15

QUALITATIVE RESPONSE REGRESSION MODELS

In all the regression models that we have considered so far, we have implicitly assumed that the regressand, the dependent variable, or the *response* variable Y is quantitative, whereas the explanatory variables are either quantitative, qualitative (or dummy), or a mixture thereof. In fact, in Chapter 9, on dummy variables, we saw how the dummy regressors are introduced in a regression model and what role they play in specific situations.

In this chapter we consider several models in which the regressand itself is qualitative in nature. Although increasingly used in various areas of social sciences and medical research, qualitative response regression models pose interesting estimation and interpretation challenges. In this chapter we only touch on some of the major themes in this area, leaving the details to more specialized books.[1]

15.1 THE NATURE OF QUALITATIVE RESPONSE MODELS

Suppose we want to study the labor force participation (LFP) decision of adult males. Since an adult is either in the labor force or not, LFP is a *yes* or *no* decision. Hence, the response variable, or regressand, can take only two

[1]At the introductory level, the reader may find the following sources very useful. Daniel A. Powers and Yu Xie, *Statistical Methods for Categorical Data Analysis,* Academic Press, 2000; John H. Aldrich and Forrest Nelson, *Linear Probability, Logit, and Probit Models,* Sage Publications, 1984; Tim Futing Liao, *Interpreting Probability Models: Logit, Probit and Other Generalized Linear Models,* Sage Publications, 1994. For a very comprehensive review of the literature, see G. S. Maddala, *Limited-Dependent and Qualitative Variables in Econometrics,* Cambridge University Press, 1983.

values, say, 1 if the person is in the labor force and 0 if he or she is not. In other words, the regressand is a **binary, or dichotomous, variable.** Labor economics research suggests that the LFP decision is a function of the unemployment rate, average wage rate, education, family income, etc.

As another example, consider U.S. presidential elections. Assume that there are two political parties, Democratic and Republican. The dependent variable here is vote choice between the two political parties. Suppose we let $Y = 1$, if the vote is for a Democratic candidate, and $Y = 0$, if the vote is for a Republican candidate. A considerable amount of research on this topic has been done by the economist Ray Fair of Yale University and several political scientists.[2] Some of the variables used in the vote choice are growth rate of GDP, unemployment and inflation rates, whether the candidate is running for reelection, etc. For the present purposes, the important thing to note is that the regressand is a qualitative variable.

One can think of several other examples where the regressand is qualitative in nature. Thus, a family either owns a house or it does not, it has disability insurance or it does not, both husband and wife are in the labor force or only one spouse is. Similarly, a certain drug is effective in curing an illness or it is not. A firm decides to declare a stock dividend or not, a senator decides to vote for a tax cut or not, a U.S. President decides to veto a bill or accept it, etc.

We do not have to restrict our response variable to yes/no or dichotomous categories only. Returning to our presidential elections example, suppose there are three parties, Democratic, Republican, and Independent. The response variable here is **trichotomous.** In general, we can have a **polychotomous** (or **multiple-category**) response variable.

What we plan to do is to first consider the dichotomous regressand and then consider various extensions of the basic model. But before we do that, it is important to note a fundamental difference between a regression model where the regressand Y is quantitative and a model where it is qualitative.

In a model where Y is quantitative, our objective is to estimate its expected, or mean, value given the values of the regressors. In terms of Chapter 2, what we want is $E(Y_i \mid X_{1i}, X_{2i}, \ldots, X_{ki})$, where the X's are regressors, both quantitative and qualitative. In models where Y is qualitative, our objective is to find the probability of something happening, such as voting for a Democratic candidate, or owning a house, or belonging to a union, or participating in a sport etc. Hence, qualitative response regression models are often known as *probability models*.

In the rest of this chapter, we seek answers to the following questions:

1. How do we estimate qualitative response regression models? Can we simply estimate them with the usual OLS procedures?

[2]See, for example, Ray Fair, "Econometrics and Presidential Elections," *Journal of Economic Perspective*, Summer 1996, pp. 89–102, and Machael S. Lewis-Beck, *Economics and Elections: The Major Western Democracies,* University of Michigan Press, Ann Arbor, 1980.

2. Are there special inference problems? In other words, is the hypothesis testing procedure any different from the ones we have learned so far?

3. If a regressand is qualitative, how can we measure the goodness of fit of such models? Is the conventionally computed R^2 of any value in such models?

4. Once we go beyond the dichotomous regressand case, how do we estimate and interpret the polychotomous regression models? Also, how do we handle models in which the regressand is **ordinal,** that is, an ordered categorical variable, such as schooling (less than 8 years, 8 to 11 years, 12 years, and 13 or more years), or the regressand is **nominal** where there is no inherent ordering, such as ethnicity (Black, White, Hispanic, Asian, and other)?

5. How do we model phenomena, such as the number of visits to one's physician per year, the number of patents received by a firm in a given year, the number of articles published by a college professor in a year, the number of telephone calls received in a span of 5 minutes, or the number of cars passing through a toll booth in a span of 5 minutes? Such phenomena, called **count data,** or **rare event** data, are an example of the **Poisson** (probability) process.

In this chapter we provide answers to some of these questions at the elementary level, for some of the topics are quite advanced and require more background in mathematics and statistics than assumed in this book. References cited in the various footnotes may be consulted for further details.

We start our study of qualitative response models by first considering the **binary response** regression model. There are three approaches to developing a probability model for a binary response variable:

1. The **linear probability model (LPM)**
2. The **logit model**
3. The **probit model**

Because of its comparative simplicity, and because it can be estimated by OLS, we will first consider the LPM, leaving the other two models for subsequent sections.

15.2 THE LINEAR PROBABILITY MODEL (LPM)

To fix ideas, consider the following regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(15.2.1)}$$

where X = family income and $Y = 1$ if the family owns a house and 0 if it does not own a house.

Model (15.2.1) looks like a typical linear regression model but because the regressand is binary, or dichotomous, it is called a **linear probability model (LPM).** This is because the conditional expectation of Y_i given

X_i, $E(Y_i \mid X_i)$, can be interpreted as the *conditional probability* that the event will occur given X_i, that is, Pr $(Y_i = 1 \mid X_i)$. Thus, in our example, $E(Y_i \mid X_i)$ gives the probability of a family owning a house and whose income is the given amount X_i.

The justification of the name LPM for models like (15.2.1) can be seen as follows: Assuming $E(u_i) = 0$, as usual (to obtain unbiased estimators), we obtain

$$E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i \tag{15.2.2}$$

Now, if P_i = probability that $Y_i = 1$ (that is, the event occurs), and $(1 - P_i)$ = probability that $Y_i = 0$ (that is, that the event does not occur), the variable Y_i has the following (probability) distribution.

Y_i	Probability
0	$1 - P_i$
1	P_i
Total	1

That is, Y_i follows the **Bernoulli probability distribution.**

Now, by the definition of mathematical expectation, we obtain:

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i \tag{15.2.3}$$

Comparing (15.2.2) with (15.2.3), we can equate

$$E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i = P_i \tag{15.2.4}$$

that is, the conditional expectation of the model (15.2.1) can, in fact, be interpreted as the conditional probability of Y_i. In general, the expectation of a Bernoulli random variable is the probability that the random variable equals 1. In passing note that if there are n independent trials, each with a probability p of success and probability $(1 - p)$ of failure, and X of these trials represent the number of successes, then X is said to follow the **binomial distribution.** The mean of the binomial distribution is np and its variance is $np(1 - p)$. The term *success* is defined in the context of the problem.

Since the probability P_i must lie between 0 and 1, we have the restriction

$$0 \le E(Y_i \mid X_i) \le 1 \tag{15.2.5}$$

that is, the conditional expectation (or conditional probability) must lie between 0 and 1.

From the preceding discussion it would seem that OLS can be easily extended to binary dependent variable regression models. So, perhaps there is nothing new here. Unfortunately, this is not the case, for the LPM poses several problems, which are as follows:

Non-Normality of the Disturbances u_i

Although OLS does not require the disturbances (u_i) to be normally distributed, we assumed them to be so distributed for the purpose of statistical inference.[3] But the assumption of normality for u_i is not tenable for the LPMs because, like Y_i, the disturbances u_i also take only two values; that is, they also follow the Bernoulli distribution. This can be seen clearly if we write (15.2.1) as

$$u_i = Y_i - \beta_1 - \beta_2 X_i \tag{15.2.6}$$

The probability distribution of u_i is

	u_i	Probability
When $Y_i = 1$	$1 - \beta_1 - \beta_2 X_i$	P_i
When $Y_i = 0$	$-\beta_1 - \beta_2 X_i$	$(1 - P_i)$

(15.2.7)

Obviously, u_i cannot be assumed to be normally distributed; they follow the Bernoulli distribution.

But the nonfulfillment of the normality assumption may not be so critical as it appears because we know that the OLS point estimates still remain unbiased (recall that, if the objective is point estimation, the normality assumption is not necessary). Besides, as the sample size increases indefinitely, statistical theory shows that the OLS estimators tend to be normally distributed generally.[4] As a result, in large samples the statistical inference of the LPM will follow the usual OLS procedure under the normality assumption.

Heteroscedastic Variances of the Disturbances

Even if $E(u_i) = 0$ and $\operatorname{cov}(u_i, u_j) = 0$ for $i \neq j$ (i.e., no serial correlation), it can no longer be maintained that in the LPM the disturbances are

[3]Recall that we have recommended that the normality assumption be checked in an application by suitable normality tests, such as the Jarque–Bera test.

[4]The proof is based on the central limit theorem and may be found in E. Malinvaud, *Statistical Methods of Econometrics,* Rand McNally, Chicago, 1966, pp. 195–197. If the regressors are deemed stochastic and are jointly normally distributed, the F and t tests can still be used even though the disturbances are non-normal. Also keep in mind that as the sample size increases indefinitely, the binomial distribution converges to the normal distribution.

homoscedastic. This is, however, not surprising. As statistical theory shows, for a Bernoulli distribution the theoretical mean and variance are, respectively, p and $p(1-p)$, where p is the probability of success (i.e., something happening), showing that the variance is a function of the mean. Hence the error variance is heteroscedastic.

For the distribution of the error term given in (15.2.7), applying the definition of variance, the reader should verify that (see exercise 15.10)

$$\operatorname{var}(u_i) = P_i(1 - P_i) \tag{15.2.8}$$

That is, the variance of the error term in the LPM is heteroscedastic. Since $P_i = E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i$, the variance of u_i ultimately depends on the values of X and hence is not homoscedastic.

We already know that, in the presence of heteroscedasticity, the OLS estimators, although unbiased, are not efficient; that is, they do not have minimum variance. But the problem of heteroscedasticity, like the problem of non-normality, is not insurmountable. In Chapter 11 we discussed several methods of handling the heteroscedasticity problem. Since the variance of u_i depends on $E(Y_i \mid X_i)$, one way to resolve the heteroscedasticity problem is to transform the model (15.2.1) by dividing it through by

$$\sqrt{E(Y_i/X_i)[1 - E(Y_i/X_i)]} = \sqrt{P_i(1 - P_i)} = \text{say } \sqrt{w_i}$$

that is,

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_1}{\sqrt{w_i}} + \beta_2 \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \tag{15.2.9}$$

As you can readily verify, the transformed error term in (15.2.9) is homoscedastic. Therefore, after estimating (15.2.1), we can now estimate (15.2.9) by OLS, which is nothing but the *weighted least squares* (WLS) with w_i serving as the weights.

