Chapter 10
Qualitative Variables

William H. Greene

Abstract Qualitative dependent variables are nonquantitative indicators of the occurrence of economic events or decisions or behavioral outcomes, such as commuter choices of travel mode, insurance purchase decisions, choices of product brands, movie ratings, opinion surveys or self-assessments of health or wellbeing. Unobservable latent variables represent ideas, tendencies or concepts, such as permanent income, health, preferences about public transportation or strength of preference for environmental stewardship.

Ragnar Frisch (1933) introduced an Econometrics methodology in the inaugural issue of *Econometrica*. Econometric models for qualitative variables have been developed since the 1960's with the publication of many social science data sets, such as the U.S. *National Longitudinal Survey* (NLS) and the Michigan *Panel Study of Income Dynamics* (PSID). Applications that involve qualitative variables are modeled with a specific econometric methodology and set of empirical methods.

There has been a robust stream of theoretical research on methods and nonlinear models for qualitative outcomes that has produced a standard approach that has guided empirical research. Recently, the 'Credibility Revolution' (Angrist & Pischke, 2010) has called tightly specified nonlinear models into question and recommended simple statistics and basic linear regression. This chapter recounts the history and surveys the ebb and flow of the development of econometric models for qualitative variables.

William H. Greene
University of South Florida, Tampa, Florida, USA and New York University (ret.), New York, USA,
e-mail: wgreene@stern.nyu.edu

10.1 Introduction

The theme of this book is the history of econometric methods and models, including why some approaches have remained in the canon ('aged well'), while others have lost their appeal (or never gained it) as the methodology evolved. The subject of this chapter is econometric models for qualitative variables, defined here as indicators of discrete outcomes and unobservable latent attitudes or tendencies. This type of mathematical model building originated in *Bioassay* in the 1930's. Chester Bliss (1934) proposed a function to relate a 'stimulus,' the dosage level of a pesticide, to a pest's random discrete 'response,' to die or survive.[1] Bliss points toward the modern literature. Thurstone (1927) and Luce (1959) are early contributors to the maximum random utility theory that underlies the modern form of discrete choice modeling.

Discrete Choice Analysis originated in ca. 1965-1970 (e.g., Theil, 1971). The field has evolved and matured, and techniques and model forms have come and gone – for example, the Linear Probability Model (which has gone and recome) and the Multinomial Probit Model. The progression of discrete choice modeling has seen a stream of refinement and sharpening of ever more detailed models. With an abrupt reconsideration, the Credibility Revolution, ca. 2000 (Angrist & Pischke, 2010) advocated avoiding nonlinear models in favor of simple statistics such as OLS and 2SLS. The probit binary choice model is the standard framework for pedagogy about nonlinear econometric modeling. Some of the received econometric theory originates in discrete choice modeling. For example, what little is known in general about the Incidental Parameters problem in fixed effects models is anchored by the binary logit model.[2]

This chapter examines econometric models for qualitative variables.[3] Section 10.1 defines qualitative variables. Section 10.2 focuses on the early history of the probit model that is the keystone of discrete choice modeling. We review the early developments that led to the formation of a standard model for binary choices. This section introduces the range of model specifications in the modern literature. Section 10.3 gives a broad timeline of developments in discrete choice modeling since the mid-1960's. Section 10.4 describes large segments of qualitative dependent variables interests. Section 10.5 presents a timeline of specific innovations that together form modern methods of qualitative choice modeling. Sections 10.6 and 10.7 discuss some issues and trends in econometric methodology relating to qualitative choice modeling including a turning point ca. 2005. Section 10.8 concludes.

---

[1] See, also, Gaddum (1933), Fisher (1935), Amemiya (1973) and Finney (1947, 1971).

[2] See, e.g., Wooldridge (2010), pp. 612 and 619-620 and Abrevaya (1997).

[3] Advanced theoretical treatments can be found in Wooldridge (2010) and Hansen (2022). Technical surveys of most of the models examined in this chapter can be found in Greene and Hensher (2010) on Ordered Choice Models, Hensher, Rose and Greene (2015) on Multinomial Discrete Choice Models and Greene and Hensher (2026) on Hybrid Choice Models. See also, C. Cameron and Trivedi (2005) and Wooldridge (2010) for general references on Microeconometrics including treatments of discrete choice models. McFadden (1976) provides an exhaustive survey of the first decade of qualitative choice modeling with a focus on multinomial choice theory and models.

10.2 The Method of Probits

Some foundations of current discrete choice modeling were laid by bioassayist Chester Bliss (1934) in his note, *The Method of Probits*. Studying the dosage response of a pest population to an insecticide, Bliss noted: "*[T]he result of an investigation of the action of a toxic agent upon the mortality of an organism is usually expressed as an asymmetrical S-shaped curve in which the percentage mortality of each set of individuals is related to the dosage to which it has been exposed ...*" (page 38). Dosages approaching a 100% kill (denoted Lethal Dose 100 or LD100) are of interest. Some of Bliss's results appear in Figure 10.1.[4] The relationship is flattening above 75% and below 25%, suggesting a 'sigmoidal' shape.

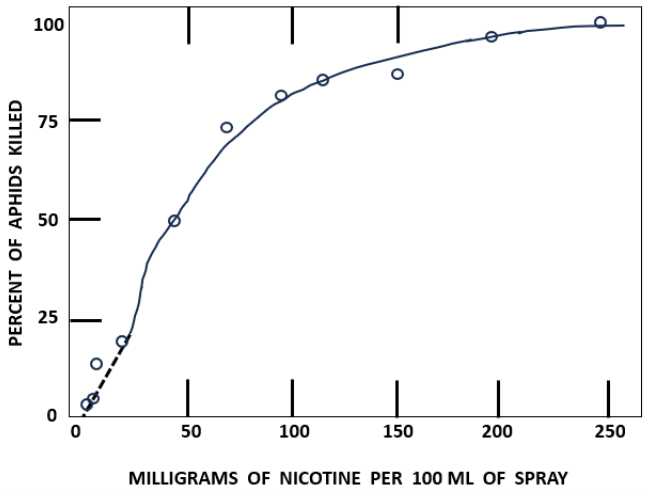

Fig. 10.1: Bliss Dosage Response Curve

Bliss expressed some skepticism of the accuracy of the then-current practice of hand drawing the curve by adjusting it in small segments of the range of dosages rather than to the data set as a whole. It would be preferable to transform the percentage Kill Rate and Dosage to "*units which may be plotted as straight lines on ordinary cross-section paper and hence permit fitting by the customary technique of least squares or of the straight-line regression equation*" (p. 34). The experimental data consist of batches of N_i insects treated with dosage D_i, number killed, K_i and kill rate, $R_i = K_i/N_i$. The coefficients are computed by least squares regression of the

[4] We have redrawn Bliss's Figures 1 and 2 in Figures 10.1 and 10.2. LD50 and LD99 are added to Figure 10.2.

probits of R_i on a constant term and $\log D_i$.[5] The plotted function is the estimated 'dosage response curve.'[6]

Bliss was interested in fitting a curve to the dosage/response data. The suggested method contains elements of an underlying theory and an econometric framework. The shape of the curve suggested a normal CDF based on other observed experiments. The unsystematic variation in the response data arose from heterogeneity among individuals in a population in their susceptibility to the insecticide. Bliss surmised that the shape of the response function followed from the normality of the distribution of resistance among the members of the underlying population. He also noted that dosage minus resistance acted as a threshold for whether the pest would survive or die. The probit-linear index function form of the estimated curve is the same as the one we will use later. Bliss noted the possibility of using other functions, but Pearson's table for the normal distribution was convenient. (The table was reproduced in the note.) He did note that the curvature of the function was essential.

10.2.1 A Dosage Response Function Based on Individual Data

Daniel Finney (1943, 1947) considered how to extend the method of probits to individual data. The context was a biological measurement involving 39 patients who underwent a treatment and either 'responded,' with y = 1 or 'did not respond,' with y = 0. He began with a model for the 'quantal' (binary) response of a single individual, He devised an iterative algorithm for computing the parameters that he expected to approximate Maximum Likelihood (Finney, 1947, p.34). Finney remarks that his development appears to be the first to approach probit analysis with individual data. If so, this 1947 method with individual data and two dosage factors would be a precursor to modern probit (discrete choice) analysis. We note, though, this is still curve fitting for a dosage response. Finney has not yet identified a source of random variation in the data. The objective function minimizes the approximation error in the fitted function.

There is an important distinction between Bliss's and Finney's dosage response curves. Bliss's function does not predict that any individual will die. It does predict that a higher dosage will be associated with a higher kill rate - more pests will die when the dosage is increased. Finney's response curve appears to relate individual outcomes to the stimulus. But, the curve should be interpreted to state only that with increased stimulus, any individual is more likely to die. The model makes the same prediction as Bliss's. In Bliss's case, the ordinate of the response curve represents the actual proportion of a group of insects that will die. For Finney's case, the ordinate is interpreted as the likelihood that an individual will die. One could invoke a law

[5] The normit of R_i is the inverse normal transformation. The probit (probability unit) is computed as the normit+5 to avoid dealing with negative numbers. R_i must be inside (0,1) for all i. The normit cannot be computed for $R_i \leq 0$ or $R_i \geq 1$.

[6] Maritz (1965) suggested the fit could be improved by expanding to a polynomial function of the dosage.

of large numbers to suggest that Bliss's curve would 'converge' to Finney's with a sufficient number of batches (if Finney were examining the same pest). But, in fact, it was the physical (finite sample) response data that were of interest.

10.2.2 Fitting the Logit and Probit Models

Finney did not state a particular reliance on the normal distribution. He could equally have used a logistic distribution. The counterpart to the normit would have been the simpler logit of the probability[7] (see Berkson, 1951). The normal distribution was sufficiently convenient. Finney's algorithm did not rely directly on the probit likelihood function. It did rely on some computations based on the normal density found in the Fisher and Yates tables (Fisher & Yates, 1938). We re-estimated the model using Finney's data and a modern maximum likelihood probit estimator and obtained values that differed by only 0.4%. Walker and Duncan (1967) developed a minimum chi-squared estimator for the (α, β) of a simple logit model. Goldberger (1964) and Theil (1971) suggested a generic log likelihood that showed how the current estimator for the probit model would appear.

Based on the empirical shape of the observed data, Bliss and Finney found that the response curve resembled a normal CDF. Bliss transformed the data to produce a linear figure that could be fit using least squares (see Figure 10.2 below). Finney's fitting method minimized the estimation error using a technique that resembled maximum likelihood. Walker and Duncan (1967) used a 'Generalized Linear Model' approach. For any individual, the binary outcome is governed by a Bernoulli process with a conditional mean function using the normal CDF. They devised a minimum chi-squared estimator. Berkson (1951) suggested that the calculations would be more convenient if they were based on the logistic function rather than the normal CDF. The numerical differences between the two were small.

Theil (1971) relied on the notion of a behavioral outcome. His analysis began by applying Bliss's logic to the individual data. He proposed the probit log-likelihood as a fitting criterion. Further thought suggested a 'potential purchase' argument based on Tobin (1958)'s censored regression model. Theil based his suggestions on the probit model, but suggested that the logit (and infinitely many other) models would work as well. He did note "*It is easily seen that (5.1)* [the linear model] *cannot be an attractive specification*" (p. 628). Several other applications were based on the logit model, including Theil (1969, 1970).

10.2.3 Thresholds and Random Utility

Elements of the current standard model for binary dependent variables appear in Bliss's and Finney's method of probits. The response function is a conditional

[7] The logit of P equals log[P/(1-P)].

probability. The data generating mechanism is precisely the threshold model that underlies all of the specifications of modern discrete choice models. In modern terms, the dosage would be a treatment or stimulus. The individual responds by making a choice. In Bliss's application, nature endows each pest with their resistance. If the dosage is sufficient, the insect 'responds' to the treatment by dying. In "*Shedding Light on the Light Bulb Puzzle: The Role of Attitudes and Perceptions in the Adoption of Energy Efficient Light Bulbs*," Di Maria, Ferreira and Lazarova (2010) analyzed Irish household data on adoption of a new technologys. The random element that enters the choice is 'preference' for the environment. The choice to 'Adopt' the CFL lightbulb technology occurs if the intrinsic preference for the environment exceeds the reluctance to adopt a new technology.

Applications that involve random choice between two alternatives are ubiquitous. The model is typically used for describing behavior. That requires some assumption about the choice mechanism. The most common approach is maximum utility; The individual chooses alternative 1 if net utility is positive and alternative 0 otherwise. This arrangement is labelled a 'threshold model.' The difference of the two utilities is compared to the threshold value. The modeling framework is commonly called 'Random Utility Modeling' or RUM. Two crucial aspects of the framework are: (1) Individuals with the same characteristics facing the same alternatives may make different choices. The random elements of the utility functions produce this result. (2) The utilities are unobserved. The difference is only partially observed. The sign of the difference is revealed through the choice made. Data that are based on ex post observed choices are 'Revealed Preference data.' A large segment of the literature involves surveys in which individuals are asked questions of the sort "If you were offered these two alternatives, which would you choose?" These are 'Stated Preference data.'

10.2.4 Normality: Is Probit the Right Model?

When the behavioral assumptions (RUM) were added to the specification and individual coefficients and partial effects became of interest, the role of functional form became an issue. There is considerable discussion around whether the normal distribution is 'right' or 'wrong' for an application. Di Maria et al. (2010) chose, without apology, to fit a probit model. Couture, Cross and Wu (2024) fit a linear probability model and reported in a footnote that they also fit a logit model and found similar qualitative and quantitative results.

Theil (1971), p. 632 presages the current discussion: "*The theoretical background of probit analysis is rather complicated, and the justification of the normality assumption in this particular context is not very strong in econometric applications. Therefore it is worthwhile to consider another approach, that of logit analysis.*" He goes on to discuss how odds can be easily computed for the logit model. The 'justification' for the logit model appears (again) to be its mathematical convenience. Based on experience, arguments that one parametric model or another (including the linear model) is

'correct' (or not) appear to exaggerated (see further Section 10.7). It might be argued that one should use a linear specification because either the logit or probit might be the 'wrong model.' It is difficult to sustain this argument. The linear form is not less parametric than the others. But, it alone does have a potential to distort the estimated probabilities precisely because of its lack of curvature. However, notwithstanding the obvious logic, the linear approach dominates the recent applications of binary choice models with endogenous treatment effects.

10.2.5 Specifications of the Binary Choice Model

As suggested by Figure 10.3, there are thousands of binary choice applications in the applied literature. The binary choice model is also the laboratory for generations of theoretical contributions. Binary choice models can be divided by types, model-free nonparametric approaches, semiparametric models and parametric models. The maximum random utility, threshold basis is essential to the parametric forms - the model emphasizes random choice. Some regularity conditions must be assumed for the distribution of the source of randomness, ε, in the data generation process. Smoothness, continuity and monotonicity would be typical. Notwithstanding the prominent role of the linear probability model in the current literature, it is generally assumed that the model specifies probabilities that are bounded in (0,1). Finally, ultimately the analyst is interested in conditional probabilities that define a regression-like treatment. The typical specification is built around an index function, $\boldsymbol{\beta}'\boldsymbol{x}$ for a vector of strictly exogenous variables. The dependent variable in these threshold crossing cases is $y = \mathbf{1}[\boldsymbol{\beta}'\boldsymbol{x}+\varepsilon > 0]$.

A catalog of models includes the following: **Fully parametric**. A specific distribution is assumed for ε, typically normal or logistic. In spite of some impressions, the linear probability model is also fully parametric. However, since it is not bounded by (0,1), the index function in the linear probability is not actually a probability. **Semiparametric**. The index function remains typical. But, no particular distribution is assumed for the random term in the model. Contributions in this branch of the literature include Manski (1975, 1985); Manski and Thompson (1986), Klein and Spady (1993), Ichimura (1987), Cosslett (1983), Han (1987), Yan and Zhan (2024) and Horowitz (1992). Semiparametric approaches are generally robust to heteroskedasticity. The observed data contain no information on the variance of the response, so it is unclear how helpful this is. Nonetheless, see Zhi (2025). **Nonparametric** approaches are completely unspecified. No distribution is assumed and no index function is specified. Hansen, 2022's, p. 830 example suggests a limited application. In Gautier & Kitamura, 2013's formulation, variation arises from the assumed random parameters. In this survey, we are mainly interested in parametric models that guide the applied literature.

