高等学校经济类双语教学推荐教材 PEARSON

ECONOMICS CLASSICS

经济学经典教材 · 核心课系列

Economics Classics

计量经济分析 下册

Econometric Analysis

（英文版 · 第七版）

(Seventh Edition)

威廉 · H · 格林（William H. Greene） 著

中国人民大学出版社

高等学校经济学双语教学推荐教材

经济学经典教材·核心课系列

计量经济分析 下册

Econometric Analysis

(英文版·第七版)

(Seventh Edition)

威廉·H·格林(William H. Greene) 著

中国人民大学出版社
·北京·

图书在版编目（CIP）数据

计量经济分析（第 7 版）：英文 / 格林著．—北京：中国人民大学出版社，2013.11
高等学校经济类双语教学推荐教材
经济学经典教材·核心课系列
ISBN 978-7-300-18340-4

Ⅰ.①计…　Ⅱ.①格…　Ⅲ.①计量经济学－双语教学－高等学校－教材－英文　Ⅳ.①F224.0

中国版本图书馆 CIP 数据核字（2013）第 257326 号

高等学校经济类双语教学推荐教材
经济学经典教材·核心课系列
计量经济分析（英文版·第七版）
威廉·H·格林　著
Jiliang Jingji fenxi

出版发行	中国人民大学出版社		
社　　址	北京中关村大街31号	**邮政编码**	100080
电　　话	010—62511242（总编室）		010—62511398（质管部）
	010—82501766（邮购部）		010—62514148（门市部）
	010—62515195（发行公司）		010—62515275（盗版举报）
网　　址	http://www.crup.com.cn		
	http://www.ttrnet.com（人大教研网）		
经　　销	新华书店		
印　　刷	涿州市星河印刷有限公司		
规　　格	215 mm×275 mm　16 开本	**版　　次**	2013 年12月第 1 版
印　　张	62.25　插页2	**印　　次**	2013 年12月第 1 次印刷
字　　数	1 565 000	**定　　价**	125.00 元

尊敬的老师:

您好!

为了确保您及时有效地申请培生整体教学资源,请您务必完整填写如下表格,加盖学院的公章后传真给我们,我们将会在 2~3 个工作日内为您处理。

请填写所需教辅的开课信息:

采用教材			□中文版 □英文版 □双语版
作　者		出版社	
版　次		**ISBN**	
课程时间	始于　年　月　日	学生人数	
	止于　年　月　日	学生年级	□专 科　□本科 **1/2** 年级 □研究生　□本科 **3/4** 年级

请填写您的个人信息:

学　校			
院系/专业			
姓　名		职　称	□助教 □讲师 □副教授 □教授
通信地址/邮编			
手　机		电　话	
传　真			
official email(必填) **(eg:XXX@ruc.edu.cn)**		**email** **(eg:XXX@163.com)**	
是否愿意接受我们定期的新书讯息通知:　□是　□否			

系 / 院主任: ___________ (签字)

(系 / 院办公室章)

___年___月___日

资源介绍:

--教材、常规教辅(PPT、教师手册、题库等)资源:请访问 **www.pearsonhighered.com/educator**;　(免费)

--MyLabs/Mastering 系列在线平台:适合老师和学生共同使用;访问需要 **Access Code**;　(付费)

100013　北京市东城区北三环东路 36 号环球贸易中心 D 座 1208 室

电话: (8610)57355169　　传真: (8610)58257961

Please send this form to: **Service.CN@pearson.com**

目　录

第一部分　线性回归模型

第二部分　广义回归模型与联立方程模型

第三部分　估计方法

第四部分　横截面、面板数据及微观计量学

第五部分　时间序列与宏观计量学

第六部分　附录

CONTENTS

17
DISCRETE CHOICE

17.1 INTRODUCTION

This is the first of three chapters that will survey models used in **microeconometrics**. The analysis of individual choice that is the focus of this field is fundamentally about modeling discrete outcomes such as purchase decisions, for example whether or not to buy insurance, voting behavior, choice among a set of alternative brands, travel modes or places to live, and responses to survey questions about the strength of preferences or about self-assessed health or well-being. In these and any number of other cases, the "dependent variable" is not a quantitative measure of some economic outcome, but rather an indicator of whether or not some outcome occurred. It follows that the regression methods we have used up to this point are largely inappropriate. We turn, instead, to modeling probabilities and using econometric tools to make probabilistic statements about the occurrence of these events. We will also examine models for counts of occurrences. These are closer to familiar regression models, but are, once again, about discrete outcomes of behavioral choices. As such, in this setting as well, we will be modeling probabilities of events, rather than conditional mean functions.

