Chapter 24

ECONOMETRIC ANALYSIS OF QUALITATIVE RESPONSE MODELS

DANIEL L. McFADDEN

Massachusetts Institute of Technology

Contents

Handbook of Econometrics, Volume II, Edited by Z. Griliches and M.D. Intriligator

1. The problem

The subject of this chapter is the econometric analysis of qualitative endogenous variables. Such variables may result from economic behavior which is intrinsically categorical, such as choice of occupation, marriage partner, or entry into a product market. Alternatively, they may come from classification during observation, such as the coding of housing choices into "sub-standard" or "standard". Qualitative variables may be binomial (yes/no) or multinomial, and multinomial responses may be naturally ordered (number of telephone calls) or unordered (freight shipment mode). There may also be multivariate combinations of discrete and continuous variables (appliance portfolio and energy consumption). Binomial and multinomial models are the primary subject of this chapter. A final section considers extensions.

The most suitable statistical model for qualitative response data will depend on the nature of the economic behavior governing the response and on the objectives of the analysis. Consistency with utility or profit maximization may be imposed for consumer or firm decisions such as educational level or plant location. Other data such as resolution of labor disputes (strike/settlement) may involve behavior of multiple agents. In some cases such as industrial accident data, the response model may not have a behavioral, or even a causal, interpretation. The discrete response will be of primary interest in many problems, but may also indicate self-selection into a target population requiring correction of sampling bias. An example is potential bias in analysis of housing expenditure in a self-selected population of renters.

2. Binomial response models

2.1. *Latent variable specification*

The starting point for econometric analysis of a continuous response variable y is often a linear regression model:

$$y_t = x_t\beta - \varepsilon_t, \tag{2.1}$$

where x is a vector of exogenous variables, ε is an unobserved disturbance, and $t = 1,\dots,T$ indexes sample observations. The disturbances are usually assumed to have a convenient cumulative distribution function $F(\varepsilon|x)$ such as multivariate normal. The model is then characterized by the conditional distribution

$F(y - x\beta | x)$, up to the unknown parameters β and parameters of the distribution F. In economic applications, $x\beta$ may have a structure derived exactly or approximately from theory. For example, competitive firms may have $x\beta$ determined by Shephard's identity from a profit function.

The linear regression model is extended to binomial response by introducing an intermediate unobserved (latent) variable y^* with:

$$y_t^* = x_t\beta - \varepsilon_t, \tag{2.2}$$

and an indicator function:

$$y_t = z(y_t^*) = \begin{cases} 0, & \text{if } y_t^* < 0, \\ 1, & \text{if } y_t^* \geq 0. \end{cases} \tag{2.3}$$

If $F(\varepsilon | x)$ is the cumulative distribution function of the disturbances, then just as in the continuous case the model is characterized by the conditional distribution of y given x:

$$\begin{aligned} P_1 &= P(z(y^*) = 1 | x) \\ &= P(y^* = x\beta - \varepsilon \geq 0) \\ &= F(x\beta | x), \end{aligned} \tag{2.4}$$

also termed the *response probability*.

2.2. *Functional forms*

The most common binomial models, which assume ε independent of x, are *logit* with

$$F(x\beta) = 1/(1 + e^{-x\beta}), \tag{2.5}$$

probit with

$$F(x\beta) = \Phi(x\beta), \tag{2.6}$$

where Φ is the standard cumulative normal, the *linear probability model* with

$$F(x\beta) = x\beta \qquad (0 \leq x\beta \leq 1), \tag{2.7}$$

and the *log linear model* with

$$F(x\beta) = e^{x\beta} \qquad (x\beta \leq 0). \tag{2.8}$$

The last two models require restrictions on the domain of the latent variable which may be difficult to enforce in estimation or forecasting.

The preceding models are derived from distribution functions with thin tails. Alternatives in which the response probabilities approach zero or one less rapidly can be constructed from the Student-t or Cauchy distributions; the latter yields the *arctan model*:

$$F(x\beta) = \tfrac{1}{2} + \frac{1}{\pi}\tan^{-1}(x\beta). \tag{2.9}$$

For a given latent variable model $y^* = x\beta + \varepsilon$, specification of the distribution function F for ε may change substantially the model's ability to fit data, particularly if restrictions are imposed on the domain of $x\beta$.[1] However respecification of the latent variable model can circumvent this problem. Suppose $F(\varepsilon)$ is any continuous cumulative distribution function, and $\tilde{x}\tilde{\beta} \doteq \ln(F(\beta x)/(1-F(\beta x)))$ is a linear (in parameters $\tilde{\beta}$) global approximation on a compact set[2] of βx satisfying $0 < F(\beta x) < 1$. Then to any desired level of accuracy, the response probability is logistic in the transformed latent variable model $\tilde{y}^* = \tilde{x}\tilde{\beta} + \varepsilon$:

$$F(x\beta) = 1/(1+\mathrm{e}^{-\tilde{x}\tilde{\beta}}). \tag{2.10}$$

Thus, the question of the appropriate F is recast as the question of the appropriate specification of arithmetic transformations $\tilde{x}$ of the data x in a logit model.[3]

2.3. *Estimation*

Consider a sample (y_t, x_t) with observations indexed $t = 1,\dots,T$, and a binomial model $P_{1t} = F(x_t\beta)$. Assume the sample is random[4] with independent observations. Then the log-likelihood normalized by sample size is:

$$L = \frac{1}{T}\sum_{t=1}^{T}\left[y_t \ln P_{1t} + (1-y_t)\ln P_{0t}\right], \tag{2.11}$$

[1] The logit and probit models however are rarely distinguishable empirically.

[2] The existence of such an approximation is guaranteed by the Weierstrauss approximation theorem. A constructive approximation theorem with explicit error bounds is given in McFadden (1981).

[3] Obviously, the logit base *cdf* could be replaced by any other continuous invertible *cdf* $G(\varepsilon)$, with $\tilde{x}\tilde{\beta} \doteq G^{-1}(F(x\beta))$.

[4] Specifically, the probability of being sampled is assumed independent of response; stratification with respect to x_t is permitted.

with $P_{1t} = F(x_t\beta)$ and $P_{0t} = 1 - P_{1t}$. The gradient of this function is:

$$L_\beta = \frac{1}{T}\sum_{t=1}^{T} w_t x_t (y_t - F(x_t\beta)), \tag{2.12}$$

with $w_t = F'(x_t\beta)/P_{0t}P_{1t}$, and the hessian is:

$$L_{\beta\beta} = -J_T + \frac{1}{T}\sum_{t=1}^{T} u_t x_t x_t' (y_t - F(x_t\beta)), \tag{2.13}$$

where

$$J_T = \frac{1}{T}\sum_{t=1}^{T} w_t^2 P_{0t} P_{1t} x_t x_t' = -EL_{\beta\beta} \tag{2.14}$$

is the information matrix, and

$$u_t = \left(F''(x_t\beta) + (P_{1t} - P_{0t})(F'(x_t\beta))^2\right)/P_{0t}P_{1t}. \tag{2.15}$$

Under mild regularity conditions, detailed in Section 3.2 below, the maximum likelihood estimator $\hat{\beta}$ of β is consistent, and $\sqrt{T}(\hat{\beta} - \beta)$ is asymptotically normal with mean zero and covariance matrix $J^{-1} = \lim_{T\to\infty} J_T^{-1}$. Solution of the normal equation (2.12) usually requires an iterative procedure. Optimizers such as Newton–Raphson, quadratic hill-climbing, or BHHH[5] work well if three cautions are observed:

(1) Accurate numerical approximations for $\ln F(x_t\beta)$ and $\ln(1 - F(x_t\beta))$ are needed in the tails of the distribution.

(2) There is a small (and vanishing) probability, in models where the domain of F is unbounded, that the maximum likelihood estimator will fail to exist and response is perfectly correlated with the sign of an index $x\bar{\beta}$. Adding a test for this condition during iteration permits detection of this case and estimation of the relative weights $\bar{\beta}$. For sample sizes of a few hundred, this outcome is extremely improbable unless the analyst has entered misspecified x variables which depend on y.

[5]See Berndt–Hausman–Hall–Hall (1974) and Goldfeld and Quandt (1972) for discussions of these algorithms. The largest component of computation cost in maximum likelihood estimation is usually evaluation of the response probabilities. Consequently, for maximum efficiency, the number of function evaluations and passes through the data should be minimized. This is usually achieved by using analytic derivatives calculated jointly with the likelihood for each observation. For initial search, it may be advantageous to calculate the hessian matrix required for the Newton–Raphson search direction rather than use the BHHH approximation. Methods such as Davidon–Fletcher–Powell which use numerical updates of the hessian matrix are not usually efficient for these problems. A careful interpolation along the direction of search (e.g. Davidon's linear search method which uses cubic interpolation) usually speeds convergence.

(3) The log-likelihood L need not be concave in the general case, and there may be local maxima. However, the logit, probit, and linear probability models for binomial response have strictly concave log-likelihood functions, provided the explanatory variables are linearly independent. A check of the condition number of the information matrix J_T during iteration should detect linear dependencies.

A family of consistent estimators of β can be derived by replacing w_t in (2.12) with other weight functions, which may depend on x_t and β but not the response y_t; for example $w_t = F'(x_t\beta)$ corresponds to non-linear least squares. These alternatives are usually inferior to maximum likelihood estimators in both computation and asymptotic statistical properties.

2.4. *Contingency table analysis*

In some economic applications, the number of configurations of explanatory variables is finite, and the data can be displayed in a contingency table with counts of responses in each cell. A variety of statistical methods are available for contingency table analysis; Goodman (1971) and Fienberg (1977) are general introductions. A common approach is to adopt a log-linear model of the *joint* distribution of (y, x) without imposing any structure of cause and response. The conditional probability of y given x will then have a logit form.

Log-linear models of contingency tables can be estimated by simple analysis-of-variance, and are often the most convenient method of obtaining a logit response probability when the dimension of x is not too large. It is difficult within this framework to impose prior restrictions from economic theory on the form of the response probability, a feature that most econometricians would consider a disadvantage.

2.5. *Minimum chi-square method*

Suppose the configurations of x in a contingency table are indexed $n = 1,\dots,N$, and let m_{in} denote the count in the cell with $y = i$ and configuration x_n. The log-likelihood function (2.1) in this notation becomes:

$$L = \frac{1}{T}\sum_{n=1}^{N}\left[m_{1n}\ln F(x_n\beta) + m_{0n}\ln(1 - F(x_n\beta)) + \ln C(m_{\cdot n}, m_{1n})\right], \quad (2.16)$$

with $m_{\cdot n} = m_{0n} + m_{1n}$, $C(m, r) = m!/r!(m-r)!$, and $T = \sum_{n=1}^{N} m_{\cdot n}$. Consistency of maximum likelihood estimates will follow whenever $T \to \infty$, provided a rank condition on the hessian is met. This can be accomplished by letting $N \to \infty$, all $m_{\cdot n} \to \infty$, or both, as long as N is at least the dimension of β.

When the $m_{\cdot n} \to \infty$, the cell frequencies $m_{1n}/m_{\cdot n}$ converge in probability to $P_{1n} = F(x_n\beta)$, suggesting an alternative method of estimating β. Let G denote the inverse of F, so that $G(P_{1n}) = x_n\beta$. A Taylor's expansion of $G(m_{1n}/m_{\cdot n})$ about P_{1n} yields:

$$G(m_{1n}/m_{\cdot n}) = x_n\beta + \varepsilon_n, \tag{2.17}$$

with

$$\varepsilon_n = G'(\tilde{P}_{1n})(m_{1n}/m_{\cdot n} - P_{1n}) \tag{2.18}$$

and $\tilde{P}_{1n}$ an intermediate point. If G' is continuous, then $m_{\cdot n} \to \infty$ implies:

$$\sqrt{m_{\cdot n}}\,\varepsilon_n \xrightarrow{\text{d}} N\big(0, (G'(P_{1n}))^2 P_{0n} P_{1n}\big). \tag{2.19}$$

Then β can be estimated consistently by applying least squares to (2.17). Large sample efficiency is improved by correcting for heteroscedasticity; note that $\sigma_n^2 = (G'(P_{1n}))^2 P_{0n} P_{1n}$ is estimated consistently by $(G'(m_{1n}/m_{\cdot n}))^2 m_{0n} m_{1n}/m_{\cdot n}^2$. This estimator was proposed for the logit model by Berkson (1955); further discussion of the approach is in Cox (1970), Goldberger (1964), Theil (1969), and McFadden (1973, 1976). Amemiya (1980) compares the second order asymptotic properties of the Berkson and maximum likelihood estimators.

Economic surveys seldom yield the large cell counts necessary for Berkson's method to have good statistical properties unless data are grouped. However, grouping introduces biases which in many cases are of unacceptable magnitude; see Domencich and McFadden (1975).

2.6. *Discriminant analysis*

When the response probability is deduced from an economic theory implying a causal relationship from x to y, then it is natural to parameterize this form directly as a function $P_1 = F(x\beta)$ consistent with the theory. In other cases it may be more natural to parameterize the joint distribution $H(y, x)$ of (y, x) or the conditional distribution $Q(x|y)$ of x given y. Letting $p(x)$ denote the marginal distribution of x and q_1 the marginal distribution of y, Bayes rule implies that the conditional distribution of y given x is:

$$P_1 = H(1,x)/(H(0,x) + H(1,x)) = Q(x|1)q_1/(Q(x|0)q_0 + Q(x|1)q_1). \tag{2.20}$$

This probability may in some cases have a parametric form commonly assumed for response models, and it may be tempting to give it a causal interpretation. However, a key property of a true causal response $P_1 = F(x\beta)$ is invariance with respect to the marginal distribution $p(x)$ of the explanatory variables. This invariance condition will be satisfied by (2.20) only if the parameterization of $H(y,x)$ or $Q(x|y)$ is "saturated" in x.[6]

Discriminant models parameterize the conditional distributions $Q(x|y)$, and may be motivated by an assumption of causality from y (subpopulation) to x (attributes of subpopulation members). For example, y may index subpopulations of sterile and fecund insects; then $Q(x|y)$ characterizes the distribution of observable attributes of these subpopulations and P_1 in (2.20) gives the probability that an insect with attributes x belongs to population 1. The commonly used normal linear discriminant model assumes the $Q(x|y)$ are normal with means μ_y and common covariance matrix Ω. This requires the x variables to be continuous and range over the real line. The conditional probability of y given x, from (2.20), then has a logit form:

$$P_1 = 1/(1+\mathrm{e}^{-\alpha - x\beta}), \tag{2.21}$$

with $\beta = \Omega^{-1}(\mu_1 - \mu_0)$ and $\alpha = \frac{1}{2}(\mu_0'\Omega^{-1}\mu_0 - \mu_1'\Omega^{-1}\mu_1) + \ln(q_1/q_0)$. The parameters μ_y and Ω can be estimated using sub-sample means and pooled sample covariance, $\hat{\mu}_y$ and $\hat{\Omega}$. Alternatively, ordinary least squares applied to the "linear probability model",

$$y = a + xb + \nu, \tag{2.22}$$

yields an estimator $b = \lambda\hat{\Omega}^{-1}(\hat{\mu}_1 - \hat{\mu}_0) = \lambda\hat{\beta}$, where $\lambda = r_0 r_1/(1+(\hat{\mu}_1 - \hat{\mu}_0)'\hat{\Omega}^{-1}(\hat{\mu}_1 - \hat{\mu}_0))$ and r_i is the proportion of sub-population i in the pooled sample. This relation between logit and linear model parameters under the normality assumptions of discriminant analysis was noted by Fisher (1939); other references are Ladd (1966), Anderson (1958) and Chung and Goldberger (1982). It should be emphasized that the relations (2.21) and (2.22) obtained from the discriminant model do not imply a causal response structure despite the familiarity of the forms. Also, if there is in truth a logistic causal response model, it will be coincidental if the distribution of x is the precise mixture of normals consistent with the normal conditional distributions $Q(x|y)$ assumed in discriminant analysis. Otherwise, use of the discriminant sample moments will not yield consistent estimates of the logit model parameters. There is some evidence, however, that the

[6]A model is "saturated" in x if it has enough parameters to completely characterize the marginal distribution $p(x)$ without prior restrictions on $p(x)$. A full log-linear model for $H(y,x)$ has this property.

discriminant estimates of (2.21) are relatively robust with respect to some departures from the normality assumptions; see Domencich and McFadden (1975) and Amemiya (1981).

