

## **Binary Response Models: Logits, Probits and Semiparametrics**

Joel L. Horowitz and N.E. Savin

**A** binary response model is a regression model in which the dependent variable  $Y$  is a binary random variable that takes on only the values zero and one. In many economic applications of this model, an agent makes a choice between two alternatives: for example, a commuter chooses to drive a car to work or to take public transport. Another example is the choice of a worker between taking a job or not. Driving to work and taking a job are choices that correspond to  $Y = 1$ , and taking public transport and not taking a job to  $Y = 0$ . The model gives the probability that  $Y = 1$  is chosen conditional on a set of explanatory variables. In the transportation example, common explanatory variables include the time and the cost of travel; in the worker example, common explanatory variables include age, education and experience.

The econometric problem is to estimate the conditional probability that  $Y = 1$  considered as a function of the explanatory variables. The most commonly used approach, notably logit and probit models, assumes that the functional form of the dependence on the explanatory variables is known. The logit and probit models have been used almost exclusively in econometric applications in the leading journals.

However, the functional form is seldom known in practice. If the functional form is misspecified, then the estimates of the coefficients and the inferences based on them can be highly misleading. It is possible to relax the restrictive assumption that the functional form is known by using either semiparametric or nonparametric models. In these types of models, the functional form is unknown. The problems of estimating semiparametric and nonparametric binary response models have

■ *Joel L. Horowitz is Professor of Economics, Northwestern University, Evanston, Illinois. N.E. Savin is Professor of Economics, University of Iowa, Iowa City, Iowa.*

generated considerable interest in recent years. This paper traces the evolution of estimation methods for binary response models.

## Review of Linear, Logit and Probit Models

This section reviews the binary response models most commonly used in applications. For simplicity, consider the case where the probability that  $Y$  takes on the value zero or one is conditional on a single explanatory variable  $X$ .

Suppose that the true conditional probability of  $Y = 1$  given  $X = x$  is  $P(Y = 1|X = x) = F(\beta_0 + x\beta_1)$ . In the parametric approach to modeling, the function  $F$  is known and the values of the parameters  $\beta_0$  and  $\beta_1$  are unknown.

The *linear probability model* specifies that the conditional probability is a linear function of  $X$ :  $P(Y = 1|X = x) = \beta_0 + x\beta_1$ . This model says that the probability of  $Y = 1$  increases by  $\beta_1$  for each one unit increase in  $X$ . This implies that the probability is negative if  $X$  is small enough and greater than one if  $X$  is large enough. Hence, this model has the defect that the conditional probability is not constrained to lie between zero and one.

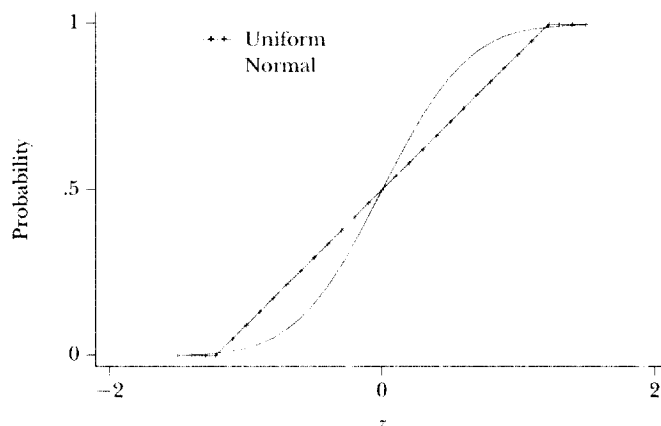
This defect can be corrected by replacing the linear function with one with a lower kink that keeps the conditional probability from being less than zero and an upper kink that keeps it from being greater than one. The two-kink function is illustrated in Figure 1; it is the cumulative distribution function of a uniform distribution. The cumulative distribution function of a probability distribution is an S-shaped curve that has a lower bound of zero and an upper bound of one. Thus, if  $F$  is a cumulative distribution function, the conditional probabilities are automatically constrained to lie between zero and one.

The kinks in the two-kink function are unrealistic, however. The commonly used cumulative distribution functions are smooth S-shaped curves. Such a curve is illustrated in Figure 1. It is the cumulative distribution function of a normal distribution.

A binary response model is referred to as a *probit model* if  $F$  is the cumulative normal distribution function. It is called a *logit model* if  $F$  is the cumulative logistic distribution function. The logistic and normal distributions are both symmetrical around zero and have very similar shapes, except that the logistic distribution has fatter tails. As a result, the conditional probability functions are very similar for both models, except in the extreme tails.