The models that are analyzed in this and the next chapter are built on a platform of preferences of decision makers. We take a **random utility** view of the choices that are observed. The decision maker is faced with a situation or set of alternatives and reveals something about their underlying preferences by the choice that he or she makes. The choice(s) made will be affected by observable influences—this is, of course, the ultimate objective of advertising—and by unobservable characteristics of the chooser. The blend of these fundamental bases for individual choice is at the core of the broad range of models that we will examine here.[1]

This chapter and Chapter 18 will describe four broad frameworks for analysis:

Binary Choice: The individual faces a pair of choices and makes that choice between the two that provides the greater utility. Many such settings involve the choice between taking an action and not taking that action, for example the decision whether or not to purchase health insurance. In other cases, the decision might be between two distinctly different choices, such as the decision whether to travel to and from work via public or private transportation. In the binary choice case, the 0/1 outcome is merely a label for "no/yes"—the numerical values are a mere convenience.

Multinomial Choice: The individual chooses among more than two choices, once again, making the choice that provides the greatest utility. In the previous example, private travel might involve a choice of being a driver or passenger while public transport might involve a choice between bus and train. At one level, this is a minor variation of the binary choice case—the latter is, of course, a special case of the former. But, more elaborate models of multinomial choice allow a rich specification of consumer preferences. In the multinomial case, the observed response is simply a label for the selected choice; it might be a brand, the name of a place, or the type of travel mode. Numerical assignments are not meaningful in this setting.

Ordered Choice: The individual reveals the strength of his or her preferences with respect to a single outcome. Familiar cases involve survey questions about strength of feelings about a particular commodity such as a movie, or self-assessments of social outcomes such as health in general or self-assessed well-being. In the ordered choice

[1] See Greene and Hensher (2010, Chapter 4) for an historical perspective on this approach to model specification.

setting, opinions are given meaningful numeric values, usually $0, 1, \ldots, J$ for some upper limit, J. For example, opinions might be labelled 0, 1, 2, 3, 4 to indicate the strength of preferences, for example, for a product, a movie, a candidate or a piece of legislation. But, in this context, the numerical values are only a ranking, not a quantitative measure. Thus a "1" is greater than a "0" in a qualitative sense, but not by one unit, and the difference between a "2" and a "1" is not the same as that between a "1" and a "0".

In these three cases, although the numerical outcomes are merely labels of some nonquantitative outcome, the analysis will nonetheless have a regresson-style motivation. Throughout, the models will be based on the idea that observed "covariates" are relevant in explaining the observed choices. For example, in the binary outcome "did or did not purchase health insurance," a conditioning model suggests that covariates such as age, income, and family situation will help to explain the choice. This chapter will describe a range of models that have been developed around these considerations. We will also be interested in a fourth application of discrete outcome models:

Event Counts: The observed outcome is a count of the number of occurrences. In many cases, this is similar to the preceding three settings in that the "dependent variable" measures an individual choice, such as the number of visits to the physician or the hospital, the number of derogatory reports in one's credit history, or the number of visits to a particular recreation site. In other cases, the event count might be the outcome of some natural process, such as incidence of a disease in a population or the number of defects per unit of time in a production process. In this setting, we will be doing a more familiar sort of regression modeling. However, the models will still be constructed specifically to accommodate the discrete nature of the observed response variable.

We will consider these four cases in turn. The four broad areas have many elements in common; however, there are also substantive differences between the particular models and analysis techniques used in each. This chapter will develop the first topic, models for binary choices. In each section, we will begin with an overview of applications and then present the single basic model that is the centerpiece of the methodology, and, finally, examine some recently developed extensions of the model. This chapter contains a very lengthy discussion of models for binary choices. This analysis is as long as it is because, first, the models discussed are used throughout microeconometrics—the central model of binary choice in this area is as ubiquitous as linear regression. Second, all the econometric issues and features that are encountered in the other areas will appear in the analysis of binary choice, where we can examine them in a fairly straightforward fashion.