3. Multinomial response models

3.1. Foundations

The latent variable model (2.2) which generated a binomial response probability generalizes readily to systems in which a vector of latent variables is determined by the explanatory variables, and a generalized indicator function maps the latent variable vector into a vector indexing observed response. Suppose there are m mutually exclusive and exhaustive possible responses, indexed $i=1,\dots,m$. Let y be an m-vector with $y_i=1$ if response i is observed and $y_i=0$ otherwise, and S_m the set of such vectors. Let

$$y^* = x\beta - \varepsilon \tag{3.1}$$

be a multivariate latent variable model with domain R^h for y^* and a cumulative distribution function $F(\varepsilon|x)$. The generalized indicator function is defined by a partition of R^h into subsets $A_1,\dots,A_m$, and $z: R^h \to S_m$ with $y=z(y^*)$ satisfying $y_i=1$ if and only if $y^* \in A_i$. Then the response probability is:

$$P_i = P(y_i=1|x,\beta) = P(x\beta-\varepsilon \in A_i) = F(\{x\beta\}-A_i|x), \tag{3.2}$$

where $\{x\beta\}-A_i = \{\varepsilon|x\beta-\varepsilon\in A_i\}$. The latent variable model (3.1) may be a reduced form from a simultaneous equation system, and x may include lagged values of y and y^*, permitting a rich dynamic structure. The relationship between the dimensions h and m, the structure of the sets A_i, and the form of F can all be varied to fit the application. We give three examples.

(a) *Multinomial choice*. Suppose $h=m$, and y_i^* is interpreted as a measure of the utility of alternative i. This may be an index of desirability for consumers, or profit for firms. Suppose A_i is defined to be the set of y^* which have $y_i^* \geq y_j^*$ for $j=1,\dots,m$, with some rules for breaking ties so that $A_1,\dots,A_m$ define a partition. Then the response probabilities can be interpreted as the proportions of agents maximizing utility at each alternative when faced with a decision characterized by x. For example, i may index the brand of automobile purchased and y_i^* the maximum utility given brand i, achieved by optimizing on all remaining dimensions of choice.

(b) *Ordered choice*. Suppose $h=1$, and $y^* = x\beta+\varepsilon$ is the "ideal" demand for a commodity if it were available in continuous quantities. Suppose the commodity is actually available in discrete quantities λ_i, and $U(y^* - \lambda_i)$ is the utility of λ_i when the ideal is y^*. Define a_i so that $U(a_i - \lambda_i) = U(a_i - \lambda_{i-1})$ and $A_i = [a_i, a_{i+1})$. Then the response probability:

$$P_i = P(x\beta + \varepsilon \in A_i) = F(a_{i+1} - x\beta | x) - F(a_i - x\beta | x), \tag{3.3}$$

gives the proportion of agents for which quantity λ_i is optimal. This model might be appropriate for describing the choice of number of children or frequency of shopping trips.

(c) *Multivariate binomial choice.* Suppose a vector of h binomial choices $y = (y^1, \ldots, y^h)$ is observed, with $y^j = 1$ if $y_j^* \geq 0$ and $y^j = 0$ otherwise. There are $m = 2^h$ possible observable vectors. In the general terminology, A_y is a cartesian product of half-lines, with term j equal to $(-\infty, 0)$ if $y^j = 0$, $[0, +\infty]$ otherwise, and $P_y = P(x\beta - \varepsilon \in A_y)$. If $\sum_{j=1}^{h} y_j^*$ is interpreted as an additively separable utility, with y_j^* the relative desirability of $y^j = 1$ over $y^j = 0$, then P_y gives the proportion of agents for which y is optimal. Dependence in the joint distribution $F(\varepsilon | x)$ generates dependence among the binomial choices. This model might be appropriate for describing holdings in a portfolio of household appliances, or for describing a sequence of binomial decisions over time such as participation in the labor force.

These examples should make clear that there is a rich variety of qualitative response models, drawing upon alternative latent variable structures and generalized indicator functions, which can be tailored for appropriateness and convenience in various applications. Multinomial, ordered, and multivariate responses can appear in any combination. In the third example above, multivariate binomial responses are rewritten as a single multinomial response. Conversely, a multinomial response can always be represented as a sequence of binomial responses. When observations extend over time, the system can be enriched further by treating ε as a stochastic process and permitting lagged responses ("state dependence") among the explanatory variables. With these elaborations, the full panoply of econometric techniques for linear models and time series problems can be brought to bear on qualitative response data. This development of the latent variable formulation of qualitative response models is due to Goldberger (1971), Heckman (1976), Amemiya (1976), and Lee (1981). The last paper also generalizes these systems to combinations of discrete, continuous, censored, and truncated variables. The examples above have been phrased in terms of optimizing behavior by economic agents. We shall develop this connection further to establish the link between stochastic factors surrounding agent decision-making and the structure of response probabilities. However, it should be noted that there are applications of qualitative response models where this framework is inappropriate, or where the analyst may not wish to impose it a priori. This will in general relax prior restrictions on the structure of $x\beta$ or the distribution $F(\varepsilon|x)$ in the latent variable model, but otherwise leave unchanged the latent variable system determining qualitative response. For example, the ordered response model (b) with the latent variable y^* interpreted as susceptibility and the a_i as thresholds for onset of a disease at varying degrees of severity is the Bradley–Terry model widely used in toxicology. Another example is the multivariate binomial model (c) applied to a sequence of outcomes of a collective bargaining process, with y_h^* interpreted as a measure of the relative strength of the opposing agents in period h.

Returning to the problem of qualitative response generated by optimization on the part of economic agents, consider the multinomial choice example (a). For concreteness, suppose the agent is a profit-maximizing firm deciding what product markets to enter or where to locate plants. Given a qualitative alternative i, the firm faces a technology T^i describing its feasible production plans. Maximization of profit subject to T^i yields a restricted profit function Π^i. The technology will depend on attributes t of the firm; the restricted profit function will consequently depend on t and on characteristics w of the firm's market environment, $\Pi^i(t, w)$. The firm will choose the alternative i which maximizes $\Pi^i(t, w)$.

The form of the restricted profit function Π^i will depend on prior assumptions on the technology and on the nature of the markets the firm faces. If, for example, the firm faces competitive markets and w is the vector of prices, then Π^i is a closed, convex, conical[7] function of w; see McFadden (1978a). In non-competitive markets, w summarizes the information available to the firm on strategies of other agents, and the form of Π^i is determined by a theory of non-competitive market behavior.

In empirical application, (t, w) will contain both observed and unobserved components, and the unobserved components will have some distribution over the population of firms. Let z denote the observed components of (t, w), and ν the unobserved components, and let $G(\nu|z)$ denote the distribution of the unobserved components, given z, in the population. Let $\overline{\Pi}^i(z)$ be the expectation of $\Pi^i(z, \nu)$ with respect to $G(\nu|z)$, or some other measure of location for the random function $\Pi^i(z, \cdot)$. Finally, let $x_i\beta$ be a linear-in-parameters global approximation to $\overline{\Pi}^i(z)$, where x is a vector of arithmetic functions of z, and define $\varepsilon_i = x_i\beta - \Pi^i(z, \nu)$. Then ε has a distribution $F(\varepsilon|x)$ induced by ν, and $y_i^* = x_i\beta - \varepsilon_i$ equals the maximum profit obtainable given discrete alternative i, written in the latent variable model notation. If all prices are observed and the function $\Pi^i(t, w)$ is closed, convex, and conical in prices, then the expectation $\overline{\Pi}^i(z)$ will have these properties. The approximation $x_i\beta$ to $\overline{\Pi}^i$ must then approximate these properties, although it need not have them exactly unless the family of functions $x(z)$ used in the approximation is selected to achieve this result. For example, a

[7]A function is conical if it is homogeneous of degree one; closed if the epigraph of the function is a closed set.

convex function $\overline{\Pi}^i$ can be approximated globally by a nonnegative linear combination of convex functions, or alternatively by a polynomial which may fail to be convex over some range; see McFadden (1978a). If it is important to the analysis to impose on the response model all the prior restrictions implied by the theory, as would be the case, for example, if the objective of the study were to test these restrictions, then an approximation should be chosen which inherits the prior restrictions and which does not in itself restrict the ability of the model to fit the data. Given the approximation $x_i\beta$, note that as a consequence of the definition of ε_i, the distribution $F(\varepsilon|x)$ will inherit some properties from the theory. For example, if Π^i and x_i are conical in prices, then $F(\varepsilon|x)$ must have a scale which is conical in prices.

The preceding paragraphs have described a path from the economic theory of behavior of a firm to properties of the latent variable model and associated response probability it generates. In applications it is often useful to reverse this path, writing down a convenient response probability model and then establishing that it meets sufficient conditions for derivation from the theory of the profit-maximizing firm. For the competitive case, a quite general sufficient condition is that $x_i\beta$ be closed, convex, and conical in prices and that ε be linear in prices; see Duncan (1980a) and McFadden (1979a).

Problems involving utility-maximizing consumers can be analyzed by methods paralleling the treatment of the firm, with Π^i replaced by the indirect utility function achieved for given i by optimizing in all remaining dimensions. However, this case is more complex since the expectation with respect to unobservables of the indirect utility function given i does not in general inherit all the properties of an indirect utility function. Consequently, known sufficient conditions for a specified response probability model to be derivable from a population of utility maximizers are quite restrictive, bearing a close relation to the sufficient conditions for individual preferences to aggregate to a social utility consistent with market demands; see McFadden (1981). Whether there is a practical general characterization of the response probability models consistent with a population of utility maximizers, analogous to the integrability theory for individual demand functions, remains an open question.

3.2. *Statistical analysis*

Consider a general multinomial response model with m alternatives, indexed $i=1,\dots,m$,

$$P_i = f^i(x,\theta), \tag{3.4}$$

generated by some latent variable model and generalized indicator function as in

(3.1) and (3.2). The x are observed explanatory variables, and θ is a vector of parameters. Consider an independent random sample with observations (y_t, x_t) for $t = 1,\dots,T$. As indicated for the binomial case, maximum likelihood estimation is the most generally applicable and usually the most satisfactory approach to estimation of θ. Let

$$l(y_t, x_t, \theta) = \sum_{i=1}^{m} y_{it} \ln f^i(x_t, \theta) \tag{3.5}$$

denote the log-likelihood of observation t, and

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} l(y_t, x_t, \theta) \tag{3.6}$$

the sample log-likelihood normalized by sample size. The following regularity conditions will be shown to imply that the maximum likelihood estimator is consistent and asymptotically normal.

(1) The domain of the explanatory variables is a measurable set X with a probability $p(x)$.

(2) The parameter space Θ is a subset of R^k, and the true parameter vector θ^* is in the interior of Θ.

(3) The response model $P_i = f^i(x, \theta)$ is measurable in x for each θ, and for x in a set X_1 with $p(X_1) = 1$, $f^i(x, \theta)$ is continuous in θ.

(4) The model satisfies a global identification condition: given $\varepsilon > 0$, there exists $\delta > 0$ such that $|\theta - \theta^*| \geq \varepsilon$ implies:

$$\psi(\theta) \equiv \int \mathrm{d}p(x) \sum_{i=1}^{m} f^i(x, \theta^*) \ln[f^i(x, \theta^*)/f^i(x, \theta)] \geq \delta \tag{3.7}$$

(5) For $x \in X_1$ with $p(X_1) = 1$, and some neighborhood Θ_0 of θ^*, the derivative $\partial f^i(x, \theta)/\partial\theta$ exists and is measurable in x.

(6) For some neighborhood Θ_0 of θ^* and measurable functions $\alpha^i(x)$, $\beta^i(x)$, $\gamma^i(x)$, the following bounds hold:

(i) $f^i(x, \theta) \leq \alpha^i(x)$,

(ii) $|\partial \ln f^i(x, \theta)/\partial\theta| \leq \beta^i(x)$,

(iii) $|\partial \ln f^i(x, \theta)/\partial\theta - \partial \ln f^i(x, \theta')/\partial\theta| \leq \gamma^i(x)|\theta - \theta'|$,

(iv) $\int \mathrm{d}p(x)\alpha^i(x)\gamma^i(x)^2 < \infty$,

(v) $\int \mathrm{d}p(x)\alpha^i(x)\beta^i(x)\gamma^i(x) < \infty$,

(vi) $\int \mathrm{d}p(x)\alpha^i(x)\beta^i(x)^3 < \infty$.

(7) The information matrix $J(\theta^*)$, given by

$$\int \mathrm{d}p(x) \sum_{i=1}^{m} f^i(x,\theta^*)[\partial \ln f^i(x,\theta^*)/\partial\theta][\partial \ln f^i(x,\theta^*)/\partial\theta]', \tag{3.8}$$

is non-singular.

The main results are given by the following theorems.

Theorem 1

If conditions (1)–(4) hold, and $\bar{\theta}_T$ is any sequence of measurable estimators which satisfy

$$L_T(\bar{\theta}_T) \geq \sup_{\Theta} L_T(\theta) - 1/T \tag{3.9}$$

with probability one, then $\bar{\theta}_T$ converges almost surely to θ^*.

Theorem 2

If conditions (1)–(5) hold, then almost surely a unique maximum likelihood estimator $\hat{\theta}_T$ eventually exists and satisfies $\partial L_T(\hat{\theta}_T)/\partial\theta = 0$ and $\hat{\theta}_T \to \theta^*$.

Theorem 3

If conditions (1)–(7) hold, then $\sqrt{T}(\hat{\theta}_T - \theta^*)$ converges in distribution to a normal random vector with mean zero and covariance matrix $J(\theta^*)^{-1}$.