In theory, what we have just described is fine. But in practice the true $E(Y_i \mid X_i)$ is unknown; hence the weights w_i are unknown. To estimate w_i, we can use the following two-step procedure[5]:

Step 1. Run the OLS regression (15.2.1) despite the heteroscedasticity problem and obtain $\hat{Y}_i$ = estimate of the true $E(Y_i \mid X_i)$. Then obtain $\hat{w}_i = \hat{Y}_i(1 - \hat{Y}_i)$, the estimate of w_i.

[5]For the justification of this procedure, see Arthur S. Goldberger, *Econometric Theory,* John Wiley & Sons, New York, 1964, pp. 249–250. The justification is basically a large-sample one that we discussed under the topic of feasible or estimated generalized least squares in the chapter on heteroscedasticity (see Sec. 11.6).

Step 2. Use the estimated w_i to transform the data as shown in (15.2.9) and estimate the transformed equation by OLS (i.e., weighted least squares).

We will illustrate this procedure for our example shortly. But there is another problem with LPM that we need to address first.

Nonfulfillment of $0 \le E(Y_i \mid X) \le 1$

Since $E(Y_i \mid X)$ in the linear probability models measures the conditional probability of the event Y occurring given X, it must necessarily lie between 0 and 1. Although this is true a priori, there is no guarantee that $\hat{Y}_i$, the estimators of $E(Y_i \mid X_i)$, will necessarily fulfill this restriction, *and this is the real problem with the OLS estimation of the LPM*. There are two ways of finding out whether the estimated $\hat{Y}_i$ lie between 0 and 1. One is to estimate the LPM by the usual OLS method and find out whether the estimated $\hat{Y}_i$ lie between 0 and 1. If some are less than 0 (that is, negative), $\hat{Y}_i$ is assumed to be zero for those cases; if they are greater than 1, they are assumed to be 1. The second procedure is to devise an estimating technique that will guarantee that the estimated conditional probabilities $\hat{Y}_i$ will lie between 0 and 1. The logit and probit models discussed later will guarantee that the estimated probabilities will indeed lie between the logical limits 0 and 1.

Questionable Value of R^2 as a Measure of Goodness of Fit

The conventionally computed R^2 is of limited value in the dichotomous response models. To see why, consider the following figure. Corresponding to a given X, Y is either 0 or 1. Therefore, all the Y values will either lie along the X axis or along the line corresponding to 1. Therefore, generally no LPM is expected to fit such a scatter well, whether it is the *unconstrained LPM* (Figure 15.1*a*) or the *truncated* or *constrained LPM* (Figure 15.1*b*), an LPM estimated in such a way that it will not fall outside the logical band 0–1. As a result, the conventionally computed R^2 is likely to be much lower than 1 for such models. In most practical applications the R^2 ranges between 0.2 to 0.6. R^2 in such models will be high, say, in excess of 0.8 only when the actual scatter is very closely clustered around points A and B (Figure 15.1*c*), for in that case it is easy to fix the straight line by joining the two points A and B. In this case the predicted Y_i will be very close to either 0 or 1.

For these reasons John Aldrich and Forrest Nelson contend that "use of the coefficient of determination as a summary statistic should be avoided in models with qualitative dependent variable."[6]

[6]Aldrich and Nelson, op. cit., p. 15. For other measures of goodness of fit in models involving dummy regressands, see T. Amemiya, "Qualitative Response Models," *Journal of Economic Literature*, vol. 19, 1981, pp. 331–354.

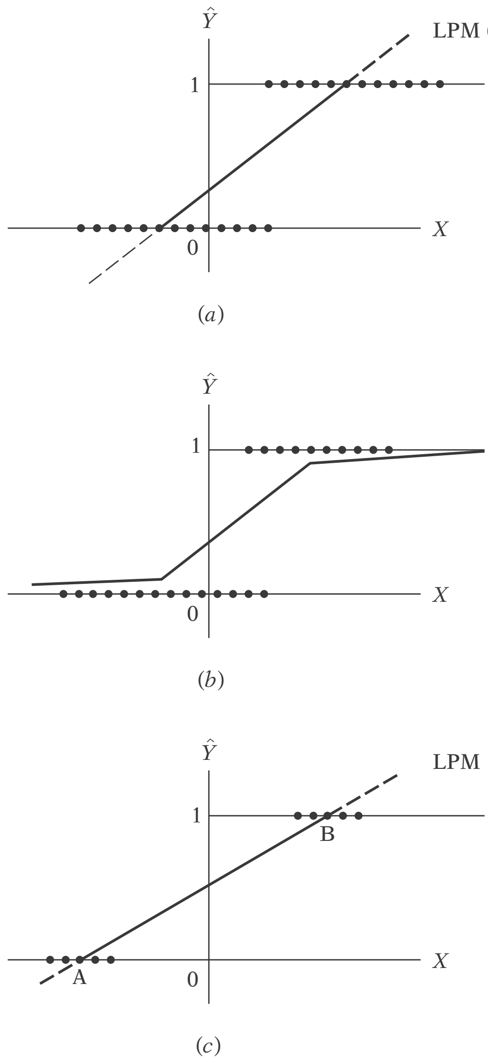


FIGURE 15.1 Linear probability models.

LPM: A NUMERICAL EXAMPLE

To illustrate some of the points made about the LPM in the preceding section, we present a numerical example. Table 15.1 gives invented data on home ownership Y (1 = owns a house, 0 = does not own a house) and family income X (thousands of dollars) for 40 families. From these data the LPM estimated by OLS was as follows:

$$\hat{Y}_i = -0.9457 + 0.1021X_i$$
$$(0.1228) \quad (0.0082) \qquad \textbf{(15.2.10)}$$
$$t = (-7.6984) \quad (12.515) \qquad R^2 = 0.8048$$

(*Continued*)

LPM: A NUMERICAL EXAMPLE (*Continued*)

TABLE 15.1
HYPOTHETICAL DATA ON HOME OWNERSHIP ($Y = 1$ IF OWNS HOME, 0 OTHERWISE) AND INCOME X (THOUSANDS OF DOLLARS)

Family	Y	X	Family	Y	X
1	0	8	21	1	22
2	1	16	22	1	16
3	1	18	23	0	12
4	0	11	24	0	11
5	0	12	25	1	16
6	1	19	26	0	11
7	1	20	27	1	20
8	0	13	28	1	18
9	0	9	29	0	11
10	0	10	30	0	10
11	1	17	31	1	17
12	1	18	32	0	13
13	0	14	33	1	21
14	1	20	34	1	20
15	0	6	35	0	11
16	1	19	36	0	8
17	1	16	37	1	17
18	0	10	38	1	16
19	0	8	39	0	7
20	1	18	40	1	17

First, let us interpret this regression. The intercept of −0.9457 gives the "probability" that a family with zero income will own a house. Since this value is negative, and since probability cannot be negative, we treat this value as zero, which is sensible in the present instance.[7] The slope value of 0.1021 means that for a unit change in income (here $1000), on the average the probability of owning a house increases by 0.1021 or about 10 percent. Of course, given a particular level of income, we can estimate the actual probability of owning a house from (15.2.10). Thus, for $X = 12$ ($12,000), the estimated probability of owning a house is

$$\begin{aligned}(\hat{Y}_i \mid X = 12) &= -0.9457 + 12(0.1021)\\ &= 0.2795\end{aligned}$$

That is, the probability that a family with an income of $12,000 will own a house is about 28 percent. Table 15.2 shows the estimated probabilities, $\hat{Y}_i$, for the various income levels listed in the table. The most noticeable feature of this table is that six estimated values are negative and six values are in excess of 1, demonstrating clearly the point made earlier that, although $E(Y_i \mid X)$ is positive and less than 1, their estimators, $\hat{Y}_i$, need not be necessarily positive or less than 1. This is one reason that the LPM is not the recommended model when the dependent variable is dichotomous.

Even if the estimated Y_i were all positive and less than 1, the LPM still suffers from the problem of heteroscedasticity, which can be seen readily from (15.2.8). As a consequence, we cannot trust the estimated standard errors reported in (15.12.10). (Why?) But we can use the weighted least-squares (WLS) procedure discussed earlier to obtain more efficient estimates of the standard errors. The necessary weights, $\hat{w}_i$, required for the application of WLS are also shown in Table 15.2. But note that since some Y_i are negative and some are in excess of one, the $\hat{w}_i$ corresponding to these values will be negative. Thus, we cannot use these observations in WLS (why?), thereby reducing the number of

(*Continued*)

[7] One can loosely interpret the highly negative value as near improbability of owning a house when income is zero.

LPM: A NUMERICAL EXAMPLE (*Continued*)

TABLE 15.2
ACTUAL *Y*, ESTIMATED *Y*, AND WEIGHTS w_i FOR THE HOME OWNERSHIP EXAMPLE

Y_i	$\hat{Y}_i$	$\hat{w}_i$‡	$\sqrt{\hat{w}_i}$	Y_i	$\hat{Y}_i$	$\hat{w}_i$‡	$\sqrt{\hat{w}_i}$
0	−0.129*			1	1.301†		
1	0.688	0.2146	0.4633	1	0.688	0.2147	0.4633
1	0.893	0.0956	0.3091	0	0.280	0.2016	0.4990
0	0.178	0.1463	0.3825	0	0.178	0.1463	0.3825
0	0.280	0.2016	0.4490	1	0.688	0.2147	0.4633
1	0.995	0.00498	0.0705	0	0.178	0.1463	0.3825
1	1.098†			1	1.097†		
0	0.382	0.2361	0.4859	1	0.893	0.0956	0.3091
0	−0.0265*			0	0.178	0.1463	0.3825
0	0.076	0.0702	0.2650	0	0.076	0.0702	0.2650
1	0.791	0.1653	0.4066	1	0.791	0.1653	0.4055
1	0.893	0.0956	0.3091	0	0.382	0.2361	0.4859
0	0.484	0.2497	0.4997	1	1.199†		
1	1.097†			1	1.097†		
0	−0.333*			0	0.178	0.1463	0.3825
1	0.995	0.00498	0.0705	0	−0.129*		
1	0.688	0.2147	0.4633	1	0.791	0.1653	0.4066
0	0.076	0.0702	0.2650	1	0.688	0.2147	0.4633
0	−0.129*			0	−0.231*		
1	0.893	0.0956	0.3091	1	0.791	0.1653	0.4066

*Treated as zero to avoid probabilities being negative.
†Treated as unity to avoid probabilities exceeding one.
‡$\hat{Y}_i(1 - \hat{Y}_i)$.

observations, from 40 to 28 in the present example.[8] Omitting these observations, the WLS regression is

$$\frac{\hat{Y}_i}{\sqrt{\hat{w}_i}} = -1.2456\frac{1}{\sqrt{\hat{w}_i}} + 0.1196\frac{X_i}{\sqrt{\hat{w}_i}} \qquad \textbf{(15.2.11)}$$
$$(0.1206) \qquad (0.0069)$$
$$t = (-10.332) \qquad (17.454) \qquad R^2 = 0.9214$$

These results show that, compared with (15.12.10), the estimated standard errors are smaller and, correspondingly, the estimated *t* ratios (in absolute value) larger. But one should take this result with a grain of salt since in estimating (15.12.11) we had to drop 12 observations. Also, since w_i are estimated, the usual statistical hypothesis-testing procedures are, strictly speaking, valid in the large samples (see Chapter 11).