It will emerge that, at least in econometric terms, the models for multinomial and ordered choice considered in Chapter 18 can be built from the two fundamental building blocks, the model of random utility and the translation of that model into a description of binary choices. There are relatively few new econometric issues that arise here. Chapter 18 will be largely devoted to suggesting different approaches to modeling choices among multiple alternatives and models for ordered choices. Once again, models of preference scales, such as movie or product ratings, or self-assessments of health or well-being, can be naturally built up from the fundamental model of random utility. Finally, Chapter 18 will develop the well-known Poisson regression model for counts of events. We will then extend the model to demonstrate some recent applications and innovations.

Chapters 17 and 18 are a lengthy but far from complete survey of topics in estimating **qualitative response (QR)** models. None of these models can consistently be estimated with linear regression methods. In most cases, the method of estimation is **maximum likelihood**. Therefore, readers interested in the mechanics of estimation may want to review the material in Appendices D and E before continuing. The various properties of maximum likelihood estimators are discussed in Chapter 14. We shall assume throughout these chapters that the necessary conditions behind the optimality properties of maximum likelihood estimators are met and, therefore, we will not derive or establish these properties specifically for the QR models. Detailed proofs for most of these models can be found in surveys by Amemiya (1981), McFadden (1984), Maddala

(1983), and Dhrymes (1984). Additional commentary on some of the issues of interest in the contemporary literature is given by Manski and McFadden (1981) and Maddala and Flores-Lagunes (2001). Agresti (2002) and Cameron and Trivedi (2005) contain numerous theoretical developments and applications. Greene (2008) and Hensher and Greene (2010) provide, among many others, general surveys of discrete choice models and methods.[2]

17.2 MODELS FOR BINARY OUTCOMES

For purposes of studying individual behavior, we will construct models that link the decision or outcome to a set of factors, at least in the spirit of regression. Our approach will be to analyze each of them in the general framework of probability models:

$$\text{Prob}(\text{event } j \text{ occurs}) = \text{Prob}(Y = j) = F[\text{relevant effects, parameters}]. \quad \textbf{(17-1)}$$

The study of qualitative choice focuses on appropriate specification, estimation, and use of models for the probabilities of events, where in most cases, the "event" is an individual's choice among a set of two or more alternatives.

Example 17.1 Labor Force Participation Model

In Example 5.2 we estimated an earnings equation for the subsample of 428 married women who participated in the formal labor market taken from a full sample of 753 observations. The semilog earnings equation is of the form

$$\ln earnings = \beta_1 + \beta_2\, age + \beta_3\, age^2 + \beta_4\, education + \beta_5\, kids + \varepsilon,$$

where *earnings* is *hourly wage* times *hours worked, education* is measured in years of schooling, and *kids* is a binary variable which equals one if there are children under 18 in the household. What of the other 325 individuals? The underlying labor supply model described a market in which labor force participation was the outcome of a market process whereby the demanders of labor services were willing to offer a wage based on expected marginal product and individuals themselves made a decision whether or not to accept the offer depending on whether it exceeded their own reservation wage. The first of these depends on, among other things, education, while the second (we assume) depends on such variables as age, the presence of children in the household, other sources of income (husband's), and marginal tax rates on labor income. The sample we used to fit the earnings equation contains data on all these other variables. The models considered in this chapter would be appropriate for modeling the outcome $y = 1$ if in the labor force, and 0 if not.

Models for explaining a binary (0/1) dependent variable are typically motivated in two contexts. The labor force participation model in Example 17.1 describes a process of individual choice between two alternatives in which the choice is influenced by observable effects (children, tax rates) and unobservable aspects of the preferences of the individual. The relationship between voting behavior and income is another example. In other cases, the **binary choice model** arises in a setting in which the nature of the observed data dictate the special treatment of a binary dependent variable model. In these cases, the analyst is essentially interested in a regression-like model of the sort considered in Chapters 2 through 7. With data on the variable of interest and a set of covariates, they are interested in specifying a relationship between the former and the latter, more or less along the lines of the models we have already studied. For example, in a model of the demand for tickets for sporting events, in which the variable of interest is number of tickets, it could happen that the observation consists only of whether the sports facility was filled to capacity (demand greater than or equal to capacity so $Y = 1$) or not ($Y = 0$). It will generally turn out that the models and techniques used in both cases are the same. Nonetheless, it is useful to examine both of them.