The following paragraphs discuss the regularity conditions and theorems; proof outlines are deferred to the Appendix. Note first that the theorems assume the explanatory variables are independently identically distributed for each observation. This is appropriate for sample survey data, but not necessarily for time-series data. Analogous theorems hold for the case of non-stochastic or jointly distributed explanatory variables, but require stronger bounds and a more complicated definition of the information matrix.

Conditions (1)–(3) are very mild and easily verified in most models. Note that the parameter space Θ is not required to be compact, nor is $\ln f^i(x,\theta)$ required to be bounded. Condition (4) is a substantive identification requirement which states that no parameter vector other than the true one can achieve as high a limiting value of the log-likelihood. Theorem 1 specializes a general consistency theorem of Huber (1965, theorem 1). It is possible to weaken conditions (1)–(4) further, with some loss of simplicity, and still utilize Huber's argument. Note that $L_T(\theta) \leq 0$ and, since $y_{it} = 1$ implies $f^i(x_t, \theta^*) > 0$ almost surely, $L_T(\theta^*) > -\infty$ almost surely. Hence, a sequence of estimators $\bar{\theta}_T$ satisfying (3.9) almost surely exists.

Condition (5), requiring differentiability of $L_T(\theta)$ in a neighborhood of θ^*, will be satisfied by most models. With this condition, Theorem 2 implies that a unique maximum likelihood estimator almost surely eventually exists and satisfies the first-order condition for an interior maximum. This result does *not* imply that every solution of the first-order conditions is consistent. Note that any strongly consistent estimator of θ^* almost surely eventually stays in any specified compact neighborhood of θ^*.

Condition (6) imposes uniform (in θ) bounds on the response probabilities and their first derivatives in a neighborhood of θ^*. Condition (6) (iii) requires that $\partial \ln f^i(x,\theta)/\partial\theta$ be Lipschitzian in a neighborhood of θ^*.

Condition (4) combined with (5) and (6) implies $J(\theta)$ is non-singular at some point in the intersection of each neighborhood of θ^* and line segment extending from θ^*. Hence, condition (7) excludes only pathological irregularities.

Theorem 3 establishes asymptotic normality for maximum likelihood estimates of discrete response models under substantially weaker conditions than are usually imposed. In particular, no assumptions are made regarding second or third derivatives. Theorem 3 extends an asymptotic normality argument of Rao (1972, 5e2) for the case of a multinomial model without explanatory variables.

To illustrate the use of these theorems, consider the multinomial logit model:

$$P_i = \mathrm{e}^{x_i\theta} / \sum_{j=1}^{m} \mathrm{e}^{x_j\theta}, \tag{3.10}$$

with $x = (x_1,\dots,x_m) \in R^{mk}$ and $\theta \in R^k$. This model is continuous in x and θ, and twice continuously differentiable in θ for each x. Hence, conditions (1)–(3) and (5) are immediately satisfied. Since

$$\partial \ln f^i(x,\theta)/\partial\theta = x_i - \sum_j x_j f^j(x,\theta) \equiv x_i - \bar{x}(\theta), \tag{3.11}$$

$\mathrm{E}|x|^3 < \infty$ is sufficient for condition (6). The information matrix is:

$$J(\theta^*) = \int \mathrm{d}p(x) \sum_{i=1}^{m} f^i(x,\theta)(x_i - \bar{x}(\theta^*))(x_i - \bar{x}(\theta^*))'; \tag{3.12}$$

its non-singularity in (7) is equivalent to a linear independence condition on $(x_1 - \bar{x}(\theta^*),\dots,x_m - \bar{x}(\theta^*))$. The function $\ln f^i(x,\theta)$ is strictly concave in θ if condition (7) holds, implying that condition (4) is satisfied. Then Theorems 1–3 establish for this model that the maximum likelihood estimator $\hat{\theta}_T$ almost surely eventually exists and converges to θ^*, and $\sqrt{T}(\hat{\theta}_T - \theta^*)$ is asymptotically normal with covariance matrix $J(\theta^*)^{-1}$.

Since maximum likelihood estimators of qualitative response models fit within the general large sample theory for non-linear models, statistical inference is completely conventional, and Wald, Lagrange multiplier, or likelihood ratio statistics can be used for large sample tests. It is also possible to define summary measures of goodness of fit which are related to the likelihood ratio. Let g_t^i and f_t^i be two sequences of response probabilities for the sample points $t=1,\dots,T$, and define

$$I_T(g,f)=\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}g_t^i\ln\left(\frac{g_t^i}{f_t^i}\right) \tag{3.13}$$

to be the "average information in g beyond that in f". If g is the empirical distribution of the observed response and f is a parametric response model, then $I(g,f)$ is monotone in the likelihood function, and maximum likelihood estimation minimizes the average unexplained information. The better the model fits, the smaller $I_T(g,f)$. Note that for two models f_0 and f_1, the difference in average information $I_T(g,f_0)-I_T(g,f_1)$ is proportional to a likelihood ratio statistic. Goodness-of-fit measures related to (3.13) have been developed by Theil (1970); see also Judge et al. (1981). Related goodness of fit measures are discussed in Amemiya (1982). It is also possible to assess qualitative response models in terms of predictive accuracy; McFadden (1979b) defines prediction success tables and summary measures of predictive accuracy.

3.3. *Functional form*

The primary issues in choice of a functional form for a response probability model are computational practicality and flexibility in representing patterns of similarity across alternatives. Practical experience suggests that functional forms which allow similar patterns of inter-alternative substitution will give comparable fits to existing economic data sets. Of course, laboratory experimentation or more comprehensive economic observations may make it possible to differentiate the fit of function forms with respect to characteristics other than flexibility.

Currently three major families of concrete functional forms for response probabilities have been developed in the literature. These are multinomial logit models, based on the work of Luce (1959), multinomial probit models, based on the work of Thurstone (1927), and elimination models, based on the work of Tversky (1972). Figure 3.1 outlines these families; the members are defined in the following sections. We argue in the following sections that the multinomial logit model scores well on simplicity and computation, but poorly on flexibility. The multinomial probit model is simple and flexible, but scores poorly on computation. Variants of these models, the nested multinomial logit model and the factorial multinomial probit model, attempt to achieve both flexibility and computational practicality.

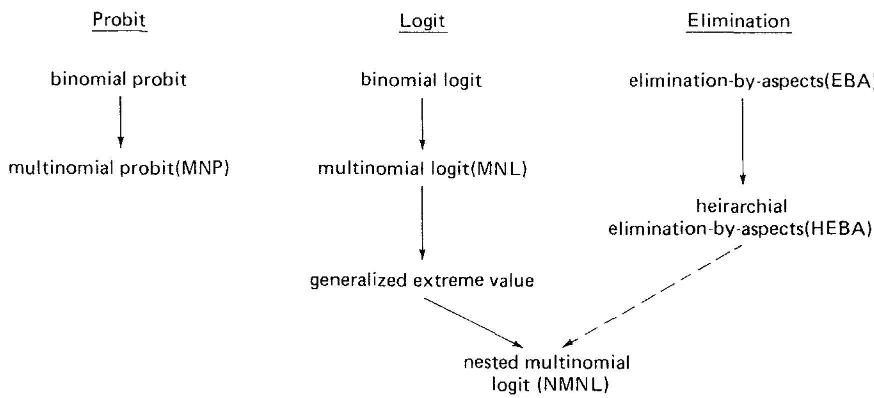

Figure 3.1. Functional forms for multinomial response probabilities.

In considering probit, logit, and related models, it is useful to quantify the hypothesis of an optimizing economic agent in the following terms. Consider a choice set $B = \{1,\dots,m\}$. Alternative i has a column vector of observed attributes x_i, and an associated utility $y_i^* = \alpha' x_i$, where α is a vector of taste weights. Assume α to have a parametric probability distribution with parameter vector θ, and let $\beta = \beta(\theta)$ and $\Omega = \Omega(\theta)$ denote the mean and covariance matrix of α. Let $x_B = (x_1,\dots,x_m)$ denote the array of observed attributes of the available alternatives. Then the vector of utilities $y_B^* = (y_1^*,\dots,y_m^*)$ has a multivariate probability distribution with mean $\beta' x_B$ and covariance matrix $x_B' \Omega x_B$. The response probability $f^i(x_B,\theta)$ for alternative i then equals the probability of drawing a vector y_B^* from this distribution such that $y_i^* \geq y_j^*$ for $j \in B$. For calculation, it is convenient to note that $y_{B-i}^* = (y_1^* - y_i^*,\dots,y_{i-1}^* - y_i^*, y_{i+1}^* - y_i^*,\dots,y_m^* - y_i^*)$ has a multivariate distribution with mean $\beta' x_{B-i}$ and covariance matrix $x_{B-i}' \Omega x_{B-i}$, where $x_{B-i} = (x_1 - x_i,\dots, x_{i-1} - x_i, x_{i+1} - x_i,\dots, x_m - x_i)$, and that $f^i(z_B,\theta)$ equals the non-positive orthant probability for this $(m-1)$-dimensional distribution.

The following sections review a series of concrete probabilistic choice models which can be derived from the structure above.

3.4. *The multinomial logit model*

The most widely used model of multinomial response is the multinomial logit (MNL) form:

$$f^i(x_B,\theta) = \mathrm{e}^{x_i\theta} / \sum_{j \in B} \mathrm{e}^{x_j\theta}. \tag{3.14}$$

This model permits easy computation and interpretation, but has a restrictive pattern of inter-alternative substitutions.

The MNL model can be derived from the latent variable model given in (3.1) and (3.2) by specifying the distribution of the disturbances $\varepsilon_t = (\varepsilon_{1t},\ldots,\varepsilon_{mt})$ to be independent identical type I extreme value:

$$F(\varepsilon_t|x_t) = e^{-e^{-\varepsilon_{1t}}} \cdots e^{-e^{-\varepsilon_{mt}}}. \tag{3.15}$$

This result is demonstrated by a straightforward integration; see McFadden (1973) and Yellot (1977). Note that this case is a specialization of the model $y_i^* = \alpha x_i$ in which only the coefficients α of alternative-specific dummy variables are stochastic.

The disturbance ε_t in the latent variable model yielding the MNL form may have the conventional econometric interpretation of the impact of factors known to the decision-maker but not to the observer. However, it is also possible that a disturbance exists in the decision protocol of the economic agent, yielding stochastic choice behavior. These alternatives cannot ordinarily be distinguished unless the decision protocol is observable or individuals can be confronted experimentally with a variety of decisions.

Interpreted as a stochastic choice model, the MNL form is used in psychometrics and is termed the Luce strict utility model. In this literature, $v_{it} = x_{it}\beta$ is interpreted as a scale value associated with alternative i. References are Luce (1959, 1977) and Marschak (1960).

The vector of explanatory variables x_{it} in the MNL model can be interpreted as attributes of alternative i. Note that components of x_{it} which do not vary with i cancel out of the MNL formula (3.13), and the corresponding component of the parameter vector θ cannot be identified from observation on discrete response.

Some components of x_{it} may be alternative-specific, resulting from the interaction of a variable with a dummy variable for alternative i. This is meaningful if the alternatives are naturally indexed. For example, in a study of durable ownership the alternative of not holding the durable is naturally distinguished from all the alternatives where the durable is held. On the other hand, if there is no link between the true attributes of an alternative and its index i, as might be the case for the set of available dwellings in a study of housing purchase behavior, alternative dummies are meaningless.

Attributes of the respondent may enter the MNL model in interaction with attributes of alternatives or with alternative specific dummies. For example, income may enter a MNL model of the housing purchase decision in interaction with a dwelling attribute such as price, or with a dummy variable for the non-ownership alternative.

A case of the MNL model frequently encountered in sociometrics is that in which the variables in x_{it} are all interactions of respondent attributes and alternative-specific dummies. Let z_t be a $1 \times s$ vector of respondent attributes and δ_{im} be a dummy variable which is one when $i = m$, zero otherwise. Define the $1 \times sM$ vector of interactions,

$$x_{it} = (\delta_{i1} z_t, \dots, \delta_{iM} z_t),$$

and let $\theta' = (\theta_1', \dots, \theta_m')$ be a commensurate vector of parameters. Then

$$\begin{aligned} f^i(x_t, \theta) &= \frac{e^{x_{it}\theta}}{e^{x_{1t}\theta} + \cdots + e^{x_{mt}\theta}} \\ &= \frac{e^{z_t\theta_i}}{e^{z_t\theta_1} + \cdots + e^{z_t\theta_m}}. \end{aligned} \tag{3.16}$$

An identifying normalization, say $\theta_1 = 0$, is required. This model is analyzed further by Goodman (1972) and Nerlove and Press (1976).

A convenient feature of the MNL model is that the hessian of the log-likelihood is everywhere negative definite (barring linear dependence of explanatory variables), so that any stationary value is a global maximum.

3.5. *Independence from irrelevant alternatives*

Suppose in the MNL model (3.13) that the vector x_{it} of explanatory variables associated with alternative i depends solely on the attributes of i, possibly interacted with attributes of the respondent. That is, x_{it} does not depend on the attributes of alternatives other than i. Then the MNL model has the Independence from Irrelevant Alternatives (IIA) property, which states that the odds of i being chosen over j is independent of the availability or attributes of alternatives other than i and j. In symbols, this property can be written:

$$\ln \frac{f^i(x_t, \theta)}{f^j(x_t, \theta)} = (x_{it} - x_{jt})\theta, \tag{3.17}$$

independent of x_{mt} for $m \neq i, j$. Equivalently, for $i \in A = \{1, \dots, J\} \subseteq C = \{1, \dots, M\}$:

$$f^i(x_{1t}, \dots, x_{Mt}, \theta) = f^i(x_{1t}, \dots, x_{Jt}, \theta) \cdot f^A(x_{1t}, \dots, x_{Mt}, \theta),$$

where

$$f^A(x_{1t}, \dots, x_{Mt}, \theta) \equiv \sum_{j \in A} f^j(x_{1t}, \dots, x_{Mt}, \theta).$$

An implication of the IIA property is that the cross-elasticity of the probability of response i with respect to a component of x_{jt} is the same for all i with $i \neq j$. This property is theoretically implausible in many applications. Nevertheless, empirical experience is that the MNL model is relatively robust, as measured by goodness of fit or prediction accuracy, in many cases where the IIA property is theoretically implausible.

When the IIA property is valid, it provides a powerful and useful restriction on model structure. One of its implications is that response probabilities for choice in restricted or expanded choice sets are obtained from the basic MNL form (3.14) simply by deleting or adding terms in the denominator. Thus, for example, one can use the model estimated on existing alternatives to forecast the probability of a new alternative so long as no parameters unique to the new alternative are added.