15.3 APPLICATIONS OF LPM

Until the availability of readily accessible computer packages to estimate the logit and probit models (to be discussed shortly), the LPM was used quite extensively because of its simplicity. We now illustrate some of these applications.

[8]To avoid the loss of the degrees of freedom, we could let $\hat{Y}_i = 0.01$ when the estimated Y_i are negative and $\hat{Y}_i = 0.99$ when they are in excess of or equal to 1. See exercise 15.1.

EXAMPLE 15.1

COHEN–REA–LERMAN STUDY[9]

In a study prepared for the U.S. Department of Labor, Cohen, Rea, and Lerman were interested in examining the labor-force participation of various categories of labor as a function of several socioeconomic–demographic variables. In all their regressions, the dependent variable was a dummy, taking a value of 1 if a person is in the labor force, 0 if he or she is not. In Table 15.3 we reproduce one of their several dummy-dependent variable regressions.

Before interpreting the results, note these features: The preceding regression was estimated by using the OLS. To correct for heteroscedasticity, the authors used the two-step procedure outlined previously in some of their regressions but found that the standard errors of the estimates thus obtained did not differ materially from those obtained without correction for heteroscedasticity. Perhaps this result is due to the sheer size of the sample, namely, about 25,000. Because of this large sample size, the estimated t values may be tested for statistical significance by the usual OLS procedure even though the error term takes dichotomous values. The estimated R^2 of 0.175 may seem rather low, but in view of the large sample size, this R^2 is still significant on the basis of the F test given in Section 8.5. Finally, notice how the authors have blended quantitative and qualitative variables and how they have taken into account the interaction effects.

Turning to the interpretations of the findings, we see that each slope coefficient gives the rate of change in the conditional probability of the event occurring for a given unit change in the value of the explanatory variable. For instance, the coefficient of −0.2753 attached to the variable "age 65 and over" means, holding all other factors constant, the probability of participation in the labor force by women in this age group is smaller by about 27 percent (as compared with the base category of women aged 22 to 54). By the same token, the coefficient of 0.3061 attached to the variable "16 or more years of schooling" means, holding all other factors constant, the probability of women with this much education participating in the labor force is higher by about 31 percent (as compared with women with less than 5 years of schooling, the base category).

Now consider the **interaction term** marital status and age. The table shows that the labor-force participation probability is higher by some 29 percent for those women who were never married (as compared with the base category) and smaller by about 28 percent for those women who are 65 and over (again in relation to the base category). But the probability of participation of women who were never married and are 65 or over is smaller by about 20 percent as compared with the base category. This implies that women aged 65 and over but never married are likely to participate in the labor force more than those who are aged 65 and over and are married or fall into the "other" category.

Following this procedure, the reader can easily interpret the rest of the coefficients given in Table 15.3. From the given information, it is easy to obtain the estimates of the conditional probabilities of labor-force participation of the various categories. Thus, if we want to find the probability for married women (other), aged 22 to 54, with 12 to 15 years of schooling, with an unemployment rate of 2.5 to 3.4 percent, employment change of 3.5 to 6.49 percent, relative employment opportunities of 74 percent and over, and with FILOW of $7500 and over, we obtain

$$0.4368 + 0.1523 + 0.2231 - 0.0213 + 0.0301 + 0.0571 - 0.2455 = 0.6326$$

In other words, the probability of labor-force participation by women with the preceding characteristics is estimated to be about 63 percent.

(*Continued*)

[9]Malcolm S. Cohen, Samuel A. Rea, Jr., and Robert I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, U.S. Department of Labor, 1970.

EXAMPLE 15.1 (*Continued*)

TABLE 15.3 LABOR-FORCE PARTICIPATION
Regression of women, age 22 and over, living in largest 96 standard metropolitan statistical areas (SMSA) (dependent variable: in or out of labor force during 1966)

Explanatory variable		Coefficient	*t* ratio
Constant		0.4368	15.4
Marital status			
Married, spouse present		—	—
Married, other		0.1523	13.8
Never married		0.2915	22.0
Age			
22–54		—	—
55–64		−0.0594	−5.7
65 and over		−0.2753	−9.0
Years of schooling			
0–4		—	—
5–8		0.1255	5.8
9–11		0.1704	7.9
12–15		0.2231	10.6
16 and over		0.3061	13.3
Unemployment rate (1966), %			
Under 2.5		—	—
2.5–3.4		−0.0213	−1.6
3.5–4.0		−0.0269	−2.0
4.1–5.0		−0.0291	−2.2
5.1 and over		−0.0311	−2.4
Employment change (1965–1966), %			
Under 3.5		—	—
3.5–6.49		0.0301	3.2
6.5 and over		0.0529	5.1
Relative employment opportunities, %			
Under 62		—	—
62–73.9		0.0381	3.2
74 and over		0.0571	3.2
FILOW, $			
Less than 1,500 and negative		—	—
1,500–7,499		−0.1451	−15.4
7,500 and over		−0.2455	−24.4
Interaction (marital status and age)			
Marital status	Age		
Other	55–64	−0.0406	−2.1
Other	65 and over	−0.1391	−7.4
Never married	55–64	−0.1104	−3.3
Never married	65 and over	−0.2045	−6.4
Interaction (age and years of schooling completed)			
Age	Years of schooling		
65 and over	5–8	−0.0885	−2.8
65 and over	9–11	−0.0848	−2.4
65 and over	12–15	−0.1288	−4.0
65 and over	16 and over	−0.1628	−3.6
	$R^2 = 0.175$		
No. of observations = 25,153			

Note: — indicates the base or omitted category.
FILOW: family income less own wage and salary income.
Source: Malcolm S. Cohen, Samuel A. Rea, Jr., and Robert I. Lerman, *A Micro Model of Labor Supply,* BLS Staff Paper 4, U.S. Department of Labor, 1970, Table F-6, pp. 212–213.

EXAMPLE 15.2

PREDICTING A BOND RATING

Based on a pooled time series and cross-sectional data of 200 Aa (high-quality) and Baa (medium-quality) bonds over the period 1961–1966, Joseph Cappelleri estimated the following bond rating prediction model.[10]

$$Y_i = \beta_1 + \beta_2 X_{2i}^2 + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$

where Y_i = 1 if the bond rating is Aa (Moody's rating)
= 0 if the bond rating is Baa (Moody's rating)
X_2 = debt capitalization ratio, a measure of leverage

$$= \frac{\text{dollar value of long-term debt}}{\text{dollar value of total capitalization}} \cdot 100$$

X_3 = profit rate

$$= \frac{\text{dollar value of after-tax income}}{\text{dollar value of net total assets}} \cdot 100$$

X_4 = standard deviation of the profit rate, a measure of profit rate variability
X_5 = net total assets (thousands of dollars), a measure of size

A priori, β_2 and β_4 are expected to be negative (why?) and β_3 and β_5 are expected to be positive.

After correcting for heteroscedasticity and first-order autocorrelation, Cappelleri obtained the following results[11]:

$$\begin{aligned}\hat{Y}_i = \ &0.6860 - 0.0179 X_{2i}^2 + 0.0486 X_{3i} + 0.0572 X_{4i} + 0.378(E\text{-}7) X_5 \\ &(0.1775) \quad (0.0024) \qquad (0.0486) \qquad (0.0178) \qquad (0.039)(E\text{-}8) \\ &R^2 = 0.6933\end{aligned} \tag{15.3.1}$$

Note: 0.378 E-7 means 0.0000000378, etc.

All but the coefficient of X_4 have the correct signs. It is left to finance students to rationalize why the profit rate variability coefficient has a positive sign, for one would expect that the greater the variability in profits, the less likely it is Moody's would give an Aa rating, other things remaining the same.

The interpretation of the regression is straightforward. For example, 0.0486 attached to X_3 means that, other things being the same, a 1 percentage point increase in the profit rate will lead on average to about a 0.05 increase in the probability of a bond getting the Aa rating. Similarly, the higher the squared leveraged ratio, the lower by 0.02 is the probability of a bond being classified as an Aa bond per unit increase in this ratio.

[10]Joseph Cappelleri, "Predicting a Bond Rating," unpublished term paper, C.U.N.Y. The model used in the paper is a modification of the model used by Thomas F. Pogue and Robert M. Soldofsky, "What Is in a Bond Rating?" *Journal of Financial and Quantitative Analysis*, June 1969, pp. 201–228.

[11]Some of the estimated probabilities before correcting for heteroscedasticity were negative and some were in excess of 1; in these cases they were assumed to be 0.01 and 0.99, respectively, to facilitate the computation of the weights w_i.

EXAMPLE 15.3

PREDICTING BOND DEFAULTS

To predict the probability of default on their bond obligations, Daniel Rubinfeld studied a sample of 35 municipalities in Massachusetts for the year 1930, several of which did in fact default. The LPM model he chose and estimated was as follows[12]:

$$\hat{P} = \underset{(0.29)}{1.96} - \underset{(0.009)}{0.029}\,\text{TAX} - \underset{(2.13)}{4.86}\,\text{INT} + \underset{(0.028)}{0.063}\,\text{AV} + \underset{(0.003)}{0.007}\,\text{DAV} - \underset{(0.88)}{0.48}\,\text{WELF} \tag{15.3.2}$$

$$R^2 = 0.36$$

Where $P = 0$ if the municipality defaulted and 1 otherwise, TAX = average of 1929, 1930, and 1931 tax rates: INT = percentage of current budget allocated to interest payments in 1930; AV = percentage growth in assessed property valuation from 1925 to 1930; DAV = ratio of total direct net debt to total assessed valuation in 1930; and WELF = percentage of 1930 budget allocated to charities, pensions, and soldiers' benefits.

The interpretation (15.3.2) is again fairly straightforward. Thus, other things being the same, an increase in the tax rate of $1 per thousand will raise the probability of default by about 0.03, or 3 percent. The R^2 value is rather low but, as noted previously, in LPMs the R^2 values generally tend to be lower and are of limited use in judging the goodness of fit of the model.

15.4 ALTERNATIVES TO LPM

As we have seen, the LPM is plagued by several problems, such as (1) non-normality of u_i, (2) heteroscedasticity of u_i, (3) possibility of $\hat{Y}_i$ lying outside the 0–1 range, and (4) the generally lower R^2 values. But these problems are surmountable. For example, we can use WLS to resolve the heteroscedasticity problem or increase the sample size to minimize the non-normality problem. By resorting to restricted least-squares or mathematical programming techniques we can even make the estimated probabilities lie in the 0–1 interval.

But even then the fundamental problem with the LPM is that it is not logically a very attractive model because it assumes that $P_i = E(Y = 1 \mid X)$ increases linearly with X, that is, the marginal or incremental effect of X remains constant throughout. Thus, in our home ownership example we found that as X increases by a unit ($1000), the probability of owning a house increases by the same constant amount of 0.10. This is so whether the income level is $8000, $10,000, $18,000, or $22,000. This seems patently unrealistic. In reality one would expect that P_i is nonlinearly related to X_i:

[12] D. Rubinfeld, "An Econometric Analysis of the Market for General Municipal Bonds," unpublished doctoral dissertation, Massachusetts Institute of Technology, 1972. The results given in this example are reproduced from Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 2d ed., McGraw-Hill, New York, 1981, p. 279.

