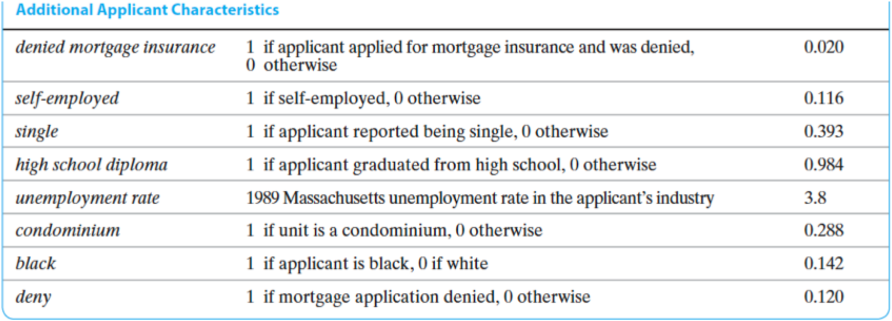


TABLE 11.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
Financial Variables		
P/I ratio	Ratio of total monthly debt payments to total monthly income	0.331
housing expense-to-income ratio	Ratio of monthly housing expenses to total monthly income	0.255
loan-to-value ratio	Ratio of size of loan to assessed value of property	0.738
consumer credit score	1 if no "slow" payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
mortgage credit score	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
public bad credit record	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074

图 1: pic

Example: Mortgage Applications(long regression)



Additional Applicant Characteristics

denied mortgage insurance	1 if applicant applied for mortgage insurance and was denied, 0 otherwise	0.020
self-employed	1 if self-employed, 0 otherwise	0.116
single	1 if applicant reported being single, 0 otherwise	0.393
high school diploma	1 if applicant graduated from high school, 0 otherwise	0.984
unemployment rate	1989 Massachusetts unemployment rate in the applicant's industry	3.8
condominium	1 if unit is a condominium, 0 otherwise	0.288
black	1 if applicant is black, 0 if white	0.142
deny	1 if mortgage application denied, 0 otherwise	0.120

图 **2:** pic

Example: Mortgage Applications(long regression)

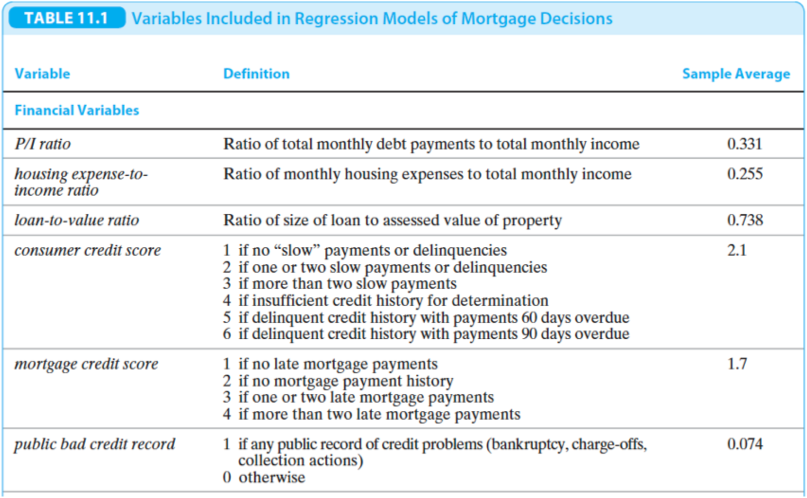
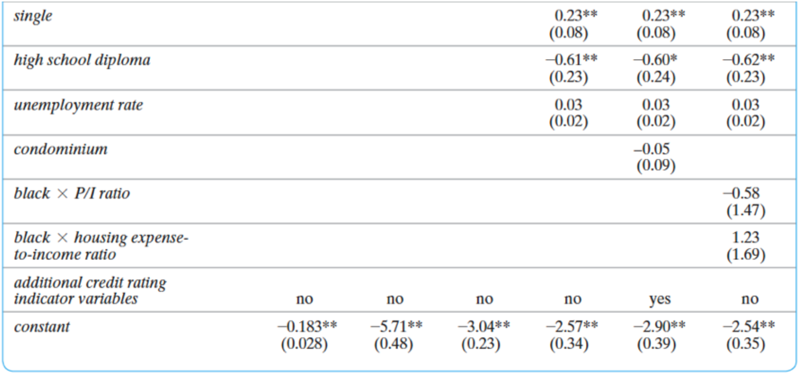