The estimation problem is to estimate the unknown parameters  $\beta_1$  and  $\beta_2$ . In practice, the linear probability model is estimated by fitting a straight line to the observations on  $X$  and  $Y$  by ordinary least squares. The ordinary least squares-based predictions of the conditional probability can be greater than one or less than zero. The logit and probit models are typically estimated by maximum likelihood. This is because the maximum likelihood estimator has good properties in large samples. In particular, it is asymptotically efficient; that is, it is the most precise estimator in large samples.

Figure 1  
Cumulative Probability Distributions



In the past, the main motivation for using ordinary least squares was computational effort. Software for ordinary least squares regression became widely available much earlier than software for maximum likelihood estimation of the logit and probit models. Now, maximum likelihood can be performed by any one of several well-known commercial statistical computing packages. There is no longer any computational advantage to using ordinary least squares.

## Biometric Background

The archetypical biometric study is one that analyzes the effect of a pesticide on an insect. The two outcomes are dead and alive. The dependent variable,  $Y$ , takes on the value one if the insect dies ("success") and zero if it survives ("failure"). The explanatory variable,  $X$ , is the dose of a pesticide. The data used in the study are produced by an experiment in which the experimenter controls the dose applied to each insect. Hence, the experimenter controls the values of  $X$ . A small number of  $X$  values are used: the number typically ranges between three and eight.

An important feature of the experiment is that many insects receive the same dose. This situation, which is referred to as one *with many observations per cell*, is illustrated by a classic pesticide experiment. In the experiment, batches of about fifty *Macrosiphoniella sanborni*, the chrysanthemum aphid, were sprayed with a series of concentrations of pesticide rotenone. The results are reported in Table 1. Given many observations per cell, the sample proportion of the dead insects can be computed for each cell, that is, for each value of  $X$  in the experiment. (The insects apparently dead, moribund or so badly affected as to be unable to walk more than a few steps were classified as dead for the purpose of the analysis.) These sample

Table 1

**Results of the Rotenone-*Macrosiphoniella sanborni* Experiment**

<i>Dose (in logarithms)</i>	<i>Number of Insects</i>	<i>Number Dead</i>	<i>Percentage Dead</i>
0.41	50	6	12
0.58	48	16	33
0.71	46	24	52
0.89	49	42	86
1.01	50	44	88

Source: Finney (1971).

proportions are natural estimates of the conditional probabilities that  $Y = 1$  for those  $X$  values that occur in the sample data.

The focus in biometric applications, however, is to obtain the conditional probability that  $Y = 1$  given  $X = x$  for *all* values of  $X$ , not just for the values of  $X$  in the data. One of the main objectives of pesticide studies is to estimate the dose that kills a certain proportion of the insects, say, 50 percent. To obtain a good estimate of this dose, the biometrician needs to have more information than is supplied by the sample proportions for the  $X$  values in the data. Specifically, a model is needed to estimate the probability that  $Y = 1$  at points between the values of  $X$  in the data.

The probit model has been the dominant model in biometrics. The leading textbook in biometrics for many years was *Probit Analysis* by Finney (1971). The experience of one of us (Savin) with pesticide studies suggests that both logit and probit models provide good fits to samples from laboratory-reared colonies. The two models give similar predictions except for extreme values of the dose. There is no compelling biological reason, however, to adopt either the logit or the probit specification.

A key question in economics is whether models are “structural models,” that is, whether the values of the parameters do not change across certain environments, such as policy regimes. A similar problem arises with biometric data. Savin and his colleagues reviewed a series of studies in which the probit model was estimated for different generations of laboratory colonies of the same species and similarly for the logit model. Their findings overwhelmingly rejected the hypothesis that the parameter values were the same for different generations (Savin, Robertson and Russell, 1977). This is consistent with Finney (1964, pp. 91–92): “In general, the assumption that a response curve once determined can be used in further assays [analysis] is inadmissible.”

The assumption that the binary response function is structural should be treated with caution. If it does not hold up in laboratory-controlled colonies, then there is reason to believe that it may not hold up in the real-world evidence studied by economists. However, binary response studies in economics are seldom repeated, so it is more difficult to detect if there is parameter constancy. One notable exception is mode choice for travel to work. The transferability of the parameters of mode choice models among cities was studied intensively in the 1970s.