At very low income a family will not own a house but at a sufficiently high level of income, say, X^*, it most likely will own a house. Any increase in income beyond X^* will have little effect on the probability of owning a house. Thus, at both ends of the income distribution, the probability of owning a house will be virtually unaffected by a small increase in X.

Therefore, what we need is a (probability) model that has these two features: (1) As X_i increases, $P_i = E(Y = 1 \mid X)$ increases but never steps outside the 0–1 interval, and (2) the relationship between P_i and X_i is nonlinear, that is, "one which approaches zero at slower and slower rates as X_i gets small and approaches one at slower and slower rates as X_i gets very large."[13]

Geometrically, the model we want would look something like Figure 15.2. Notice in this model that the probability lies between 0 and 1 and that it varies nonlinearly with X.

The reader will realize that the sigmoid, or S-shaped, curve in the figure very much resembles the **cumulative distribution function** (CDF) of a random variable.[14] Therefore, one can easily use the CDF to model regressions where the response variable is dichotomous, taking 0–1 values. The practical question now is, which CDF? For although all CDFs are S shaped, for each random variable there is a unique CDF. For historical as well as practical reasons, the CDFs commonly chosen to represent the 0–1 response

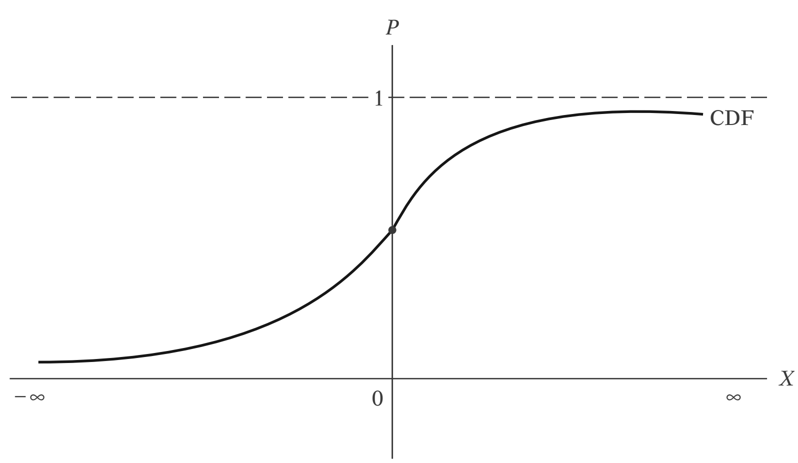


FIGURE 15.2 A cumulative distribution function (CDF).

[13]John Aldrich and Forrest Nelson, op. cit., p. 26.

[14]As discussed in **App. A**, the CDF of a random variable X is simply the probability that it takes a value less than or equal to x_0, where x_0 is some specified numerical value of X. In short, $F(X)$, the CDF of X, is $F(X = x_0) = P(X \le x_0)$.

models are (1) the logistic and (2) the normal, the former giving rise to the **logit** model and the latter to the **probit** (or **normit**) model.

Although a detailed discussion of the logit and probit models is beyond the scope of this book, we will indicate somewhat informally how one estimates such models and how one interprets them.

15.5 THE LOGIT MODEL

We will continue with our home ownership example to explain the basic ideas underlying the logit model. Recall that in explaining home ownership in relation to income, the LPM was

$$P_i = E(Y = 1 \mid X_i) = \beta_1 + \beta_2 X_i \tag{15.5.1}$$

where X is income and $Y = 1$ means the family owns a house. But now consider the following representation of home ownership:

$$P_i = E(Y = 1 \mid X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} \tag{15.5.2}$$

For ease of exposition, we write (15.5.2) as

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^z}{1 + e^z} \tag{15.5.3}$$

where $Z_i = \beta_1 + \beta_2 X_i$.

Equation (15.5.3) represents what is known as the (cumulative) **logistic distribution function.**[15]

It is easy to verify that as Z_i ranges from $-\infty$ to $+\infty$, P_i ranges between 0 and 1 and that P_i is nonlinearly related to Z_i (i.e., X_i), thus satisfying the two requirements considered earlier.[16] But it seems that in satisfying these requirements, we have created an estimation problem because P_i is nonlinear not only in X but also in the β's as can be seen clearly from (15.5.2). This means that we cannot use the familiar OLS procedure to estimate the parameters.[17] But this problem is more apparent than real because (15.5.2) can be linearized, which can be shown as follows.

---

[15]The logistic model has been used extensively in analyzing growth phenomena, such as population, GNP, money supply, etc. For theoretical and practical details of logit and probit models, see J. S. Kramer, *The Logit Model for Economists*, Edward Arnold Publishers, London, 1991; and G. S. Maddala, op. cit.

[16]Note that as $Z_i \to +\infty$, e^{-Z_i} tends to zero and as $Z_i \to -\infty$, e^{-Z_i} increases indefinitely. Recall that $e = 2.71828$.

[17]Of course, one could use nonlinear estimation techniques discussed in Chap. 14. See also Sec. 15.8.

If P_i, the probability of owning a house, is given by (15.5.3), then $(1 - P_i)$, the probability of not owning a house, is

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \tag{15.5.4}$$

Therefore, we can write

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \tag{15.5.5}$$

Now $P_i/(1 - P_i)$ is simply the **odds ratio** in favor of owning a house—the ratio of the probability that a family will own a house to the probability that it will not own a house. Thus, if $P_i = 0.8$, it means that odds are 4 to 1 in favor of the family owning a house.

Now if we take the natural log of (15.5.5), we obtain a very interesting result, namely,

$$\begin{aligned} L_i = \ln\left(\frac{P_i}{1 - P_i}\right) &= Z_i \\ &= \beta_1 + \beta_2 X_i \end{aligned} \tag{15.5.6}$$

that is, L, the log of the odds ratio, is not only linear in X, but also (from the estimation viewpoint) linear in the parameters.[18] L is called the **logit,** and hence the name **logit model** for models like (15.5.6).

Notice these features of the logit model.

1. As P goes from 0 to 1 (i.e., as Z varies from $-\infty$ to $+\infty$), the logit L goes from $-\infty$ to $+\infty$. That is, although the probabilities (of necessity) lie between 0 and 1, the logits are not so bounded.

2. Although L is linear in X, the probabilities themselves are not. This property is in contrast with the LPM model (15.5.1) where the probabilities increase linearly with X.[19]

3. Although we have included only a single X variable, or regressor, in the preceding model, one can add as many regressors as may be dictated by the underlying theory.

4. If L, the logit, is positive, it means that when the value of the regressor(s) increases, the odds that the regressand equals 1 (meaning some event of interest happens) increases. If L is negative, the odds that the regressand equals 1 decreases as the value of X increases. To put it differently, the logit

[18] Recall that the linearity assumption of OLS does not require that the X variable be necessarily linear. So we can have X^2, X^3, etc., as regressors in the model. For our purpose, it is linearity in the parameters that is crucial.

[19] Using calculus, it can be shown that $dP/dX = \beta_2 P(1 - P)$, which shows that the rate of change in probability with respect to X involves not only β_2 but also the level of probability from which the change is measured (but more on this in Sec. 15.7). In passing, note that the effect of a unit change in X_i on P is greatest when $P = 0.5$ and least when P is close to 0 or 1.

becomes negative and increasingly large in magnitude as the odds ratio decreases from 1 to 0 and becomes increasingly large and positive as the odds ratio increases from 1 to infinity.[20]

5. More formally, the interpretation of the logit model given in (15.5.6) is as follows: β_2, the slope, measures the change in L for a unit change in X, that is, it tells how the log-odds in favor of owning a house change as income changes by a unit, say, $1000. The intercept β_1 is the value of the log-odds in favor of owning a house if income is zero. Like most interpretations of intercepts, this interpretation may not have any physical meaning.

6. Given a certain level of income, say, X^*, if we actually want to estimate not the odds in favor of owning a house but the probability of owning a house itself, this can be done directly from (15.5.3) once the estimates of $\beta_1 + \beta_2$ are available. This, however, raises the most important question: How do we estimate β_1 and β_2 in the first place? The answer is given in the next section.

7. Whereas the LPM assumes that P_i is linearly related to X_i, the logit model assumes that the log of the odds ratio is linearly related to X_i.

15.6 ESTIMATION OF THE LOGIT MODEL

For estimation purposes, we write (15.5.6) as follows:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(15.6.1)}$$

We will discuss the properties of the stochastic error term u_i shortly.

To estimate (15.6.1), we need, apart from X_i, the values of the regressand, or logit, L_i. This depends on the type of data we have for analysis. We distinguish two types of data: (1) *data at the individual, or micro, level,* and (2) *grouped or replicated data*.

Data at the Individual Level

If we have data on individual families, as in the case of Table 15.1, OLS estimation of (15.6.1) is infeasible. This is easy to see. In terms of the data given in Table 15.1, $P_i = 1$ if a family owns a house and $P_i = 0$ if it does not own a house. But if we put these values directly into the logit L_i, we obtain:

$$L_i = \ln\left(\frac{1}{0}\right) \qquad \text{if a family own a house}$$

$$L_i = \ln\left(\frac{0}{1}\right) \qquad \text{if a family does not own a house}$$

Obviously, these expressions are meaningless. Therefore, if we have data at the micro, or individual, level, we cannot estimate (15.6.1) by the standard

[20]This point is due to David Garson.

OLS routine. In this situation we may have to resort to the **maximum-likelihood (ML)** method to estimate the parameters. Although the rudiments of this method were discussed in the appendix to Chapter 4, its application in the present context will be discussed in Appendix 15A, Section 15A.1, for the benefit of readers who would like to learn more about it.[21] Software packages, such as Microfit, Eviews, Limdep, Shazam, PcGive, and Minitab, have built-in routines to estimate the logit model at the individual level. We will illustrate the use of the ML method later in the chapter.

Grouped or Replicated Data

Now consider the data given in Table 15.4. This table gives data on several families *grouped* or *replicated* (repeat observations) according to income level and the number of families owning a house at each income level. Corresponding to each income level X_i, there are N_i families, n_i among whom are home owners ($n_i \le N_i$). Therefore, if we compute

$$\hat{P}_i = \frac{n_i}{N_i} \tag{15.6.2}$$

that is, the *relative frequency*, we can use it as an estimate of the true P_i corresponding to each X_i. If N_i is fairly large, $\hat{P}_i$ will be a reasonably good estimate of P_i.[22] Using the estimated P_i, we can obtain the estimated logit as

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{15.6.3}$$

TABLE 15.4 HYPOTHETICAL DATA ON X_i (INCOME), N_i (NUMBER OF FAMILIES AT INCOME X_i), AND n_i (NUMBER OF FAMILIES OWNING A HOUSE)

X (thousands of dollars)	N_i	n_i
6	40	8
8	50	12
10	60	18
13	80	28
15	100	45
20	70	36
25	65	39
30	50	33
35	40	30
40	25	20

[21] For a comparatively simple discussion of maximum likelihood in the context of the logit model, see John Aldrich and Forrest Nelson, op. cit., pp. 49–54. See also, Alfred Demarsi, *Logit Modeling: Practical Applications,* Sage Publications, Newbury Park, Calif., 1992.