The overwhelming majority of applications of binary choice models use the parametric linear, normal or logistic distribution (the probit or logit model). There are many other fringe candidates, such as the Burr distribution (skewed logit model - see

Nagler, 1994). There are special models developed for particular cases, for examples, models with discrete regressors (see Manski (1988), Komarova (2013) and Lewbel, 2000). Alban Thomas (2006) proposes a panel data model with fixed individual specific trends. Some topics that have received attention are endogenous variables (continuous and binary), treatment effects and many models for panel data.

10.3 Modeling Individual Responses

The Method of Probits assumes that the LD100 (minimum dose for 100% kill) is a measurable characteristic of the larger population and that the sample data can reveal useful information about it. Bliss (1934) suggested a behavioral explanation for the shape of the curve. A pest is endowed with normally distributed resistance and is exposed to the stimulus (dosage). It responds by dying if the dosage exceeds the resistance, and by not dying otherwise. A regression-like model describes the conditional relationship between dosage and response. The Method of Probits fits a regression function. The threshold theory underlying the statistical methodology is the characteristic feature of the microeconometric approach.

Logically, the econometric approach would model the binary response, survive or die, of each individual pest in the experiment. Theil (1971) derived the probit log-likelihood for this case. The microeconometric approach to individual behavioral modeling took shape in the mid 1960's. Goldberger (1964) suggested a regression approach to the binary dependent variable. His proposal of weighted least squares rather than OLS sought to remedy a prominent defect of the linear regression approach, its heteroskedasticity. Worse, the linear index could predict the probability to be less than zero or greater than one. Theil (1971), pp. 628-635), writing of "*Frontiers of Econometrics*" hypothesized a logit model in which Car Purchase (yes/no) by a household responds to Income. Theil, 1971's p. 631, eqn. (5.4) logit log likelihood mimics Fisher, 1935's. As an extension, Theil proposed a second variable, age of current car. Cragg and Uhler (1970) built a logit model for the demand for auatomobiles. Econometric modeling for discrete outcomes begins here, in the late 1960's and early 1970's. Heckman (1974) *Shadow Prices, Market Wages and Labor Supply* was a landmark application that built on the binary choice models. Heckman (1974, 1975, 1978, 1979) built on Gronau (1974) and Lewis (1974). The modern forms of the probit model were pioneering works on labor supply and sample selection. There have been many modifications and extensions of the basic binary choice model built around new applications, some of which are shown in Section 10.4.

A parallel thread of research began on models for choice among multiple alternatives. Building on theories of individual choice, Nerlove and Press (1973) and McFadden (1974) built multinomial choice models of career choice and of transport mode choice. Heckman and McFadden shared the 2000 Nobel prize in Economic Science for their pioneering work on discrete choice models, multinomial choice and sample selection and for their revolutionary work in Microeconometrics. Microeconometric applications have surged with interest in social science research enabled

by the creation of many large social science research data sets. The University of Michigan's Panel Study of Income Dynamics (PSID) was begun in 1968 with data on 18,000 individuals in 5,000 families. The PSID advertises that over 7,600 refereed publications have been based on the PSID data. Government agencies in several countries have produced data sets for social science research. Some of these are:

British Household Panel Survey, (BHPS became *Understanding Society*, 1991;
European Community Household Panel, (ECHP) 1994-2001;
German Socioeconomic Panel (GSOEP), 1984;
National Longitudinal Surveys (NLS), 1979;
Medical Expenditures Panel Survey (MEPS), 1996;
Household Income and Labor Dynamics in Australia (HILDA), 2001;
Job Training Partnership Act (JTPA), 1993;
Seattle and Denver Income Maintenance Experiments (SIME/DIME), 1983;
Oregon Health Insurance Experiment (OHIE), 2008.

Microeconometrics emphasizes the analysis of cross-section and panel data. The starting point in the early 1970's was binary choice modeling. There have been numerous extensions of the probit model, some listed below in Section 10.5. Multinomial unordered choice (McFadden, 1974) is another focal point. The literature contains thousands of applications in Transportation, Labor Economics, Health, and elsewhere throughout the social sciences. The recent literature on causal inference is focused on Microeconometrics. Canonical reference works with a focus on Microeconometrics include Wooldridge (2010), Hansen (2022) and, especially, C. Cameron and Trivedi (2005).

10.3.1 Infrastructure: Hardware and Software

Widespread application of Microeconometric methods was aided by the development of computers with sufficient speed and power to handle large nonlinear models and tens of thousands (then hundreds of thousands, now millions) of observations. Walker and Duncan (1967) provide a useful benchmark. They report estimating a model using 5,202 observations and 19 independent variables using an IBM 7090 mainframe computer (the predecessor to the 360 series). Computation of two iterations required 4 to 8 minutes (roughly 360 seconds). Using a 2025 vintage desktop computer, we used 5,202 observations to fit a logit model with 19 regressors. Computing the 6 iterations took 0.14 seconds, a speed improvement of nearly 8,000 fold.

IBM's 360 series dominated the market from 1964 to 1979. Econometric software was built using Fortran compilers created by IBM in 1957 and subroutine libraries such as IBM's *Scientific Subroutine Package* (SSP). Modern toolkits of Fortran, C, C++ and Python subroutines include:

IMSL, International Mathematics and Statistics Library;
LAPACK, Linear Algebra Package;
NAG Library, Numerical Analysis Group;
Numerical Recipes, Flannery, Press, Teukolski and Vetterling (2007).

Development studios include Fortran and C++ programming environments, Ox-Metrics, Mathematica, MatLab, Maple, Gauss and R. Stata's Mata language (2005) integrates a matrix oriented language into their large package. Some of the early integrated Econometrics and Statistics packages are:[8]

TSP, Equation Systems, Time Series Processor (1965);
SPSS, Social Science Statistics (1968);
SAS, Social Science, Statistical Analysis System; (1969)
RATS, Time Series, General Econometrics (1980);
Shazam, Econometrics and Statistics (1977);
Stata, Econometrics and Statistics (1984).

Microeconometrics entered the development stream at scale in the early 1980's with the publication of the large social science panel data sets noted earlier and following interest in applications in labor markets (Heckman & Willis, 1975), health and multinomial choice (McFadden, 1974). By the mid-1980's, all of these integrated packages included probit, logit and Tobit estimators in their menus of procedures.[9] Other areas that attracted interest were multinomial logit models, Nerlove & Press, 1973 and McFadden, 1974), and models for duration and counts of events. Stata, ca. 1984, brought an organized focus on discrete and limited data methods that greatly expanded the spread of Microeconometrics. C. Cameron and Trivedi (2005) is a major reference work for practitioners. A recent (since 1996) specialization continues the developments of packages focused on binary and multinomial choice including ALOGIT, Apollo, Biogeme, NLOGIT, Sawtooth and Latent Gold.[10]

10.3.2 Estimators

Bliss's interest in the method of probits was to estimate the $LD100$, or the minimum dosage level at which 100% of the pest population would be killed. Figure 10.2 shows Bliss's results. Bliss's interest was in fitting a linearized ("rectilinear") function of R to $\log(D)$. He examined data and studies of other species of pests. There would be different values of the response curve for different pests and different insecticides (with different effectiveness) even for the same pest. The exercise presumes that the parameters are stable features of the ecosystem for a particular pest species and poison. Thus, the computations were 'estimation,' in spirit. This implies that the computed values can be carried into a new population to calculate the dosage needed for a specific application. Further, the shape of the function was suggested by an underlying theory of the distribution of resistance, and a random thresholds interpretation of the DGP.

---

[8] Ooms and Doornik (2006) is an extensive survey of Econometric software.

[9] Footnote 4 in Greene, 1981's comment on Heckman, 1979's sample selection model offers: "*A program which computes all of the estimators and this asymptotic covariance matrix is available from the author at nominal cost.*" This was LIMDEP. Manski, 1975, p. 225, fn 21) offers to provide a copy of the Fortran code for MSCORE on request (see Section 10.7.7).

[10] See McFadden, 1976, footnote 4, p. 375 for the state of this market as of 1976.

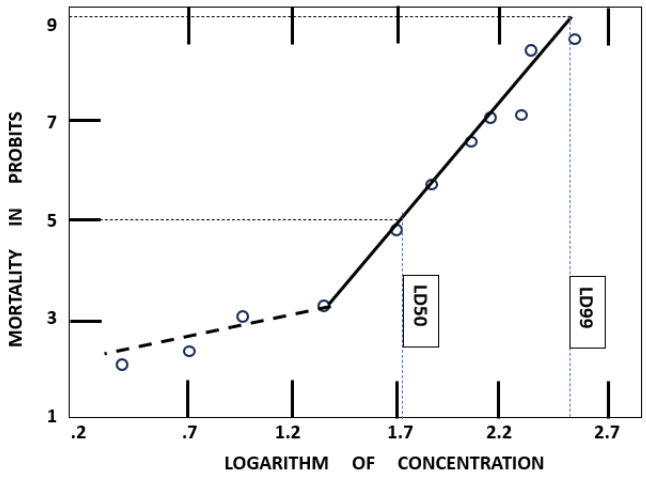

Fig. 10.2: Bliss Least Squares Plot of *Probit(R)*

Based on Figure 10.2, we have Bliss's Method of Probits for plotting the dosage response function. The value that corresponds to LD100 is slightly above 9.3. There were other methods that might have been used for obtaining the parameters, such as Minimum Chi-Squared or Maximum Likelihood, but not with the then-available computing capability. The modern researcher with the same limited objective would use maximum likelihood, carrying out this and the post estimation calculations with ease based on logits or normits instead of probits). Contemporary microeconometric applications of binary choice modeling would start from a behavioral model of a binary decision, and individual data on the binary response indicator, a set of attributes of the choices and a set of characteristics of the chooser. The lightbulbs application Di Maria et al. (2010) is a straightforward example. The modern counterpart to The Method of Probits would revolve around the log-likelihood function. Estimation and inference is a template application of maximum likelihood estimation. Post-estimation involves prediction of probabilities, analysis of partial effects and so on. We note, the last fifty years of research have produced a huge stream of papers on alternatives to the likelihood based formulations of binary choice modeling. Some of these are described below. There are a variety of applications in many fields and many different forms of the model. The probit and logit models are the most common in social science research. Other model forms include the Type 1 Extreme Value, Complementary Log-Log and the Burr-10 (Scobit) model. In practice, these different functional forms tend to show little difference in the estimates of marginal effects (see Table 10.1 below). Each has its own implicit scaling, so that estimates of $\boldsymbol{\beta}$ will vary widely. For example, an estimated logit coefficient vector will reliably equal roughly 1.6 times its probit counterpart, while estimates of partial effects will be essentially the same. It is difficult to make a case for one functional form or another. The logit model is mathematically convenient though with modern software, that is a small virtue. Odds

ratios are a useful feature of the logit model. However, when the model is extended to more than one equation, the bivariate normal distribution gains a convincing advantage. There is no convenient form of a bivariate logit model. The target of the model estimation effort has some variation as well. Social science applications are typically focused on margins. In Transport Research, marginal valuation, such as willingness to pay, is often a focus. In machine learning and credit and banking settings, a common application is 'classification'. Predicting the outcome variable given the inputs and an algorithm is a frequent use, for example, forecasting loan default or separating spam from email or separating credit fraud from valid charges.

10.3.3 Censoring and a Missing Variable Model

The threshold probit model holds that latent unobserved utility and the observed binary indicator are governed by the standard normal distribution. This is a missing data model. If the underlying utility were observed, the specified model would be an ordinary linear regression model, amenable to least squares regression and conventional analysis.[11] The pivotal part of the model, as stated thus far, is inestimable. The data contain no information on the latent outcome variable. Missing variables taint the results of models that involve them. General results about missing (omitted) regressors are central in the theory of the regression model. In the binary choice case considered here, utility is partially observed; the sign is observed. The latent utility is censored. Values in specified ranges are masked by collapsing them into a single value or a few values. The crucial question for the modeler is, what can be learned about the probit model, when data on the outcome variable are censored? The generic threshold (random utility) binary choice model is fundamentally unidentified. A normalization is needed. Any element of the structural parameter vector can be fixed. The most common approach is to fix the variance of the random component at one.[12] Censoring is a central feature of all of the qualitative models that we consider in this chapter. Tobin (1958) is an early contribution to this aspect of the methodology. The random utility approach underlies most of the discrete choice and other behavioral models in microeconometric applications.

10.4 Qualitative Dependent Variable Models

The majority of recent discrete choice applications consist of binary choice models and bivariate or multivariate extensions of binary choice models. The foundation is random utility. The model form generally accommodates the observation mechanism. The estimator of choice in most cases is the Maximum Likelihood Estimator. Five

[11] The binary dependent variable would still be observable but would be redundant.

[12] But, see Lewbel (2000) where one of the slopes is fixed at one and Manski (1975) where the length of the slope vector is fixed at one.

main cases are:
Univariate Binary Choice;
Recursive Bivariate Probit;
Univariate Multinomial Ordered Choice;
Multinomial Unordered Choices;
Hybrid Multinomial Unordered/Ordered Choices.
This section identifies the main frameworks used in empirical analyses of discrete outcomes. We note for completeness the nonstandard approaches. Almost all of these, in theory and practice, relate to binary choice outcomes.

10.4.1 Non-Parametric and Semi-Parametric Approaches

The most general form of the binary choice model specifies only that the probability of the outcome has a handful of regularity conditions; it is an unknown smooth, continuous, monotonic function that is bounded in (0,1) over the range of the data. (It is coherent.) In this most general form, the data can reveal points that can be plotted using kernel methods (see, e.g., Klein & Spady, 1993). The data in Figure 10.1 or 10.2 would be natural candidates. See, as well, Hansen, 2022, p. 830, Figure 25.1. The approach avoids any restriction imposed by a specific functional form. The lack of structure comes at a price, however. Not much information is derived and inference is difficult because of the curse of dimensionality. Manski's Maximum Score estimator is a second flexible approach. The estimator is based on a criterion function and normalized so that the length of the coefficient vector equals one. Estimation of the parameters requires non-standard methods. Inference would be based on bootstrapping. However, in this case, the atypically slow rate of convergence of the estimator hinders the standard bootstrapping approach. A smoothed estimator is developed by Horowitz (1992). The MSCORE estimator is analyzed by Patra, Seijo and Sen (2018). The appeal of MSCORE is that it can be based on multiple x's without a formal structure. Cosslett (1983) devised a similar approach based on an empirical likelihood function.

The main appeal of non-parametric and semi-parametric estimators is avoiding the "strong" distributional assumptions. Among the suggested benefits is robustness to heteroskedasticity in the randomness of the binary choice mechanism. There is no information about scaling of the underlying utility contained in the observed binary choices. In the end, the payoff to the increased flexibility of the non- and semi-parametric approaches seems unclear.

Advocacy for the Linear Probably Model rests on avoiding a possibly restrictive functional form. Applications have suggested that the 'strength' of the distributional assumptions tends to be exaggerated. Moreover, it is easy to visualize how a linear model would distort the estimation based on data like Bliss's in Figure 10.1. The fact that the model is incoherent undercuts inference. Moreover, the linear probability model is not semiparametric, as it might appear. The assumption of a uniform distribution is not less parametric than the normal distribution underlying the probit

model. Heckman and Snyder (1996) derived the linear probability model analytically. The parametric model specifies that the net utility is distributed U[0,1].

10.4.2 Parametric Discrete Choice Models

The cases of main interest in this chapter are models with non-quantitative discrete and latent 'dependent' variables. The most familiar case will be the binary outcome. The dependent variable will equal one if a condition is true and zero if not. This is a qualitative dependent variable model, often misleadingly called a binary regression model. As discussed earlier, the methodology typically employed is the Bernoulli outcome model; Prob[$y = 1|\mathbf{x}$] is a model of the probability. Broadly, while in regression model settings, the outcome being modeled is a quantity, here the function being modeled is a probability.

The applied econometrics literature from the mid-1960's onward presents many forms of discrete choice models that are developed along the lines discussed above. We can collect many of these in a few basic model forms.