The major controversy in biometrics has not been over logit versus probit modeling, but over estimation methods. For some decades, there was controversy over whether to use the minimum chi-square method proposed by Berkson (1944) or to use maximum likelihood. The controversy, which was resolved in favor of maximum likelihood, is briefly summarized in Amemiya (1985).

## Econometric Applications

The 1970s witnessed an upsurge of models involving discrete dependent variables in econometrics, as documented by Amemiya (1981). In economic applications, the data typically come from surveys rather than controlled experiments. As will be explained, the nature of survey data has tipped the balance in favor of maximum likelihood estimation and the logit model, at least until recently.

In survey data, there are often few observations on the dependent variable per cell. The reason for few observations per cell is that survey data often involve several explanatory variables, some of which are continuous. In the case of a continuous variable, it is easy to see that there may be a large number of sparsely populated cells. But this can happen even if the explanatory variables take on a modest number of discrete values. For example, if there are four explanatory variables that can each take on ten values, there are  $10^4 = 10,000$  possible cells. Hence, it is not surprising that there are often few observations per cell for data with multiple explanatory variables.

An advantage of maximum likelihood estimation is that it is feasible when there are few observations per cell, which includes the case of no observations in some cells. The logit work trip models considered by Domencich and McFadden (1975) illustrate the case of few observations per cell. The models explain the probability that an individual chooses auto over mass transit. The transit mode is primarily bus, but there are also some streetcar trips. In one version of the model, the explanatory variables are transit walk time (in minutes), auto-in-vehicle time (in minutes) minus transit station-to-station time (in minutes), auto parking charges (in dollars) minus transit fare (in dollars), and autos per worker in the household. The sample employed to estimate this version of the model consists of 115 observations on individual trip makers. In this example, there are many cells with no observations and some with only one observation per cell. The maximum likelihood estimates of the parameters and the  $t$ -ratios are presented in Table 2. The estimated coefficients indicate, among other things, that the value of transit walk time is \$3.94/hour ( $60 \times 0.147/2.24$ ), whereas the value of transit station-to-station time is only \$1.10/hour ( $60 \times 0.0411/2.24$ ). In fact, most analyses of urban travel behavior find that travelers place a higher value on walk time than on in-vehicle time.

In survey data, there are often three or more possible responses. Models in which a single dependent variable takes on three or more discrete and unordered values are usually called *multinomial* response models. Multinomial

Table 2

**Maximum Likelihood Estimates of Logit Work Trip Model**

<i>Variable</i>	<i>Parameter</i>	<i>Asymptotic t-ratio</i>
Constant	-4.76	-3.79
Transit walk time	0.147	-2.62
Auto-in-vehicle time minus transit station-to-station time	-0.0411	-2.05
Auto parking charges minus transit fare	-2.24	-4.53
Autos per worker in the household	3.78	4.06

Source: Domencich and McFadden (1975).

models have been extensively used in urban travel demand studies; for example, a work trip may be taken using one of three travel modes: bus, carpool or driving alone. They have also been used in many other applications. For example, in an ownership study, the household may own an electric dryer, a gas dryer or no dryer.

Multinomial versions of the logit and probit models have been the subject of much attention in the econometric literature; see, for example, Amemiya (1985), Mittelhammer, Judge and Miller (2000) and Ruud (2000). The multinomial logit specification is attractive on analytical grounds because of an important and elegant result by McFadden (1974), which shows that the multinomial logit model can be derived from utility maximization under certain conditions and that the probabilities have simple closed-form expressions.<sup>1</sup> The logit specification also became attractive in the 1970s and 1980s on computational grounds because maximum likelihood estimation of the logit model was feasible with the computing technology available at that time.

One danger in using the multinomial logit model is that it can produce misleading inferences when some of the alternatives are close substitutes. This arises because the multinomial logit specification imposes the restriction that the odds (the ratios of the probabilities) of choosing the  $j$ th alternative over the  $i$ th depend only on the characteristics of those two alternatives. In other words, the characteristics of any other alternatives in the choice set have no influence on the odds between the  $i$ th and  $j$ th alternatives. This feature is called the *independence from irrelevant alternatives* property.