[22] From elementary statistics recall that the probability of an event is the limit of the relative frequency as the sample size becomes infinitely large.

which will be a fairly good estimate of the true logit L_i if the number of observations N_i at each X_i is reasonably large.

In short, given the *grouped* or *replicated* data, such as Table 15.4, one can obtain the data on the dependent variable, the logits, to estimate the model (15.6.1). Can we then apply OLS to (15.6.3) and estimate the parameters in the usual fashion? The answer is, not quite, since we have not yet said anything about the properties of the stochastic disturbance term. It can be shown that if N_i is fairly large and if each observation in a given income class X_i is distributed independently as a binomial variable, then

$$u_i \sim N\left[0, \frac{1}{N_i P_i(1 - P_i)}\right] \tag{15.6.4}$$

that is u_i follows the normal distribution with zero mean and variance equal to $1/[N_i P_i(1 - P_i)]$.[23]

Therefore, as in the case of the LPM, the disturbance term in the logit model is heteroscedastic. Thus, instead of using OLS we will have to use the weighted least squares (WLS). For empirical purposes, however, we will replace the unknown P_i by $\hat{P}_i$ and use

$$\hat{\sigma}^2 = \frac{1}{N_i \hat{P}_i(1 - \hat{P}_i)} \tag{15.6.5}$$

as estimator of σ^2.

We now describe the various steps in estimating the logit regression (15.6.1):

1. For each income level X, compute the probability of owning a house as $\hat{P}_i = n_i/N_i$.

2. For each X_i, obtain the logit as[24]

$$\hat{L}_i = \ln[\hat{P}_i/(1 - \hat{P}_i)]$$

3. To resolve the problem of heteroscedasticity, transform (15.6.1) as follows[25]:

$$\sqrt{w_i} L_i = \beta_1\sqrt{w_i} + \beta_2\sqrt{w_i} X_i + \sqrt{w_i} u_i \tag{15.6.6}$$

---

[23]As shown in elementary probability theory, $\hat{P}_i$, the proportion of successes (here, owning a house), follows the binomial distribution with mean equal to true P_i and variance equal to $P_i(1 - P_i)/N_i$; and as N_i increases indefinitely the binomial distribution approximates the normal distribution. The distributional properties of u_i given in (15.6.4) follow from this basic theory. For details, see Henry Theil, "On the Relationships Involving Qualitative Variables," *American Journal of Sociology*, vol. 76, July 1970, pp. 103–154.

[24]Since $\hat{P}_i = n_i/N_i$, L_i can be alternatively expressed as $\hat{L}_i = \ln n_i/(N_i - n_i)$. In passing it should be noted that to avoid $\hat{P}_i$ taking the value of 0 or 1, in practice $\hat{L}_i$ is measured as $\hat{L}_i = \ln(n_i + \frac{1}{2})/(N_i - n_i + \frac{1}{2}) = \ln(\hat{P}_i + 1/2N_i)/(1 - \hat{P}_i + 1/2N_i)$. It is recommended as a rule of thumb that N_i be at least 5 at each value of X_i. For additional details, see D. R. Cox, *Analysis of Binary Data*, Methuen, London, 1970, p. 33.

[25]If we estimate (15.6.1) disregarding heteroscedasticity, the estimators, although unbiased, will not be efficient, as we know from Chap. 11.

which we write as

$$L_i^* = \beta_1\sqrt{w_i} + \beta_2 X_i^* + v_i \qquad \textbf{(15.6.7)}$$

where the weights $w_i = N_i\hat{P}_i(1-\hat{P}_i)$; L_i^* = transformed or weighted L_i; X_i^* = transformed or weighted X_i; and v_i = transformed error term. It is easy to verify that the transformed error term v_i is homoscedastic, keeping in mind that the original error variance is $\sigma_u^2 = 1/[N_i P_i(1-P_i)]$.

4. Estimate (15.6.6) by OLS—recall that WLS is OLS on the transformed data. Notice that in (15.6.6) there is no intercept term introduced explicitly (why?). Therefore, one will have to use the regression through the origin routine to estimate (15.6.6).

5. Establish confidence intervals and/or test hypotheses in the usual OLS framework, *but keep in mind that all the conclusions will be valid strictly speaking if the sample is reasonably large* (why?). Therefore, in small samples, the estimated results should be interpreted carefully.

15.7 THE GROUPED LOGIT (GLOGIT) MODEL: A NUMERICAL EXAMPLE

To illustrate the theory just discussed, we will use the data given in Table 15.4. Since the data in the table are grouped, the logit model based on this data will be called a grouped logit model, glogit, for short. The necessary raw data and other relevant calculations necessary to implement glogit are given in Table 15.5. The results of the weighted least-squares regression (15.6.7) based on the data given in Table 15.5 are as follows: Note that there is no intercept in (15.6.7); hence the regression-through-origin procedure is appropriate here.

$$\begin{aligned}
\hat{L}_i^* &= \quad -1.59474\sqrt{w_i} + \quad 0.07862X_i^* \\
\text{se} &= \quad (0.11046) \qquad\qquad (0.00539) \\
t &= (-14.43619) \qquad (14.56675) \qquad R^2 = 0.9642
\end{aligned} \qquad \textbf{(15.7.1)}$$

The R^2 is the squared correlation coefficient between actual and estimated L_i^*. L_i^* and X_i^* are weighted L_i and X_i, as shown in (15.6.6).

Interpretation of the Estimated Logit Model

How do we interpret (15.7.1)? There are various ways, some intuitive and some not:

Logit Interpretation. As (15.7.1) shows, the estimated slope coefficient suggests that for a unit ($1000) increase in weighted income, the weighted log of the odds in favor of owning a house goes up by 0.08 units. This mechanical interpretation, however, is not very appealing.

TABLE 15.5 DATA TO ESTIMATE THE LOGIT MODEL OF OWNERSHIP

X (thousands of dollars) (1)	N_i (2)	n_i (3)	$\hat{P}_i$ (4) = (3) ÷ (2)	$1-\hat{P}_i$ (5)	$\frac{\hat{P}_i}{1-\hat{P}_i}$ (6)	$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)$ (7)	$N_i\hat{P}_i(1-\hat{P}_i) = w_i$ (8)	$\sqrt{w_i} = \sqrt{N_i\hat{P}_i(1-\hat{P}_i)}$ (9) = $\sqrt{(8)}$	$\hat{L}_i^* = \hat{L}_i\sqrt{w_i}$ (10) = (7)(9)	$\hat{X}_i^* = X_i\sqrt{w_i}$ (11) = (1)(9)
6	40	8	0.20	0.80	0.25	−1.3863	6.40	2.5298	−3.5071	15.1788
8	50	12	0.24	0.76	0.32	−1.1526	9.12	3.0199	−3.4807	24.1592
10	60	18	0.30	0.70	0.43	−0.8472	12.60	3.5496	−3.0072	35.4960
13	80	28	0.35	0.65	0.54	−0.6190	18.20	4.2661	−2.6407	55.4593
15	100	45	0.45	0.55	0.82	−0.2007	24.75	4.9749	−0.9985	74.6235
20	70	36	0.51	0.49	1.04	0.0400	17.49	4.1825	0.1673	83.6506
25	65	39	0.60	0.40	1.50	0.4054	15.60	3.9497	1.6012	98.7425
30	50	33	0.66	0.34	1.94	0.6633	11.20	3.3496	2.2218	100.4880
35	40	30	0.75	0.25	3.0	1.0986	7.50	2.7386	3.0086	95.8405
40	25	20	0.80	0.20	4.0	1.3863	4.00	2.000	2.7726	80.0000

Odds Interpretation. Remember that $L_i = \ln[P_i/(1 - P_i)]$. Therefore, taking the antilog of the estimated logit, we get $P_i/(1 - P_i)$, that is, the odds ratio. Hence, taking the antilog of (15.7.1), we obtain:

$$\frac{\hat{P}_i}{1-\hat{P}_i} = e^{-1.59474\sqrt{w_i}+0.07862X_i^*} = e^{-1.59474\sqrt{w_i}} \cdot e^{0.07862X_i^*} \tag{15.7.2}$$

Using a calculator, you can easily verify that $e^{0.07862} = 1.0817$. This means that for a unit increase in weighted income, the (weighted) odds in favor of owing a house increases by 1.0817 or about 8.17%. *In general, if you take the antilog of the jth slope coefficient (in case there is more than one regressor in the model), subtract 1 from it, and multiply the result by 100, you will get the percent change in the odds for a unit increase in the jth regressor.*

Incidentally, if you want to carry the analysis in terms of unweighted logit, all you have to do is to divide the estimated L_i^* by $\sqrt{w_i}$. Table 15.6 gives the estimated weighted and unweighted logits for each observation and some other data, which we will discuss shortly.

Computing Probabilities. Since the language of logit and odds ratio may be unfamiliar to some, we can always compute the probability of owning a house at a certain level of income. Suppose we want to compute this probability at $X = 20$ ($20,000). Plugging this value in (15.7.1), we obtain: $\hat{L}_i^* = -0.09311$ and dividing this by $\sqrt{w_i} = 4.2661$ (see Table 15.5), we obtain $\hat{L}_i = -0.02226$. Therefore, at the income level of $20,000, we have

$$-0.02226 = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)$$

TABLE 15.6 LSTAR, XSTAR, ESTIMATED LSTAR, PROBABILITY, AND CHANGE IN PROBABILITY*

Lstar	Xstar	ELstar	Logit	Probability, $\hat{P}$	Change in probability†
−3.50710	15.1788	−2.84096	−1.12299	0.24545	0.01456
−3.48070	24.15920	−2.91648	−0.96575	0.27572	0.01570
−3.48070	35.49600	−2.86988	−0.80850	0.30821	0.01676
−2.64070	55.45930	−2.44293	−0.57263	0.36063	0.01813
−0.99850	74.62350	−2.06652	−0.41538	0.39762	0.01883
0.16730	83.65060	−0.09311	−0.02226	0.49443	0.01965
1.60120	98.74250	1.46472	0.37984	0.59166	0.01899
2.22118	100.48800	2.55896	0.76396	0.68221	0.01704
3.00860	95.84050	3.16794	1.15677	0.76074	0.01431
2.77260	80.00000	3.10038	1.55019	0.82494	0.01135

*Lstar and Xstar are from Table 15.5. ELstar is the estimated Lstar. Logit is the unweighted logit. Probability is the estimated probability of owning a house. Change in probability is the change per unit change in income.

†Computed from $\hat{\beta}_2\hat{P}(1-\hat{P}) = 0.07862\hat{P}(1-\hat{P})$.

Therefore,

$$\frac{\hat{P}}{1-\hat{P}_i} = e^{-0.02226} = 1.0225$$

Solving this for

$$\hat{P}_i = \frac{e^{-0.02226}}{1+e^{-0.02226}}$$

the reader can see that the estimated probability is 0.4944. That is, given the income of $20,000, the probability of a family owning a house is about 49 percent. Table 15.6 shows the probabilities thus computed at various income levels. As this table shows, the probability of house ownership increases with income, but not linearly as with the LPM model.