[2]There are dozens of book length surveys of discrete choice models. Two others that are heavily oriented to application of the methods are Train (2003) and Hensher, Rose, and Greene (2005).

17.2.1 RANDOM UTILITY MODELS FOR INDIVIDUAL CHOICE

An interpretation of data on individual choices is provided by the random utility model. Let U_a and U_b represent an individual's utility of two choices. For example, U_a might be the utility of rental housing and U_b that of home ownership. The observed choice between the two reveals which one provides the greater utility, but not the unobservable utilities. Hence, the observed indicator equals 1 if $U_a > U_b$ and 0 if $U_a \leq U_b$. A common formulation is the linear random utility model,

$$U_a = \mathbf{w}'\boldsymbol{\beta}_a + \mathbf{z}_a{}'\boldsymbol{\gamma}_a + \varepsilon_a \quad \text{and} \quad U_b = \mathbf{w}'\boldsymbol{\beta}_b + \mathbf{z}_b{}'\boldsymbol{\gamma}_b + \varepsilon_b. \tag{17-2}$$

In (17-2), the observable (measurable) vector of **characteristics** of the individual is denoted $\mathbf{w}$; this might include gender, age, income, and other demographics. The vectors $\mathbf{z}_a$ and $\mathbf{z}_b$ denote features (**attributes**) of the two choices that might be choice specific. In a voting context, for example, the attributes might be indicators of the competing candidates' positions on important issues. The random terms, ε_a and ε_b represent the stochastic elements that are specific to and known only by the individual, but not by the observer (analyst). To continue our voting example, ε_a might represent an intangible, general "preference" for candidate a.

The completion of the model for the determination of the observed outcome (choice) is the revelation of the ranking of the preferences by the choice the individual makes. Thus, if we denote by $Y = 1$ the consumer's choice of alternative a, we infer from $Y = 1$ that $U_a > U_b$. Since the outcome is ultimately driven by the random elements in the utility functions, we have

$$\begin{aligned}
\text{Prob}[Y = 1|\mathbf{w}, \mathbf{z}_a, \mathbf{z}_b] &= \text{Prob}[U_a > U_b] \\
&= \text{Prob}[(\mathbf{w}'\boldsymbol{\beta}_a + \mathbf{z}_a{}'\boldsymbol{\gamma}_a + \varepsilon_a) - (\mathbf{x}'\boldsymbol{\beta}_b + \mathbf{z}_b{}'\boldsymbol{\gamma}_b + \varepsilon_b) > 0 \,|\, \mathbf{w}, \mathbf{z}_a, \mathbf{z}_b] \\
&= \text{Prob}[(\mathbf{w}'(\boldsymbol{\beta}_a - \boldsymbol{\beta}_b) + \mathbf{z}_a{}'\boldsymbol{\gamma}_a - \mathbf{z}_b{}'\boldsymbol{\gamma}_b + \varepsilon_a - \varepsilon_b) > 0 \,|\, \mathbf{w}, \mathbf{z}_a, \mathbf{z}_b] \\
&= \text{Prob}[\mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0 \,|\, \mathbf{x}],
\end{aligned}$$

where $\mathbf{x}'\boldsymbol{\beta}$ collects all the observable elements of the difference of the two utility functions and ε denotes the difference between the two random elements.