One useful application of the IIA property is to data where preference rankings of alternatives are observed, or can be inferred from observed purchase order. If the probabilities for the most preferred alternatives in each choice set satisfy the IIA property, then they must be of the MNL form [see McFadden (1973)], and the probability of an observed ranking $1 > 2 > \cdots > m$ of the alternatives is the product of conditional probabilities of choice from successively restricted subsets:

$$P(1 > 2 > \cdots > m) = \frac{e^{x_1\beta}}{\sum_{i=1}^{m} e^{x_i\beta}} \cdot \frac{e^{x_2\beta}}{\sum_{i=2}^{m} e^{x_i\beta}} \cdot \cdots \cdot \frac{e^{x_{m-1}\beta}}{e^{x_{m-1}\beta} + e^{x_m\beta}}.$$

Thus, each selection of a next-ranked alternative from the subset of alternatives not previously ranked can be treated as an independent observation of choice from a MNL model. This formulation of ranking probabilities is due to Marschak (1960). An econometric application has been made by Beggs, Cardell and Hausman (1981); these authors use the method to estimate individual taste parameters and investigate the heterogeneity of these parameters across the population.

The restrictive IIA feature of the MNL model is present only when the vector x_{it} for alternative i is independent of the attributes of alternatives other than i. When this restriction is dropped, the MNL form is sufficiently flexible to approximate any continuous positive response probability model on a compact set of the explanatory variables. Specifically, if $f^i(x_t, \theta)$ is continuous, then it can be approximated globally to any desired degree of accuracy by a MNL model of the form:

$$\tilde{f}^i(x_t, \theta) = \frac{e^{z_{it}\theta}}{e^{z_{1t}\theta} + \cdots + e^{z_{mt}\theta}}, \tag{3.18}$$

where $z_{it} = z_{it}(x_t)$ is an arithmetic function of the attributes of *all* available alternatives, not just the attributes of alternative i. This approximation has been termed the universal logit model. The result follows easily from a global approximation of the vector of logs of choice probabilities by a multivariate Bernstein polynomial; details are given in McFadden (1981).

The universal logit model can describe any pattern of cross-elasticities. Thus, it is not the MNL form per se, but rather the restriction of x_{it} to depend only on attributes of i, which implies IIA restrictions. In practice, the global approximations yielding the universal logit model may be computationally infeasible or inefficient. In addition, the approximation makes it difficult to impose or verify consistency with economic theory. The idea underlying the universal logit model does suggest some useful specification tests; see McFadden, Tye and Train (1976).

3.6. *Limiting the number of alternatives*

When the number of alternatives is large, response probability models may impose heavy burdens of data collection and computation. The special structure of the MNL model permits a reduction in problem scale by either aggregating alternatives or by analyzing a sample of the full alternative set. Consider first the aggregation of relatively homogeneous alternatives into a smaller number of primary types.

Suppose elemental alternatives are doubly indexed ij, with i denoting primary type and j denoting alternatives within a type. Let M_i denote the number of alternatives which are of type i. Suppose choice among all alternatives is described by the MNL model. Then choice among primary types is described by MNL probabilities of the form:

$$f^i(x_t, \theta) = \frac{\exp(x_{it}\theta + \ln M_i + w_{it})}{\sum_k \exp(x_{kt}\theta + \ln M_k + w_{kt})}, \tag{3.19}$$

where x_{it} is the mean within type i of the vectors x_{ijt} of explanatory variables for the alternative ij, and w_{it} is a correction factor for heterogeneity within type i which satisfies:

$$w_{it} = \ln \frac{1}{M_i} \sum_{j=1}^{M_i} \exp\left[(x_{ijt} - x_{it})\theta\right]. \tag{3.20}$$

If the alternatives within a type are homogeneous, then $w_i = 0$.

A useful approximation to w_i can be obtained if the deviations $x_{ijt} - x_{it}$ within type i can be treated as independent random drawings from a multivariate distribution which has a cumulant generating function $W_{it}(\cdot)$. If the number of alternatives M_i is large, then the law of large numbers implies that w_i converges almost surely to $w_i = W_{it}(\theta)$. For example, if $x_{ijt} - x_{it}$ is multivariate normal with covariance matrix Ω_{it}, then $w_i \approx W_{it}(\theta) \equiv \theta'\Omega_{it}\theta/2$.

A practical method for estimation is to either assume within-type homogeneity, or to use the normal approximation to w_i, with Ω_{it} either fitted from data or treated as parameters with some identifying restrictions over i and t. Then θ can be estimated by maximum likelihood estimation of (3.19). The procedure can be iterated using intermediate estimates of θ in the exact formula for w_i. Data collection and processing can be reduced by sampling elemental alternatives to estimate w_i. However, it is then necessary to adjust the asymptotic standard errors of coefficients to include the effect of sampling errors on the measurement of w_i. Further discussion of aggregation of alternatives in a MNL model can be found in McFadden (1978b).

A second method of reducing the scale of data collection and computation in the MNL model when it has the IIA property is to sample a sub-set of the full set of alternatives. The IIA property implies that the *conditional* probabilities of choosing from a restricted subset of the full choice set equal the choice probabilities when the choice set equals the restricted set. Then the MNL model can be estimated from data on alternatives sampled from the full choice set. In particular, the MNL model can be estimated from data on binary conditional choices. Furthermore, subject to one weak restriction, biased sampling of alternatives can be compensated for within the MNL estimation.

Let $C = \{1,\dots,M\}$ denote the full choice set, and $D \subseteq C$ a restricted subset. The protocol for sampling alternatives is defined by a probability $\pi(D|i_t, x_t)$ that D will be sampled, given observed explanatory variables x_t and choice i_t. For example, the sampling protocol of selecting the chosen alternative plus one non-chosen alternative drawn at random satisfies

$$\pi(D|i_t, x_t) = \begin{cases} 1/(M-1), & \text{if } D = \{i_t, j\} \subseteqq C, i_t \neq j, \\ 0, & \text{otherwise.} \end{cases} \tag{3.21}$$

Let D_t denote the subset for case t. The weak regularity condition is:

Positive conditioning property

If an alternative $j \in D_t$ were the observed choice, there would be a positive probability that the sampling protocol would select D_t; i.e. if $j \in D_t$, then $\pi(D_t|j, x_t) > 0$.

If the positive conditioning property and a standard identification condition hold, then maximization of the modified MNL log-likelihood function:

$$\frac{1}{T}\sum_{t=1}^{T}\ln\frac{\exp\left[x_{i_t}\theta+\ln\pi(D_t|i_t,x_t)\right]}{\sum\limits_{j\in D_t}\exp\left[x_j\theta+\ln\pi(D_t|j,x_t)\right]}\tag{3.22}$$

yields consistent estimates of θ. This result is proved by showing that (3.22) converges in probability uniformly in θ to an expression which has a unique maximum at the true parameter vector; details are given in McFadden (1978). When π is the same for all $j \in D_t$, the terms involving π cancel out of the above expression. This is termed the *uniform* conditioning property; the example (3.21) satisfies this property.

Note that the modified MNL log-likelihood function (3.22) is simply the conditional log-likelihood of the i_t, given the D_t. The inverse of the information matrix for this conditional likelihood is a consistent estimator of the covariance matrix of the estimated coefficients, as usual.

3.7. *Specification tests for the MNL model*

The MNL model in which the explanatory variables for alternative i are functions solely of the attributes of that alternative satisfies the restrictive IIA property. An implication of this property is that the model structure and parameters are unchanged when choice is analyzed conditional on a restricted subset of the full choice set. This is a special case of uniform conditioning from the section above on sampling alternatives.

The IIA property can be used to form a specification test for the MNL model. Let C denote the full choice set, and D a proper subset of C. Let β_C and V_C denote parameter estimates obtained by maximum likelihood on the full choice set, and the associated estimate of the covariance matrix of the estimators. Let β_D and V_D be the corresponding expressions for maximum likelihood applied to the restricted choice set D. (If some components of the full parameter vector cannot be identified from choice within D, let β_C, β_D, V_C, and V_D denote estimates corresponding to the identifiable sub-vector.) Under the null hypothesis that the IIA property holds, implying the MNL specification, $\beta_D - \beta_C$ is a consistent estimator of zero. Under alternative model specifications where IIA fails, $\beta_D - \beta_C$ will almost certainly not be a consistent estimator of zero. Under the null hypothesis, $\beta_D - \beta_C$ has an estimated covariance matrix $V_D - V_C$. Hence, the statistic

$$S=(\beta_D-\beta_C)'(V_D-V_C)^{-1}(\beta_D-\beta_C) \tag{3.23}$$

is asymptotically chi-square with degrees of freedom equal to the rank of V_D-V_C.

This test is analyzed further in Hausman and McFadden (1984). Note that this is an omnibus test which may fail because of misspecifications other than IIA. Empirical experience and limited numerical experiments suggest that the test is not very powerful unless deviations from MNL structure are substantial.

3.8. *Multinomial probit*

Consider the latent variable model for discrete response, $y_t^*=x_t\theta+\varepsilon_t$ and $y_{mt}=1$ if $y_{mt}^*\geq y_{nt}^*$ for $n=1,\dots,M$, from (3.1) and (3.2). If ε_t is assumed to be multivariate normal, the resulting discrete response model is termed the multinomial probit (MNP) model. The binary case has been used extensively in biometrics; see Finney (1971). The multivariate model has been investigated by Bock and Jones (1968), McFadden (1976), Hausman and Wise (1978), Daganzo (1980), Manski and Lerman (1981), and McFadden (1981).

A form of the MNP model with a plausible economic interpretation is $y_t^*=x_t\alpha_t$, where α_t is multivariate normal with mean β and covariance matrix Ω, and represents taste weights which vary randomly in the population. Note that this form implies $\mathrm{E}\varepsilon_t=0$ and $\operatorname{cov}(\varepsilon_t)=x_t\Omega x_t'$ in the latent variable model formulation. If x_t includes alternative dummies, then the corresponding components of α_t are additive random contributions to the latent values of the alternatives. Some normalizations are required in this model for identification.

When correlation is permitted between alternatives, so $\operatorname{cov}(\varepsilon_t)$ is not diagonal, the MNP model does not have the IIA or related restrictive properties, and permits very general patterns of cross-elasticities. This is true in particular for the random taste weight version of the MNP model when there are random components of α_t corresponding to attributes which vary across alternatives.

Evaluation of MNP probabilities for M alternatives generally requires evaluation of $(M-1)$-dimensional orthant probabilities. In the notation of subsection 3.3, $f^1(x_B;\beta,\Omega)$ is the probability that the $(M-1)$-dimensional normal random vector y_{B-1}^* with mean βx_{B-1} and covariance matrix $x_{B-1}\Omega X_{B-1}'$ is non-positive. For $M\leq 3$, the computation of these probabilities is comparable to that for the MNL model. However, for $M\geq 5$ and Ω unrestricted, numerical integration to obtain these orthant probabilities is usually too costly for practical application in iterative likelihood maximization for large data sets. An additional complication is that the hessian of the MNP model is not known to be negative definite; hence a search may be required to avoid secondary maxima.

For a multivariate normal vector $(y_1^*,\dots,y_m^*)$, one can calculate the mean and covariance matrix of $(y_1^*,\dots,y_{m-2}^*,\max(y_{m-1}^*, y_m^*))$; these moments involve only binary probits and can be computed rapidly. A quick, but crude, approximation to MNP probabilities can then be obtained by writing:

$$f^1(x,\beta,\Omega) = P(y_1^* > \max(y_2^*,\max(y_3^*,\dots)\dots)) \tag{3.24}$$

and approximating the maximum of two normal variates by a normal variate; see Clark (1961) and Daganzo (1980). This approximation is good for non-negatively correlated variates of comparable variance, but is poor for negative correlations or unequal variances. The method tends to overestimate small probabilities. For assessments of this method, see Horowitz, Sparmann and Daganzo (1981) and McFadden (1981).

A potentially rapid method of fitting MNP probabilities is to draw realizations of α_t repeatedly and use the latent variable model to calculate relative frequencies, starting from some approximation such as the Clark procedure. This requires a large number of simple computer tasks, and can be programmed quite efficiently on an array processor. However, it is difficult to compute small probabilities accurately by this method; see Lerman and Manski (1980).

One way to reduce the complexity of the MNP calculation is to restrict the structure of the covariance matrix Ω by adopting a "factor-analytic" specification of the latent variable model $y_i^* = \beta x_i + \varepsilon_i$. Take

$$\varepsilon_i = \eta_i + \sum_{j=1}^{J} \gamma_{ij}\nu_j, \tag{3.25}$$

with η_i and ν_j independent normal variates with zero means and variances σ_i^2 and 1 respectively. The "factor loading" γ_{ij} is in the most general case a parametric function of the observed attributes of alternatives, and can be interpreted as the level in alternative i of an unobserved characteristic j. With this structure, the response probability can be written:

$$f^1(x,\beta,\gamma,\sigma) = \int_{\eta_1=-\infty}^{+\infty}\int_{\nu_1=-\infty}^{+\infty}\cdots\int_{\nu_J=-\infty}^{+\infty}\frac{1}{\sigma_1}\phi\left(\frac{\eta_1}{\sigma_1}\right)\prod_{j=1}^{J}\phi(\nu_j)$$
$$\times\prod_{i=2}^{M}\Phi\left(\frac{\beta(x_1-x_i)+\sum_{j=1}^{J}(\gamma_{1j}-\gamma_{ij})\nu_j+\eta_1}{\sigma_i}\right)\mathrm{d}\eta_1,\mathrm{d}\nu_1\cdots\mathrm{d}\nu_J. \tag{3.26}$$

Numerical integration of this formula is easy for $J \leq 1$, but costly for $J \geq 3$. Thus, this approach is generally practical only for one or two factor models. The independent MNP model ($J = 0$) has essentially the same restrictions on cross-alternative substitutions as the MNL model; there appears to be little reason to prefer one of these models over the other. However, the one and two factor models permit moderately rich patterns of cross-elasticities, and are attractive practical alternatives in cases where the MNL functional form is too restrictive.

Computation is the primary impediment to widespread use of the MNP model, which otherwise has the elements of flexibility and ease of interpretation desirable in a general purpose qualitative response model. Implementation of a fast and accurate approximation to the MNP probabilities remains an important research problem.

3.9. *Elimination models*

An elimination model views choice as a process in which alternatives are screened from the choice set, using various criteria, until a single element remains. It can be defined by the probability of transition from a set of alternatives to any subset, $Q(D|C)$. If each transition probability is stationary throughout the elimination process, then the choice probabilities satisfy the recursion formula:

$$f^i(C) = \sum_D Q(D|C) f^i(D), \tag{3.27}$$

where $f^i(C)$ is the probability of choosing i from set C.

Elimination models were introduced by Tversky (1972) as a generalization of the Luce model to allow dependence between alternatives. An adaptation of Tversky's elimination by aspects (EBA) model suitable for econometric work takes transition probabilities to have a MNL form:

$$Q(D|C) = \mathrm{e}^{x_D\beta_D} \Big/ \sum_{\substack{A \subseteq C \\ A \neq C}} \mathrm{e}^{x_A\beta_A}, \tag{3.28}$$

where x_D is a vector of attributes common to and unique to the set of alternatives in D. When x_B is a null vector and by definition $\mathrm{e}^{x_B\beta_B} = 0$ for sets B of more than one element, this model reduces to the MNL model. Otherwise, it does not have restrictive IIA-like properties.