McFadden's red bus and blue bus example illustrates the problem caused by independence from irrelevant alternatives. Suppose the initial transportation choice is between driving and taking a red bus. Assume individuals are split fifty-fifty between driving and taking the red bus, which implies the odds of taking the red

<sup>1</sup> As an example, suppose an individual has three alternatives, and the utility associated with the  $j$ th alternative is  $U_j = \mu_j + \varepsilon_j$ ,  $j = 0, 1, 2$ , where  $\mu_j$  is a nonstochastic function of the explanatory variables and unknown parameters and  $\varepsilon_j$  is an unobservable random variable that follows the Type I extreme value (log Weibull) distribution. Then the probability of an individual choosing the first alternative,  $P(Y = 1|X = x)$ , is  $e^{\mu_1}$  divided by  $e^{\mu_1} + e^{\mu_2} + e^{\mu_3}$ , and similarly for the other alternatives.

bus over driving are 1:1. Now, suppose a new, rival bus company introduces a blue bus, which aside from color is indistinguishable from the red bus. Then it is reasonable to expect that the bus riders split their trips evenly between the two buses and that car drivers will continue to drive.

If the independence from irrelevant alternatives property holds, however, then the odds of taking the red bus over driving still would be 1:1. The odds of taking the red bus over the blue bus must also be 1:1, because the two types of bus are indistinguishable except for color. Therefore, because the probabilities of the three alternatives have to sum to one, independence from irrelevant alternatives implies that the probabilities of driving, taking the red bus and taking the blue bus must be equal to one-third. Hence, the multinomial logit model gives the counter-intuitive prediction that one-third  $((\frac{1}{2} - \frac{1}{3})/(\frac{1}{2}) = \frac{1}{3})$  of the car drivers will switch to the bus without any real change in the alternatives simply because there are now both blue and red buses.

The practical message here is that multinomial logit specification should only be used in applications where the alternatives are dissimilar. When the multinomial logit specification is not plausible, other models can be used. One is the so-called nested multinomial logit model (McFadden, 1977, 1981). Another model, which is more flexible than the nested logit model, is the multinomial probit model. Until recently, the application of the multinomial probit model with more than three or four choices was computationally difficult; it involves evaluating multiple integrals, which is computer intensive. But due to dramatic increases in computer speed and the recent development of new computing algorithms, simulation-based estimation of multinomial probit models is now quite feasible.<sup>2</sup> Currently available software for implementing these methods is cited in Geweke and Keane (2001).

## An Earnings Example

This section uses an earnings example to illustrate the differences between a linear probability model and a logit model. The example consists of estimating the probability that an individual's weekly wages are below \$280 a week, conditional on education and experience. Following a commonly used approach in labor economics, the general specification is

$$P(Y = 1 | ed, ex) = F(\beta_0 + \beta_1 ed + \beta_2 ex + \beta_3 ex^2),$$

<sup>2</sup> The ideas motivating the simulation estimation methods are due to Lerman and Manski (1981), McFadden (1989) and Pakes and Pollard (1989). Several approaches that combine estimation and simulation are described in Hajivassiliou and Ruud (1994) and Geweke and Keane (2001). Moreover, Geweke, Keane and Runkle (1994, 1997) show that several competitive simulation-based methods yield reliable inference in finite samples.

where  $F$  is a function that depends on the model being estimated,  $Y = 1$  if the individual's weekly wage is less than or equal to the \$280,  $ed$  is years of education, and  $ex$  is years of experience. A weekly wage of \$280 is close to the tenth percentile of weekly wages of all individuals in the data set. The data are taken from the 1993 Current Population Survey (CPS). They consist of observations of weekly wages, years of education and years of experience for 3,123 full-time, full-year white males who were employed in a metropolitan area in the north central region of the United States.

According to the linear probability model,

$$Y = \beta_0 + \beta_1 ed + \beta_2 ex + \beta_3 ex^2 + U,$$

where  $U$  is an unobserved random variable whose mean is zero. The binary logit model is

$$P(Y = 1 | ed, ex) = F_L(\beta_0 + \beta_1 ed + \beta_2 ex + \beta_3 ex^2),$$

where  $F_L$  is the cumulative logistic distribution function,  $F_L(z) = 1/(1 + e^{-z})$ , and

$$z = \beta_0 + \beta_1 ed + \beta_2 ex + \beta_3 ex^2.$$

We estimated the linear probability model by ordinary least squares and the binary logit model by maximum likelihood.