Binary Outcome (probit and logit): Di Maria et al. (2010) is a straightforward example of a binary choice model. The core of the model is a conditional Bernoulli distribution.

Bivariate Binomial and Recursive Bivariate Binomial: Many settings can be modeled around two (or more) corrrelated binary outcomes.[13] For example, we might suggest that in Di Maria et al. (2010), *Adopt the new technology* and *Knowledge about the environment* are determined jointly by demographic factors and by housing attributes. The appropriate model might hold that Prob[*Adopt,Knowledge*] might be codetermined. Alternatively, the model might specify that both sets of covariates determine *Adopt*, but only the demographics underlie *Knowledge*. Further, a natural application might specify that the outcomes of interest are the (marginal) variable, *Knowledge* and the conditional outcome *Adopt*|*Knowledge*. This would constitute the

Recursive Bivariate Probit (RBP) model (see Scott, Schurer, Jensen & Sivey, 2009). Another modification of the the bivariate probit model is the Poirier (1980) and C. Meng and Schmidt (1985) model of **Partial Observability**. In this case, two probit processes operate simultaneously. The observed outcome is the product of the two. An example is two individuals (e.g., couples) making a decision that requires unanimity. Only the final outcome is observed.

[13] There is an ambiguity about the correlation of two binary variables. The ordinary Pearson correlation applies to continuous variables. The common counterpart for binary variables is the tetrachoric correlation, which conveniently can be obtained as the correlation coefficient in a bivariate probit model (with or without covariates).

Ordered Multinomial Choice: The ordered choice model extends the binary probit or logit model to scale variables such as the ubiquitous Likert scale, [*poor, fair, good, very good, excellent*], usually coded [0,1,2,3,4] or [1,2,3,4,5]. This specification typically builds on the Random Utility interpretation. This case resembles a regression model, but the linear regression would be misspecified. It is inappropriate to specify that the intervals are equally spaced on an opinion scale In the lightbulbs application, the *Support Kyoto* scale is measured [1,2,3,4]. The more natural specification treats the observed variable as a censored observation on the underlying random utility (see Zavoina & McKelvey, 1975).

Cell Inflation: In some cases, a particular outcome or cell (probability) may be larger in the data than would naturally be predicted by the model. Harris and Zhao (2007) examined data on smoking behavior that appeared, based on related data, to contain too many zeros. A **Zero Inflated Ordered Probit** or *ZIOP* model is connected to a second model that inflates a particular cell. Another application is a 'two inflation model of fertility' (T. Meng & Lyu, 2022). The first application of cell inflation was Lambert (1992)'s specification of a Zero Inflated Poisson (ZIP) model for industrial quality control.

Misclassification: This is a type of measurement error. The dependent and/or a binary independent variable in a binary choice model are miscoded with ones as zeros or the reverse with a certain probability. Applications are usually based on parametric models with estimation based on maximum likelihood. Some empirical applications involve government programs such as Food Stamps or social statistics such as crime. Developments have been provided by Meyer and Mittag (2017), Bollinger and Martin (1997), Hausman, Abrevaya and Scott-Morton (1998) and Lewbel (2000).

Unordered Multinomial Choice: Many choice situations involve choosing the most preferred one in a set of unranked alternatives. Political candidates, travel modes, medical treatments, food or clothing brands and entertainments would be applications. Observed choices are driven by underlying, unobservable preferences. A regression model for this situation would be inappropriate. The most common specification is a "multinomial" choice model such as the **Multinomial Logit**. (McFadden, 1974). Implicitly, the multinomial choice among J alternatives reveals J preference outcomes, one of which indicates the most favored and J-1 of which indicate the less favored alternatives. An intermediate case between this and the ordered choice would be the rank ordered choice. The partial revelation of preferences indicates a sequence of indicators of the most preferred outcome among the remaining ones. These basic frameworks form the platform for a variety of extensions that include most of the

qualitative dependent variable models in the contemporary literature.[14] The next section will introduce a class of models that integrates some of these in a single multi-equation framework.

10.4.3 Hybrid Choices and Latent Variables

The **Hybrid Choice Model** (HCM) in the form considered here is a relatively recent (ca. 2012) development in this literature. The Model includes latent attitudes in a discrete choice model. In Di Maria et al. (2010), the authors modeled the influence of latent "Preferences for the Environment" in a binary choice model of adoption of CFL light bulbs. The HCM integrates a proxy for the unobserved attitude in the choice model. A partner submodel uses observed information and structural assumptions to create the proxy. Early work on HCMs begins with applications in Marketing such as McFadden (1983) and Swait (1994). Many recent applications use the framework proposed by Daly, Hess, Patruni, Potoglou and Rohr (2012).

The essential form of the HCM is as follows:

1. Structural model for a latent variable;
2. Measurement of a set of observable indicators;
3. Data reduction to create the proxy for the attitude;
4. Latent utilities for choice model;
5. Choice model including proxy for latent attitude.

The first generation of HCMs employed **Structural Equations Models** including Factor Analysis and Principal Components. Steps 1-3 were denoted collectively the *Multiple Indicators and Multiple Causes Model for a Latent Variable* (MIMIC model) in reference to Jöreskog and Goldberger (1975a). The body of techniques included *LISREL, ACOVS, SEM*, /hybrid choice modeletc.[15] Step 2 in this methodology assumes that the observed indicators, and the inputs are multivariate normally distributed. Daly et al. (2012) noted that HCM applications typically collected the information on surveys in which the responses were discrete, binary or ordered, such as the three in the light bulbs application. Their proposal replaced the data reduction in Step 3 with aggregation of binary choice and ordered choice models. The literature on choice modeling contains hundreds of applications in Transportation, Health, Marketing and other areas. (Greene and Hensher (2026) is a survey.)

[14] We have omitted one class of discrete dependent variable models, the count data models. These are used, e.g., for number of hospital visits, doctor visits, items purchased, vehicles owned, patents, number of children, length of stay and others. The dependent variables in these models are not qualitative; they are quantities that are amenable to conventional (albeit nonlinear) regression methods.

[15] Joreskog, Gruvaeus and Thillo (1970) and Joreskog and Sorbom (1993).

10.5 Milestones in Discrete Choice Modeling

There has been a stream of modern developments since a beginning around 1970. Most of the current range of models were in mainstream use by 1990. Many recent innovations are more finely grained, such as ongoing improvements in random parameter methods for multinomial choice models. The catalogue of methods grows with current research interests. The Hybrid Choice Model in Section 10.4.3 is a recent major innovation.

10.5.1 Departure and Ending Points

An early development in choice modeling is Tobin, 1958's pioneering work on the *censored regression model*. He proposed the elements of the canonical probit model that entered the literature ten years later. Tobin's censored regression idea laid out a framework for the random utility/threshold interpretation that supports most of the model forms in current use. Most of the tools that followed are extensions or modifications of the familiar probit model.

Daly et al., 2012's extension of the Hybrid Choice Model is a major recent development. This specification wraps a scaffold of latent variable modeling, mixed logit modeling and ordered choice modeling around a core of McFadden, 1974's Multinomial Choice model (see Greene & Hensher, 2026).

Contemporary methodological proposals have redirected much of the common adoption of the probit model. Angrist and Pischke (2010) have argued that formal parametric models, notably the probit model, produce 'model driven' results that are not 'credible'. Their proposed remedy is the linear probability model that Goldberger (1964) first considered.

10.5.2 Milestones

A small set of general specifications forms the platform for qualitative choice modeling:

Univariate binary choice models;
Bivariate and multivariate binary choice models;
Multinomial ordered choice models;
Multinomial unordered choice models (with latent variable models);

Accommodating endogenous variables and panel data are major extensions for all four topics. Uses of latent variable models are extracted from social science applications in Psychology, Education, Economics, Transportation, Political Science, Sociology and so on. These focus on *Data Reduction* such as *Factor Analysis* and *Principal Component Analysis* and the *MIMIC model*. (Jöreskog & Goldberger,

1975a.) These enter the *Hybrid Choice Models*. Some milestones in the development of microeconometric qualitative choice models are as follows.

10.5.2.1 Tobit Model - Limited Dependent Variables. (1958)

Tobin (1958) proposed a model for a partially observed dependent variable. In the application, gross investment is normally distributed. If the capital depreciates and is not fully replaced, then net investment is negative. Negative investment is unobservable. The observed investment is the maximum of zero and the net investment. The underlying variable is '*censored*'. The regression model that links observed investment and **x** is a 'censored regression', or 'Tobit' model (Goldberger, 1964). The type of model is a '*Limited Dependent Variable model*' (see Maddala, 1983 and Amemiya, 1973, 1984). Recent applications have redefined the observation mechanism to be a two part structure that produces zeros or positive values depending on some criterion, for example, a model for individual charitable donations (Brown, Greene & Harris, 2015). Wooldridge (2010) labels this a '*Corner Solution model*'. Di Maria et al., 2010's lightbulb application might describe a model for expenditure on efficientlightbulbs as zero or some positive value. Several variants of the two part structure have been proposed, such as Cragg (1971) and Fin and Schmidt (1984) who allow the participation equation to be a probit model and the regression to obey a separate '*truncated regression model*'. Models with a two part structure such as these are an innovation. Amemiya, 1984, worked through the theoretical issues.

Several important theoretical results are found in Tobin's paper, including the full log likelihoods for the probit and Tobit models and for the (then) new **Truncated Regression** model.[16] Some foundational results can be inferred from Tobin's derivations. For the observed dependent variable, $y = \text{Max}[0, \boldsymbol{\beta}'\boldsymbol{x} + \varepsilon]$, Tobin presented (1) the latent regression model, (2) the truncated regression model, (3) the censored regression model and (4) results for Heckman, 1979's sample selection model. Detailed treatments of limited dependent variable models appear in Wooldridge (2010), Amemiya (1984) and C. Cameron and Trivedi (2005). Rigorous theoretical treatments appear in Amemiya (1973, 1981, 1984) and Maddala (1983).

10.5.2.2 Nonlinear Models for Binary Choices. (1964, 1969)

The probit model was developed by Bliss (1934) and Finney (1947).[17] Berkson (1944, 1951) advocated for the logit model (see Cramer, 2002). These early researchers modeled shares and proportions. Estimation of a regression-style model for individual data with a binary dependent variable by least squares was suggested by Goldberger (1964). A microeconometric approach based on maximum likelihood

[16] The truncated regression model applies to a data set that is generated by discarding, rather than censoring, the negative observations.

[17] Cramer (2002) credits Fechner (1860) for the inverse normal transformation.

was suggested by Goldberger, 1964, pp. 250-251, Theil (1970, 1971) and Cragg and Uhler (1970). By 1970, MLE became the estimator of choice. Generalized linear models (Walker & Duncan, 1967) were often fit by minimum chi-squared. The linear least squares approach was largely abandoned for empirical work.

10.5.2.3 Modern Regression Style Form, Software (1970-1985)

Maximum likelihood emerged as the usual estimator in the early 1970s. Other intricate models and functional forms were proposed, such as Heckman and Willis (1975)'s 'beta-logistic' model. General software for limited and qualitative dependent variables included (among others) SAS (1966-1976), RATS (1980), Stata (1984), Shazam (1977) and LIMDEP (1981). Important innovations in software development included: nonlinear optimization by the DFP and BFGS algorithms (LISREL by Joreskog & Thillo, 1972); MLE for nonlinear equation systems and a general routine for using the delta method (TSP/`Analyze` by Robert Hall and a handbook by Robert Pindyck and Daniel Rubinfeld, ca. 1975); and a package of subprograms for analysing partial effects in a wide variety of nonlinear models (`Margins` in Stata Version 11).

The `Margins` command in Stata (ca 2010) is a major innovation for empirical microeconometric research. Reported results changed from ambiguously interpreted structural coefficients to informative partial effects. The delta method is employed to great advantage. Some applications include Di Maria et al. (2010) and Scott et al. (2009).

10.5.2.4 Computation for Fixed and Random Effects (1980-1982)

The unconditional fixed effects (FE) probit model has two undesirable features: (1) There is no counterpart to the within transformation to remove the fixed effects and (2) Because of the "Incidental Parameters" (IP) problem, the unconditional probit and logit estimators are inconsistent. The Chamberlain (1980) conditional estimator for the logit model solves both problems. The solution is of limited usefulness, however, because the constants are not estimated and not estimable, so partial effects and predicted probabilities cannot be computed. It is also complicated by the extreme practical complexity when the number of periods, T, is greater than 5. Some milestones include: The Krailo and Pike (1984) algorithm for the conditional probabilities allows far larger T.[18] Abrevaya (1997) demonstrated that the bias of the IP problem in the logit model is proportional, 100% when T = 2. Monte Carlo results in Greene (2004) suggest the same result applies for the probit model though a formal proof remains to be provided. The proportional bias appears to persist when T > 2, diminishing considerably when T > 15 or so.

[18] With T = 50, say, the denominator term in the C-Logit probability has about 10^{25} terms. The model with thousands of observations will take a few seconds of computing time.

A random effects treatment for the probit model (and other nonlinear models) based on Gauss-Hermite quadrature was proposed by Butler and Moffitt (1982). By treating the random effects model as a random constant term model, then extending to the other parameters, a vast range of "mixed," or random parameters models emerges. McFadden and Train (2000) show the great utility of the mixed model approach. An early application is Revelt and Train (1998). Stata implemented quadrature based mixed modeling with `GLAMM`. Bhat (2001) and Train (2001) showed how dramatic improvement in estimation efficiency is achieved by using Monte Carlo simulation and intelligent draws such as Halton sequences instead of quadrature.

10.5.2.5 A Finite Mixture of Parametric Models

Heckman and Singer (1984) proposed a finite mixture approach to unobserved heterogeneity. Their suggestion was for duration models. But, the method is portable and well suited for the parametric models considered here. The starting point would be a random effects RUM such as the Butler and Moffitt-style random effects probit model. With the usual single index approach, the utility function is $U_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + \sigma_v v_i$ with v_i distributed N(0,1). The authors argued that the normal distribution was too much structure. They proposed a nonparametric approach, with the heterogeneity represented by a draw from a distribution with a finite number, M, of mass points α_m and probabilities π_m. The only constraints are that the αs have population mean zero and the probabilities sum to one. The joint density is the probability weighted mixture of the conditional distributions.

There are difficult problems of identification that depend partly on the data. For example, applications that contain many dummy variables can run into trouble. It is extremely difficult to be specific about the identification criteria. The authors note that in applications, problems can reveal themselves by producing extreme results such as (essentially) infinite standard errors, which might suggest M is too large. We note, this approach can be transported to a cross section without modification. The identification problems get worse, however.

The Heckman/Singer model is intricate. Implementation in the base case is surprisingly simple. The model can be estimated as an M-class latent class model in which the heterogeneity is carried by class specific constant terms. The remaining index function parameters are constrained to be equal across the classes. M must be chosen in advance. It is not necessary to fix the mean of the αs at zero. There are M-1 free αs in the mixture and an Mth is the constant term in the index. To compute them, all M constants are treated as free parameters. The probabilities are constrained to sum to one by, for example, using a multinomial logit format.[19] S. Cameron and Heckman (2001) is an application.

[19] See Greene (2018) on latent class models.

10.5.2.6 Mixed Logit: Random Parameters (1998, 2001)

The random parameters multinomial logit model encompasses most of the model forms currently in use for multinomial choice modeling (see Hensher & Greene, 2003). A pioneering application is Revelt and Train (1998). Simulation based estimators for mixed modeling occupy a major share of recent applications (see McFadden & Train, 2000). Pioneering work by Bhat (2001) and Train (2001) showed the dramatic gains in speed and simulation quality brought by using intelligent draws such as Halton and Sobol sequences instead of quasi-random draws. Software for Mixed Logit estimation appeared in ca. 2001 (NLOGIT, Stata, Biogeme) (see Hensher and Greene (2003) and Mabit, 2026). Generalizations of the Random Effects Logit or Error Components Logit model in NLOGIT (`RPLogit, ECLogit`) obviated the cumbersome multinomial probit formulation.[20]

10.5.2.7 Machine Learning – Support Vector Machines (2000)

Machine Learning is generally associated with large (big, by current standards) data sets and algorithmic specification searches. The large N aspect evolves over time. A data set with a million observations would have been impossible in the 1960's; it would be 'moderately' large today. The signature feature of big data is large K - many variables. Techniques for finding regression specification such as ridge regression and pre-test estimation have been brought forward and extended. The connection to our interests relates to analyses of binary outcomes. Machine learning has paid particular attention to what statisticians and econometricians call the 'classification problem.' An underlying theory that is consistent with the methodology assumes that a large population of individuals with many features (characteristics) contains two sharply defined types of individuals, called, say, good and bad - for example, mail and spam. An algorithm is used to form a function of the features that divides the sample into two groups that it is hoped match the underlying classification. A logit binary choice model, called a *support vector machine*, is often used to form a function of features like a probability in (0,1). Using 0.5 as a threshold the machine is used to partition the data based on the predicted probability. Other than the logit form, no particular specification is assumed at the outset - the computed index function is discovered algorithmically. Hastie, Tibshirani and Friedman (2008) is a reference.