Computing the Rate of Change of Probability. As you can gather from Table 15.6, the probability of owning a house depends on the income level. How can we compute the rate of change of probabilities as income varies? As noted in footnote 19, that depends not only on the estimated slope coefficient β_2 but also on the level of the probability from which the change is measured; the latter of course depends on the income level at which the probability is computed.

To illustrate, suppose we want to measure the change in the probability of owning a house at the income level $20,000. Then, from footnote 19 the change in probability for a unit increase in income from the level 20 (thousand) is: $\hat{\beta}(1-\hat{P})\hat{P} = 0.07862(0.5056)(0.4944) = 0.01965$.

It is left as an exercise for the reader to show that at income level $40,000, the change in probability is 0.01135. Table 15.6 shows the change in probability of owning a house at various income levels; these probabilities are also depicted in Figure 15.3.

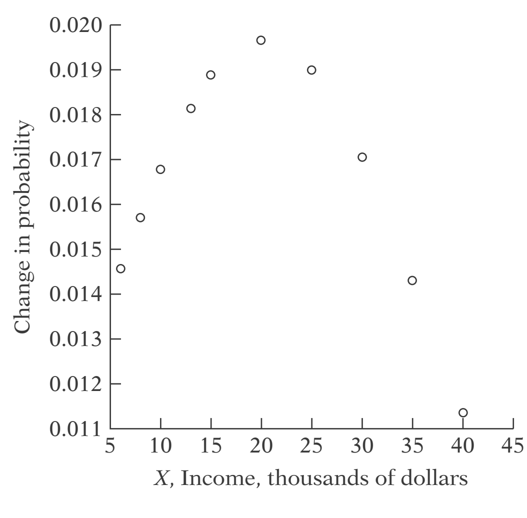


FIGURE 15.3 Change in probability in relation to income.

To conclude our discussion of the glogit model, we present the results based on OLS, or unweighted regression, for the home ownership example:

$$\begin{aligned} \hat{L}_i &= \quad -1.6587 + 0.0792X_i \\ \text{se} &= \quad (0.0958) \quad (0.0041) \\ t &= (-17.32) \quad (19.11) \qquad r^2 = 0.9786 \end{aligned} \tag{15.7.3}$$

We leave it to the reader to compare this regression with the weighted least-squares regression given by (15.7.1).

15.8 THE LOGIT MODEL FOR UNGROUPED OR INDIVIDUAL DATA

To set the stage, consider the data given in Table 15.7. Letting $Y = 1$ if a student's final grade in an intermediate microeconomics course was A and $Y = 0$ if the final grade was a B or a C, Spector and Mazzeo used grade point average (GPA), TUCE, and Personalized System of Instruction (PSI) as the

TABLE 15.7 DATA ON THE EFFECT OF PERSONALIZED SYSTEM OF INSTRUCTION (PSI) ON COURSE GRADES

Observation	GPA grade	TUCE grade	PSI	Grade	Letter grade	Observation	GPA grade	TUCE grade	PSI	Grade	Letter grade
1	2.66	20	0	0	C	17	2.75	25	0	0	C
2	2.89	22	0	0	B	18	2.83	19	0	0	C
3	3.28	24	0	0	B	19	3.12	23	1	0	B
4	2.92	12	0	0	B	20	3.16	25	1	1	A
5	4.00	21	0	1	A	21	2.06	22	1	0	C
6	2.86	17	0	0	B	22	3.62	28	1	1	A
7	2.76	17	0	0	B	23	2.89	14	1	0	C
8	2.87	21	0	0	B	24	3.51	26	1	0	B
9	3.03	25	0	0	C	25	3.54	24	1	1	A
10	3.92	29	0	1	A	26	2.83	27	1	1	A
11	2.63	20	0	0	C	27	3.39	17	1	1	A
12	3.32	23	0	0	B	28	2.67	24	1	0	B
13	3.57	23	0	0	B	29	3.65	21	1	1	A
14	3.26	25	0	1	A	30	4.00	23	1	1	A
15	3.53	26	0	0	B	31	3.10	21	1	0	C
16	2.74	19	0	0	B	32	2.39	19	1	1	A

Notes: Grade $Y = 1$ if the final grade is A
$= 0$ if the final grade is B or C
TUCE = score on an examination given at the beginning of the term to test entering knowledge of macroeconomics
PSI = 1 if the new teaching method is used
= 0 otherwise
GPA = the entering grade point average

Source: L. Spector and M. Mazzeo, "Probit Analysis and Economic Education," *Journal of Economic Education,* vol. 11, 1980, pp. 37–44.

grade predictors. The logit model here can be written as:

$$L_i = \left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 \text{GPA}_i + \beta_3 \text{TUCE}_i + \beta_4 \text{PSI}_i + u_i \qquad \textbf{(15.8.1)}$$

As we noted in Section 15.6, we cannot simply put $P_i = 1$ if a family owns a house, and zero if it does not own a house. Here neither OLS nor weighted least squares (WLS) is helpful. We have to resort to nonlinear estimating procedures using the method of maximum likelihood. The details of this method are given in Appendix 15A, Section 15A.1. Since most modern statistical packages have routines to estimate logit models on the basis of ungrouped data, we will present the results of model (15.8.1) using the data given in Table 15.7 and show how to interpret the results. The results are given in Table 15.8 in tabular form and are obtained by using Eviews 4. Before interpreting these results, some general observations are in order.

1. Since we are using the method of maximum likelihood, which is generally a large-sample method, the estimated standard errors are *asymptotic*.

2. As a result, instead of using the t statistic to evaluate the statistical significance of a coefficient, we use the (standard normal) Z statistic. So inferences are based on the normal table. Recall that if the sample size is reasonably large, the t distribution converges to the normal distribution.

3. As noted earlier, the conventional measure of goodness of fit, R^2, is not particularly meaningful in binary regressand models. Measures similar to R^2, called **pseudo R^2,** are available, and there are a variety of them.[26] Eviews presents one such measure, the McFadden R^2, denoted by R^2_{McF}, whose

TABLE 15.8 REGRESSION RESULTS OF (15.8.1)

Dependent Variable: Grade
Method: ML-Binary Logit
Convergence achieved after 5 iterations

Variable	Coefficient	Std. error	Z statistic	Probability
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255

McFadden R^2 = 0.3740   LR statistic (3 df) = 15.40419

[26]For an accessible discussion, see J. Scott Long, *Regression Models for Categorical and Limited Dependent Variables,* Sage Publications, Newbury Park, California, 1997, pp. 102–113.

value in our example is 0.3740.[27] Like R^2, R^2_{McF} also ranges between 0 and 1. Another comparatively simple measure of goodness of fit is the **count R^2,** which is defined as:

$$\textbf{Count } \boldsymbol{R^2} = \frac{\text{number of correct predictions}}{\text{total number of observations}} \qquad \textbf{(15.8.2)}$$

Since the regressand in the logit model takes a value of 1 or zero, if the predicted probability is greater than 0.5, we classify that as 1, but if it is less than 0.5, we classify that as 0. We then count the number of correct predictions and compute the R^2 as given in (15.7.2). We will illustrate this shortly.

It should be noted, however, that in binary regressand models, goodness of fit is of secondary importance. What matters is the expected signs of the regression coefficients and their statistical and/or practical significance.

4. To test the null hypothesis that all the slope coefficients are simultaneously equal to zero, the equivalent of the F test in the linear regression model is the **likelihood ratio (LR) statistic.** Given the null hypothesis, the LR statistic follows the χ^2 distribution with df equal to the number of explanatory variables, three in the present example. (*Note:* Exclude the intercept term in computing the df).

Now let us interpret the regression results given in (15.8.1). Each slope coefficient in this equation is a *partial slope* coefficient and measures the change in the estimated logit for a unit change in the value of the given regressor (holding other regressors constant). Thus, the GPA coefficient of 2.8261 means, with other variables held constant, that if GPA increases by a unit, on average the estimated logit increases by about 2.83 units, suggesting a positive relationship between the two. As you can see, all the other regressors have a positive effect on the logit, although statistically the effect of TUCE is not significant. However, together all the regressors have a significant impact on the final grade, as the LR statistic is 15.40, whose p value is about 0.0015, which is very small.

As noted previously, a more meaningful interpretation is in terms of odds, which are obtained by taking the antilog of the various slope coefficients. Thus, if you take the antilog of the PSI coefficient of 2.3786 you will get 10.7897 ($\approx e^{2.3786}$). This suggests that students who are exposed to the new method of teaching are more than 10 times likely to get an A than students who are not exposed to it, other things remaining the same.

Suppose we want to compute the actual probability of a student getting an A grade. Consider student number 10 in Table 15.7. Putting the actual data for this student in the estimated logit model given in Table 15.8, the reader can check that the estimated logit value for this student is 0.8178.

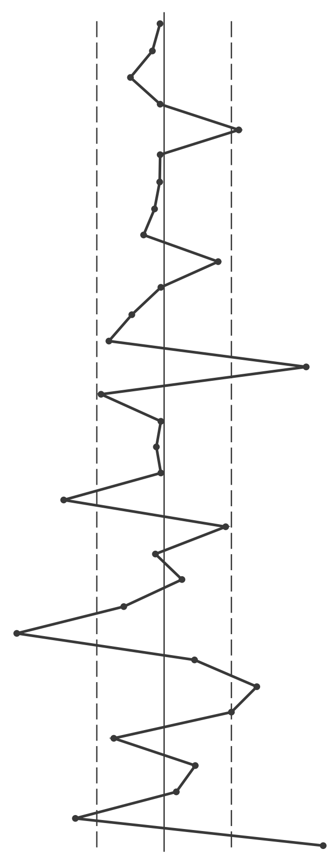
[27]Technically, this is defined as: $1 - (\text{LLF}_{\text{ur}}/\text{LLF}_r)$, where LLF_{ur} is the unrestricted log likelihood function where all regressors are included in the model and LLF_r is the restricted log likelihood function where only the intercept is included in the model. Conceptually, LLF_{ur} is equivalent to RSS and LLF_r is the equivalent to TSS of the linear regression model.

Using Eq. (15.5.2), the reader can easily check that the estimated probability is 0.69351. Since this student's actual final grade was an A, and since our logit model assigns a probability of 1 to a student who gets an A, the estimated probability of 0.69351 is not exactly 1 but close to it.

Recall the count R^2 defined earlier. Table 15.9 gives you the actual and predicted values of the regressand for our illustrative example. From this table you can observe that, out of 32 observations, there were 6 incorrect predictions (students 14, 19, 24, 26, 31, and 32). Hence the count R^2 value is $26/32 = 0.8125$, whereas the McFadden R^2 value is 0.3740. Although these two values are not directly comparable, they give you some idea about the orders of magnitude. Besides, one should not overplay the importance of goodness of fit in models where the regressand is dichotomous.