The estimated coefficients and standard errors are reported in Table 3. For both models, the estimated coefficients are different from zero at the usual significance levels. The models can be easily compared by plotting the predicted probabilities of earning less than \$280. Figure 2 shows the predictions of  $P(Y = 1 | ed, ex)$  by the two models as a function of experience,  $ex$ , which is on the horizontal axis, when the level of education is fixed. For each model, there are two curves: one curve shows the probability of earning less than \$280 for those with a high school education only (12 years of education,  $ed = 12$ ) and the other for those with a college education only (16 years of education,  $ed = 16$ ).

The predictions from the linear probability and logit models are very different, so the two models are not substitutes for one another, and the linear model is not a good approximation to the logit model. Most notably, the linear model gives negative predicted probabilities for  $ed = 16$  over a wide range of experience levels, namely,  $19 \leq ex \leq 40$ . This implies that the linear model is badly misspecified and provides a poor fit to the data.

In many applications, the change in the conditional probability function with respect to a change in the explanatory variables (the derivative) is of interest, especially when analyzing the effects of policy changes. Given a fixed value for education, the derivative with respect to experience is

$$\partial F / \partial ex = \beta_2 + 2\beta_3 ex$$

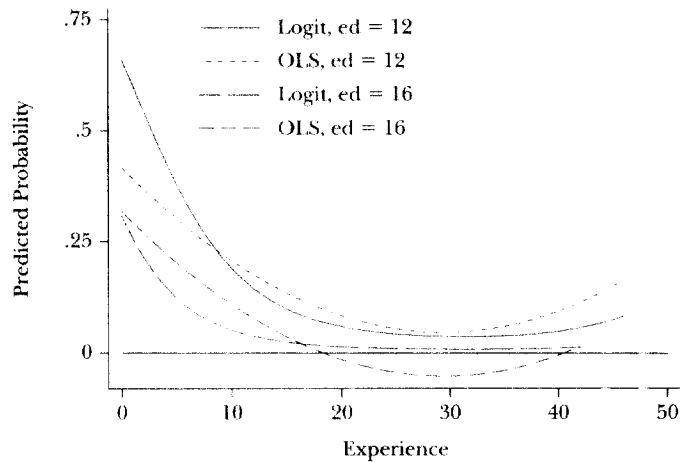


Table 3  
Coefficients of Linear and Logit Models

Model	Variable	Coefficient	Std. Error
Linear	Intercept	0.715	0.031
	Experience	-0.025	0.002
	Experience squared	0.0004	0.00001
	Education	-0.025	0.002
Logit	Intercept	5.068	0.422
	Experience	-0.251	0.020
	Experience squared	0.004	0.0004
	Education	-0.367	0.029

Figure 2

Predicted Probability that Weekly Wages are Less Than or Equal to \$280



Source: 1993 Current Population Survey.

for the linear probability model and

$$\partial F / \partial ex = P(Y = 1 | ed, ex)P(Y = 0 | ed, ex)(\beta_2 + 2\beta_3 ex)$$

for the logit model. Figure 3 shows the derivatives of the linear probability and logit models. The derivative function of the linear probability model is an upward sloping straight line, a linear function of experience. By contrast, the derivative function of the logit model is a curve, a nonlinear function of experience.

The derivative of the linear probability model and the derivative of the logit model have the same value only where the straight line and the curve intersect. Thus, it is clear from Figure 3 that the straight line is not a good approximation to the curve for  $ed = 12$ ,  $ed = 13.5$  or  $ed = 16$ . Table 4 reports the derivatives evaluated

Figure 3

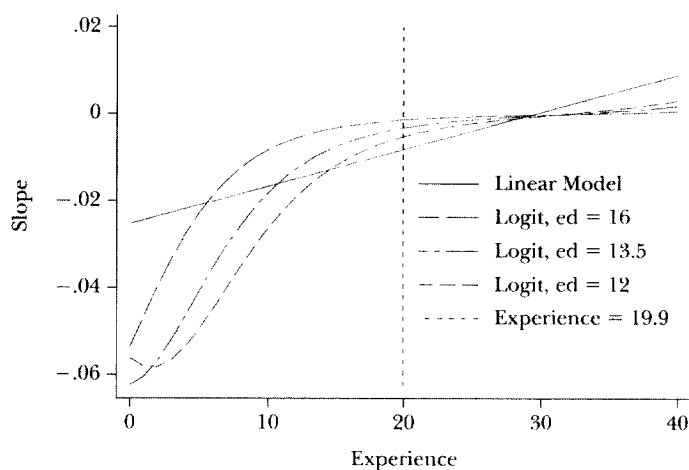
**Derivatives of Probability Function with Respect to Experience for Linear and Logit Models**