N. Chen, Gallego and Tang (2019) developed a random forest approach to building a binary choice model. They applied the method to several revealed preference data sets, including a hotel choice setting. The authors argue that the combination of binary choice models performs better than single binary choice models.

[20] Lechner, Lollivier & Magnac, 2008 discuss this issue at length.

10.5.2.8 Binary Choice with Endogenous Variables (1979)

Intricate functional forms for binary data were developed in the mid- 1970's, such as Heckman and Willis (1975), 'beta-logistic' model. Another direction of development was simultaneous equations and endogenous regressors - for example, a health outcome model that contains endogenous income on the right hand side. Modelers considered including an endogenous continuous variable in a linear simultaneous equation model for the random utility. Early approaches proposed two step approaches. The first step is linear LS regression for the reduced form of the continuous covariate. The second step is the inclusion of predictions (mimicking 2SLS) or residuals (control functions) in the structural probit equation. Applications included (Heckman, 1979) and Rivers and Vuong (1988). These avoid the complexity of full information maximum likelihood. A full MLE was devised by Newey (1987). The FIML 'IV probit' is now a standard technique. Earliest applications were individual codes. It was added as a command option in Stata in (1999, 2000), `ivprobit` in Stata and R. Both two step and FIML estimators are now provided.

10.5.2.9 Bivariate Probit, Recursive Bivariate Probit (1989)

Heckman (1979), Greene (2018) and Boyes, Hoffman and Low (1989) note that the typical application of IV-probit would involve an endogenous binary variable, which would naturally suggest a bivariate probit model - a two equation model. The starting point is a pair of probit models tied together by correlation across the random parts of the utility functions, like a seemingly unrelated regression model. The first approach for a bivariate probit model would be to estimate the two equations individually, consistently, ignoring the correlation - a LIML approach. If the correlation is of interest, the FIML approach requires a method of evaluating the integrals for the bivariate normal distribution.[21] Fast algorithms using Hermite quadrature are available.[22]

Boyes et al. (1989) extended Heckman, 1979's selection model to a probit model by using a two step approach like the one used for a regression model. This adds the inverse Mills ratio to the probit model. There is no theory to justify this approach. But, a FIML estimator for the same model is straightforward. (See Ven and van Praag (1981) and the Stata documentation for `HeckProbit`.) This is also now available in R.)

Boyes et al. (1989) was an application of IV-Probit, though the endogenous variable was binary. The probit model with an endogenous RHS variable is the *Recursive Bivariate Probit* (RBP) model (see Greene, 2018). This has recently been reinterpreted as a probit model with endogenous treatment effect (see Angrist & Pischke, 2010). Well-known applications include Evans and Schwab (1995) and Scott

[21] Whether the LIML estimator is more or less efficient than the FIML estimator (asymptotic variances) remains to be settled.

[22] The FIML approach with only constant terms provides an estimator of the *tetrachoric correlation* of the two binary variables.

et al. (2009). The likelihood for the RBP model is constructed by treating the two equations as a marginal probit model for the treatment variable times a conditional (on the binary treatment variable) probit model for the outcome variable. This is estimated as the bivariate probit model with two correlated equations.

The RBP model exposed the subtle problem of *identification by functional form.* If the two equations were linear, exclusion restrictions would be required for identification. The RBP model will generally 'work' without such restrictions. Analysts typically add instruments to the treatment equation anyway.

10.5.2.10 Binary Choice with Endogenous Binary Regressor (2000)

This is the probit model with an embedded probit model. The underlying RUM structure is a recursive simultaneous equations model. The conditioning variable in the main equation is the binary treatment, not the utility under the treatment equation. The logical specification implies that the treatment is predetermined, i.e., in reduced form. The simultaneous equations model based on the two unobserved utilities is incoherent. There does not exist a reduced form. That model is inestimable without aditional assumptions. Because of the censoring transformation, the simple logic and results of the basic recursive linear model do not extend to this case. However, it is a straightforward recursive parametric bivariate probit model. The recursive bivariate probit model is the endogenous binary treatments case. Details appear in Section 10.6.8 below.

10.5.2.11 Multivariate and Multinomial Probit (1995)

The multivariate probit model extends the bivariate model to three or more equations. The difference in the computational complexity is stark. Until the 1990's, no suitable method was available for computing multivariate normal integrals fast enough for an econometric application of MLE.[23] The GHK simulator was developed in 1991-1994 (see Geweke, Keane and Runkle (1994) and a 1994 symposium in the Review of Economics and Statistics). The most immediate application appeared to be the multinomial probit model. The symposium suggested optimism. Ultimately it was found that multinomial choice models with correlated normal utilities could be calibrated far more efficiently (with no loss of generality) using Monte Carlo Simulation (see Hensher et al., 2015). An intermediate step in this development was Bertschek and Lechner (1998) 'Panel Probit Model.' Theirs was a five period probit model with common coefficient vector and unrestricted correlation matrix. The authors devised a GMM estimator in order to obviate the multivariate normal integrations. By 2000, incorporating their advances, estimating their model (with their data) was routine using Halton sequences for the simulations and took only a

[23] McFadden, 1976, p. 376 discusses this problem as of 1976 and reports early developments.

few minutes.

10.5.2.12 Ordered Multinomial Choice Models (1975)

The base probit model provides the platform for accommodating the data generating mechanism for a binary outcome. In Zavoina and McKelvey (1975), an **Ordered Choice Model** (probit or logit) was used to model survey responses. Their application modeled Congressional opinion (dislike = 0, neutral = 1, favor = 2) on a piece of legislation. This new model class extends the threshold model to interval censoring in RUM specifications. The usual estimator is MLE. OLS would be inappropriate for this case. However, the model is preserved after combining adjacent cells. Aggregating all nonzero values into one is 'binarization.' This collapses the model into the probit or logit model. This would be a LIML approach. Whether fewer cells (LIML) is more efficient than greater (FIML) remains to be settled. There are many applications to self-reported health (Contoyannis, Rice & Jones, 2004), well being (Helliwell et al., 2026), etc. Aitcheson and Silvey (1957) showed an antecedent to this model. They devised a method to estimate the cutpoints using a method of moments technique similar to Bliss (1934) for the binary case.

10.5.2.13 Unordered Multinomial Choice (1974)

Theory for discrete choice among a set of unordered choices is based on Thurstone, 1927's proposal for binary choices and Luce, 1956, 1957, 1959's suggestions for how the 'Independence from Irrelevant Alternatives' can produce a coherent multiple equation model.[24] Pioneering applications are McFadden (1974) conditional logit model (applied to a stated preference survey about travel mode choices) and the Nerlove and Press (1973) multinomial logit model (applied to labor market career choices). Generalized formulations such as the nested logit model were proposed by Ben-Akiva (1973) and Hensher (1986). Full and limited information ML are the estimators of choice. The multinomial probit model is a natural application (Keane, 1992). Direct ML using the GHK simulator is used initially. An *Error Components Logit* approach based on Monte Carlo simulation (Hensher et al., 2015) proves to be far more efficient. The multinomial logit model with random parameters is a large expansion in model capabilities (McFadden & Train, 2000 and Revelt & Train, 1999). An extension of the random parameters approach to market share data estimated using GMM was proposed by Berry, Levinsohn and Pakes (1995). The pioneering application was a study of the automobile market. Comprehensive software for estimation of multinomial choice models includes NLOGIT, ca.1996, Apollo, ca. 2005, Biogeme, ca. 2008), Alogit, ca. 1986, Latent Gold, ca. 2000. Some particular

[24] See McFadden (1976) for a detailed survey of the background theory behind multinomial choice models.

capabilities have been developed in Stata (ca 2015) (`mlogit, nlogit, cmclogit, cmmprobit, cmmixlogit`) and in R, implemented in Apollo.

10.5.2.14 Hybrid Choice (2012)

The Hybrid Choice Model for multinomial choices (e.g., travel mode) incorporates latent variable(s) such as preferences and attitudes. In the efficient lightbulbs application described in Section 10.2.3, the adoption decision is influenced by attitudes and perceptions about environmental stewardship. Latent variable methods are used to aggregate observed information about latent variables. The earliest work was McFadden (1983). The current leading specification was proposed by Daly et al. (2012).

This framework incorporates MIMIC models (Jöreskog & Goldberger, 1975a) and Ordered indicators (survey data). The model for "*Multiple Indicators and Multiple Causes of a Single Latent Variable* (MIMIC)" is a specific application in the broader form of '*Linear Structural Relationships*' (LISREL), pioneered by Joreskog et al. (1970). Software includes ACOVS, PrinComp, Latent Variable techniques focus on '*Data Reduction*' (see Jöreskog & Goldberger, 1975a). The contemporary literature is surveyed in Greene and Hensher (2026).Greene and Hensher (2026).

10.5.2.15 The Linear Probability Model: Demise (1975) and Return (2000)

Early treatment of the linear model with a binary dependent variable cast it as an ordinary regression model. Formal treatment appears in Goldberger (1964) and Amemiya (1975). The model is heteroskedastic, non-normal and incoherent - it allows negative probabilities. Goldberger proposed weighted least squares to deal with the heteroskedasticity. McGillivray (1970) proposed exponential heteroskedasticity and MLE as in Harvey (1976) and Godfrey (1978). Estimation machinery (theory and software) for the MLE is no longer a constraint (see `hetregress` in Stata).

The LPM has an obvious defect in that the CDF for this model is not a proper distribution function; it is not constrained to lie between 0 and 1. This defect can be 'corrected' by simply defining the probability to equal zero or one if the index strays below zero or above one. But, "*the procedure produces unrealistic kinks at the truncation points*".[25] With the nonlinear probit and logit models firmly established and software widely distributed, there was no need to patch the linear model. It was largely abandoned by 1975. Linear probability models produced raised eyebrows in seminars.

Angrist and Pischke (2010), announcing the 'Credibility Revolution' in microeconometrics, disputed the use of nonlinear models. A large fraction of applications

[25] See Amemiya, 1975, 1981 and Amemiya, 1985, p. 268-269.

of treatment effects involve binary outcomes with binary treatment effects. The LPM was resurrected ca. 2000; thousands of applications have followed (see Figure 10.3).

An ongoing discussion among practitioners argues for or against the LPM for empirical research. K. Chen, Martin and Wooldridge (2025) explore the issue theoretically. Although the discussion about strong assumptions in nonlinear models is general, the conversation applies mostly to the probit model. The linear model would be transparently misspecified for the ordered or multinomial choice models. Moreover, the issue of the basic probit model is misleading. The setting in which OLS is advocated involves an endogenous treatment effect and 2SLS, not OLS. There is no evidence yet that 2SLS is a reliable substitute for the recursive bivariate probit model - the prevailing view carries the motivation for the LPM forward. (See Li, Poskitt, Windmeijer & Zhao, 2022.) The LPM is currently the standard approach for applications involving endogenous treatment effects (binary or continuous) in binary choice models.

10.6 Econometric Methodology for Discrete Choice Modeling

The development of a methodology for qualitative dependent variables focuses on binary choice. Most of the specifications can be cast as extensions of the canonical probit model. Models for endogenous regressors, multinomial unordered choices and panel data models add additional complications.

The probit and logit models have been the standard frameworks for analysis of binary dependent variables since their introduction in the early 1970's. Proponents of the 'Credibility Revolution,' e.g., Angrist & Pischke, 2009, have argued that formal nonlinear models drive empirical results and hide influential assumptions. The remedy (ca. 2000) is the linear probability model. The LPM has (re)emerged as the method of choice for many applications.

10.6.1 An Empirical Econometric Model for Binary Outcomes

The formal treatment of individual data appears in the late 1950's, e.g., with Aitcheson and Silvey (1957), and in the late-1960's with, for examples, Goldberger (1964), Walker and Duncan (1967) and Theil (1971). Goldberger suggested an ordinary linear regression mechanism, with a binary dependent variable and with the usual assumptions for regression. The proposed treatment had four significant shortcomings: (1) The implied disturbance term is inherently heteroskedastic in a manner that depends on the structural parameters. (2) The disturbance could only take two values and could not be normally distributed. (This was problematic at the time.) (3) The marginal effects in the LPM are constant , which is not reasonable outside the central range (say, between 0.1 and 0.9). (4) Without unreasonable constraints on the data, the model is incoherent — it allows probabilities to be negative or exceed one. Despite

attempts to redeem it (Amemiya, 1975 and Horrace & Oaxaca, 2006) – the LPM was largely abandoned.

Probit and Logit become the natural choices for binary choice modeling. The two approaches are generally viewed as interchangeable; there are usually no discernible differences in estimated probabilities. The logit model is often preferred for its mathematical convenience. The Walker and Duncan (1967) approach to estimation is a 'minimum chi-squared' estimator. Logically, the estimator is the same as Theil's 'Maximum Likelihood Estimator.' In modern terms, both are 'M-Estimators' (see Wooldridge, 2010, pp. 400-401). Under ordinary regularity conditions, both Walker/Duncan's and Theil's approaches can be shown to to be root-n consistent and asymptotically normally distributed estimators of the index function parameters. Finally, Theil, 1971's maximum likelihood (logit) approach is precisely the contemporary threshold approach. Attention to partial effects and robust inference methods comes later.

10.6.2 The Influence of Functional Form

Angrist & Pischke, 2009, p. xii state: "*Most econometrics texts appear to take econometric models very seriously. We take a more forgiving and less literal-minded approach. The estimators in common use almost always have a simple interpretation that is not heavily model dependent*" (p. xii). In practical terms, this argument is a suggestion that in spite of its flawed reputation, comfortable simple OLS should be preferred to analysis of tightly specified models of economic behavior. They emphasize the heavy influence of model specification on empirical results.

Does the choice of functional form matter? To what extent are the results model dependent? Researchers rarely find empirical differences between probit and logit models large enough to attribute to anything other than finite sample variation. The difference is not testable. This is part of a broader recent conversation (Angrist & Pischke, 2010) about using experimental credible design based approaches to inference rather than model-dependent procedures. For current purposes, the methodological issue is the extent to which the model results are influenced ('driven') by the choice of functional form, since it is necessary to make an assumption. The following experiment echoes the common experience. Table 10.1 below lists the partial effects computed from seven parametric binary choice models.[26] The data are the 1994 wave of the German Socioeconomic Panel (GSOEP) data used in Riphahn, Wambach and Million (2003). The dependent variable is one if the number of doctor visits is positive. There are 3,377 observations.

The results echo the typical experience. The widely varying parametric functions do not vary enough to suggest a systematic difference in the ultimate functional form. There is no general theorem at work. This is the observed empirical regularity. The distribution-free Maximum Score estimates are included for comparison. It is unclear

[26] See Nagler (1994) and the Stata manual.

Table 10.1: Estimated Partial Effects for Binary Choice Models

Model	Female	Age	Education	Married	Income
Probit	0.1354	0.0051	-0.0046	-0.0127	-0.0776
Logit	0.1360	0.0051	-0.0048	-0.0139	-0.0760
Burr (Scobit)	0.1348	0.0051	-0.0043	-0.0103	-0.0824
Comp. Log Log	0.1335	0.0051	-0.0040	-0.0079	-0.0854
Extreme Value	0.1367	0.0052	-0.0050	-0.0175	-0.0677
ArcTangent	0.1366	0.0051	-0.0050	-0.0150	-0.0747
Linear Probability	0.1357	0.0050	-0.0048	-0.0103	-0.0744
Maximum Score*	0.4874	0.1671	-0.2690	-0.6846	-0.4112

*Coefficients. $\beta'\beta = 1$

what is estimated by the Maximum Score coefficients; they are normalized by $\boldsymbol{\beta}'\boldsymbol{\beta}$=1. In principle, ratios of coefficients should match. Within a row, b_{Female}/b_{Age} should be the same irrespective of the scaling. The ratio is 27 for the probit model (and the others) and 3 for MSCORE. We will return to this issue in Section 10.8.