TABLE 15.9 ACTUAL AND FITTED VALUES BASED ON REGRESSION IN TABLE 15.8

Observation	Actual	Fitted	Residual	Residual plot
1	0	0.02658	−0.02658	
2	0	0.05950	−0.05950	
3	0	0.18726	−0.18726	
4	0	0.02590	−0.02590	
5	1	0.56989	0.43011	
6	0	0.03486	−0.03486	
7	0	0.02650	−0.02650	
8	0	0.05156	−0.05156	
9	0	0.11113	−0.11113	
10	1	0.69351	0.30649	
11	0	0.02447	−0.02447	
12	0	0.19000	−0.19000	
13	0	0.32224	−0.32224	
*14	1	0.19321	0.80679	
15	0	0.36099	−0.36099	
16	0	0.03018	−0.03018	
17	0	0.05363	−0.05363	
18	0	0.03859	−0.03859	
*19	0	0.58987	−0.58987	
20	1	0.66079	0.33921	
21	0	0.06138	−0.06138	
22	1	0.90485	0.09515	
23	0	0.24177	−0.24177	
*24	0	0.85209	−0.85209	
25	1	0.83829	0.16171	
*26	1	0.48113	0.51887	
27	1	0.63542	0.36458	
28	0	0.30722	−0.30722	
29	1	0.84170	0.15830	
30	1	0.94534	0.05466	
*31	0	0.52912	−0.52912	
*32	1	0.11103	0.88897	



*Incorrect predictions.

15.9 THE PROBIT MODEL

As we have noted, to explain the behavior of a dichotomous dependent variable we will have to use a suitably chosen CDF. The logit model uses the cumulative logistic function, as shown in (15.5.2). But this is not the only CDF that one can use. In some applications, the normal CDF has been found useful. The estimating model that emerges from the normal CDF[28] is popularly known as the **probit model,** although sometimes it is also known as the **normit model.** In principle one could substitute the normal CDF in place of the logistic CDF in (15.5.2) and proceed as in Section 16.5. Instead of following this route, we will present the probit model based on utility theory, or rational choice perspective on behavior, as developed by McFadden.[29]

To motivate the probit model, assume that in our home ownership example the decision of the ith family to own a house or not depends on an *unobservable* **utility index** I_i (also known as a **latent variable**), that is determined by one or more explanatory variables, say income X_i, in such a way that the larger the value of the index I_i, the greater the probability of a family owning a house. We express the index I_i as

$$I_i = \beta_1 + \beta_2 X_i \qquad \textbf{(15.9.1)}$$

where X_i is the income of the ith family.

How is the (unobservable) index related to the actual decision to own a house? As before, let $Y = 1$ if the family owns a house and $Y = 0$ if it does not. Now it is reasonable to assume that there is a **critical** or **threshold level** of the index, call it I_i^*, such that if I_i exceeds I_i^*, the family will own a house, otherwise it will not. The threshold I_i^*, like I_i, is not observable, but if we assume that it is normally distributed with the same mean and variance, it is possible not only to estimate the parameters of the index given in (15.9.1) but also to get some information about the unobservable index itself. This calculation is as follows.

Given the assumption of normality, the probability that I_i^* is less than or equal to I_i can be computed from the standardized normal CDF as[30]:

$$P_i = P(Y = 1 \mid X) = P(I_i^* \le I_i) = P(Z_i \le \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i) \qquad \textbf{(15.9.2)}$$

---

[28]See **App. A** for a discussion of the normal CDF. Briefly, if a variable X follows the normal distribution with mean μ and variance σ^2, its PDF is

$$f(X) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

and its CDF is

$$F(X) = \int_{-\infty}^{X_0} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

where X_0 is some specified value of X.

[29]D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics,* Academic Press, New York, 1973.

[30]A normal distribution with zero mean and unit (= 1) variance is known as a standard or standardized normal variable (see **App. A**).

where $P(Y = 1 \mid X)$ means the probability that an event occurs given the value(s) of the X, or explanatory, variable(s) and where Z_i is the standard normal variable, i.e., $Z \sim N(0, \sigma^2)$. F is the standard normal CDF, which written explicitly in the present context is:

$$\begin{aligned} F(I_i) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2}\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-z^2/2}\, dz \end{aligned} \tag{15.9.3}$$

Since P represents the probability that an event will occur, here the probability of owning a house, it is measured by the area of the standard normal curve from $-\infty$ to I_i as shown in Figure 15.4*a*.

Now to obtain information on I_i, the utility index, as well as on β_1 and β_2, we take the inverse of (15.9.2) to obtain:

$$\begin{aligned} I_i &= F^{-1}(I_i) = F^{-1}(P_i) \\ &= \beta_1 + \beta_2 X_i \end{aligned} \tag{15.9.4}$$

where F^{-1} is the inverse of the normal CDF. What all this means can be made clear from Figure 15.4. In panel *a* of this figure we obtain from the ordinate the (cumulative) probability of owning a house given $I_i^* \le I_i$, whereas in panel *b* we obtain from the abscissa the value of I_i given the value of P_i, which is simply the reverse of the former.

But how do we actually go about obtaining the index I_i as well as estimating β_1 and β_2? As in the case of the logit model, the answer depends on whether we have grouped data or ungrouped data. We consider the two cases individually.

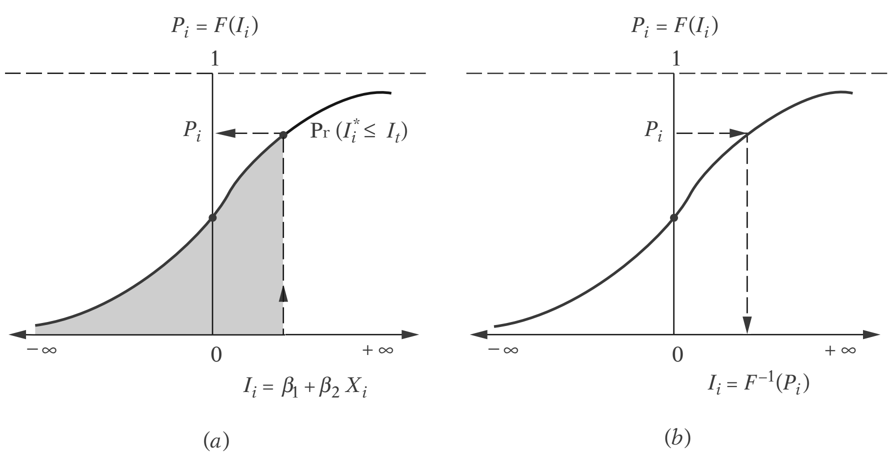


FIGURE 15.4 Probit model: (*a*) given I_i, read P_i from the ordinate; (*b*) given P_i, read I_i from the abscissa.

Probit Estimation with Grouped Data: gprobit

We will use the same data that we used for glogit, which is given in Table 15.4. Since we already have $\hat{P}_i$, the relative frequency (the empirical measure of probability) of owning a house at various income levels as shown in Table 15.5, we can use it to obtain I_i from the normal CDF as shown in Table 15.10, or from Figure 15.5.

Once we have the estimated I_i, estimating β_1 and β_2 is relatively straightforward, as we show shortly. In passing, note that in the language of probit analysis the unobservable utility index I_i is known as the **normal equivalent deviate** (n.e.d.) or simply **normit.** Since the n.e.d. or I_i will be negative whenever $P_i < 0.5$, in practice the number 5 is added to the n.e.d. and the result is called a probit.

TABLE 15.10 ESTIMATING THE INDEX I_i FROM THE STANDARD NORMAL CDF

$\hat{P}_i$	$I_i = F^{-1}(\hat{P}_i)$
0.20	−0.8416
0.24	−0.7063
0.30	−0.5244
0.35	−0.3853
0.45	−0.1257
0.51	0.0251
0.60	0.2533
0.66	0.4125
0.75	0.6745
0.80	0.8416

Notes: (1) $\hat{P}_i$ are from Table 15.5; (2) I_i are estimated from the standard normal CDF.

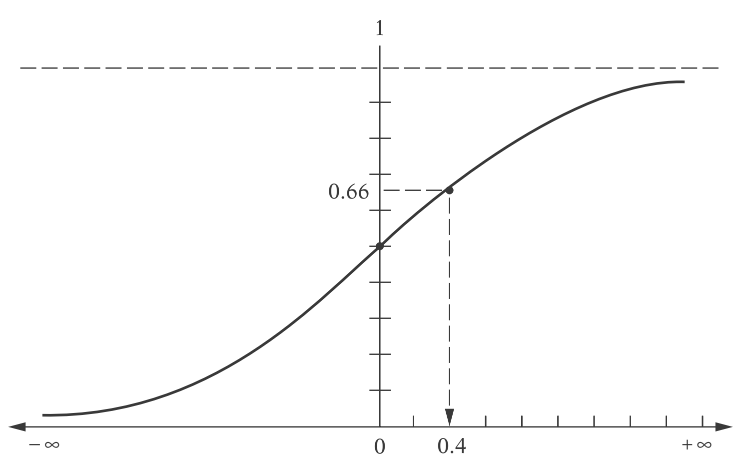


FIGURE 15.5 Normal CDF.

ILLUSTRATION OF GPROBIT USING HOUSING EXAMPLE

Let us continue with our housing example. We have already presented the results of the glogit model for this example. The grouped probit (gprobit) results of the same data are as follows:

Using the n.e.d. ($= I$) given in Table 15.10, the regression results are as shown in Table 15.11.[31] The regression results based on the probits ($=$ n.e.d. $+ 5$) are as shown in Table 15.12.

Except for the intercept term, these results are identical with those given in the previous table. But this should not be surprising. (Why?)

TABLE 15.11
Dependent Variable: I

Variable	Coefficient	Std. error	t statistic	Probability
C	-1.0166	0.0572	-17.7473	1.0397E-07
Income	0.04846	0.00247	19.5585	4.8547E-08

$R^2 = 0.97951$ Durbin–Watson statistic $= 0.91384$

TABLE 15.12
Dependent Variable: Probit

Variable	Coefficient	Std. error	t statistic	Probability
C	3.9833	0.05728	69.5336	2.03737E-12
Income	0.04846	0.00247	19.5585	4.8547E-08

$R^2 = 0.9795$ Durbin–Watson statistic $= 0.9138$

Note: These results are not corrected for heteroscedasticity (see exercise 15.12).

Interpretation of the Probit Estimates in Table 15.11. How do we interpret the preceding results? Suppose we want to find out the effect of a unit change in X (income measured in thousands of dollars) on the probability that $Y = 1$, that is, a family purchases a house. To do this, look at Eq. (15.9.2). We want to take the derivative of this function with respect to X (that is, the rate of change of the probability with respect to income). It turns out that this derivative is:

$$\frac{dP_i}{dX_i} = f(\beta_1 + \beta_2 X_i)\beta_2 \tag{15.9.5}$$ [32]

[31]The following results are not corrected for heteroscedasticity. See exercise 15.12 for the appropriate procedure to correct heteroscedasticity.

[32]We use the chain rule of derivatives:

$$\frac{dP_i}{dX_i} = \frac{dF(t)}{dt} \cdot \frac{dt}{dX}$$

where $t = \beta_1 + \beta_2 X_i$.

where $f(\beta_1 + \beta_2 X_i)$ is the standard normal probability density function evaluated at $\beta_1 + \beta_2 X_i$. As you will realize, this evaluation will depend on the particular value of the X variables. Let us take a value of X from Table 15.5, say, $X = 6$ (thousand dollars). Using the estimated values of the parameters given in Table 15.11, we thus want to find the normal density function at $f[-1.0166 + 0.04846(6)] = f(-0.72548)$. If you refer to the normal distribution tables, you will find that for $Z = -0.72548$, the normal density is about 0.3066.[33] Now multiplying this value by the estimated slope coefficient of 0.04846, we obtain 0.01485. This means that starting with an income level of \$6000, if the income goes up by \$1000, the probability of a family purchasing a house goes up by about 1.4 percent. (Compare this result with that given in Table 15.6.)