Example 17.2 Structural Equations for a Binary Choice Model

Nakosteen and Zimmer (1980) analyzed a model of migration based on the following structure:[3] For a given individual, the market wage that can be earned at the present location is

$$y_p^* = \mathbf{w}_p'\boldsymbol{\beta}_p + \varepsilon_p.$$

Variables in the equation include age, sex, race, growth in employment, and growth in per capita income. If the individual migrates to a new location, then his or her market wage would be

$$y_m^* = \mathbf{w}_m'\boldsymbol{\beta}_m + \varepsilon_m.$$

Migration entails costs that are related both to the individual and to the labor market:

$$C^* = \mathbf{z}'\boldsymbol{\alpha} + u.$$

Costs of moving are related to whether the individual is self-employed and whether that person recently changed his or her industry of employment. They migrate if the benefit $y_m^* - y_p^*$ is greater than the cost, C. The net benefit of moving is

[3] A number of other studies have also used variants of this basic formulation. Some important examples are Willis and Rosen (1979) and Robinson and Tomes (1982). The study by Tunali (1986) examined in Example 17.6 is another application. The now standard approach, in which "participation" equals one if wage offer $(\mathbf{x}_w'\boldsymbol{\beta}_w + \varepsilon_w)$ minus reservation wage $(\mathbf{x}_r'\boldsymbol{\beta}_r + \varepsilon_r)$ is positive, is also used in Fernandez and Rodriguez-Poo (1997). Brock and Durlauf (2000) describe a number of models and situations involving individual behavior that give rise to binary choice models.

$$\begin{aligned} M^* &= y_m^* - y_p^* - C^* \\ &= \mathbf{w}_m'\boldsymbol{\beta}_m - \mathbf{w}_p'\boldsymbol{\beta}_p - \mathbf{z}'\boldsymbol{\alpha} + (\varepsilon_m - \varepsilon_p - u) \\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon. \end{aligned}$$

Because M^* is unobservable, we cannot treat this equation as an ordinary regression. The individual either moves or does not. After the fact, we observe only y_m^* if the individual has moved or y_p^* if he or she has not. But we do observe that $M = 1$ for a move and $M = 0$ for no move.

17.2.2 A LATENT REGRESSION MODEL

Discrete dependent-variable models are often cast in the form of **index function models**. We view the outcome of a discrete choice as a reflection of an underlying regression. As an often-cited example, consider the decision to make a large purchase. The theory states that the consumer makes a marginal benefit/marginal cost calculation based on the utilities achieved by making the purchase and by not making the purchase and by using the money for something else. We model the difference between benefit and cost as an unobserved variable y^* such that

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon.$$

Note that this is the result of the "net utility" calculation in the previous section and in Example 17.2. We assume that ε has mean zero and has either a standardized logistic with variance $\pi^2/3$ or a standard normal distribution with variance one or some other specific distribution with known variance. We do not observe the net benefit of the purchase (i.e., net utility), only whether it is made or not. Therefore, our observation is

$$\begin{aligned} y &= 1 \quad \text{if } y^* > 0, \\ y &= 0 \quad \text{if } y^* \le 0. \end{aligned} \qquad \textbf{(17-3)}$$

In this formulation, $\mathbf{x}'\boldsymbol{\beta}$ is called the index function. The assumption of known variance of ε is an innocent normalization. Suppose the variance of ε is scaled by an unrestricted parameter σ^2. The **latent regression** will be $y^* = \mathbf{x}'\boldsymbol{\beta} + \sigma\varepsilon$. But, $(y^*/\sigma) = \mathbf{x}'(\boldsymbol{\beta}/\sigma) + \varepsilon$ is the same model with the same data. The observed data will be unchanged; y is still 0 or 1, depending only on the sign of y^* not on its scale. This means that there is no information about σ in the sample data so σ cannot be estimated. The parameter vector $\boldsymbol{\beta}$ in this model is only "identified up to scale." The assumption of zero for the threshold in (17-3) is likewise innocent if the model contains a constant term (and not if it does not).[4] Let a be the supposed nonzero threshold and α be the unknown constant term and, for the present, $\mathbf{x}$ and $\boldsymbol{\beta}$ contain the rest of the index not including the constant term. Then, the probability that y equals one is

$$\text{Prob}(y^* > a \mid \mathbf{x}) = \text{Prob}(\alpha + \mathbf{x}'\boldsymbol{\beta} + \varepsilon > a \mid \mathbf{x}) = \text{Prob}[(\alpha - a) + \mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0 \mid \mathbf{x}].$$