Table 4

**Derivatives of Conditional Probability Function**

Model	Education	Experience = 19.9	Mean Experience Conditional on Education	Experience = 5	Experience = 10
Derivatives with Respect to Experience					
Linear	12		-0.07	-0.021	-0.017
Logit	12		-0.004	-0.050	-0.027
Linear	16		-0.012	-0.021	-0.017
Logit	16		-0.003	-0.023	-0.008
Linear	13.5	-0.008			
Logit	13.5	-0.003			
Nonparametric	12		-0.003	-0.058	-0.031
Derivatives with Respect to Education					
Linear	13.5	-0.025			
Logit	13.5	-0.013			

*Notes:* The mean values of experience conditional on education are 21.4 and 16.3, respectively, for 12 and 16 years of education. The unconditional mean values of experience and education are 13.5 years and 19.9 years, respectively.

at three values of experience, including the mean value of experience conditional on the educational level. The table also shows derivatives of the probability function at the unconditional mean values of education and experience. Derivatives are often evaluated at the mean values of the explanatory variables. The results in Table 4 show that the derivative at the mean value may be a poor approximation of the

derivative evaluated at other values and that the derivative of the linear probability model at the mean of experience may be a poor approximation to the derivative of the logit model at the mean.

## Nonparametrics to Semiparametrics

As noted in the introduction, there is usually little justification for assuming that the conditional probability function is known. Again, consider estimating the probability that weekly wages are below \$280 a week, conditional on education and experience. The conditional probability function was previously estimated assuming the logit model is correct. Another approach is to estimate the conditional probability function nonparametrically. In nonparametric estimation, no assumptions are made about the form of the function or the nature of the dependence on the explanatory variables (Härdle, 1990). Figure 4 displays the predictions as a function of  $ex$  for  $ed = 12$  for the logit model estimated by maximum likelihood and for the nonparametric alternative.<sup>3</sup> The predictions of the logit and nonparametric models differ noticeably, especially when  $ex < 20$ . A formal test based on Horowitz and Spokoiny (2001) confirms that the predictions are, in fact, significantly different. This implies that the logit model is misspecified.

Of course, the fit of the logit model may be improved by adding powers of experience. Searching over the powers amounts to informal nonparametric fitting. The performance of such an informal specification search depends on the search rule (what powers to add) and the stopping rule (how many powers to add). With the search and stopping rules typically employed in practice but seldom explicitly cited, informal nonparametric fitting is inconsistent; it does not recover the true conditional probability function as the sample size increases.

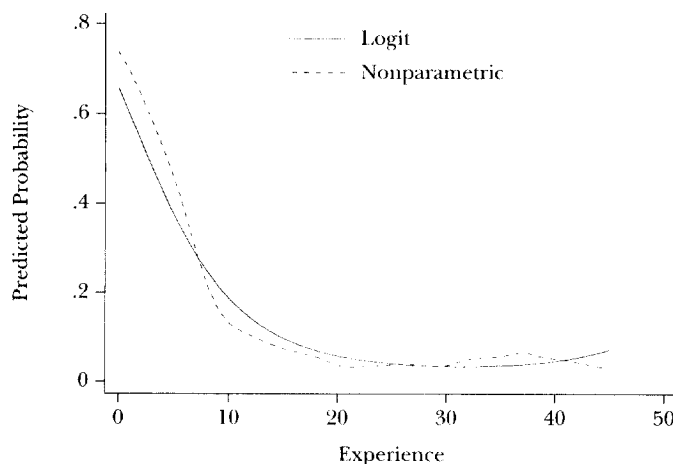
The nonparametric approach maximizes flexibility in that it imposes no distributional assumptions on  $F$ . Matzkin (1992, 1993) has studied nonparametric estimation of structural binary response models. However, the price of the flexibility of the nonparametric models can be high for several reasons. One is that estimation precision decreases rapidly as the number of explanatory variables increases. This is due to the so-called curse of dimensionality. As a result, impractically large samples may be needed to obtain acceptable estimation precision when there are multiple explanatory variables. A second problem with nonparametric estimation is that its results can be difficult to display, report and interpret when there are multiple explanatory variables. A third problem is that nonparametric estimation does not permit extrapolation.