The elimination model is not known to have a latent variable characterization. However, it can be characterized as the result of maximization of random lexicographic preferences. The model defined by (3.27) and (3.28) has not been applied in economics. However, if the common unique attributes x_D can be defined in an application, this should be a straightforward and flexible functional form.

One elimination model which can be expressed in latent variable form is the *generalized extreme value* (GEV) model introduced by McFadden (1978, 1981). Let $H(w_1,\ldots,w_m)$ be a non-negative, linear homogeneous function of non-negative $w_1,\ldots,w_m$ which satisfies

$$\lim_{w_i \to \infty} H(w_1,\ldots,w_m) = +\infty, \tag{3.29}$$

and has mixed partial derivatives of all orders, with non-positive even and non-negative odd mixed derivatives. Then,

$$F(\varepsilon_1,\ldots,\varepsilon_m) = \exp\{-H(\mathrm{e}^{-\varepsilon_1},\ldots,\mathrm{e}^{-\varepsilon_m})\} \tag{3.30}$$

is a multivariate extreme value cumulative distribution function. The latent variable model $y_i^* = x_i\beta + \varepsilon_i$ for $i \in B = \{1,\ldots,m\}$ with $(\varepsilon_1,\ldots,\varepsilon_m)$ distributed as (3.30) has response probabilities:

$$f^i(x,\beta) = \partial \ln H(\mathrm{e}^{x_1\beta},\ldots,\mathrm{e}^{x_m\beta})/\partial(x_i\beta). \tag{3.31}$$

The GEV model reduces to the MNL model when

$$H(w_1,\ldots,w_m) = \left(\sum_{i=1}^{M} w_i^{1/\lambda}\right)^{\lambda}, \tag{3.32}$$

with $0 < \lambda \leq 1$. An example of a more general GEV function is:

$$H(w_1,\ldots,w_m) = \sum_{C \subseteq B} a(C)\left\{\sum_{D \subseteq C} b(D,C)\left[\sum_{i \in D} w_i^{1/\lambda_{DC}\lambda_C}\right]^{\lambda_{DC}}\right\}^{\lambda_C}, \tag{3.33}$$

where $0 < \lambda_{DC}, \lambda_C \leq 1$ and a and b are non-negative functions such that each i is contained in a D and C with $a(C), b(D,C) > 0$. The response probability for (3.33) can be written:

$$f^i(x,\theta) = \sum_{C \subseteq B}\sum_{D \subseteq C} Q(i|D,C)Q(D|C)Q(C|B), \tag{3.34}$$

where $i \in D \subseteq C \subseteq B$,

$$Q(i|D,C) = \exp(x_i\beta/\lambda_{DC}\lambda_C) / \sum_{j \in D} \exp(x_j\beta/\lambda_{DC}\lambda_C), \tag{3.35}$$

$$J(D,C) = \ln \sum_{j \in D} \exp(x_j\beta/\lambda_{DC}\lambda_C), \tag{3.36}$$

$$Q(D|C) = b(D,C)\exp[J(D,C)\lambda_{DC}] / \sum_{D' \subseteq C} b(D',C)\exp[J(D',C)\lambda_{D'C}], \tag{3.37}$$

$$I(C) = \ln \sum_{D' \subseteq C} b(D',C)\exp[J(D',C)\lambda_C], \tag{3.38}$$

$$Q(C) = a(C)\exp[I(C)\lambda_C] / \sum_{C' \subseteq B} a(C')\exp[I(C')\lambda_{C'}]. \tag{3.39}$$

This can be interpreted as an elimination model in which $a(C)$ and $b(D,C)$ determine the probability of various chains of sets of non-eliminated alternatives, and λ_{DC} and λ_C measure the degree of independence of the ε_i within the set D obtained from C, and within the set C, respectively. The expressions in (3.36) and (3.38) are termed *inclusive values* of the associated sets of alternatives.

When all the λ's are one, this model reduces to a simple MNL model. Alternatively, when λ_{DC} is near zero, the elimination model treats D essentially as if it contained a single alternative with a scale value equal to the maximum of the scale values of the elements in D.

Inspection of the two elimination models described above suggests that they are comparable in terms of flexibility and complexity. Other things equal, the GEV model will tend to imply sharper discrimination among similar alternatives than the EBA model. Limited numerical experiments suggest that the two models will be difficult to distinguish empirically.

3.10. Hierarchical response models

When asked to describe the decision process leading to qualitative choice, individuals often depict a hierarchical structure in which alternatives are grouped into clusters which are "similar". The decision protocol is then to eliminate clusters, proceeding until a single alternative remains. An example of a decision tree is given in Figure 3.2. Alternatives $a-e$ are in one primary cluster, f and g in a second, and $a-c$ are in a secondary cluster. Either of the elimination models described in the preceding section can be specialized to describe hierarchical response by permitting transitions from a node only to one of the nodes

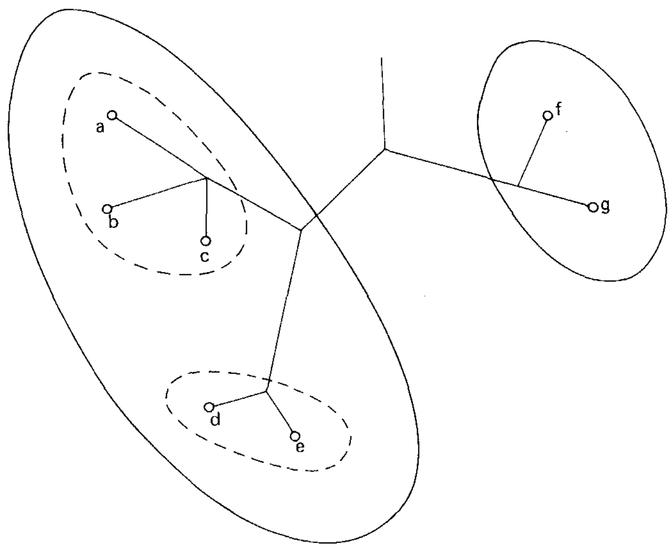

Figure 3.2. A hierarchical decision tree.

immediately below it in the tree. Hierarchical decision models are discussed further in Tversky and Sattath (1979), and McFadden (1981).

A hierarchical elimination model based on the generalized extreme value structure described earlier generalizes the MNL model to a nested multinomial logit (NMNL) structure. Each transition in the tree is described by a MNL model with one of the variables being an "inclusive value" which summarizes the attributes of alternatives below a node. An "independence" parameter at each node in the tree discounts the contribution to value of highly similar alternatives.

We shall discuss the structure of the NMNL model using an example of consumer choice of housing. As illustrated in Figure 3.3, the decision can be described in hierarchical form: first whether to own or rent, second if renting whether to be the head of household or to sublet from someone else (non-head), and finally what dwelling unit to occupy within the chosen cluster. Let $C = \{1,\dots,12\}$ index the final alternatives, $r = 0,1$ index the primary cluster for own and rent, and $h = 0,1$ index the secondary clusters for head and non-head. Define A_{rh} to be the set of final alternatives contained in the subcluster rh, and A_r to be the set of subclusters contained in the cluster r. For example, A_0 contains the (trivial) subcluster $h=1$; A_1 contains two subclusters $h=0$ and $h=1$; and $A_{11} = \{10,11,12\}$.

The response probability for the NMNL model can be written as a product of transition probabilities. For $i \in A_{rh}$ and $h \in A_r$:

$$f^i(x,\theta) = Q(i|A_{rh})Q(A_{rh}|A_r)Q(A_r). \tag{3.40}$$

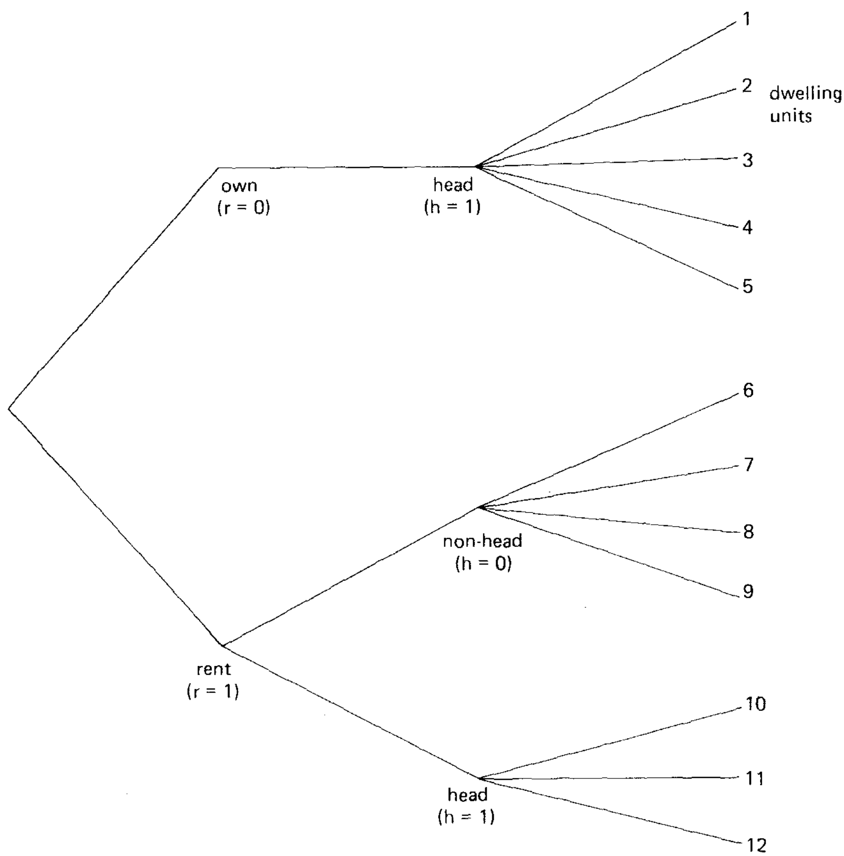

Figure 3.3. Housing choice.

Each transition probability has a NMNL form:

$$Q(i|A_{rh}) = \mathrm{e}^{z_i\alpha} \Big/ \sum_{j \in A_{rh}} \mathrm{e}^{z_j\alpha}, \tag{3.41}$$

$$Q(A_{rh}|A_r) = \exp(w_{rh}\beta + J_{rh}\kappa_{rh}) \Big/ \sum_{c \in A_r} \exp(w_{rc}\beta + J_{rc}\kappa_{rc}), \tag{3.42}$$

$$Q(A_r) = \exp(u_r\gamma + I_r\lambda_r) \Big/ \sum_{s=0}^{1} \exp(u_s\gamma + I_s\lambda_s). \tag{3.43}$$

Here, $x_i = (z_i, w_{rh}, u_r)$ is the vector of attributes associated with alternative $i \in A_{rh}$ and $h \in A_r$, with w_{rh} and u_r denoting components which are common within the clusters A_{rh} or A_r, respectively; $(\alpha, \beta, \gamma, \kappa_{rs}, \lambda_r) \equiv \theta$ are parameters; and J_{rh} and I_h

are inclusive values satisfying:

$$J_{rh} = \ln \sum_{i \in A_{rh}} e^{z_i \alpha}, \tag{3.44}$$

$$I_r = \ln \sum_{h \in A_r} \exp(w_{rh}\beta + J_{rh}\kappa_{rh}). \tag{3.45}$$

Note that J_{rh} and I_r are logs of the denominators in (3.41) and (3.42), respectively. For this example, note that $Q(A_{01}|A_0) = 1$ and $I_0 = w_{01}\beta + J_{01}\kappa_{01}$.

Consider the function:

$$H(e^{v_1}, \dots, e^{v_{12}}; \theta) = \sum_{r=0}^{1} \left[\sum_{h \in A_r} \left(\sum_{i \in A_{rh}} \exp(v_i / \kappa_{rh}\lambda_r) \right)^{\kappa_{rh}} \right]^{\lambda_r}, \tag{3.46}$$

with

$$v_i = z_i \alpha \kappa_{rh} \lambda_r + w_{rh}\beta\lambda_r + u_r\delta, \tag{3.47}$$

for $i \in A_{rh}$ and $h \in A_r$. This is a generating function for the response probabilities, satisfying $\partial \ln H / \partial v_i = f^i(x, \theta)$, and can be interpreted as a measure of social utility; see McFadden (1981). The parameters κ_{rh} and λ_r are measures of the "independence" of alternatives within subclusters and clusters respectively.

If $\kappa_{rh} = \lambda_r = 1$, then the NMNL model reduces to a simple MNL model. When $0 < \kappa_{rh}, \lambda_r \leq 1$, the NMNL model is consistent with a latent variable model with generalized extreme value distributed disturbances: $y_i^* = v_i + \varepsilon_i$ and

$$F(\varepsilon_1, \dots, \varepsilon_{12}) = \exp[-H(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_{12}}; \theta)], \tag{3.48}$$

and is therefore consistent with an assumption of optimizing economic agents. It should be obvious that this structure generalizes to any number of alternatives and levels of clustering.

To interpret the impact of the independence parameters κ_{rh} and λ_r on cross-alternative substitutability, consider the cross-elasticity of the response probability for $i \in A_{rh}$ and $h \in A_r$ with respect to component k of the vector z_j of attributes of alternative $j \in A_{r'h'}$ and $h' \in A_{r'}$:

$$\begin{aligned}
\partial \ln f^i(x, \theta) / \partial \ln z_{jk} &= \kappa_{r'h'}\lambda_{r'}\alpha_k z_{jk} f_i(x, \theta)^{-1} \partial^2 H / \partial v_i \partial v_j \\
&= \alpha_k z_{jk} \big\{ \delta_{ij} - \kappa_{r'h'}\lambda_{r'} f^j(x, \theta) \\
&\quad + \kappa_{rh'}(\lambda_r - 1)\delta_{rr'} Q(j|A_{rh'}) Q(A_{rh'}|A_r) \\
&\quad + (\kappa_{rh} - 1)\delta_{rr'}\delta_{hh'} Q(j|A_{rh}) \big\}.
\end{aligned} \tag{3.49}$$

For $0 < \kappa_{rh}, \lambda_r < 1$, one obtains the plausible property that cross-elasticities are largest in magnitude for alternatives in the same r and h cluster, and smallest in magnitude when both r and h clusters differ. Note that values of κ_{rh} or λ_r outside the unit interval imply that one of the expected magnitude rankings is violated. Therefore, estimates of κ_{rh} or λ_r outside the unit interval may indicate a misspecified hierarchical structure, and the fitted cross-elasticity magnitude may identify a more appropriate structure.

It is of interest to compare the complexity and flexibility of the NMNL model, say in the form (3.40) corresponding to Figure 3.3, to a MNP model with a factorial structure which has the same pattern of similarities. This is achieved in the MNP model by introducing one factor for each node in the decision tree between the stem and the final "twigs". Thus, the clustering in Figure 3.3 requires four factors—own, rent, rent/head, and rent/non-head. An MNP model with this structure can be specified with a number of parameters comparable to the NMNL model by making the factor loadings uniform within each cluster, or can be made more flexible by allowing intra-cluster heterogeneity in loadings. However, as noted in the discussion of (3.26) computation of a four-factor model will be too costly in most applications. We conclude that the NMNL and factorial MNP are comparable in complexity, with some advantage to the latter in terms of flexibility and ease of interpretation. However, computational barriers currently limit use of the factorial MNP model to simple trees with one to three nodes.