Because α is unknown, the difference $(\alpha - a)$ remains an unknown parameter. The end result is that if the model contains a constant term, it is unchanged by the choice of the threshold in (17-3). The choice of zero is a normalization with no significance. With the two normalizations, then,

$$\text{Prob}(y^* > 0 \mid \mathbf{x}) = \text{Prob}(\varepsilon > -\mathbf{x}'\boldsymbol{\beta} \mid \mathbf{x}).$$

A remaining detail in the model is the choice of the specific distribution for ε. We will consider several. The overwhelming majority of applications are based either on the normal or the logistic distribution. If the distribution is symmetric, as are the normal and logistic, then

$$\text{Prob}(y^* > 0 \mid \mathbf{x}) = \text{Prob}(\varepsilon < \mathbf{x}'\boldsymbol{\beta} \mid \mathbf{x}) = F(\mathbf{x}'\boldsymbol{\beta}), \qquad \textbf{(17-4)}$$

[4]Unless there is some compelling reason, binomial probability models should not be estimated without constant terms.

where $F(t)$ is the cdf of the random variable, ε. This provides an underlying structural model for the probability.

17.2.3 FUNCTIONAL FORM AND REGRESSION

Consider the model of labor force participation suggested in Example 17.1. The respondent either works or seeks work ($Y = 1$) or does not ($Y = 0$) in the period in which our survey is taken. We believe that a set of factors, such as age, marital status, education, and work history, gathered in a vector **x**, explain the decision, so that

$$\begin{aligned}\text{Prob}(Y = 1 \mid \mathbf{x}) &= F(\mathbf{x}, \boldsymbol{\beta})\\ \text{Prob}(Y = 0 \mid \mathbf{x}) &= 1 - F(\mathbf{x}, \boldsymbol{\beta}).\end{aligned} \qquad \textbf{(17-5)}$$

The set of parameters $\boldsymbol{\beta}$ reflects the impact of changes in **x** on the probability. For example, among the factors that might interest us is the marginal effect of marital status on the probability of labor force participation. The problem at this point is to devise a suitable model for the right-hand side of the equation. One possibility is to retain the familiar linear regression,

$$F(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}'\boldsymbol{\beta}.$$

Because $E[y \mid \mathbf{x}] = 0[1 - F(\mathbf{x}, \boldsymbol{\beta})] + 1[F(\mathbf{x}, \boldsymbol{\beta})] = F(\mathbf{x}, \boldsymbol{\beta})$, we can construct the regression model,

$$\begin{aligned} y &= E[y \mid \mathbf{x}] + y - E[y \mid \mathbf{x}]\\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon.\end{aligned} \qquad \textbf{(17-6)}$$

The **linear probability model** has a number of shortcomings. A minor complication arises because ε is heteroscedastic in a way that depends on $\boldsymbol{\beta}$. Because $\mathbf{x}'\boldsymbol{\beta} + \varepsilon$ must equal 0 or 1, ε equals either $-\mathbf{x}'\boldsymbol{\beta}$ or $1 - \mathbf{x}'\boldsymbol{\beta}$, with probabilities $1 - F$ and F, respectively. Thus, you can easily show that in this model,

$$\text{Var}[\varepsilon \mid \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}(1 - \mathbf{x}'\boldsymbol{\beta}). \qquad \textbf{(17-7)}$$

We could manage this complication with an FGLS estimator in the fashion of Chapter 9, though this only solves the estimation problem, not the theoretical one. A more serious flaw is that without some ad hoc tinkering with the disturbances, we cannot be assured that the predictions from this model will truly look like probabilities. We cannot constrain $\mathbf{x}'\boldsymbol{\beta}$ to the 0–1 interval. Such a model produces both nonsense probabilities and negative variances. For these reasons, the linear probability model is becoming less frequently used except as a basis for comparison to some other more appropriate models.[5]

Our requirement, then, is a model that will produce predictions consistent with the underlying theory in (17-4). For a given regressor vector, we would expect

$$\begin{aligned}\lim_{\mathbf{x}'\boldsymbol{\beta} \to +\infty} \text{Prob}(Y = 1 \mid \mathbf{x}) &= 1\\ \lim_{\mathbf{x}'\boldsymbol{\beta} \to -\infty} \text{Prob}(Y = 1 \mid \mathbf{x}) &= 0.\end{aligned} \qquad \textbf{(17-8)}$$