The distinction between semiparametric and nonparametric models is that a semiparametric model makes assumptions that avoid the curse of dimensionality. Often, the resulting model includes an unknown finite-dimensional parameter.

<sup>3</sup> We used the Nadaraya-Watson kernel estimation method with the standard normal density function for the kernel. The bandwidth was chosen by least squares cross-validation, which yielded a bandwidth value of 1.6.

Figure 4

**Predicted Probability that Weekly Wages are Less than or Equal to \$280,  $ed = 12$**



Thus, the semiparametric approach is a halfway house between the parametric and nonparametric approaches. It imposes less structure than does the parametric approach, but more structure than the nonparametric approach imposes, and thus it is a compromise between restrictive distributional assumptions on the one hand and flexibility on the other.

This section briefly introduces two approaches to semiparametric estimation, one based on *single-index modeling* and the other on the binary response version of the *median* regression model. These two approaches are the consequence of different sets of identification conditions. The identification problem is to find conditions under which the parameters are uniquely determined by the population distribution of the data and auxiliary assumptions. Manski (1988) has shown that in semiparametric settings the conditions required for identification are often subtle.

In the case of the single index model, the conditional probability that  $Y = 1$  given  $X = x$  has the form  $P(Y = 1|X = x) = G(x\beta)$ , where  $\beta$  is an unknown  $K \times 1$  constant vector and  $G$  is an *unknown* distribution function. The quantity  $x\beta$  is called an *index*, where, in general,  $X$  is a  $1 \times K$  random vector. If there are two explanatory variables, the index is  $x_1\beta_1 + x_2\beta_2$ . The semiparametric estimation problem is to estimate both  $\beta$  and  $G$  from observations on  $X$  and  $Y$ . Note that the above specification is more flexible than are the logit and probit binary response models. The logit (probit) model is a special case obtained by assuming the logistic (normal) cumulative distribution function for  $G$ . Under the appropriate conditions, the asymptotically efficient estimator of  $\beta$  in a single-index model is the semiparametric maximum likelihood estimator of Klein and Spady (1993).

Another approach specifies that  $Y = 1$  if  $X\beta + U > 0$  and  $Y = 0$  otherwise, where  $U$  is an unobserved random variable whose *median* is zero. Except in special cases, the resulting model is not a single-index model. Manski (1985) proposed the

*maximum score estimator* for estimating  $\beta$  consistently. Horowitz (1992) modified this estimator to improve its precision and other statistical properties. See Greene (2000) for a textbook presentation of the maximum score estimator and an empirical application.

This section concludes with an empirical example that illustrates the usefulness of semiparametric single-index models. The example is taken from Horowitz and Härdle (1996) and consists of the estimating of a model of product innovation by German manufactures of investment goods. The estimation of a binary probit model appeared to produce reasonable estimates for the parameters and the probability of an innovation. If the probit model is correct, the first derivative of  $G$  is the standard normal density. Semiparametric estimation revealed, however, that the first derivative of  $G$  was bimodal. This is an important feature of the data that could not easily be found by standard parametric methods. The product innovation study demonstrates that usefulness of a semiparametric single-index model in reducing the risk of obtaining misleading results.

## Concluding Comments

The focus of this paper has been on the evolution of more flexible estimation methods in binary response and its multinomial cousin. For the sake of brevity, many topics involving binary variables have been omitted. These include random coefficient models, panel data models and choice-based sampling. Moreover, the evolution in estimation methods is ongoing both in classical and Bayesian econometrics, although this paper only considers classical econometrics. Geweke, Keane and Runkle (1994, 1997) present some recent developments using the Bayesian approach. The expectation is that flexible methods will become more widely employed with the development of user-friendly software. The new methods promise to offer more insightful data analysis and, in turn, to provide a sounder basis for policy and forecasting.