The NMNL model can be estimated by direct maximum likelihood methods. The likelihood is not concave in all parameters, and is highly non-linear in the inclusive value coefficients. A simpler procedure which is consistent, but often fairly inefficient, is to estimate the transition probabilities (3.31)–(3.33) sequentially, using the data on transitions implied by the observed choices and inclusive values calculated from preceding stages. Each stage involves a concave MNL maximum likelihood problem. Beyond the first stage, standard errors are affected by the use of estimated coefficients in the calculation of inclusive values. Amemiya (1978d) and McFadden (1981) provide formulae for correcting the standard errors.

It is possible in principle to obtain asymptotically efficient estimates by carrying out one Newton–Raphson iteration on the full likelihood function, starting from the consistent sequential estimates. In practice, the strong non-linearity of the likelihood in the inclusive value coefficients and sensitivity of the estimates to model specification sometimes lead to full maximum likelihood estimates which are rather far from the initially consistent estimates, and to erratic results from the one-step procedure. Consequently, multiple iterations may be required to approximate the maximum of the full likelihood function. These problems seem to be particularly common when by other indications the decision tree is misspecified.

Under the null hypothesis that the NMNL model is correctly specified, the sequential estimator $\tilde{\theta}_T$ and full MLE $\hat{\theta}_T$ satisfy $\sqrt{T}(\tilde{\theta}_T-\theta)\xrightarrow{\mathrm{L}}N(0,\tilde{\Omega})$ and $\sqrt{T}(\hat{\theta}_T-\theta)\xrightarrow{\mathrm{L}}N(0,\Omega)$, and asymptotic efficiency implies $\sqrt{T}(\tilde{\theta}_T-\hat{\theta}_T)\xrightarrow{\mathrm{L}}$

$N(0, \tilde{\Omega} - \Omega)$. Thus the statistic $T(\tilde{\theta}_T - \hat{\theta}_T)'(\tilde{\Omega} - \Omega)^+(\tilde{\theta}_T - \hat{\theta}_T)$, where $(\tilde{\Omega} - \Omega)^+$ is a generalized inverse of rank p, is asymptotically χ_p^2 under the null. This statistic can then be used as an omnibus test of the NMNL specification.

Since the NMNL model reduces to a MNL model when the inclusive value coefficients are one, it can provide a basis for a classical Lagrange multiplier test of the IIA property of the MNL model. Consider testing $\lambda = 1$ in the model (3.30)–(3.33), with $\kappa = 1$ as a maintained hypothesis. Suppose A_1 is the set of rental alternatives. Let $\sigma = (\alpha\lambda, \beta\lambda, \gamma)$ and $\theta = (\sigma, \lambda)$, and let

$$L = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in C} y_{it} \ln f^i(x_t, \theta) \tag{3.50}$$

be the normalized log-likelihood function. The Lagrange multiplier statistic has the general form:

$$\mathrm{LM} = L'_\lambda \left[EL_\lambda L'_\lambda - (EL_\lambda L'_\sigma)(EL_\sigma L'_\sigma)^{-1} EL_\sigma L'_\lambda \right]^{-1} L_\lambda, \tag{3.51}$$

where the derivatives are evaluated at $\lambda = 1$ at which the model reduces to a MNL model. For this problem, letting $f_t^i = f^i(x_t, \theta)$:

$$L_\sigma = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in C} (y_{it} - f_t^i) x_{it}, \tag{3.52}$$

$$L_\lambda = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in A_1} (y_{it} - f_t^i)(I_t\lambda - x_{it}\sigma), \tag{3.53}$$

$$V_{\lambda\lambda} \equiv TEL_\lambda L'_\lambda = \frac{1}{T} \sum_{t=1}^{T} \left[\sum_{i \in A_1} f_t^i (I_t\lambda - x_{it}\sigma)^2 - \left(\sum_{i \in A_1} f_t^i (I_t\lambda - x_{it}\sigma) \right)^2 \right], \tag{3.54}$$

$$V_{\sigma\sigma} \equiv TEL_\sigma L'_\sigma = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in C} f_t^i (x_{it} - x_{Ct})'(x_{it} - x_{Ct}), \tag{3.55}$$

$$V_{\lambda\sigma} = TEL_\lambda L'_\sigma = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in A_1} f_t^i (I_t\lambda - x_{it}\sigma)(x_{it} - x_{Ct}), \tag{3.56}$$

$$x_{Ct} = \sum_{i \in C} f_t^i x_{it}, \tag{3.57}$$

$$I_t = \ln \sum_{i \in A_1} \mathrm{e}^{x_{it}\sigma}. \tag{3.58}$$

Then the final form of the test statistic is:

$$\mathrm{LM}=\left[\sum_{t=1}^{T}\sum_{i\in A_1}(y_{it}-f_t^i)(I_t-x_{it}\sigma)\right]^2 / T\left(V_{\lambda\lambda}-V_{\lambda\sigma}V_{\sigma\sigma}^{-1}V_{\sigma\lambda}\right). \tag{3.59}$$

Under the null hypothesis, this statistic is asymptotically chi-square with one degree of freedom. Further discussion of specification tests for the MNL model and examples are given in Hausman and McFadden (1984).

3.11. *An empirical example*

To illustrate application of the MNL and NMNL models, and associated tests, we apply the housing decision tree given in Figure 3.3 and the associated NMNL model in (3.40)–(3.43) to data on the housing status of single elderly men from the 1977 U.S. Annual Housing Survey for the Albany–Schenectady–Troy, N.Y. SMSA. These results were prepared by Axel Boersch-Supan, and are a simplified version of some models estimated by Boersch-Supan and Pitkin (1982); this reference contains a detailed description of variables and analysis of other socioeconomic groups.

The sample contains 159 single elderly men, of whom 45.9% are owners, 30.2% are renter-heads, and 23.9% are renter non-heads. Selection of dwelling unit is not modeled, and units within a cluster are treated as homogeneous and sufficiently similar to be adequately characterized as a single typical unit.

Two price variables are considered for each person in the sample, out-of-pocket costs (*OPCOST*) and expected net return on equity in owned units (*RETURN*). Out-of-pocket costs are the operating costs of the housing unit. For rental units, this is gross rent including utilities as reported in the survey. For owner-occupied units, *OPCOST* consists of mortgage and real-estate tax payments, utility costs, insurance payments, and maintenance. These direct costs are reduced by estimated savings on federal income taxes resulting from the deductability of mortgage interest and property taxes. Consequently *OPCOST* is influenced in a non-linear way by income. Costs in dwellings with more than one nuclear family unit are assumed to be apportioned according to the total number of adults, children counting as half an adult.

The *RETURN* variable for owner-occupied dwellings is defined as expected appreciation less equity cost, and is taken to be a proportion of dwelling value determined by average annual appreciation in the area since 1970, equity as a fraction of value estimated from date of purchase, and a discount factor reflecting opportunity cost of equity.

The construction of *OPCOST* and *RETURN* for chosen alternatives is based on individual reported costs, while for non-chosen alternatives these variables are based on the average experience of recent movers. Consequently, the estimated models should be interpreted as "reduced form" state models which reflect the relationship between status and costs, taking into account the inertia and non-transferable discounts associated with tenure. These models may accurately forecast future status-cost patterns provided there is no structural change in turnover rates or tenure distributions. They should not be interpreted as transition probabilities from one dwelling state to another—the latter probabilities are likely to be strongly state dependent and display less sensitivity to costs.

In addition to *OPCOST* and *RETURN*, income enters as an explanatory variable in interaction with a dummy variable for owner (*YOWN*) and a dummy variable for non-head (*YNH*). Table 3.1 illustrates the structure of the explanatory variables.

First consider a MNL model fitted to these data. Table 3.2 gives the estimates, asymptotic standard errors and t-statistics, Table 3.4 gives elasticities for each response probability calculated at sample means (by alternative) of the explanatory variables. This model excludes alternative-specific dummy variables. Consequently, the coefficients of the income variables reflect both a correlation of response with income, and unobserved features specific to the associated alternative. The model suggests a strong positive association between ownership and return, and strong negative association between choice and out-of-pocket cost and between non-headship and income. The value of the log-likelihood at the maxi-

Table 3.1
Structure of the explanatory variables.

Person	Alternative[a]	*OPCOST*[b]	*RETURN*[c]	*YOWN*[d]	*YNH*[d]	*CHOICE*
1	1	1.24	2.79	6.1	0	1
1	2	1.13	0	0	6.1	0
1	3	2.33	0	0	0	0
2	1	5.48	5.07	4.3	0	0
2	2	1.13	0	0	4.3	0
2	3	0.93	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Av.[e]	1	3.44	4.41	6.4	0	0.46
	2	0.97	0	0	6.4	0.24
	3	1.96	0	0	0	0.30

[a]Alternative 1 = own (73 cases); alternative 2 = rent/non-head (38 cases); alternative 3 = rent/head (48 cases).
[b]In thousand 1977 dollars.
[c]In thousand 1977 dollars.
[d]In thousand 1977 dollars.
[e]Sample average by alternative.

Table 3.2
Multinomial logit model of housing status

Variable	Parameter estimate	Standard error	t-Statistic
OPCOST	−4.544	1.011	−4.50
RETURN	2.506	0.747	3.35
YOWN	−0.055	0.074	−0.73
YNH	−0.838	0.202	−4.16

Auxiliary statistics
Sample size = 159
Log-likelihood = −15.91
Estimation method: maximum likelihood, model (3.14).

mum likelihood estimates is −15.9, compared with a value of −174.7 when all coefficients are zero. Using the criterion of maximum probability, the model predicts correctly 97.5% of the observed states.

The MNL model specification can be tested using the procedure described in (3.28). Let S_{own} and S_{head} denote the test statistics obtained by deleting owner and renter-head alternatives, respectively, and estimating the reduced MNL model. Under the null hypothesis that the MNL specification is correct, S_{own} and S_{head} are asymptotically chi-square with three degrees of freedom. For this sample, $S_{\text{own}} = 22.4$ and $S_{\text{head}} = 1.7$. Then the first statistic rejects the MNL specification at the 0.001 significance level, the second does not reject. We conclude, with a significance level at most 0.002, that the MNL specification should be rejected.[8]

[8]The statistics S_{own} and S_{head} are not independent. However, the inequalities $\max(P(S_{\text{own}} > c), P(S_{\text{head}} > c)) \leq P(\max(S_{\text{own}}, S_{\text{head}}) > c) \leq P(S_{\text{own}} > c) + P(S_{\text{head}} > c)$ can be used to bound the significance level and power curve of a criterion which rejects MNL if either of the statistics exceeds c. Alternatively, one can extend the analysis of Hausman and McFadden to establish an exact asymptotic joint test. Using the notation of subsection 3.7, let A and D be restricted subsets of the choice set C, β_D and β_C the restricted and full maximum likelihood estimates of the parameters identifiable from D, and α_A and α_C the restricted and full estimates of the parameters identifiable from A. There may be overlap between β_C and α_C. Let V_D and V_A denote the estimated covariance matrices of β_D and α_A respectively, and let $V_{C\alpha\alpha}$, $V_{C\alpha\beta}$, etc. denote submatrices of the estimated covariance matrix of the full maximum likelihood estimator. Define:

$$H = \frac{1}{T}\sum_{t=1}^{T}\sum_{i \in A \cap D}(x_{it} - x_{At})(z_{it} - z_{Dt})'P_{it},$$

where x and z are the variables corresponding to α and β; x_A and z_D are the probability-weighted averages of these variables within A and D; and P_{it} is the estimated response probability from the full choice set. Then $\sqrt{T}((\alpha_A - \alpha_C)', (\beta_D - \beta_C)')$ is, under the null hypothesis, asymptotically normal with mean zero and covariance matrix:

$$T\begin{bmatrix} V_A - V_{C\alpha\alpha} & V_A H V_D - V_{C\alpha\beta} \\ V_D H' V_A - V_{C\beta\alpha} & V_B - V_{C\beta\beta} \end{bmatrix}.$$

Then a quadratic form in these expressions is asymptotically chi-square under the null with degrees of freedom equal to the rank of the matrix.

Table 3.3
Nested multinomial logit model of housing status.

Variable	Parameter estimate	Standard error	t-Statistic
OPCOST	−2.334	0.622	−3.75
RETURN	0.922	0.479	1.93
YOWN	0.034	0.064	0.53
YNH	−0.438	0.121	−3.62
λ	0.179	0.109	1.64

Auxiliary statistics:
Sample size = 159
Log-likelihood = −10.79
Estimation method:
full maximum likelihood estimation, model (3.40) with lowest level of tree (unit choice) deleted

The MNL specification can be tested against a NMNL decision tree using the Lagrange multiplier statistic given in (3.59). We have calculated this test for the tree in Figure 3.3, with the rental alternatives in a cluster, and obtain a value LM = 15.73. This statistic is asymptotically chi-square with one degree of freedom under the null, leading to a rejection of the MNL specification at the .001 significance level.

We next estimate a nested MNL model of the form (3.40), but continue the simplifying assumption that dwelling units within a cluster are homogeneous. Then there are three alternatives: own, rent/non-head, and rent/head, and one inclusive value coefficient λ. Table 3.3 reports full maximum likelihood estimates of this model. Asymptotic standard errors and t-statistics are given. Table 3.4 gives elasticities calculated at the sample mean using (3.49). The estimated inclusive value coefficient $\lambda = 0.179$ is significantly less than one: the statistic $W = (1.0 - \lambda)^2/SE_\lambda^2 = 56.77$ is just the Wald statistic for the null hypothesis that the MNL model in Table 3.2 is correct, which is chi-square with one degree of freedom under the null. Hence, we again reject the MNL model at the 0.001 significance level. It is also possible to use a likelihood ratio test for the hypothesis. In the example, the likelihood ratio test statistic is LR = 10.24, leading to rejection at the 0.005 but not the 0.001 level. We note that for this example the Lagrange multiplier (LM), likelihood ratio (LR), and Wald (W) statistics differ substantially in value, with LR < LM < W. This suggests that for the sample size in the example the accuracy of the first-order asymptotic approximation to the tails of the exact distributions of these statistics may be low.

The impact of clustering of alternatives can be seen clearly by comparing elasticities between the MNL model and the NMNL model in Table 3.4. For example, the MNL model has equal elasticities of owner and renter-head response probabilities with respect to renter non-head *OPCOST*, as forced by the IIA

Table 3.4
Elasticities in the MNL and NMNL models.[a]

	NML Model			MNML model		
	Alt.1	Alt.2	Alt.3	Alt.1	Alt.2	Alt.3
Variable	Own	Rent non-head	Rent head	Own	Rent non-head	Rent head
Owner *OPCOST*	−8.44	+7.19	+7.19	−4.33	+3.69	+3.69
Rental non-head *OPCOST*	+1.06	−3.35	+1.06	+0.54	−7.48	+5.15
Rent head *OPCOST*	+2.67	+2.67	−6.06	+1.37	+13.10	−12.41
RETURN	+5.97	−5.08	−5.08	+2.20	−1.87	−1.87
INCOME	+1.10	−3.92	+1.45	+0.79	−9.37	+6.29

[a]Elasticities are calculated at sample means of the explanatory variables; see formula (3.49).

property, whereas in the NMNL model the first of these elasticities is substantially decreased and the second is substantially increased.