A misconception about the LPM suggests that it is less parametric than the probit model. Angrist and Pischke emphasize that the LPM appears to be a reliable approximation to the underlying, appropriate model. The LPM is a model, albeit a flawed one. The curvature is an essential element of the regularity conditions for the model structure. Bliss (1934) emphasized the need for curvature in the dosage response function. It is not clear what OLS estimates in the binary choice case. K. Chen et al. (2025) and Li et al. (2022) have explored this issue.

10.6.3 A Theoretical Framework for the Threshold Model Based on Maximum Random Utility

We emphasize the choice aspect of the model for discrete choice between two alternatives. Building on Luce (1959), binary choices are taken to be the outcomes of utility maximization. Other paradigms, such as minimum regret (Chorus, 2010) appear briefly in the literature but do not gain a following. There exist utility functions defined over the alternatives, conditioned on attributes of the choices and characteristics of the chooser. A 'choice' should be viewed as a context plus a bundle of attributes. E.g., [Travel mode, price, travel time, income]. Utilities obey the standard axioms of choice, transitivity, completeness, continuity and independence from irrelevant alternatives. We define utility functions over the two alternatives. Since utility is latent, questions of functional form are moot. The standard linear functional form is

assumed without loss of generality. Specific nonlinearities can be built into the index functions by transformations of variables.

Outwardly identical individuals facing identical alternatives might make different choices – utilities are heterogeneous. We thus adopt a random utility framework. Again WLOG, the random term is taken to be additive. There are boutique treatments of multiplicative random utility such as Fosgerau & Bierlaire, 2009, but these add unnecessary complication.[27] The template dgp for the binary choice model is the single index threshold model for binary choice. The model is parameterized by the index function, $\boldsymbol{\beta}'\boldsymbol{x}$ and by the distribution of the random term.

10.6.4 Nonparametric vs. Parametric Approaches

Bliss's hand-drawn dosage response curve in Figure 10.1 is a nonparametric estimate of the underlying feature of the ecosystem. With only 11 data points, it would be difficult to improve on the technique. Bliss was skeptical of the accuracy of the technique of hand drawing the curve in short segments. With more data (and computing power), the author could have used kernel density estimators to obtain a smoother, presumably more accurate, still nonparametric estimate of the dosage response curve.

It is straightforward to envision what this nonparametric approach would produce for individual treatment-response data. The raw graphical data would be two blankets of points at y = 0 and y = 1. The kernel smoother would plot an S- shaped curve if the data tended to clump at zero with low values of the treatment, or an inverted S-shaped curve if the data tended to clump at one with low values of the treatment. Results from the nonparametric approach are largely impressionistic. Specific calculations such as the dosage response are imprecise, by construction. The function in Figure 10.2 is a parametric model of the dosage response.[28] The model is $\Phi^{-1}(R_i)+5 = \alpha+\beta\times logD_i+\gamma\times d_i\times (logD_i$ - 1.3) + ε_i where $d_i = \mathbf{1}[logD_i \leq 1.3]$ and ε_i is distributed as N[0,σ^2]. The model is fully parameterized. If the normality assumption is dropped, what remains is a semiparametric model. Note that the use of the inverse normal distribution in the shape of the function is not related to the full- vs. semi-parameterization of the model. It is akin to using logs in a loglinear model. (Berkson (1951) would have used 'logit(R_i)+5.') The index function is the right hand side of the regression.

---

[27] The specification considered additive vs. multiplicative random terms in a multinomial choice model and compared the 'fits' on the basis of the log-likelihoods. This seems dubious – the test compared different non-nested functions.

[28] The kinked function is a linear spline. See Hansen, 2022, pp. 727-729.

10.6.5 Scaled Coefficients

Without information about the distribution of the disturbance, there is no information about the scale of the coefficients. The parameters are only identified up to scale. Utilities are not observed, only the sign of the difference, which is invariant to scaling of β. In choice modeling with indirect utilities, ratios of coefficients are sometimes of interest. E.g., if β_{Income} is the marginal utility of income, then the marginal value of time is measured by the willingness to pay for more or less of it, $\beta_{Time}/\beta_{Income}$. The assumption of a specific functional form reveals the intrinsic, model specific scaling parameter. The partial effects are a model specific multiple of the structural parameters. For the probit model, the scale factor is the normal density. A similar calculation appears for the other parametric distributions listed in Table 10.1. For any specific distribution, the empirical estimator is the sample average of the derivative of the probability with respect to the constant term. Average partial effects are the scale factor times the respective coefficient. A typical logit coefficient estimator will produce partial effects of roughly .5(1-.5) times β_{logit}. The probit counterpart will be roughly .3989 times β_{probit}. Comparing the structural estimates, the usual logit coefficient values will be approximately .3989/.25 = 1.6 times the probit counterpart.[29] The linear probability model is not distribution-free. It has the same implicit regularity conditions (bounded support, monotonicity, continuity, etc.). The implicit scaling for the linear model is 1.0. The derivative of the probability with respect to the constant term is 1.0, not a function of the data. The implicit density is uniform (see Heckman & Snyder, 1996). Since the probability function is incoherent, the linear function can only be viewed as an approximation. The practical appeal of the LPM is that, in principle, it sidesteps these scaling issues. The implicit scale factor is 1.0 and the partial effects are the coefficients, themselves.

In many if not most applications, the target of estimation is a partial effect or a treatment effect. With a parametric model, this requires a post-estimation calculation, multiplication of the index coefficients by the average density based on the data. This calculation is a routine part of the analysis, facilitated by modern software such as `margins` in Stata. Notably, this is explicitly not possible with the semiparametric, model-free estimators (MSCORE and MRC). The analysis is confined to 'signs and significance' of structural parameters or the differences or ratios of parameters (see, e.g., Carrasco, 2001's application in Section 10.7.1).

There is an element of a theorem that would provide what we need here. Ruud (1983) provides sufficient conditions for the general MLE's to produce the vector of scaled coefficients that we are interested in.[30] If the distribution of ε is sufficiently regular (for example, normal or logistic) and $(\mathrm{x}_i, \varepsilon_i)$ are a random sample, then under certain conditions on the data, the MLE of $\boldsymbol{\beta}$ estimates $\gamma \times \boldsymbol{\beta}$ for a scalar γ. The conditions on $\mathbf{X}$ are uncertain, but multivariate normality is sufficient. It is difficult to

[29] These approximations are closest when the average outcome is close to 0.5. For the results in Table 10.1, the average outcome is 0.66. The ratio of the *Female* coeffficients is 0.62723/0.38003 = 1.650. For *Age*, the coefficients are 0.02383/0.01449=1.644.

[30] Thanks to Professor Christopher Winship for pointing this result out to me.

characterize how close the data are to multivariate normality, but altogether Ruud's results are suggestive for Table 10.1, not just for probit and logit, but also for the LPM and, e.g., the arctangent model. Ruud's result does not restrict the assumed or true probability function.

Ruud (1983, p. 228) notes: *"This note presents a striking result: consistency of the MLE for discrete dependent variable models for general misspecifications of the probability density function. This result rests, however, on an assumption about the distribution of the explanatory variables that is too restrictive to be generally applicable. Because alternative methods of estimation to maximum likelihood such as Manski's maximum score method are still impractical however however, investigation along these lines is useful."* By 1992 MSCORE and Han's MRC estimator were at least practical.

10.6.6 Distribution-Free Semiparametric Approaches

Numerous flexible distribution-free (semi-parametric) approaches have been suggested. Manski (1975) proposed the Maximum Score (MSCORE) estimator. This requires only some regularity conditions on the latent density. MSCORE resolves the normalization problem with $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$. The index function is estimable only up to scale. The probabilities are not estimable. The signs of the estimates match the other models in Table 10.1. The magnitudes do not. The ratios should match. They do not either. Cosslett (1983) proposed a different distribution-free approach based on maximum likelihood. Like MSCORE, the constraint $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ is necessary to secure identification. Han (1987) proposed an approach based on Maximum Rank Correlation. Klein and Spady (1993) and Ichimura (1987) offered similar semiparametric, single index function approaches to the smooth continuous function Prob(y=1|x) based on kernel methods and maximum likelihood. Matzkin (1992) lists a variety of approaches that are less than parametric, including Stoker (1986) and Gallant (1981, 1982). Recent developments include Yildiz (2013), Khan (2013), Chesher, Rosen and Smolinski (2013), and Lewbel, Dong & Yang, 2012) and Zhi (2025). All of these depart from minimal regularity conditions for an underlying density. In all cases, it is necessary to assume identifiability of the parameters (up to scale) given the density of ε. Well behaved data and the linear index form are typically assumed. The remaining assumptions about the distribution of the random term include smoothness, continuity and nondecreasing. Bounded support of the probability in (0,1) (i.e., coherency) is generally assumed.

These are the minimal assumptions needed to proceed. The linear model is unreasonable. The model that specifies that the probability function crashes into the boundaries (the 'ramp' function) is an unrealistic construction (see Lee, Lee and Choi (2023) and Amemiya (1975) and, esp. Amemiya, 1985).,

With only the regularity conditions, the semiparametric estimator does gain robustness to heteroskedasticity (Khan (2013) and Zhi, 2025). Finally, Lewbel's 'special regressor method' (Lewbel et al., 2012) imposes certain binding regularity

conditions on one of the right hand side variables and secures point identification in the bargain. The received literature contains a steady stream of contributions to the theoretical binary choice literature. The applied literature is still dominated by the probit and logit forms, with the resurgence of the LPM as noted.

10.6.7 Theory and Practice For The LPM

Standard pedagogy for Econometrics treats the LPM as a flawed method. For example, Hansen, 2022, p. 831 states: "*Overall, the linear probability model is a poor choice for calculation of probabilities*."

There is a misconception about the LPM that it is less parametric than the probit model. The assumed uniform distribution secures the point estimation of the parameters and the function to use to compute the probabilities. It is unclear why a linear index, ($\boldsymbol{\beta}$'$\mathbf{x}$) should be a better approximation to the underlying density than a nonlinear function such as $\Phi(\boldsymbol{\beta}'\mathbf{x})$. In practical terms, the influence of the functional form assumption appears to be of second order. The example in Table 10.1 shows the average partial effects of five variables in a set of parametric models. As often noted, the LPM is essentially the same as the others. If we argue that there is an underlying regular density, it will be a smooth continuous monotonic CDF. The placement of the empirical result will respond to the data. Whatever form one might choose, it can be argued that it is an approximation to the true model. Modern causal inference, post revolution, has focused on this feature of the linear probability model.

> "*[W]hile a nonlinear model may fit the CEF for LDVs more closely than a linear model, when it comes to marginal effects, this probably matters little. This optimistic conclusion is not a theorem but, as in the empirical example, it seems to be fairly robustly true*" (Angrist & Pischke, 2009, p. 107).

The authors argued for using the LPM for binary choice. The OLS results seem to mimic the partial effects of other nonlinear models such as probit and logit. Further, they argue that OLS usually provides an acceptable approximation to partial effects even if the probability models aren't close. Results such as those in Table 10.1 can be persuasive.

The typical finding is that the LPM seems to give the same results as the probit or logit models. However, as has been amply documented elsewhere, e.g., Lewbel et al., 2012, the LPM does frequently go awry. The results in Table 10.2 are based on the data in Bertschek & Lechner, 1998's study of innovation by German manufacturing firms. The data are a panel of 1,270 firms observed in five years. The values in Table 10.2 are three of the estimated partial effects out of 7 estimated (by this author). Suggestions that the LPM gives a reliable approximation to the appropriate model are usually based on comparisons to the probit or logit model. In Table 10.2, two

of the three LPM results (shown in bold) are clearly aberrant. It is unclear how to proceed with results such as these.

Table 10.2: Estimated Partial Effects in German Innovation Models

Model	SP	FDIUM	PROD
Probit	0.4133	1.0993	-0.9020
Logit	0.4600	1.1371	-1.0567
Burr	0.4762	1.1386	-0.9953
Extreme Value	0.5031	1.1402	-0.8832
ArcTangent	0.4954	1.1766	-1.2046
Linear Prob. Model	**0.0949**	1.1078	**-0.5501**

The authors of the Credibility Revolution argue that simplicity is a virtue. The model should be easy to explain to the audience (harmless: Angrist & Pischke, 2009). The case for the LPM is that it is simple and seems to work the way we would hope for most of the time. Familiar parametric models such as the probit and logit do not bring helpful information beyond what can be gleaned from familiar statistics (the LPM). That leaves open how the analyst should proceed with results such as those in Table 10.2.

10.6.8 An Endogenous Dummy Variable

The model of most interest in this discussion is the binary choice model with an endogenous regressor. The endogenous dummy variable is the ubiquitous model of endogenous treatment effects. The nonlinear counterpart for this case would be the recursive bivariate probit model (see Evans & Schwab, 1995 and Scott et al., 2009. The empirical tool would be IV-Probit. The linear model for this case would be the LPM estimated using two stage least squares (see, e.g., Angrist, 2001). There are several variants suggested for this case, notably, two step control function estimators (see Rivers & Vuong, 1988). The correspondence between the 2SLS estimator and a feature of the joint probability of the two binary variables is less clear than that of the partial effects in the simple probit model and the slopes of a conditional mean of the binary outcome. Li et al. (2022) have explored the issue with mixed results.

The modern 'treatment effect' specifies the model in that fashion. Two stage least squares is the estimator of choice (see Li et al., 2022). Developers including Blundell and Powell (2004) and Yildiz (2013) treated the model semiparametrically. Carrasco (2001) took a parametric approach to this model, building it around a 'switching' bivariate probit. Carrasco's application to the effect of fertility on labor

force participation was a combination of linear probability models and parametric bivariate probit models.

10.6.9 Fixed Effects in Panel Data

To this point, general results from Econometric theory and received empirical regularities would broadly favor the parametric nonlinear models over the linear probability model. Fixed effects in panel data with or without endogenous treatment effects is another ambiguous case.

A signature feature of the fixed effects estimator of the linear regression model is the use of the Frisch-Waugh theorem to condition out the effects via the within-groups linear regression. Algebraically, this convenience allows estimation of the full model with a simple linear regression, without having to compute the dummy variable coefficients. This procedure does not work for a fixed effects probit model. The only way to proceed is to fit the entire model, with the dummy variables, by brute force. This approach is likely to exceed the capacity of conventional software.

The fixed effects logit model is estimable without the dummy variables, using Chamberlain, 1980's conditional logit estimator. Unlike the linear model, however, the fixed effects are not recoverable. This approach is of limited usefulness – it is not possible to compute probabilities or partial effects.[31] The logit model does not admit a convenient counterpart to IV-probit. The treatment effects case is out of reach.

The effective barrier to the RBP model is not the size of the estimation problem.[32] The unconditional Probit/FEM estimator is inconsistent because of the Incidental Parameters problem. The only exact result in hand is the 100% bias of the logit estimator when there are two periods (see Abrevaya, 1997). Monte Carlo evidence (Greene, 2004) suggests that the result is sharper than that. Broadly, the 100% proportional bias appears to be general for two periods, it diminishes monotonically and becomes negligible as T approaches about 20. By the point where T is 5 or more, the bias is relatively small. The treatment case with two periods (Difference-in-Differences) is manageable directly using basic algebra, i.e., differences and a post treatment dummy variable. Hansen (2022) works through some of the algebra.

10.6.10 Multinomial Unordered Choices

We have considered the case of a single choice between two alternatives. The treatment effects model (recursive bivariate probit) adds an ancillary choice in a second equation. Building on Luce, 1959's development of the independence from

[31] It is possible to derive a FE estimator for α_i given a consistent estimator of β, for the individuals who switch during the panel.