See Figure 17.1. In principle, any proper, continuous probability distribution defined over the real line will suffice. The normal distribution has been used in many analyses, giving rise to the **probit** model,

[5]The linear model is not beyond redemption. Aldrich and Nelson (1984) analyze the properties of the model at length. Judge et al. (1985) and Fomby, Hill, and Johnson (1984) give interesting discussions of the ways we may modify the model to force internal consistency. But the fixes are sample dependent, and the resulting estimator, such as it is, may have no known sampling properties. Additional discussion of weighted least squares appears in Amemiya (1977) and Mullahy (1990). Finally, its shortcomings notwithstanding, the linear probability model is applied by Caudill (1988), Heckman, and MaCurdy (1985), and Heckman and Snyder (1997). An exchange on the usefulness of the approach is Angrist (2001) and Moffitt (2001). See Angrist and Pischke (2009) for some applications.

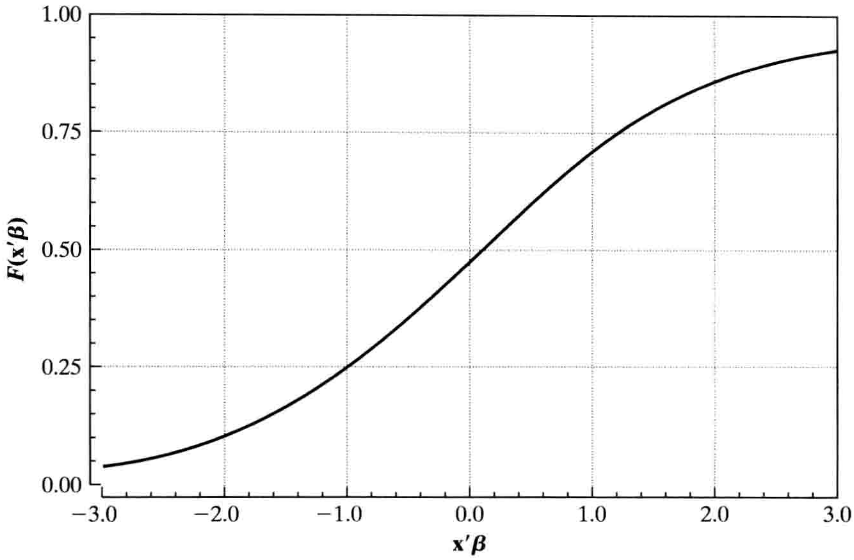

FIGURE 17.1 Model for a Probability.

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} \phi(t)dt = \Phi(\mathbf{x}'\boldsymbol{\beta}). \tag{17-9}$$

The function $\Phi(t)$ is a commonly used notation for the standard normal distribution function. Partly because of its mathematical convenience, the logistic distribution,

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} = \Lambda(\mathbf{x}'\boldsymbol{\beta}). \tag{17-10}$$

has also been used in many applications. We shall use the notation $\Lambda(.)$ to indicate the logistic cumulative distribution function. This model is called the **logit** model for reasons we shall discuss in the next section. Both of these distributions have the familiar bell shape of symmetric distributions. Other models which do not assume symmetry, such as the **Gumbel model**,

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \exp[-\exp(-\mathbf{x}'\boldsymbol{\beta})],$$

and **complementary log log model**,

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = 1 - \exp[-\exp(\mathbf{x}'\boldsymbol{\beta})],$$

have also been employed. Still other distributions have been suggested,[6] but the probit and logit models are still the most common frameworks used in econometric applications.

The question of which distribution to use is a natural one. The logistic distribution is similar to the normal except in the tails, which are considerably heavier. (It more closely resembles a t distribution with seven degrees of freedom.) Therefore, for intermediate values of $\mathbf{x}'\boldsymbol{\beta}$ (say, between -1.2 and $+1.2$), the two distributions tend to give similar probabilities. The logistic distribution tends to give larger probabilities to $Y = 1$ when $\mathbf{x}'\boldsymbol{\beta}$ is extremely small (and smaller probabilities to $Y = 1$ when $\mathbf{x}'\boldsymbol{\beta}$ is very large) than the normal distribution. It is difficult to provide practical generalities on this basis, however, as they would require knowledge of $\boldsymbol{\beta}$. We should expect different predictions from the two models, however, if the sample contains (1) very few "responses" (Y's equal to 1) or very few "nonresponses" (Y's equal to 0) and (2) very wide variation in an important independent variable, particularly if (1) is also true. There are practical