A final comment on the NMNL model concerns the efficiency of sequential estimates and their suitability as starting values for iteration to full maximum likelihood. Sequential estimation of the NMNL model starts by fitting a MNL model to headship status among renters. Only *OPCOST* and *YNH* vary among renters, so that only the coefficients of these variables are identified. Next an inclusive value for renters is calculated. Finally, a MNL model is fitted to own-rent status, with *RETURN*, *YOWN*, and the calculated inclusive value as explanatory variables. The covariance matrix associated with sequential estimates is complicated by the use of calculated inclusive values; Amemiya (1978d) and McFadden (1981) give computational formulae. One Newton–Raphson iteration from the sequential estimates yields estimates which are asymptotically equivalent to full information maximum likelihood estimates.[9] In general sequential estimation may be quite inefficient, resulting in one-step estimators which are far from the full maximum likelihood estimates. However, in this example, the sequential estimates are quite good, agreeing with the full information estimates to the third decimal place. The log-likelihood for the sequential estimates is −10.7868, compared with −10.7862 at the full maximum.

The clustering of alternatives described by the preceding NMNL model could also be captured by a one-factor MNP model. With three alternatives, it is also feasible to estimate a MNP model with varying taste coefficients on all variables. We have not estimated these models. On the basis of previous studies [Hausman

[9]One must be careful to maintain consistent parameter definitions between the sequential procedure and the NMNL model likelihood specification; namely the parameter α from the first sequential step (3.41) is scaled to $\alpha\kappa_{rh}\lambda_r$ in the NMNL generating function (3.47).

and Wise (1978), Fischer–Nagin (1981)], we would expect the one-factor MNP model to give fits comparable to the NMNL model, and the MNP model with full taste variation to capture heterogeneities which the first two models miss.

4. Further topics

4.1. Extensions

Econometric analysis of qualitative response has developed in a number of directions from the basic problem of multinomial choice. This section reviews briefly developments in the areas of dynamic models, systems involving both discrete and continuous variables, self-selection and sampling problems, and statistical methods to improve the robustness of estimators or asymptotic approximations to finite-sample distributions.

4.2. Dynamic models

Many important economic applications of qualitative response models involve observations through time, often in a combined cross-section/time-series framework. The underlying latent variable model may then have a components of variance structure, with individual effects (population heterogeneity), autocorrelation, and dependence on lagged values of latent or indicator variables (state dependence). This topic has been developed in great depth by Heckman (1974, 1978b, 1981b, 1981c), Heckman and McCurdy (1982), Heckman and Willis (1975), Flinn and Heckman (1980), and Lee (1980b).

Dynamic discrete response models are special cases of systems of non-linear simultaneous equations, and their econometric analysis can utilize the methods developed for such systems, including generalized method of moments estimators; see Hausman (1982), Hansen (1982), and Newey (1982).

Most dynamic applications have used multivariate normal disturbances so that the latent variable model has a linear structure. This leads to MNP choice probabilities. As a consequence, computation limits problem size. Sometimes the dimension of the required integrals can be reduced by use of moments estimators [Hansen (1982) and Ruud (1982)]. Alternatively, one or two factor MNP models offer computational convenience when the implied covariance structure is appropriate. For example, consider the model:

$$y_{nt}^{*} = x_{nt}\beta + \varepsilon_{nt} - \nu_{n} \tag{4.1}$$

and $y_{nt}=1$ if $y_{nt}^*\geq 0$, $y_{nt}=-1$ otherwise, where $t=1,\dots,T$; $n=1,\dots,N$; ν_n is an individual random effect which persists over time; and ε_{nt} are disturbances independent of each other and of ν. If ε_{nt} and ν_n are normal, then the probability of a response sequence has the tractable form:

$$P(y_{n1},\dots,y_{nT})=\int_{-\infty}^{+\infty}\frac{1}{\sigma_\nu}\phi\left(\frac{\nu}{\sigma_\nu}\right)\prod_{t=1}^{T}\Phi\left(\frac{y_{nt}(x_{nt}\beta+\nu)}{\sigma_\varepsilon}\right)\mathrm{d}\nu. \tag{4.2}$$

It is also possible to develop tractable dynamic models starting with extreme-value disturbances. If $(\nu_n,\varepsilon_{n1},\dots,\varepsilon_{nT})$ has the generalized extreme value distribution

$$F(\nu_n,\varepsilon_{n1},\dots,\varepsilon_{nT})=\exp\left\{-e^{-\nu_n}-\left[\sum_{t=1}^{T}\mathrm{e}^{-\varepsilon_{nt}/\lambda}\right]^{\lambda}\right\}, \tag{4.3}$$

with $0<\lambda\leq 1$, then the probability that $y_{nt}=1$ for all t in any subset A of $\{1,\dots,T\}$ is:

$$P_A=1\Big/\left(1+\left(\sum_{t\in A}\mathrm{e}^{-x_t\beta/\lambda}\right)^{\lambda}\right). \tag{4.4}$$

If A_n is the set of times with $y_{nt}=1$, and $k(B)$ is the cardinality of a set B, then

$$P(y_{n1},\dots,y_{nT})=\sum_{B\subseteq A_n^c}(-1)^{k(B)}P_{A_n\cup B}. \tag{4.5}$$

For a more general discussion of models and functional forms for discrete dynamic models, see Heckman (1981b).

4.3. *Discrete–continuous systems*

In some applications, discrete response is one aspect of a larger system which also contains continuous variables. An example is a consumer decision on what model of automobile to purchase and how many kilometers to drive (VKT). It is important to account correctly for the joint determination of discrete and continuous choices in such problems. For example, regression of VKT on socioeconomic characteristics for owners of American cars is likely to be biased by self-selection into this sub-population.

A typical discrete–continuous model is generated by the latent variable model:

$$y_{it}^* = x_{it}\beta_{it} - \varepsilon_{it} \qquad (i = 1,2,3), \tag{4.6}$$

and generalized indicator function:

$$y_{1t} = \begin{cases} 1, & \text{if } y_{1t}^* \geq 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.7}$$

$$y_{2t} = y_{1t} y_{2t}^* + (1 - y_{1t}) y_{3t}^*, \tag{4.8}$$

where $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are multivariate normal with a mean zero and covariance matrix:

$$\begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}.$$

This model is sometimes termed a switching regression with observed regime.

Discrete–continuous models such as (4.6)–(4.7) can be estimated by maximum likelihood, or a computationally easier two-step procedure. The latter method first estimates the reduced form (marginal) equation for the discrete choice. Then the fitted probabilities are either used to construct instruments for the endogenous discrete choice variables in (4.8), or else used to augment (4.8) with a "hazard rate" whose coefficient absorbs the covariance of the disturbance and explanatory variables. Developments of these methods can be found in Quandt (1972), Heckman (1974), Amemiya (1974b), Lee (1980a), Maddala and Trost (1980), Hay (1979), Duncan (1980a), Dubin and McFadden (1980), and Poirier and Ruud (1980). A comprehensive treatment of models of this type can be found in Lee (1981). Empirical experience is that estimates obtained using augmentation by the hazard rate are quite sensitive to distributional assumptions; Lee (1981a) provides results on this question and develops a specification test for the usual normality assumption.

If discrete–continuous response is the result of economic optimization, then cross-equation restrictions are implied between the discrete and continuous choice equations. These conditions may be imposed to increase the efficiency of estimation, or may be used to test the hypothesis of optimization. These conditions are developed for the firm by Duncan (1980a) and for the consumer by Dubin and McFadden (1980).

4.4. *Self-selection and biased samples*

In the preceding section, we observed that joint discrete–continuous response could introduce bias in the continuous response equation due to self-selection into a target sub-population. This is one example of a general problem where ancillary responses lead to self-selection or biased sampling.

In general, it is possible to represent self-selection phenomena in a joint latent variable model which also determines the primary response. Then models with a mathematical structure similar to the discrete-continuous response models can be used to correct self-selection biases [Heckman (1976b), Hausman and Wise (1977), Maddala (1977a), Lee (1980a)].

Self-selection is a special case of biased or stratified sampling. In general, stratified sampling can be turned to the advantage of the econometrician by using estimators that correct bias and extract maximum information from samples [Manski and Lerman (1977), Manski and McFadden (1980), Cosslett (1980a), McFadden (1979c)]. To illustrate these approaches, consider the problem of analyzing multinomial response using self-selected or biased samples. Let $f^i(x, \beta^*)$ denote the true response probability and $p(x)$ the density of the explanatory variables in the population of interest. Self-selection or stratification can be interpreted as identifying an "exposed" sub-population from which the observations are drawn; let $\pi(i, x)$ denote the conditional probability that an individual with characteristics (i, x) is selected into the exposed sub-population. For example, π may be the probability that an individual agrees to be interviewed or is able to provide complete data on x, or the probability that the individual meets the screening procedures established by the sampling protocol (e.g. "terminate the interview on rental housing costs if respondent is an owner"). The selection probability may be known, particularly in the case of deliberately biased samples (e.g. housing surveys which over-sample rural households). Alternately, the selection process may be modeled as a function of a vector of unknown parameters γ^*. An example of a latent variable model yielding this structure is the recursive system $y_i^* = x_i\beta - \varepsilon_i$ and $y_i = 1$ if $y_i^* \geq y_j^*$, $y_i = 0$, otherwise, for $i = 1,\dots,m$; $y_0^* = x\gamma_0 + \sum_{i=1}^m y_i\gamma_i - \varepsilon_0$ and $y_0 = 1$ if $y_0^* \geq 0$ and $y_0 = 0$ otherwise; where $x = (x_1,\dots,x_m)$, $\gamma = (\gamma_0,\dots,\gamma_m)$, and y_0 is an indicator for selection into the exposed sub-population.

By Bayes' law, the likelihood of an observation (i, x) in the exposed population is:

$$h(i,x) = \pi(i,x,\gamma^*)f^i(x,\beta^*)p(x)/q(\beta^*,\gamma^*), \tag{4.9}$$

where

$$q(\beta^*,\gamma^*) = \sum_x \sum_i \pi(i,x,\gamma^*)f^i(x,\beta^*)p(x) \tag{4.10}$$

is the fraction of the target population which is exposed. Note that when γ^* is unknown, it may be impossible to identify all components of (γ^*, β^*), or identification may be predicated on arbitrary restrictions on functional form. Then auxiliary information on the selection process (e.g. follow-up surveys of non-respondents, or comparison of the distributions of selected variables between the exposed population and censuses of the target population) is necessary to provide satisfactory identification.

Stratified sampling often identifies several exposed sub-populations, and draws a sub-sample from each. The sub-populations need not be mutually exclusive; e.g. a survey of housing status may draw one stratum from lists of property-owners, a second stratum by drawing random addresses in specified census tracts. If the originating strata of observations are known, then the likelihood (4.9) specific to each stratum applies, with the stratum identifier s an additional argument of π and q. However, if the observations from different strata are pooled without identification, then (4.9) applies uniformly to all observations with π a mixture of the stratum-specific selection probabilities: if μ_s is the share of the sample drawn from stratum s, then except for an inessential constant:

$$\pi(i, x) = \sum_s \pi(i, x, s, \gamma^*)\mu_s / q_s. \tag{4.11}$$

The likelihood (4.9) depends on the density of the explanatory variables $p(x)$ which is generally unknown and of high dimensionality, making direct maximum likelihood estimation impractical. When the selection probability functions are known, one alternative is to form a likelihood function for the observations as if they were drawn from a random sample, and then weight the observations to obtain a consistent maximum "pseudo-likelihood" estimator. Specifically, if the selection probability $\pi(i, x)$ is positive for all i, then, under standard regularity conditions, consistent estimates of β are obtained when observation (i, x) is assigned the log pseudo-likelihood $(1/\pi(i, x))\ln f^i(x, \beta)$. This procedure can be applied stratum by stratum when there are multiple strata, provided the positivity of π is met. However, in general it is more efficient to pool strata and use the pooled selection rate (4.11).[10] Pooling is possible if the exposure shares of q_s are known or can be estimated from an auxiliary sample which grows in size at least as rapidly as the main sample. Further discussion of this method and derivation of the appropriate covariance matrices can be found in Manski and Lerman (1977) and Manski and McFadden (1981).

[10]Additional weighting of observations with weights depending on x, but not on (i, s), will in general not affect consistency, and may be used to improve efficiency. With appropriate redefinition of $\pi(i, x)$, it is also possible to reweight strata.

A second approach to estimation is to form the pooled sample conditional likelihood of response i and stratum s, given x,

$$l(i,s|x,\beta,\gamma)=\frac{f^{i}(x,\beta)\pi(i,x,s,\gamma)\mu_s/q_s}{\sum_{j=1}^{m} f^{i}(x,\beta)\sum_{t}\pi(j,x,t,\gamma)\mu_t/q_t}. \tag{4.12}$$

When the stratum s is not identified or there is a single stratum, this reduces to

$$l(i|x,\beta,\gamma)=\frac{f^{i}(x,\beta)\pi(i,x,\gamma)}{\sum_{j=1}^{m} f^{j}(x,\beta)\pi(j,x,\gamma)}. \tag{4.13}$$

With appropriate regularity conditions, plus the requirement that $\pi(i,x,\gamma)$ be positive for all i, maximum conditional likelihood estimates of (β,γ) are consistent.[11] Further discussion of this method and derivation of the appropriate covariance matrix is given in McFadden (1979) and Manski and McFadden (1981).

When the response model has the multinomial logit functional form, conditional maximum likelihood has a simple structure. For

$$f^{i}(x,\beta)=\mathrm{e}^{x_i\beta}\Big/\sum_{j=1}^{m}\mathrm{e}^{x_j\beta}, \tag{4.14}$$

(4.12) becomes:

$$l(i,s|x,\beta,\gamma)=\frac{\exp[x_i\beta+\ln(\pi(i,x,s,\gamma)\mu_s/q_s)]}{\sum_{(j,t)\in A}\exp[x_j\beta+\ln(\pi(j,x,t,\gamma)\mu_t/q_t)]}\qquad (i,s)\in A, \tag{4.15}$$

where A is the set of pairs (i,s) with $\pi(i,x,s,\gamma)>0$. When $\pi(i,x,s,\gamma)\mu_s/q_s$ is known or can be estimated consistently from auxiliary data, or $\ln\pi(i,x,s,\gamma)$ is linear in unknown parameters, then the response and selection effects combine in a single MNL form, permitting simple estimation of the identifiable parameters.