■ *The authors thank John Geweke, Brad De Long, Alan Krueger, Jim McDonald, Forrest Nelson, Timothy Taylor, Shawn Ulrick, Michael Waldman and Allan Würtz for their help and constructive comments on earlier draft material. The research of Joel L. Horowitz was supported in part by NSF Grant SES-9910925.*

## References

- |   |   |
|---|---|
| <p><b>Amemiya, Takeshi.</b> 1981. "Qualitative Response Models: A Survey." <i>Journal of Economic Literature</i>. December, 19:4, pp. 1483–536.</p> | <p><b>Amemiya, Takeshi.</b> 1985. <i>Advanced Econometrics</i>. Cambridge, Mass.: Harvard University Press.</p> |
|   | <p><b>Berkson, Joseph.</b> 1944. "Application of the</p>  |

- Logistic Function to Bio-Assay." *Journal of the American Statistical Association*. 39, pp. 357–65.
- Domencich, Tom A. and Daniel McFadden.** 1975. *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland Publishing Co.
- Finney, David J.** 1964. *Statistical Method in Biological Assay*. New York: Hafner.
- Finney, David J.** 1971. *Probit Analysis*, Third Edition. Cambridge: Cambridge University Press.
- Geweke, John and Michael Keane.** 2001. "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics, Volume V*. James J. Heckman and Edward E. Leamer, eds. Amsterdam: North-Holland. Forthcoming.
- Geweke, John, Michael Keane and David Runkel.** 1994. "Alternative Computational Approaches to Statistical Inference in Multinomial Probit Models." *Review of Economics and Statistics*. November, 76:4, pp. 609–32.
- Geweke, John, Michael Keane and David Runkel.** 1997. "Statistical Inference in the Multinomial Multiperiod Probit Model." *Journal of Econometrics*. September, 80:1, pp. 125–65.
- Greene, William H.** 2000. *Econometric Analysis, Fourth Edition*. Upper Saddle River, N. J.: Prentice Hall.
- Hajivassiliou, Vassilis A. and Paul A. Ruud.** 1994. "Classical Estimation for LDV Models Using Simulation," in *Handbook of Econometrics, Volume IV*. Robert F. Engle and Daniel L. McFadden, eds. Amsterdam: North-Holland, pp. 2383–441.
- Härdle, Wolfgang.** 1990. *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.
- Horowitz, Joel L.** 1992. "A Smoothed Maximum Score Estimator for Binary Response Models." *Econometrica*. May, 60:3, pp. 505–31.
- Horowitz, Joel L. and Wolfgang Härdle.** 1996. "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates." *Journal of the American Statistical Association*. December, 91: 436, pp. 1632–40.
- Horowitz, Joel L. and Valdimir G. Spokoiny.** 2001. "An Adaptive, Rate Optimal Test of a Parametric Mean-Regression Model against a Nonparametric Alternative." *Econometrica*. May, 69:3, pp. 599–631.
- Klein, Roger W. and Richard H. Spady.** 1993. "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica*. March, 61:2, pp. 387–421.
- Lerman, Steven R. and Charles F. Manski.** 1981. "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis in Discrete Data with Econometric Applications*. Charles F. Manski and Daniel L. McFadden, eds. Cambridge, Mass.: MIT Press, pp. 305–19.
- Manski, Charles F.** 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator." *Journal of Econometrics*. March, 27:3, pp. 313–33.
- Manski, Charles F.** 1988. "Identification in Binary Response Model." *Journal of the American Statistical Association*. September, 83:403, pp. 729–38.
- Matzkin, Rosa L.** 1992. "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold-Crossing Models." *Econometrica*. March, 60:2, pp. 239–70.
- Matzkin, Rosa L.** 1993. "Nonparametric Identification and Estimation of Polychotomous Choice Models." *Journal of Econometrics*. July, 58: 1-2, pp. 137–68.
- McFadden, Daniel L.** 1974. "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*. Paul Zarembka, ed. New York: Academic Press, pp. 105–42.
- McFadden, Daniel L.** 1977. "Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments." Cowles Foundation Discussion Paper No. 474.
- McFadden, Daniel L.** 1981. "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*. Charles F. Manski and Daniel L. McFadden, eds. Cambridge, Mass.: MIT Press, pp. 198–272.
- McFadden, Daniel L.** 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica*. 57:5, pp. 995–1026.
- Mittelhammer, Ron C., George G. Judge and Douglas J. Miller.** 2000. *Econometric Foundations*. Cambridge: Cambridge University Press.
- Pakes, Ariel A. and David Pollard.** 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*. 57:5, pp. 1027–57.
- Ruud, Paul A.** 2000. *An Introduction to Classical Econometric Theory*. Oxford: Oxford University Press.
- Savin, N.E., Jaqueline L. Robertson and Robert M. Russell.** 1977. "A Critical Evaluation of Bioassay in Insecticide Research: Likelihood Ratio Tests of Dose-Mortality Regression." *Bulletin of the Entomological Society of America*. 23:4, pp. 257–66.