TABLE 11.1 Variables Included in Regression Models of Mortgage Decisions

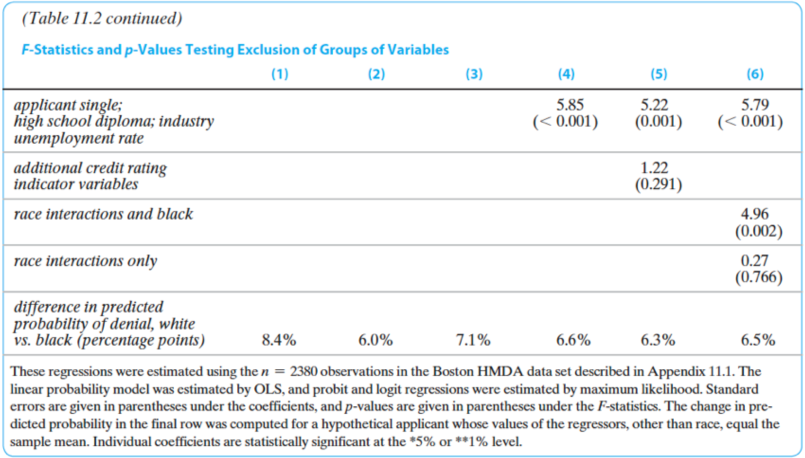
Variable	Definition	Sample Average
Financial Variables		
P/I ratio	Ratio of total monthly debt payments to total monthly income	0.331
housing expense-to-income ratio	Ratio of monthly housing expenses to total monthly income	0.255
loan-to-value ratio	Ratio of size of loan to assessed value of property	0.738
consumer credit score	1 if no "slow" payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
mortgage credit score	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
public bad credit record	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074

Example: Mortgage Applications(long regression)



single				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
high school diploma				–0.61** (0.23)	–0.60* (0.24)	–0.62** (0.23)
unemployment rate				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
condominium					–0.05 (0.09)	
black × *P/I ratio*						–0.58 (1.47)
black × *housing expense-to-income ratio*						1.23 (1.69)
additional credit rating indicator variables	no	no	no	no	yes	no
constant	–0.183** (0.028)	–5.71** (0.48)	–3.04** (0.23)	–2.57** (0.34)	–2.90** (0.39)	–2.54** (0.35)

Example: Mortgage Applications(long regression)



(Table 11.2 continued)

F-Statistics and p-Values Testing Exclusion of Groups of Variables

	(1)	(2)	(3)	(4)	(5)	(6)
applicant single; high school diploma; industry unemployment rate				5.85 (< 0.001)	5.22 (0.001)	5.79 (< 0.001)
additional credit rating indicator variables					1.22 (0.291)	
race interactions and black						4.96 (0.002)
race interactions only						0.27 (0.766)
difference in predicted probability of denial, white vs. black (percentage points)	8.4%	6.0%	7.1%	6.6%	6.3%	6.5%

These regressions were estimated using the $n = 2380$ observations in the Boston HMDA data set described in Appendix 11.1. The linear probability model was estimated by OLS, and probit and logit regressions were estimated by maximum likelihood. Standard errors are given in parentheses under the coefficients, and p-values are given in parentheses under the F-statistics. The change in predicted probability in the final row was computed for a hypothetical applicant whose values of the regressors, other than race, equal the sample mean. Individual coefficients are statistically significant at the *5% or **1% level.

Threats to internal validity

- 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?
   - Is there measurement error?
   - Is there sample selection bias?
   - is there a problem of simultaneous causality?
2. External validity

More Extension: Categoried and Limited Dependent Variables families

- Binary outcomes: LPM, logit and probit
- Mulitnomial outcomes: Discrete Choice Models(multi-logit,nested-logit)
- Ordered outcomes: Ordered Response Models(order probit and logit)
- Count outcomes: (possion model)
- Limited Dependent Varaible(Censored, Tobit and Selection Models)
- Time: (Duration Model)