[32] See Greene (2004).

irrelevant alternatives (IIA), McFadden, 1974 proposed the multinomial logit (MNL) model as a model for the probability that alternative j is the most preferred among a set of $J > 2$ alternatives. In order to reach a workable functional form, assumptions are random utility functions with IID type 1 extreme value (EV1) distributions. An extension that added a scale factor to the model produced the Generalized Extreme Value model and later facilitated development of the nested logit model.

The MNL model exhibits the IIA property. In practical terms, it implies that the odds of any two choices are independent of the probabilities of all the other alternatives. A peculiar mathematical result is that the elasticity of the probability of alternative j with respect to an attribute in any other choice probability is the same for all of the other choices.

McFadden suggested the MNL as a workable functional form for choice modeling – not because it mimicked behavior. Generations of researchers have proposed alternative functional forms that escape the IIA assumption. Gaudry and Dagenais (2005) was the first with the suggested DOGIT ("dodging the IIA assumption") model. The nested logit model was the next major innovation. The nested logit model specifies groups (nests) of alternatives. Within a nest, an MNL governs the DGP. Members of the nest are correlated by their common membership. The nested logit model was a major step away from IIA. We note, the common empirical regularity is that the aforementioned elasticities typically move only slightly away from their MNL counterparts. The next major innovation was the mixed (random parameters) logit model that was enabled by the development of fast, efficient simulation methods. Several forms of the MNL model are extended by treatments of the random effects.

The hybrid choice model (Daly et al., 2012) brings the functional form for multinomial choice modeling to its current state. We note that the progression of multinomial choice models follows a path of intricately specified extensions of the multinomial logit model

10.7 Ebb and Flow of Econometric Ideas

A few questionable notions have punctuated the discussion of the *Credibility Revolution*.[33] Some model specifications have lost traction in the applied literature. Some proposals (apparently) never gained traction. Many others are basic research that never appeared in the received applications.

10.7.1 On Model Building

All Models Are Wrong. The aphorism is attributed to George Box. The full quote, "*All models are wrong, but some are useful*", appears on p. 424 of Box and Draper

[33] Angrist and Pischke (2009, 2010). The latter borrows from Leamer (1983), *Let's Take the Con Out of Econometrics.*

(1987). By "wrong," Box meant that all models are approximations, some are better than others and none are perfect. As it appears here, the assertion is that basic tools like the LPM and 2SLS are model-free and thus more 'credible.' As suggested earlier, the LPM is not model-free. Broadly, advocates of the Credibility Revolution have suggested an aversion to formal parametric models that they argue are likely to affect empirical results.

The Wrong Model. It is argued that the particular choice of probit or logit for the functional form might lead to a specification error. This argues for a model-free platform (the linear probability model) against a model specification (probit or logit) because the model choice might be 'wrong.' In practice, this greatly exaggerates the differences in results produced by the formal models. With well behaved benign data, the probit, logit, other nonlinear and linear models usually produce indistinguishable partial effects and predicted probabilities. However, the linear probability model is not as model free as suggested. The linear functional form can drive the results when the data are strongly unbalanced (see Jacob and Levitt (2003) where 99% of the observations on y equal zero). Overall, a nonlinear functional form can mimic the linear model if needed, but the linear model cannot mimic the nonlinear one.

The Right Model. The theoretical argument that there is a **right model** for passively observed real data is specious. The parametric functional forms represent different ways to capture the essential curvature of the function, such as it is (see Bliss, 1934). There is no convincing argument for a sharply defined DGP, such as the logit model, outside the laboratory where data are simulated. Moreover, using the LPM at this point merely substitutes a particular functional form that is known not to be 'right' (because it is incoherent). A natural argument might appeal to a central limit theorem to suggest that the observed outcomes are a heterogeneous mix of data generating processes that can probably be approximated by a normal distribution. Bliss (1934) made this argument.

Model Driven Results. The Credibility Revolution in Econometrics emphasizes design-based empirical strategies over model-driven strategies. This suggests an influence of empirical models as frameworks for empirical analysis. This is argued to favor linear regression for binary outcomes, rather than probit models. In contradiction, defense of the LPM is usually based on how well it replicates the results of a corresponding probit or logit model (see the recent application in Section 10.8). Taken together, the empirical regularity suggests that the influence of the functional form is exerted over the unscaled coefficients. Without the scale, the index coefficients are not meaningful. Based on results like Table 10.1, it does appear that the influence of the specification on the scaled coefficients (partial effects) is not substantial.

This suggestion raises a dilemma for the model builder. The parametric models, such as probit and logit and the others in Table 10.1 are suspect because of their parameterization. Partial effects and probabilities are likely to be model driven. The distribution-free approaches avoid that problem, but they do not produce estimates of partial effects or probabilities because of the elusive scaling issue. It is hoped that the

LPM lands between these choices. But, the LPM turns out not to be model free and, moreover, appears to mimic the parametric models that seem to mimic each other. It has the conspicuous flaw that it does not properly constrain the probabilities to (0,1). The dilemma is that the estimates from the parametric approaches are tainted by their fixed structures. But, the semiparametric estimates are tainted because of their lack of structure.

Carrasco (2001) examined the effect of the endogeneity assumption on the estimated treatment (fertility) coefficient in a model of labor force participation. Two forms of the model were reported (in their Table 5). In the parametric nonlinear form, the endogeneity is embodied in two correlations. In the (also parametric) linear form, the model is estimated by OLS then 2SLS to reveal the effect of the endogeneity. The difference between the endogenous form and the exogenous form is (-.644,-.410) for the MLEs and (-.216,-.089) for the method of moments OLS/2SLS results. The author concludes (on page 391) "*when endogeneity is accounted for, the effect of fertility becomes more negative.*"

10.7.2 Empirical Models

Discrete choice modeling has generally not evolved by updating ideas that turned out to be flawed. Rather, new models, such as the ordered probit model (1975) and the mixed logit model (2000), accommodate new applications or types of applications. A few specifications have been proposed that have not gained much traction. For examples:

The multinomial and multivariate probit models require multivariate normal integration for estimation. Even with the GHK simulator (Geweke et al., 1994), the computational effort involved for large models and large sample sizes is extreme. For applications of this sort, various forms of the mixed logit model and the error components logit model have achieved the same generality much more easily and have been found preferable. The multivariate normal integration can be done satisfactorily using Monte Carlo simulation or an equivalent. Different forms of the mixed logit model have been simpler to apply.

The heteroskedastic probit model is observationally equivalent to a homoskedastic probit model with a heterogeneous scaling of the index function (the mean). The sample data contain no information about the variance of the random term in the random utility model. Model parameters are identified only under the assumptions about the disturbance variances that are not verifiable. There are relatively few applications of this model, though hard coded software is available (e.g., `hetprob` in Stata (see, e.g., Zhi, 2025).

The **nested logit model**. The nested logit model (Ben-Akiva, 1973; Hensher, 1986) was developed in a stream of proposals for generalizations of the multinomial logit

model in the 1970's and 1980's. The model does break the IIA assumption. However, the specific nesting scheme is ultimately ad hoc. Worse, different nesting schemes generally cannot be tested against one another. (A Wald, LR or Hausman test can be used to test IIA (the MNL model) against the broader model.) The logical equivalent of nesting can be produced by a form of error components MNL model. (For example, a four alternative MNL model, having one error component in utilities 1 and 2 and another in utilities 3 and 4 produces the functional equivalent of a four choice nested logit model with two branches. The nested logit model is now less common than the mixed model. Estimation of the mixed logit model is much simpler than the nested logit model.

10.7.3 Panel Data Treatments

A random effects (RE) probit model was operationalized by Butler & Moffitt, 1982. The two step FGLS procedure (and any GLS approach) is inappropriate. The common effect must be integrated out of the log-likelihood. The authors demonstrated how Hermite quadrature could be used to maximize the log-likelihod. By more recent treatments, the RE model would be eschewed for the pooled model with a cluster correction to the standard errors.[34]

The **Hermite quadrature method** of evaluating an integral in open form likelihoods remains a useful device. Evaluations of bivariate normal probabilities are faster (require less computation) with Hermite quadrature than with Owen's method or simulation. Notably, Stata developed a kit of random parameter models with their `GLAMM` command (see Rabe-Hesketh, Skrondal & Pickles, 2005). However, with more than one dimension of integration, the quadrature method becomes intolerably slow. With two dimensions, for example, the iterations within iterations require enormous amounts of computation.[35] At the same time, estimation times based on simulation methods are less than linear in the number of dimensions. Simulation based methods have largely replaced quadrature for econometric applications of random effects and random parameters.

The Heckman and Singer (1984) Method is a nonparametric approach to accommodating unobserved random effects in a parametric model. The method is described in Section 10.5.2.5. Estimation of the model is straightforward with the necessary software for finite mixtures in place. The amount of computation is comparable to Butler and Moffitt. However, identification is a more prominent complication in this setting. S. Cameron and Heckman (2001) is an application to

[34] Once again from Angrist & Pischke, 2009, p. xiii) "*we are not much concerned with asymptotic efficiency.*" Rather, robustness is preferred (see Stock, 2010).

[35] Lechner et al. (2008) put the practical limit for quadrature at four dimensions.

binary and multinomial choices.[36] A **Nonparametric/Parametric Hybrid Approach.** The Heckman and Singer approach could be folded into a parametric model to relax the constraint of the model's functional form. Post-estimation of the finite mixture of probit (or logit) models, estimates of $(\boldsymbol{\alpha}, \boldsymbol{\pi})$ are used to compute the expected constant term as $E[\alpha] = \boldsymbol{\pi}'\boldsymbol{\alpha}$. With this and the rest of $\boldsymbol{\beta}$ in hand, probabilities and partial effects can be computed for the probit (or other) model as usual.

Fixed Effects treatments for binary choice models have attracted a great deal of research. The starting point is the inconvenience of the basic fixed effects model. The within transformation does not condition out the effects. There are two ways to proceed:

Conditional Fixed Effects Estimator. Chamberlain (1980) devised a conditional likelihood function that does remove the effects. The joint probability for each group is conditioned on the sum of the outcomes. The resulting estimator of the slope parameters has all the desirable asymptotic properties. However, the fixed effects estimates are not recoverable, which implies that neither fitted probabilities nor partial effects can be computed. This severely limits the usefulness of the procedure.[37] The Chamberlain approach is extremely widely cited, primarily in the literature reviews of articles. The estimator, itself, is rarely used.

Individual Specific Time Trend. Alban Thomas (2006) proposes an extension of the fixed effects logit model in which α_i is replaced with an individual-specific time trend, $\alpha_i + \beta_i t$. "*Two estimators are proposed: a logit estimator based on double conditioning and a semiparametric, smoothed Maximum Score estimator based on double differences*". The method is applied to land renting decisions of Russian households.

Unconditional Fixed Effects Estimator. The practical problem of computing possibly millions of fixed effects in a nonlinear model is a significant complication. Greene (2004) devised a convenient algorithm for computing the full (unconditional) parameter vector in a fixed effects probit or logit model. The effects are computed separately from the other slopes by partitioning the log-likelihood. That partly solves the computational problem. There is an algebraic problem common to both conditional and unconditional estimators. Groups in which the outcome is always equal to one or always equal to zero fall out of the log likelihood. So, no effect can be

[36] We applied the technique to the model in Table 10.1 using the 887 groups with all 7 waves in the data. The estimates of σ for the random effects were: 0.87445 for the Butler and Moffitt method; 0.94266 for the random constant term (simulation) approach; and 0.86218 for the Heckman and Singer technique.

[37] The textbook treatment of this model typically emphasizes the complexity of Chamberlain's computation. Hansen (2022) derives the case for T=2, then drops the subject with "*The extension to $T > 2$ is similar but is algebraically complicated.*" (p. 843). The Krailo and Pike (1984) algorithm dramatically simplifies the calculation (see Section 10.3.1).

estimated in any event. Hansen (2022) notes that in his two-period case, only "switchers" remain in the estimating sample.

Incidental Parameters. The effective obstacle to the unconditional estimator - not to the conditional estimator - is the Incidental Parameters Problem. Abrevaya (1997) and Hsiao (2014) showed that for the unconditional fixed effects logit estimator with two periods, the bias is exactly 100%. There are no comparable exact theoretical results for other distributions or for $T > 2$.[38] Received Monte Carlo evidence suggests that the force of the IP bias might be proportional for the entire parameter vector, in the range from $2 \downarrow 1$ for $T = 2, \ldots$ and reaches tolerable levels in the range of $T = 10$ or 15. The exact result for $T = 2$ appears to be correct for the probit model as well.

Jackknife Estimator. Arellano & Carrasco, 2003 suggested a large T approximation. They developed a bias reduction procedure using a jackknife approach. (See Lechner et al. (2008) for details.) The approach is developed for a logit model and for an LPM. The models are applied to fertility data from the PSID data.

Mundlak Estimator. The within estimator for the linear model can be computed by adding the group means of the time varying variables to the pooled estimator. It has not been shown either way if this Mundlak approach, perhaps augmented with a random effect or some other correction could produce an (almost) consistent estimator of the slopes and the fixed effects. The second complication is estimating the mean of the distribution of the effects, post estimation. The estimates in hand exclude observations with number of ones equal to zero or T. Ultimately, what is needed is a useable estimator of $E[\alpha_i]$.

10.7.4 Aging Ideas

Discrete choice modeling is a subject that has evolved by adding ideas and techniques to the current platform. The contributions that age well last long enough to become part of the standard set of techniques. The probit and logit models have stood the test of time. Notwithstanding its obvious defects, the linear probability model has grown in popularity for 25 years (see Figure 10.3 below). Manski's (1975,1985,1986) Maximum Score estimator and Chamberlain, 1980's Conditional Logit approach have informed the development of other estimation methods.

Discrete choice modeling in its current form begins from 1965-1975 with Goldberger (1964), Theil (1970, 1971), Jöreskog and Goldberger (1975b) McFadden (1974), Nerlove and Press (1973) and the host of others who discovered the probit and logit models for individual choice and developed a robust stream of extensions. Ideas that aged well are those that persist in the methodology. In our calculation,

[38] There is no Incidental Parameters Problem for the unconditional Poisson regression estimator. See C. Cameron & Trivedi, 2005.

some ideas that aged less well might include some that have become obsolete, such as "common factor random effects models in panel data" and some that will be the subjects of ongoing research, such as 'fixed effects in panel data,' 'The Incidental Parameters Problem' and 'endogenous binary regressors (treatment effects)'.

10.7.5 Specification Errors in Model Coefficients

The main engines of empirical work involving discrete choice are the parametric probit, logit and linear models. Concern with making an arbitrary choice of distribution has driven contemporary researchers to the linear probability model, from the probit and logit models. The LPM is incoherent and ambiguous about what it estimates except in special cases (see Li et al., 2022). On the other hand, the LPM does seem reliably able to replicate the partial effects of other parametric models.

The wisdom that motivates the LPM suggests that the logit estimator is inconsistent if the probit model actually governs the data generating process, and vice versa. Hence the danger of specification error for using the 'wrong' model. Another possible interpretation is that regular parametric estimators are reliably robust estimators of the intrinsic vector of partial effects in the same way that any set of positive weights that sum to the sample size produces a consistent linear weighted least squares estimator, so long as the weights are uncorrelated with the disturbances and well-behaved. The feared specification error remains to be shown analytically. It is possible that the parametric estimators are collectively driving the results in the right direction. It is simple to generate experiments such as that in Table 10.1 that seem to suggest that the parametric estimators are all consistent. They might be considered quasi-MLEs for the template regular model, whatever it is. This possibility might help to focus ideas on what is meant by inconsistency in this context. The commonality in all the models in Table 10.1 are the partial effects, not the structural coefficients. Specification error is not an issue.

That leaves open the semiparametric estimators including MSCORE and MRC (Maximum Rank Correlation). It is unlear what these criterion based estimators estimate as no scale factor is produced to enable a comparison with the parametric models or with each other.