As you can see from the preceding discussion, compared with the LPM and logit models, the computation of changes in probability using the probit model is a bit tedious.

Instead of computing changes in probability, suppose you want to find the estimated probabilities from the fitted gprobit model. This can be done easily. Using the data in Table 15.11 and inserting the values of X from Table 15.5, the reader can check that the estimated n.i.d. values (to two digits) are as follows:

X	6	8	10	13	15	20	25	30	35	40
Estimated n.i.d.	−0.72	−0.63	−0.53	−0.39	−0.29	−0.05	0.19	0.43	0.68	0.92

Now statistical packages such as Minitab can easily compute the (cumulative) probabilities associated with the various n.i.d.'s. For example, corresponding to an n.i.d. value −0.63, the estimated probability is 0.2647 and, corresponding to an n.i.d. value of 0.43, the estimated probability is 0.6691. If you compare these estimates with the actual values given in Table 15.5, you will find that the two are fairly close, suggesting that the fitted model is quite good. Graphically, what we have just done is already shown in Figure 15.4.

The Probit Model for Ungrouped or Individual Data

Let us revisit Table 15.7, which gives data on 32 individuals about their final grade in intermediate microeconomics examination in relation to the variables GPA, TUCE, and PSI. The results of the logit regression are given in Table 15.8. Let us see what the probit results look like. Notice that as in the case of the logit model for individual data, we will have to use a nonlinear estimating procedure based on the method of maximum likelihood. The regression results calculated by Eviews 4 are given in Table 15.13.

[33]Note that the standard normal Z can range from $-\infty$ to $+\infty$, but the density function $f(Z)$ is always positive.

TABLE 15.13

Dependent Variable: grade
Method: ML—binary probit
Convergence achieved after 5 iterations

Variable	Coefficient	Std. error	*Z* statistic	Probability
C	-7.4523	2.5424	-2.9311	0.0033
GPA	1.6258	0.6938	2.3430	0.0191
TUCE	0.0517	0.0838	0.6166	0.5374
PSI	1.4263	5950	2.3970	0.0165

LR statistic (3 df) = 15.5458   McFadden R^2 = 0.3774
Probability (LR stat) = 0.0014

TABLE 15.14

Dependent Variable: grade

Variable	Coefficient	Std. error	*t* statistic	Probability
C	-1.4980	0.5238	-2.8594	0.0079
GPA	0.4638	0.1619	2.8640	0.0078
TUCE	0.0104	0.0194	0.5386	0.5943
PSI	0.3785	0.1391	2.7200	0.0110

R^2 = 0.4159   Durbin–Watson *d* = 2.3464   *F* statistic = 6.6456

"Qualitatively," the results of the probit model are comparable with those obtained from the logit model in that GPA and PSI are individually statistically significant. Collectively, all the coefficients are statistically significant, since the value of the LR statistic is 15.5458 with a *p* value of 0.0014. For reasons discussed in the next sections, we cannot directly compare the logit and probit regression coefficients.

For comparative purposes, we present the results based on the linear probability model (LPM) for the grade data in Table 15.14. Again, qualitatively, the LPM results are similar to the logit and probit models in that GPA and PSI are individually statistically significant but TUCE is not. Also, together the explanatory variables have a significant impact on grade, as the *F* value of 6.6456 is statistically significant because its *p* value is only 0.0015.

The Marginal Effect of a Unit Change in the Value of a Regressor in the Various Regression Models

In the *linear regression model*, the slope coefficient measures the change in the average value of the regressand for a unit change in the value of a regressor, with all other variables held constant.

In the *LPM*, the slope coefficient measures directly the change in the probability of an event occurring as the result of a unit change in the value of a regressor, with the effect of all other variables held constant.

In the *logit model* the slope coefficient of a variable gives the change in the log of the odds associated with a unit change in that variable, again holding all other variables constant. But as noted previously, for the logit model the rate of change in the probability of an event happening is given by $\beta_j P_i(1 - P_i)$, where β_j is the (partial regression) coefficient of the jth regressor. But in evaluating P_i, all the variables included in the analysis are involved.

In the *probit model,* as we saw earlier, the rate of change in the probability is somewhat complicated and is given by $\beta_j f(Z_i)$, where $f(Z_i)$ is the density function of the standard normal variable and $Z_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$, that is, the regression model used in the analysis.

Thus, in both the logit and probit models all the regressors are involved in computing the changes in probability, whereas in the LPM only the jth regressor is involved. This difference may be one reason for the early popularity of the LPM model.

15.10 LOGIT AND PROBIT MODELS

Although for our grade example LPM, logit, and probit give qualitatively similar results, we will confine our attention to logit and probit models because of the problems with the LPM noted earlier. Between logit and probit, which model is preferable? In most applications the models are quite similar, the main difference being that the logistic distribution has slightly fatter tails, which can be seen from Figure 15.6. That is to say, the conditional probability P_i approaches zero or one at a slower rate in logit than in probit. This can be seen more clearly from Table 15.15. Therefore, there is no compelling reason to choose one over the other. In practice many researchers choose the logit model because of its comparative mathematical simplicity.

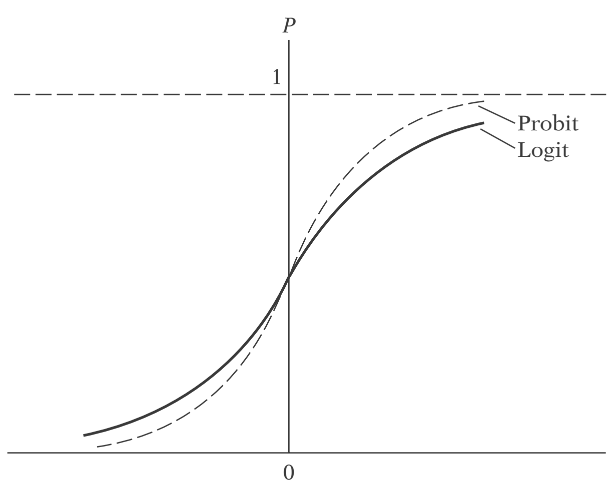


FIGURE 15.6 Logit and probit cumulative distributions.

TABLE 15.15 VALUES OF CUMULATIVE PROBABILITY FUNCTIONS

	Cumulative normal	Cumulative logistic
Z	$P_1(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-s^2/2} ds$	$P_2(Z) = \frac{1}{1+e^{-z}}$
−3.0	0.0013	0.0474
−2.0	0.0228	0.1192
−1.5	0.0668	0.1824
−1.0	0.1587	0.2689
−0.5	0.3085	0.3775
0	0.5000	0.5000
0.5	0.6915	0.6225
1.0	0.8413	0.7311
1.5	0.9332	0.8176
2.0	0.9772	0.8808
3.0	0.9987	0.9526

Though the models are similar, one has to be careful in interpreting the coefficients estimated by the two models. For example, for our grade example, the coefficient of GPA of 1.6258 of the probit model and 2.8261 of the logit model are not directly comparable. The reason is that, although the standard logistic (the basis of logit) and the standard normal distributions (the basis of probit) both have a mean value of zero, their variances are different; 1 for the standard normal (as we already know) and $\pi^2/3$ for the logistic distribution, where $\pi \approx 22/7$. Therefore, if you multiply the probit coefficient by about 1.81 (which is approximately $= \pi/\sqrt{3}$), you will get approximately the logit coefficient. For our example, the probit coefficient of GPA is 1.6258. Multiplying this by 1.81, we obtain 2.94, which is close to the logit coefficient. Alternatively, if you multiply a logit coefficient by 0.55 ($= 1/1.81$), you will get the probit coefficient. Amemiya, however, suggests multiplying a logit estimate by 0.625 to get a better estimate of the corresponding probit estimate.[34] Conversely, multiplying a probit coefficient by 1.6 ($= 1/0.625$) gives the corresponding logit coefficient.

Incidentally, Amemiya has also shown that the coefficients of LPM and logit models are related as follows:

$$\beta_{\text{LPM}} = 0.25\beta_{\text{logit}} \qquad \text{except for intercept}$$

and

$$\beta_{\text{LPM}} = 0.25\beta_{\text{logit}} + 0.5 \qquad \text{for intercept}$$

We leave it to the reader to find out if these approximations hold for our grade example.

[34] T. Amemiya, "Qualitative Response Model: A Survey," *Journal of Economic Literature,* vol. 19, 1981, pp. 481–536.

15.11 THE TOBIT MODEL

An extension of the probit model is the **tobit model** originally developed by James Tobin, the Nobel laureate economist. To explain this model, we continue with our home ownership example. In the probit model our concern was with estimating the probability of owning a house as a function of some socioeconomic variables. In the tobit model our interest is in finding out the amount of money a person or family spends on a house in relation to socioeconomic variables. Now we face a dilemma here: If a consumer does not purchase a house, obviously we have no data on housing expenditure for such consumers; we have such data only on consumers who actually purchase a house.

Thus consumers are divided into two groups, one consisting of, say, n_1 consumers about whom we have information on the regressors (say, income, mortgage interest rate, number of people in the family, etc.) as well as the regressand (amount of expenditure on housing) and another consisting of n_2 consumers about whom we have information only on the regressors but not on the regressand. A sample in which information on the regressand is available only for some observations is known as a **censored sample.**[35] Therefore, the tobit model is also known as a censored regression model. Some authors call such models **limited dependent variable regression models** because of the restriction put on the values taken by the regressand.

Statistically, we can express the tobit model as

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i && \text{if RHS} > 0 \\ &= 0 && \text{otherwise} \end{aligned} \qquad \textbf{(15.11.1)}$$

where RHS = right-hand side. *Note:* Additional X variables can be easily added to the model.

Can we estimate regression (15.11.1) using only n_1 observations and not worry about the remaining n_2 observations? The answer is no, for the OLS estimates of the parameters obtained from the subset of n_1 observations will be *biased as well as inconsistent;* that is, they are biased even asymptotically.[36]

To see this, consider Figure 15.7. As the figure shows, if Y is not observed (because of censoring), all such observations ($= n_2$), denoted by crosses, will

[35]A censored sample should be distinguished from a **truncated sample** in which information on the regressors is available only if the regressand is observed. We will not pursue this topic here, but the interested reader may consult William H. Greene, *Econometric Analysis,* Prentice Hall, 4th ed., Englewood Cliffs, N.J., Chap. 19. For an intuitive discussion, see Peter Kennedy, *A Guide to Econometrics,* The MIT Press, Cambridge, Mass., 4th ed., 1998, Chap. 16.

[36]The bias arises from the fact that if we consider only the n_1 observations and omit the others, there is no guarantee that $E(u_i)$ will be necessarily zero. And without $E(u_i) = 0$ we cannot guarantee that the OLS estimates will be unbiased. This bias can be readily seen from the discussion in App. 3A, Eqs. (4) and (5).