[6]See, for example, Maddala (1983, pp. 27–32), Aldrich and Nelson (1984), and Greene (2001).

reasons for favoring one or the other in some cases for mathematical convenience, but it is difficult to justify the choice of one distribution or another on theoretical grounds. Amemiya (1981) discusses a number of related issues, but as a general proposition, the question is unresolved. In most applications, the choice between these two seems not to make much difference. However, as seen in the following example, the symmetric and asymmetric distributions can give substantively different results, and here, the guidance on how to choose is unfortunately sparse.

The probability model is a regression:

$$E[y \mid \mathbf{x}] = F(\mathbf{x}'\boldsymbol{\beta}).$$

Whatever distribution is used, it is important to note that the parameters of the model, like those of any nonlinear regression model, are not necessarily the marginal effects we are accustomed to analyzing. In general,

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \left[\frac{dF(\mathbf{x}'\boldsymbol{\beta})}{d(\mathbf{x}'\boldsymbol{\beta})}\right] \times \boldsymbol{\beta} = f(\mathbf{x}'\boldsymbol{\beta}) \times \boldsymbol{\beta}, \qquad \textbf{(17-11)}$$

where $f(.)$ is the density function that corresponds to the cumulative distribution, $F(.)$. For the normal distribution, this result is

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \phi(\mathbf{x}'\boldsymbol{\beta}) \times \boldsymbol{\beta}, \qquad \textbf{(17-12)}$$

where $\phi(t)$ is the standard normal density. For the logistic distribution,

$$\frac{d\Lambda(\mathbf{x}'\boldsymbol{\beta})}{d(\mathbf{x}'\boldsymbol{\beta})} = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{[1+\exp(\mathbf{x}'\boldsymbol{\beta})]^2} = \Lambda(\mathbf{x}'\boldsymbol{\beta})[1-\Lambda(\mathbf{x}'\boldsymbol{\beta})],$$

so, in the logit model,

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \Lambda(\mathbf{x}'\boldsymbol{\beta})[1-\Lambda(\mathbf{x}'\boldsymbol{\beta})]\boldsymbol{\beta}. \qquad \textbf{(17-13)}$$

It is obvious that these values will vary with the values of $\mathbf{x}$. In interpreting the estimated model, it will be useful to calculate this value at, say, the means of the regressors and, where necessary, other pertinent values. For convenience, it is worth noting that the same scale factor applies to all the slopes in the model.

For computing **marginal effects**, one can evaluate the expressions at the sample means of the data or evaluate the marginal effects at every observation and use the sample average of the individual marginal effects—this produces the **average partial effects**. In large samples these generally give roughly the same answer (see Section 17.3.2). But that is not so in small- or moderate-sized samples. Current practice favors averaging the individual marginal effects when it is possible to do so.

Another complication for computing marginal effects in a binary choice model arises because $\mathbf{x}$ will often include dummy variables—for example, a labor force participation equation will often contain a dummy variable for marital status. Because the derivative is with respect to a small change, it is not appropriate to apply (17-12) for the effect of a change in a dummy variable, or a change of state. The appropriate marginal effect for a binary independent variable, say, d, would be

$$\text{Marginal effect} = \text{Prob}[Y=1 \mid \bar{\mathbf{x}}_{(d)}, d=1] - \text{Prob}[Y=1 \mid \bar{\mathbf{x}}_{(d)}, d=0], \qquad \textbf{(17-14)}$$

where $\bar{\mathbf{x}}_{(d)}$, denotes the means of all the other variables in the model. Simply taking the derivative with respect to the binary variable as if it were continuous provides an approximation that is often surprisingly accurate. In Example 17.3, for the binary variable *PSI*, the difference in the two probabilities for the probit model is $(0.5702 - 0.1057) = 0.4645$, whereas the derivative approximation reported in Table 17.1 is 0.468. Nonetheless, it might be optimistic to rely on this outcome. We will revisit this computation in the examples and discussion to follow.