10.7.6 A Finite Sample Experiment

Consider the following experiment based on the data used to compute Table 10.1. Call the data $\mathbf{y}$ and $\mathbf{X}$ including the five regressors and a constant term. There are 3,377 observations. We compute the probit estimator, $\mathbf{b}_0$, for $\mathbf{y}$ on $\mathbf{X}$. We normalize $\mathbf{b}_0$ by dividing it by $(\mathbf{b}_0'\mathbf{b}_0)^{1/2}$. The normalized $\mathbf{b}_0$ will be the true index coefficients. Compute new data $\mathbf{y}_0 = \mathbf{X}\mathbf{b}_0 + \boldsymbol{\varepsilon}$, distributed N(0,1), then binarize $\mathbf{y}_0 = \mathbf{1}[\mathbf{y}_0 > \mathbf{0}]$. Data on $\mathbf{y}_0$ and $\mathbf{X}$ conform exactly to the probit model. They also satisfy the regularity

conditions for MSCORE, though MSCORE only requires that Med(ε) equal 1/2. The underlying normal distribution is smooth and continuous.

Probit and MSCORE should both be able to estimate the same $\boldsymbol{\beta_0}$ consistently. The estimates are shown in Table 10.3.[39]

Table 10.3: Probit and MSCORE Estimates

	True	Probit	MSCORE
Constant	-0.21552	-0.35013	-0.22324
Female	0.84236	0.86154	0.24854
Age	0.03128	0.03189	0.00588
Education	-0.02904	-0.01100	0.01247
Married	-0.07982	-0.09390	0.92682
Income	-0.48553	-0.67186	-0.17092

The differences are striking. The probit estimates are what might be expected. MSCORE appears to be the maverick. Manski and Thompson note that the MSCORE coefficients are not unique in a finite sample. Moreover, the maximum likelihood estimator must be more efficient (less variable) than the semiparametric estimate. The sample is moderately large by MLE standards, but perhaps not for MSCORE, which converges to its target more slowly than root-n. The criterion function for MSCORE is $\Sigma_i(2y_i - 1)sgn(\boldsymbol{x}_i'\boldsymbol{b})$. The value is 1,909 for MSCORE and 1,892 for the probit results. Maximum Score is better at predicting y than Maximum Likelihood, at least with these data.

Horowitz (1992). notes the following: " *The asymptotic distribution of the MS estimator is too complex for use in testing hypotheses about b or constructing confidence intervals. Manski and Thompson (1986) suggested using the bootstrap to estimate the mean-square error of the MS estimator and presented Monte Carlo evidence suggesting that the bootstrap works well for this purpose. However, it is not known whether the bootstrap consistently estimates the asymptotic distribution of the MS estimator."* Inference based on MSCORE is problematic.

10.8 Conclusions

This chapter has surveyed the history and current state of econometric models with discrete dependent variables. Section 10.7 above indicates some areas that researchers are still studying, particularly panel data applications. A large part of the empirical toolkit and methodology of discrete choice modeling can be collected under an

[39] The MSCORE coefficients are computed using Manski's original Fortran code.

umbrella of models and methods for analyzing binary choices. The preceding sections detailed various specifications that form this platform. The estimator of choice for most parametric models is Maximum Likelihood, though there are Bayesian and Minimum Chi-Squared applications and Minimum Regret and Maximum Entropy have appeared in other contexts. We also encountered Maximum Score and Maximum Rank Correlation estimation in this review. Where possible (outside of this review) authors have derived (and dismissed) asymptotic properties of consistency, asymptotic normality and asymptotic efficiency. Appropriate inference procedures have been rigorously established and standardized. Consistent estimators of known features of interest such as average partial effects (scaled coefficients) have been proposed. We presented a chronology of many of the innovations that have brought this literature to its current state.

10.8.1 Rise and Fall of The Linear Probability Model

The modern form of discrete choice modeling took shape in the late 1960's. A model for binary choices represented a new form of long familiar linear regression modeling. It came to be understood early on that the linear model was an inferior approach for individual binary data. It did not confine probabilities to [0,1] and therefore was incoherent. By ca. 1975, the linear model for binary choices was abandoned and buried. The linear probability model (estimated by OLS) survived at the fringe of the literature until 2000 or so, but was generally viewed as a flawed ('wrong') model for which there was a good alternative, the probit (or logit) model estimated by maximum likelihood.

10.8.2 The Credibility Revolution and The Rise of the LPM

The linear probability model resurfaced, zombie-like, ca. 2000. The long dead model was resurrected during the *Credibility Revolution* by, e.g., Angrist & Pischke, 2010. It was argued that the nonlinear probit model is based on a 'strong' assumption that could drive the empirical results. A model-free specification based on simple common statistics (OLS) was preferable to a formal nonlinear model. Lee et al. (2023) demonstrated that with sufficient assumptions, the OLS slopes, a common statistic, generally estimated an identifiable and useful quantity. K. Chen et al. (2025) showed that the LPM could be revived with sufficient assumptions (that could not be met by a real data set). The Linear Probability Model was resurrected in spite of its incoherence. The model has been the platform of choice for thousands of recent applications (see Figure 10.3 below). Linear 2SLS is likewise incoherent (Li et al., 2022) but similarly popular for binary choice modeling. (The econometric argument for 2SLS is less persuasive.) Growth in usage of the LPM since 2000 has been robust and sustained, as shown in Figure 10.3.

10.8.3 Impressionistic Asymptotic Results

The single index models are specified so that the structural parameters appear in the form $\boldsymbol{\beta}'\boldsymbol{x}$ and the probability of interest in the model is $F(\boldsymbol{\beta}'\boldsymbol{x})$. The random utility model (RUM) is $U = \boldsymbol{\beta}'\boldsymbol{x} + \sigma\varepsilon$ and $y = \mathbf{1}[U > 0]$. The random part of the random utility is $\sigma\varepsilon$. Particular standardized distributions have an implicit scale parameter, e.g., 1.0 for the normal, $1/\sqrt{12}$ for the uniform and $\pi/\sqrt{6}$ for the logistic. Parameters are estimable (identified) only up to scale because only the sign of the utility function is observed. Any positive multiple of $(\boldsymbol{\beta}, \sigma)$ produces the same binary data. Since only $\boldsymbol{\beta}/\sigma$ are estimable, σ is normalized to 1.0. For the semiparametric cases, MSCORE and MRC, the primitive distribution is not specified so a different normalization is required. The estimated parameters are normalized by $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$.

The structural parameters are not meaningful by themselves. In principle, they are marginal utilities. For empirical purposes, the signs have direct meaning. Signs and statistical significance will be meaningful in context. Values will differ by model. Ratios of marginal utilities may be of interest. Based on the preceding, it would appear that, e.g., WTP for a choice attribute is estimable in all five settings considered. There is a particular scaled coefficient of interest, the average partial effect. The probability at the heart of the model is Prob[U $\leq 0|\boldsymbol{x}$], or (one minus) the CDF. The partial effect is the derivative. For a parametric single index model, the partial effect is the derivative of the probability with respect to the constant term times the associated structural parameter. This can be computed if the probability function is known, which is the case for the parametric models, but not for MSCORE or MRC. Table 10.4 below shows the estimated scale factor for some of the parametric models in Table 10.1.

Table 10.4: Scale Factors for Partial Effects: Income

Model	APE	Coefficient	Scale
Probit	-0.07756	-0.21904	0.35409
Logit	-0.07605	-0.35285	0.21553
ArcTangent	-0.07466	-0.28804	0.26794
Linear	-0.07442	-0.07442	1.00000

Angrist & Pischke, 2009, p. 80, assert: "*[W]hile a nonlinear model may fit the CEF for LDVs more closely than a linear model, when it comes to marginal effects, this probably matters little. This optimistic conclusion is not a theorem but, as in the empirical example, it seems to be fairly robustly true.*" The authors suggest that we can discover the partial effects of almost any model by using a generic projection or linear estimator (OLS) of the conditional mean. They concede that this is not an analytical result, but an observed regularity. We have encountered a similar, very useful empirical regularity for regular parametric models: "*While the*

several parametric estimators in use all fit different structural parameters because of the implicit scaling, when it comes to partial effects, this seems not to matter much. The partial effects seem to be anchored to the empirical CDF of the data, and all estimators appear to estimate the same vector. This is not a theorem, but it appears to be fairly robustly true" (but, see Section 10.6.5). The results in Table 10.1 are suggestive.

Two conclusions are that the probit model and the LPM achieve the same end. The probit model has the advantage of being coherent. If one has a preference for the logit model, perhaps because they prefer to obtain odds ratios in the results, the same conclusion applies. The LPM estimator occasionally goes awry in comparison to other models. The results in Table 10.2 are suggestive and consistent with other similar cases. (See, as well, Section 10.6.5.) Overall, for estimating partial effects: "*If the main purpose of estimating a binary response model is to approximate the partial effects of the explanatory variables, averaged across the distribution of x, then the LPM often does a very good job...But, there is no guarantee that the LPM provides good estimates of the partial effects*" (see Wooldridge, 2010, p. 563).

10.8.4 Specification Searches

There are a rich variety of issues to consider in specitying the binary choice model:

Specification Error Due to the Wrong Model. Concern about possible specification error from choosing probit instead of logit or vice versa is misguided. The structural coefficients will differ. But, before scaling, they are meaningless. This difference is not a specification error. The two models will estimate the same partial effects when properly scaled.

Avoiding a Parametric Model. Using the LPM instead of probit/logit to avoid a parametric model is misguided. The LPM is also parametric. See, for example, Heckman & Snyder, 1996. The assumed distribution is Uniform[0,1] and the scale factor equals 1.0. All three estimate the same partial effects. It does not seem that the estimated probabilities or partial effects are systematically model-specific.

Model-free Results. The concern about using a parametric estimator that will drive results is misguided as all parametric estimators including the LPM seem to be estimating the same things.

Maximum Score and Maximum Rank Correlation. These are established distribution-free approaches. Neither estimates probabilities or scaled coefficients. Coefficient vectors are normalized to length 1.0. The convergence rate is less than root-n. The solution is not unique. Bootstrapping doesn't work, which motivated Horowitz's (1992) estimator. The advantage to using these semiparametric estimators in an empirical application is unclear. Both estimators discard the sample

information that includes covariation of **x** and y.

RBP vs. 2SLS. None of the impressionistic asymptotics can be claimed for 2SLS. The true asymptotic properties for 2SLS are unknown. The linear 2SLS results are used in the hope of obtaining an adequate approximation. RBP should be preferable (see Li et al., 2022). These authors found that unlike OLS vs. MLE/Probit, the 2SLS and the ML estimators locate different parameters.

Causal Estimates. The validity and/or appropriateness of the LPM vs. a nonlinear alternative is not the main methodological issue. In the generic case, as in our example, the LPM is (re)emerging as a substitute for the probit or logit or a distribution free model such as MSCORE or Maximum Rank Correlation. The relevant application in the causal inference literature is a binary choice outcome with an endogenous treatment variable, either the recursive bivariate probit model for a binary treatment or the IV-probit model for a continuous treatment. The 'credible' comparison is not OLS vs. probit; it is 2SLS vs. the RBP or 2SLS vs. IV-Probit. The complication appears to arise from equating 2SLS to OLS. There is no direct theoretical underpinning for using 2SLS to estimate a recursive probit or IV-probit model. Support for the approach is "*... linear regression provides useful information about the conditional mean function regardless of the shape of this function* (see Angrist & Pischke, 2009, p. xiii)." The difficulty is that 2SLS does not estimate a conditional mean; it estimates a ratio of parameters.

10.8.5 Fin de Siècle

Econometrics has evolved steadily since Ragnar Frisch inaugurated the field in 1933. Among the clearly apparent changes is Stock, 2010's "Other Transformation" that brought an increased focus on robustness and away from efficiency. Advocates of the Credibility Revolution (Angrist & Pischke, 2009) saw a trend toward research design supported by simple generic statistics (linear regression and instrumental variables) and away from intricate nonlinear models (estimated by maximum likelihood). This is encapsulated in a rivalry between the linear probability model and the probit model.

An editorial by a coeditor of a major journal.[40]
"*[E]very once in a while, I will get admonished by an anonymous reviewer for my use of the LPM, and so I wanted to write something about it. Ultimately, I think the preference for one or the other is largely generational, with people who went to graduate school prior to the Credibility Revolution preferring the probit or logit to the LPM, and with people who went to graduate school during or after the Credibility Revolution preferring the LPM. As always, the right way to approach*

[40] See Bellemare, 2026 Visited 5/17/2026.

things is probably to estimate all three if possible, to present your preferred specification, and to explain in a footnote (or show in an appendix) that your results are robust to the choice of estimator."

Recent Applications.
Jacob and Levitt (2003), footnote 25, p. 868 reported: "*Logit and Probit models evaluated at the mean yield comparable results, so the estimates from a linear probability model are presented for ease of interpretation.*"

Couture et al. (2024) studied whether changes in legalized gambling laws affected the mental health of individuals. They reported: "*We generate a dichotomous variable that is equal to one if an individual has at least one day in the last 30 where they deemed their mental health to be poor and we estimate linear probability models ... We estimate regressions of the following form:* $Y_{ist} = \beta X_{ist} + ... + \varepsilon_{ist}$ *where ... 30 days.*[6]" Footnote 6 states "*Our results are qualitatively and quantitatively similar if we estimate a logit model instead*".

10.8.6 Zombie Econometrics

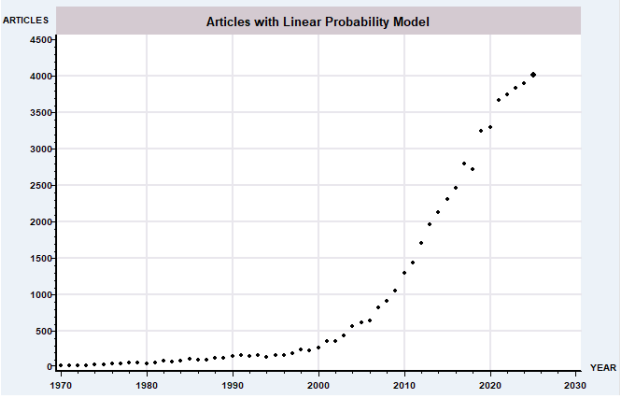

Fig. 10.3: Google Scholar Search for Linear Probability Model

Reports of the demise of the nonlinear probit model at the hands of the *Credibility Revolution* are premature. The number of applications of the reaccepted Linear Probability Model is indeed large and growing (see Figure 10.3). But, the binary logit and probit models are growing as well. A Google Scholar search for articles that mention "linear probability model," "probit model" and "logit model" in 2025 yields 4,000+, 12,000+ and 16,000+ references, respectively. The LPM is a common approach for single equation cases with endogenous treatment effects. The ML based

nonlinear models, such as multinomial and ordered choice models remain the natural approaches in other cases.

References

Abrevaya, J. (1997). The equivalence of two estimators of the fixed effects logit model. *Economics Letters*, *55*(1), 41-43.

Aitcheson, J. & Silvey, S. (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika*, *44*(2), 131-140.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, *41*(6), 997-1016.

Amemiya, T. (1975). Qualitative response models. *The annals of economic and social measurement*, *4*(3), 363-372.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, *19*(4), 1483-1536.

Amemiya, T. (1984). Tobit models: a survey. *Journal of Econometrics*, *24*(1), 3-61.

Amemiya, T. (1985). *Advanced econometrics*. Harvard University Press.

Angrist, J. (2001). Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business and Economic Statistics,*, *19*(1), 2-28.

Angrist, J. & Pischke, J. (2009). *Mostly harmless econometrics*. Princeton University Press.

Angrist, J. & Pischke, J. (2010). The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*(2), 3-30.

Arellano, M. & Carrasco, R. (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics*, *115*(1), 125-157.

Bellemare, M. (2026). *A rant on estimation with binary dependent variables (technical)*. http://marcfbellemare.com/wordpress/8951.

Ben-Akiva, M. (1973). *Structure of passenger travel demand models*. MIT Department of Transportation PhD Dissertation.

Berkson, R. (1944). Applications of the logistic function to bioassay. *Journal of the American Statistical Association*, *39*(227), 357-365.

Berkson, R. (1951). Why i prefer logits to probits. *JBiometrics*, *7*, 327-339.

Berry, S., Levinsohn, J. & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841-890.

Bertschek, I. & Lechner, M. (1998). Convenient estimators for the panel probit model. *Journal of Econometrics*, *87*(2), 329-372.

Bhat, C. (2001). Quasi-random maximum simulated likelihood estimatioin of the mixed multinomial logit model. *Transportation Research B*, *35*(7), 677-693.

Bliss, C. (1934). The method of probits. *Science*, *79*(2037), 38-39.

Blundell, R. & Powell, J. (2004). Endogeneity in semiparametric binary response models. *Review of Economics and Statistics*, *71*, 581-593.

Bollinger, C. & Martin, D. (1997). Modeling discrete choice with response errors. *Journal of the American Statistical Association*, *92*(439), 827-835.

Box, G. & Draper, N. (1987). *Empirical model building and response surfaces*. Wiley.

Boyes, W., Hoffman, D. & Low, S. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, *40*(1), 3-14.

Brown, S., Greene, W. & Harris, M. (2015). An inverse hyperbolic sine heteroskedastic latent class panel probit model: An application to modelling charitable donations. *Economic Modelling*, *50*(2), 228-236.

Butler, J. & Moffitt, R. (1982). A computationally efficient quadrature procedure for the one factor multinomial probit model. *Econometrica*, *50*(6), 761-764.

Cameron, C. & Trivedi, P. (2005). *Microeconometrics*. Cambridge University Press.

Cameron, S. & Heckman, J. (2001). The dynamics of educational attainment for black, hispanic and white males. *Journal of Political Economy*, *109*(3), 455-499.

Carrasco, R. (2001). Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labor participation. *Journal of Business and Economic Statistics*, *19*(4), 385-394.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, *47*, 325-338.

Chen, K., Martin, R. & Wooldridge, J. (2025). Another look at the linear probability and nonlinear index models. *BLS Working Papers*(WP569).

Chen, N., Gallego, G. & Tang, T. (2019). The use of binary choice forests to model and estimate discrete choices. *University of Toronto Management Dept.*, *Working Paper*, 1-49.

Chesher, A., Rosen, A. & Smolinski, K. (2013). n instrumental variable model of multiple discrete choice. *Quantitative Economics*, *42*(2), 157-196.

Chorus, C. (2010). A new model of random regret minimization. *European Journal of Transport and Infrastructure Research*, *10*, 181-196.

Contoyannis, C., Rice, N. & Jones, A. (2004). The dynamics of health in the british household panel survey. *Journal of Applied Econometrics*, *19*(4), 473-503.

Cosslett, S. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica*, *51*(3), 765-782.

Couture, C., Cross, J. & Wu, s. (2024). Impact of sports gambling on mental health. *Economics Letters*, *243*, 1-14.

Cragg, J. (1971). Some statistical models for limited dependent variables with applications to the demand for durable goods. *Econometrica*, *39*(5), 829-844.

Cragg, J. & Uhler, R. (1970). The demand for automobiles. *Canadian journal of Economics*, *3*, 386-406.

Cramer, J. (2002). The origins of logistic regression. *Tinbergen Institute, Amsterdam*, *Working paper*(WP TI 2002-119/4).

Daly, A., Hess, S., Patruni, B., Potoglou, D. & Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation*, *39*(2), 267-297.

Di Maria, C., Ferreira, S. & Lazarova, E. (2010). Shedding light on the light bulb puzzle: The role of attitudes and perceptions in the adoption of energy efficient light bulbs. *Scottish Journal of Political Eonomy*, *57*(1), 48-68.

Evans, W. & Schwab, R. (1995). Finishing high school and starting college: Do catholic schools make a difference? *The Quarterly Journal of Economics*, *110*(4), 941-974.

Fechner, G. (1860). Elemente der psychophysik. *Leipzig: Breitkopf and Hartel, Working paper*.

Fin, T. & Schmidt, P. (1984). A test for the tobit specificationversus an alternative suggested by cragg. *Review of Economics and Statistics*, *66*, 174-177.

Finney, D. (1943). *Probit analysis*. Cambridge University Press.

Finney, D. (1947). *Probit analysis*. Cambridge University Press.

Finney, D. (1971). *Probit analysis*. Cambridge University Press.

Fisher, R. (1935). *The design of experiments*. Oliver and Boyd, Edinburgh.

Fisher, R. & Yates, F. (1938). *Statistical tables for biological, agricultural and medical research*. Hafner.

Flannery, B., Press, W., Teukolski, S. & Vetterling, W. (2007). *Numerical recipes*. Cambridge University Press.

Fosgerau, M. & Bierlaire, M. (2009). Discrete choice models with multiplicative error terms. *Transportation Research Series B*, *45*(5), 494-505.

Frisch, R. (1933). Editorial. *Econometrica*, *1*(1), 1-4.

Gaddum, J. (1933). Methods of biological assay depending on the quantal response. *Medical research council*, *183*(2037), 1-33.

Gallant, R. (1981). On the bias in flexible functional forms and an essentially unbiased form. *Journal of Econometrics*, *15*, 211-245.

Gallant, R. (1982). Unbiased determination of production technologies. *Journal of Econometrics*, *20*, 285-323.

Gaudry, M. & Dagenais, M. (2005). The dogit model. *Transportation Research*, *13*, 105-112.

Gautier, E. & Kitamura, Y. (2013). Nonparametric estimation in random coefficient binary choice models. *Econometrica*, *81*(2), 581-607.

Geweke, J., Keane, M. & Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *The review of economics and statistics*, *76*(4), 609-632.

Godfrey, L. (1978). Testing for multiplicative heteroskedasticity. *Journal of Econometrics*, *8*(2), 227-236.

Goldberger, A. (1964). *Econometric theory*. John Wiley and Sons.

Greene, W. (1981). Sample selection bias as a specification error: comment. *Econometrica*, *43*(3), 795-798.

Greene, W. (2004). Fixed effects and the incidental parameters problem in the tobit model. *Econometric reviews*, *23*(2), 125-148.

Greene, W. (2018). *Econometric analysis, 8th ed.* Pearson Education,.

Greene, W. & Hensher, D. (2010). *Modeling ordered choices*. Cambridge University Press.

Greene, W. & Hensher, D. (2026). Hybrid choice models in transport research. *Foundations and Trends in Econometrics*, *35*(2), 1-175.

Gronau, R. (1974). Wage comparisons:a selectivity bias. *Journal of political economy*, *82*, 1119-1149.

Han, A. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of econometrics*, *35*(2), 303-316.

Hansen, B. (2022). *Econometrics*. Princeton University Press.

Harris, M. & Zhao, X. (2007). A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics*, *141*(2), 1073-1099.

Harvey, A. (1976). Estimating regression models with multiplicative heteroskedastidity. *Econometrica*, *44*, 461-465.

Hastie, T., Tibshirani, R. & Friedman, J. (2008). *The elements of statistical learning: data mining, inference and prediction*. Springer.

Hausman, J., Abrevaya, J. & Scott-Morton, F. (1998). Misclassification of the dependent variable in a discrete response setting. *Journal of Econometrics*, *87*(2), 239-269.

Heckman, J. (1974). Shadow prices, market wages and labor supply. *Econometrica*, *42*(4), 679-694.

Heckman, J. (1975). Estimation of income and substitution effects in a model of family labor supply. *Econometrica*, *42*(4), 73-85.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equations system. *Econometrica*, *46*(4), 931-959.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153-161.

Heckman, J. & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric mldels for duration data. *Econometrica*, *52*, 271-320.

Heckman, J. & Snyder, J. (1996). Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *NBER Working Paper*(5785).

Heckman, J. & Willis, R. (1975). A beta-logistic model for the analysis of sequential labor force participation by married women. *journal of political economy*, *85*(1), 27-58.

Helliwell, J., Layard, R., Sachs, J., DeNeve, J., Aknin, L. & Wang, S. (2026). World happiness report 2025. *University of Oxford: Wellbeing Research Centre*, *Working Paper*.

Hensher, D. (1986). Automobile-type choice: A note on alternative specifications for discrete-choice modelling. *Transportation Research Part B*, *20*(5), 429-433.

Hensher, D. & Greene, W. (2003). The mixed logit model, the state of practice. *Transportation*, *30*, 133-176.

Hensher, D., Rose, J. & Greene, W. (2015). *Applied choice analysis, 2nd ed.* Cambridge University Press.

Horowitz, J. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica*, *60*, 505-531.

Horrace, W. & Oaxaca, R. (2006). esults on the bias and inconsistency of ordinary least squares for the linear probability model,. *Economics Letters*, *90*(3), 321-327.

Hsiao, C. (2014). *Analysis of panel data*. Cambridge University Press.

Ichimura, H. (1987). *Estimation of single index models*. MIT Thesis.

Jacob, B. & Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The quarterly journal of economics*, *118*,

843-877.

Joreskog, K., Gruvaeus, G. & Thillo, M. (1970). Acovs - a general computer program for analysis of covariance structures. *ETS Research Bulletin Series*, *1970*(1).

Joreskog, K. & Sorbom, D. (1993). Lisrel 8: structural equation modeling with the simplis command language. *Scientific Software International*, *Working Paper*.

Joreskog, K. & Thillo, M. (1972). Lisrel: a general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. *ETS research bulletin 72*, *56*, 1-71.

Jöreskog, K. & Goldberger, A. (1975a). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association*, *70*(351), 631-639.

Jöreskog, K. & Goldberger, A. (1975b). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association*, *70*(351), 631-639.

Keane, M. (1992). A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics*, *10*(2), 193-200.

Khan, S. (2013). Distribution free estimation of heteroskedastic binary response models using probit/logit criterion functions. *Journal of Econometrics*, *172*(1), 168-182.

Klein, R. & Spady, R. (1993). An efficient semiparametric estimator for discrete choice models. *Econometrica*, *61*, 387-421.

Komarova, T. (2013). Binary choice models with discrete regressors: identification and misspecification. *Journal of Econometrics*, *177*(1), 14-33.

Krailo, M. & Pike, M. (1984). Conditional multivariate logistic analysis of stratified case control studies. *Applied Statistics*, *44*(1), 95-103.

Lambert, D. (1992). Zero-inflated poisson regression with an application to defects in manufacturing. *Technometrics*, *34*(2), 1-14.

Leamer, E. (1983). Let's take the con out of econometrics. *American economic Review*, *73*(1), 31-43.

Lechner, M., Lollivier, S. & Magnac, T. (2008). *Parametric binary choice models in matyas, l and sevestre, m. eds. the econometrics of panel data*. Springer.

Lee, M., Lee, G. & Choi, J. (2023). The linear probability model revisited. *Sociological methods and research*, *54*(1), 1-10.

Lewbel, A. (2000). Identification of the binary choice model with misclassification. *Econometric Theory*, *16*(4), 603-609.

Lewbel, A., Dong, Y. & Yang, M. (2012). Comparing features of convenient estimators for binary choice models with endogenous regressors. *Canadian Journal of Economics*, *45*(3), 809-829.

Lewis, H. (1974). Comments on selectivity bias. *Journal of political economy*, *82*, 1149-1155.

Li, C., Poskitt, D., Windmeijer, F. & Zhao, X. (2022). Binary outcomes, ols, 2sls and iv probit. *Econometric Reviews*, *41*(8), 859-876.

Luce, D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, *24*, 178-191.

Luce, D. (1957). *Games and decisions: Introduction and critical survey*. Wiley, New York.

Luce, D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley, New York.

Mabit, S. (2026). Mixed logit applications in travel behaviour research. *Transportation research part A*, *203*, 1-17.

Maddala, G. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge University Press.

Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics*, *3*(3), 205-228.

Manski, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, *27*, 313-333.

Manski, C. (1988). Identification of binary response models. *Journal of the American Statistical Association*, *83*(403), 729-738.

Manski, C. & Thompson, S. (1986). Operational characteristics of maximum score estimation. *Journal of Econometrics*, *32*(1), 85-108.

Maritz, J. (1965). Nonlinear probit analysis and its application to psychometric data. *Psychometrika*, *30*(1), 31-38.

Matzkin, R. (1992). Nonparametric and distribution free estimation of the binary threshold crossing and the binary choice models. *Econometrica*, *60*(2), 239-270.

McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior* (P. Zarembka, Ed.). Academic Press.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic andd Social Measurement*, *5*(4), 363-390.

McFadden, D. (1983). e choice theory approach to market research. *Marketing Science*, *5*(4), 275-297.

McFadden, D. & Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, *15*(3), 447-470.

McGillivray, R. (1970). Estimating the linear probability function. *Econometrica*, *38*(5), 775-776.

Meng, C. & Schmidt, P. (1985). On the cost of partial observability in the bivariate probit model. *International Economic Review*, *26*(1), 71-85.

Meng, T. & Lyu, S. (2022). The impact of the selective two-child policy on residents' fertility intentions in china. *Applied Economics Letters*, *29*(16), 1455-1459.

Meyer, B. & Mittag, N. (2017). Misclassification in binary choice models. *Journal of Econometrics*, *200*(2), 295-311.

Nagler, J. (1994). Scobit: An alternative estimatr to logit and probit. *American Journal of Political Science*, *38*(1), 230-255.

Nerlove, M. & Press, J. (1973). *Univariate and multivariate log-linear and logistic models*. RAND-R1306-EDA/NIH.

Newey, W. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of econometrics*, *36*, 231-250.

Ooms, M. & Doornik, J. (2006). Econometric software development: Past, present and future. *Statistica Neerlandica*, *60*(2), 206-224.

Patra, R., Seijo, E. & Sen, B. (2018). A consistent bootstrap procedure for the maximum score estimator. *Journal of econometrics*, *205*(2), 488-507.

Poirier, D. (1980). Partial observability in bivariate probit models. *Journal of Econometrics*, *12*(2), 209-217.

Rabe-Hesketh, R., Skrondal, A. & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*, 301-323.

Revelt, D. & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, *80*, 647-657.

Revelt, D. & Train, K. (1999). Customer-specific taste parameters and mixed logit. *http://eml.berkeley.edu/ train/papers/train0999.pdf*.

Riphahn, R., Wambach, A. & Million, A. (2003). Incentive effects on the demand for health care: a bivariate panel count data estimation. *Journal of Applied econometrics*, *18*(4), 387-405.

Rivers, D. & Vuong, Q. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, *39*(3), 347-366.

Ruud, P. (1983). Sufficient conditions for consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica*, *51*(1), 225-228.

Scott, A., Schurer, S., Jensen, P. & Sivey, P. (2009). The effects of an incentive program on uality of care in diabetes management. *Health Economics*, *18*, 1091-1108.

Stock, J. (2010). The other transformation in econometrics: Robust tools for inference. *Journal of economic perspectives*, *24*(2), 83-94.

Stoker, T. (1986). Consistent estimation of scaled coefficients. *Econometrica*, *54*(6), 1461-1482.

Swait, J. (1994). A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. *Journal of Retailing and Consumer Services*, *1*(2), 77-89.

Theil, H. (1969). A multivariate extension of the linear logit model. *International Economic Review*, *10*(3), 251-259.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American journal of sociology*, *76*(1), 103-154.

Theil, H. (1971). *Principles of econometrics: Frontiers of econometrics*. Wiley.

Thomas, A. (2006). Consistent estimation of binary-choice panel data models with heterogeneous linear trends. *The Econometrics Journal*, *9*(2), 177-195.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 278-286.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*(1), 24-36.

Train, K. (2001). Halton sequences for mixed logit. *Department of Economics, UC Berkeley*, *Working paper*.

Ven, W. & van Praag, B. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics*, *17*(2), 229-252.

Walker, S. & Duncan, D. (1967). Estimating the probability of an event as a function of several independent variables. *Biometrika*, *54*(1/2), 167-178.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data, 2nd ed.* MIT Press.

Yan, K. & Zhan, M. (2024). On the estimation of quantile treatment effects using a semiparametric propensity score. *Econometric Reviews*, *43*, 752-773.

Yildiz, N. (2013). Estimation of binary choice models with linear index and dummy endogenous variables. *Econometric Theory*, *29*(2), 354-392.

Zavoina, W. & McKelvey, R. (1975). A statistical model for the analysis of of legislative voting behavior. *American political science association*, *Working Paper*.

Zhi, N. (2025). Estimating a binary choice model with endogeneity in the absence of exclusion restrictions. *UFL Department of Economics*, *Working Paper*.

Index