[11]The exposure shares q_s may be known, or estimated consistently from an auxiliary sample. Alternatively, they can be estimated jointly with (β,γ). Identification in the absence of auxiliary information will usually depend on non-linearities of the functional form.

It is possible to obtain fully efficient variants of the conditional maximum likelihood method by incorporating side constraints (when q_s is known) and auxiliary information (when sample data on q_s is available); see Cosslett (1981a) and McFadden (1979).

4.5. *Statistical methods*

Econometric methods for qualitative response models have been characterized by heavy reliance on tractable but restrictive functional forms and error specifications, and on first-order asymptotic approximations. There are four areas of statistical investigation which have begun to relax these limits: development of a variety of functional forms for specialized applications, creation of batteries of specification tests, development of robust methods and identification of classes of problems where they are useful, and development of higher-order asymptotic approximations and selected finite-sample validations.

This chapter has surveyed the major lines of development of functional forms for general purpose multinomial response models. In a variety of applications with special structure such as longitudinal discrete response or serially ordered alternatives, it may be possible to develop specialized forms which are more appropriate. Consider, for example, the problem of modeling serially ordered data. One approach is to modify tractable multinomial response models to capture the pattern of dependence of errors expected for serial alternatives. This is the method adopted by Small (1982), who develops a generalized extreme value model with proximate dependence. A second approach is to generalize standard discrete densities to permit dependence of parameters on explanatory variables. For example, the Poisson density

$$P_k = e^{-\lambda}\lambda^k/k! \qquad (k = 0,\ldots), \tag{4.16}$$

with $\lambda = e^{x\beta}$, or the negative binomial density

$$P_k = \frac{\Gamma(r+k)}{\Gamma(r)k!}p^r(1-p)^k, \tag{4.17}$$

with $r > 0$ and $p = 1/(1+e^{-x\beta})$, provide relatively flexible forms. A good example of this approach and analysis of the relationships between functional forms is Griliches, Hall and Hausman (1982).

Specification tests for discrete response models have been developed primarily for multinomial response problems, using classical large-sample tests for nested hypotheses. Lagrange Multiplier and Wu–Hausman tests of the sort discussed in this chapter for testing the MNL specification clearly have much wider applicability. An example is Lee's (1981a) use of a Lagrange Multiplier test for a binomial probit model against the Pearson family. It is also possible to develop rather straightforward tests of non-nested models by applying Lagrange Multiplier tests to their probability mixtures. There has been relatively little development of non-parametric methods. McFadden (1973) and McFadden, Tye, and Train (1976) propose some tests based on standardized residuals; to date these have not proved useful. There is no finite sample theory, except for scattered Monte Carlo results, for specification tests.

The primary objective of the search for robust estimators for discrete response models is to preserve consistency when the shape of the error distribution is misspecified. This is a different, and more difficult, problem than is encountered in most discussions of linear model robust procedures where consistency is readily attained and efficiency is the issue. Consequently, results are sparse. Manski (1975) and Cosslett (1980) have developed consistent procedures for binomial response models of the form $P_1 = f^1(x\beta)$, where f^1 is known only to be monotone (with standardized location and scale); more general techniques developed by Manski (1981) may make it possible to extend these results.

To evaluate this approach, it is useful to consider the common sources of model misspecification: (1) incorrect assumptions on the error distribution, (2) reporting or coding errors in the discrete response, (3) omitted variables, and (4) measurement errors in explanatory variables. For concreteness, consider a simple binomial probit model with a latent variable representation $y^* = x^*\beta - \nu$, ν standard normal, $y = 1$ if $y^* \geq 0$ and $y = 0$ otherwise, so that in the absence of misspecification problems the response probability is $P_1 = \Phi(x^*\beta)$. Now suppose all the sources of misspecification are present: $x^* = (x_1^*, x_2^*)$ has x_2^* omitted and x_1^* measured with error, $x_1 = x_1^* + \eta_1$. Then $y^* = x_1^*\beta_1 - (\nu - x_2^*\beta_2 + \xi)$ with response error ξ. Suppose for analysis that $(x_1^*, x_2^*, \eta_1, \nu, \xi)$ is multivariate normal with mean $(\mu_1, \mu_2, 0, 0, 0)$ and covariance matrix:

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & 0 & 0 & 0 \\ \Sigma_{21} & \Sigma_{22} & 0 & 0 & 0 \\ 0 & 0 & \Omega & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}. \tag{4.18}$$

Then observations conform to the conditional probability of $y = 1$, given x_1:

$$P_1 = \Phi\left(\frac{x_1\beta_1 + \alpha + (x_1 - \mu_1)\gamma}{\lambda}\right), \tag{4.19}$$

where

$$\alpha = \mu_2\beta_2,$$
$$\gamma = (\Omega_{11}+\Sigma_{11})^{-1}(\Sigma_{12}\beta_2 - \Omega_{11}\beta_1),$$
$$\lambda^2 = 1+\sigma^2+\beta_1'\Sigma_{12}\beta_2+\beta_2'\left[\Sigma_{22}-\Sigma_{21}(\Omega_{11}+\Sigma_{11})^{-1}\Sigma_{12}\right]\beta_2$$
$$+\beta_1'\Omega_{11}(\Omega_{11}+\Sigma_{11})^{-1}[\Sigma_{11}\beta_1+\Sigma_{12}\beta_2].$$

Estimating the probit model $P_1 = \Phi(x_1\tilde{\beta}_1)$ without allowance for errors of misspecification will lead to asymptotically biased estimates of relative coefficients in β_1 if x_1 is measured with error ($\Omega_{11} \neq 0$), or omitted variables are correlated with $x_1(\Sigma_{12}\beta_2 \neq 0)$ or make a non-zero contribution to the intercept ($\mu_2\beta_2 \neq 0$). These sources of error also change the scale λ. Reporting and coding errors in the response ($\sigma^2 \neq 0$) affect the scale λ, but do not affect the asymptotic bias of relative coefficients. Misspecification of the error distribution can always be reinterpreted, in light of the discussion in Section 2 on approximating response functions, as omission of the variables necessary to make the response a probit.

Consider an estimator of the Cosslett or Manski type which effectively estimates a model $P_1 = F(x_1\tilde{\beta}_1)$ with F a monotone function which is free to conform to the data. This approach can yield consistent estimates of relative coefficients in β_1 in the presence of response coding errors or an unknown error distribution, provided there are no omitted variables or measurement errors in x. However, the Cosslett–Manski procedures are ineffective against the last two error sources.[12] Furthermore, the non-linearity of (4.19) renders inoperative the instrumental variable methods which are effective for treatment of measurement error in linear models. How to handle measurement error in qualitative response models is an important unsolved problem. This topic is discussed further by Yatchew (1980).

Most applications of qualitative response problems to date have used statistical procedures based on first-order asymptotic approximations. Scattered Monte Carlo studies and second-order asymptotic approximations suggest that in many qualitative response models with sample sizes of a few hundred or more, first-order

[12] Manski and I have considered estimators obtained by maximizing a "pseudo-log-likelihood" in which observation (y_i, x) makes the contribution:

$$(y_i+a)\left(f^i(x,\beta)+a\right)^b/b-\left(f^i(x,\beta)+a\right)^{b+1}/(b+1),$$

with a ≥ 0 and $b > -1$. In the absence of measurement errors, this method yields consistent estimators. For positive a and b, this procedure bounds the influence of extreme observations, and should reduce the impact of coding errors in y. [Note that this class defines a family of M-estimators in the terminology of Huber (1965); the cases $a = b = 0$ and $a = 0$, $b = 1$ yield maximum likelihood and non-linear least squares estimates respectively.] We find, however, that this approach does not substantially reduce asymptotic bias due to coding errors.

approximations are moderately accurate. Nevertheless, it is often worthwhile to make second-order corrections for bias of estimators and the size of tests for samples in this range. As a rule of thumb, sample sizes which yield less than thirty responses per alternative produce estimators which cannot be analyzed reliably by asymptotic methods. These issues are discussed further in Domencich and McFadden (1975), Amemiya (1980, 1981), Cavanaugh (1982), Hausman and McFadden (1982), Rothenberg (1982), and Smith, Savin, and Robertson (1982).

5. Conclusion

This chapter has surveyed the current state of econometric models and methods for the analysis of qualitative dependent variables. Several features of this discussion merit restatement. First, the models of economic optimization which are presumed to govern conventional continuous decisions are equally appropriate for the analysis of discrete response. While the intensive marginal conditions associated with many continuous decisions are not applicable, the characterization of economic agents as optimizers implies conditions at the extensive margin and substantive restrictions on functional form. Unless the tenets of the behavioral theory are themselves under test, it is good econometric practice to impose these restrictions as maintained hypotheses in the construction of discrete response models.

Second, as a formulation in terms of latent variable models makes clear, qualitative response models share many of the features of conventional econometric systems. Thus the problems and methods arising in the main stream of econometric analysis mostly transfer directly to discrete response. Divergences from the properties of the standard linear model arise from non-linearity rather than from discreteness of the dependent variable. Thus, most developments in the analysis of non-linear econometric systems apply to qualitative response models. In summary, methods for the analysis of qualitative dependent variables are part of the continuing development of econometric technique to match the real characteristics of economic behavior and data.

Appendix: Proof outlines for Theorems 1–3

Theorem 1

This result specializes a general consistency theorem of Huber (1965, theorem 1) which states that any sequence of estimators which almost surely approaches the suprema of the likelihood functions as $T \to \infty$ must almost surely converge to the true parameter vector θ^*. Assumptions (1)–(3) imply Huber's conditions A-1 and

A-2. The inequality $-e^{-1} \leq z \ln z \leq 0$ for $0 \leq z = f^i(x,\theta) \leq 1$ implies Huber's A-3, and assumption (4) and this inequality imply Huber's A-4 and A-5. It is possible to weaken assumptions (1)–(4) further and still utilize Huber's argument; the formulation of Theorem 1 is chosen for simplicity and ease in verification.

Theorem 2

Note first that $L_T(\theta) \leq 0$ and

$$\mathrm{E}L_T(\theta^*) = \int \mathrm{d}p(x) \sum_{i-1}^{m} f^i(x,\theta^*) \ln f^i(x,\theta^*) \geq -m/e$$

from the bound above on $z \ln z$ for $0 \leq z \leq 1$. Hence, $L_T(\theta^*) > -\infty$ almost surely, and a sequence of estimators satisfying (3.9) exists almost surely. Let Θ_2 be a compact subset of Θ_0 [assumption (5)] which contains a neighborhood Θ_1 of θ^*, and let $\hat{\theta}_T$ be a maximand of $L_T(\theta)$ on Θ_2. Choose $\tilde{\theta}_T \in \Theta \backslash \Theta_1$ such that $L_T(\tilde{\theta}_T) + 1/T \geq \sup_{\Theta \backslash \Theta_1} L_T(\theta)$. Define $\bar{\theta}_T = \tilde{\theta}_T$ if $L_T(\tilde{\theta}_T) \geq L_T(\hat{\theta}_T)$, and $\bar{\theta}_T = \hat{\theta}_T$ otherwise. Then $\bar{\theta}_T$ satisfies (3.9) and by Theorem 1 converges almost surely to θ^*, and therefore almost surely eventually stays in Θ_1. Hence, almost surely eventually $\bar{\theta}_T = \hat{\theta}_T \in \Theta_1$, implying $L_T(\theta_T) \geq L_T(\theta)$ on an open neighborhood of Θ_1, and therefore $\partial L_T(\hat{\theta}_T)/\partial\theta = 0$.

Theorem 3

This result extends a theorem of Rao (1973, 5e2) which establishes for a multinomial distribution without explanatory variables that maximum likelihood estimates are asymptotically normal. Assumptions (6) and (7) correspond to assumptions made by Rao, with the addition of bounds which are integrable with respect to the distribution of the explanatory variables. This proof avoids the assumption of continuous second derivatives usually made in general theorems on asymptotic normality [cf. Rao (1973, 5f.2(iii)), Huber (1965, theorem 3 corollary)].

Let Θ_1 be a neighborhood of θ^* with compact closure on which assumptions (5) and (6) hold. By Theorem 2, almost surely eventually $\hat{\theta}_T \in \Theta_1$ and

$$0 = \sum_{t=1}^{T} \sum_{i=1}^{m} y_{it} \frac{\partial \ln f^i(x_t, \hat{\theta}_T)}{\partial\theta}. \tag{A.1}$$

Noting that

$$\sum_{i=1}^{m} f^i(x_t,\theta) \frac{\partial \ln f^i(x_t,\theta)}{\partial\theta} \equiv 0, \tag{A.2}$$

one can rewrite (A.1) as $0 = A_T + B_T + C_T - D_T$ with:

$$A_T = \sum_{t=1}^{T} \sum_{i=1}^{m} \left(y_{it} - f^i(x_t,\theta^*)\right) \frac{\partial \ln f^i(x_t,\theta^*)}{\partial \theta},$$

$$B_T = \sum_{t=1}^{T} \sum_{i=1}^{m} \left(y_{it} - f^i(x_t,\theta^*)\right) \left[\frac{\partial \ln f^i(x_t,\hat\theta_T)}{\partial \theta} - \frac{\partial \ln f^i(x_t,\theta^*)}{\partial \theta} \right],$$

$$C_T = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{m} \left(f^i(x_t,\theta^*) - f^i(x_t,\hat\theta_T)\right) \frac{\partial \ln f^i(x_t,\hat\theta_T)}{\partial \theta} + J(\theta^*)\left(\hat\theta_T - \theta^*\right) \right\},$$

$$D_T = TJ(\theta^*)\left(\hat\theta_T - \theta^*\right). \tag{A.3}$$

The steps in the proof are (1) show $B_T/\sqrt{T}(1+\sqrt{T}\,|\hat\theta_T - \theta^*|) \to 0$ in probability as $T \to \infty$, (2) show $C_T/\sqrt{T}(1+\sqrt{T}\,|\hat\theta_T - \theta^*|) \to 0$ in probability, (3) show $J(\theta^*)^{-1}(A_T - D_T)/\sqrt{T} \to 0$ in probability, and (4) show $J(\theta^*)^{-1}A_T/\sqrt{T}$ converges in distribution to a normal random vector with mean zero and covariance matrix $J(\theta^*)^{-1}$. With the result of Step 3, this proves Theorem 3.

Step 1. We use a fundamental lemma of Huber (1965, lemma 3). Define

$$\begin{aligned} \psi(y,x,\theta) &= \sum_{i=1}^{m} \left(y_i - f^i(x,\theta^*)\right) \frac{\partial \ln f^i(x,\theta)}{\partial \theta} + \theta - \theta^*, \\ \lambda(\theta) &= \mathrm{E}\psi(y,x,\theta) = \theta - \theta^* \\ u(y,x,\theta,d) &= \sup_{|\theta'-\theta| \leq d} |\psi(y,x,\theta') - \psi(y,x,\theta)|. \end{aligned} \tag{A.4}$$

Then assumption (6) (iii) implies:

$$u(y,x,\theta,d) \leq d\left(1 + \sum_{i=1}^{m} |y_i - f^i(x,\theta^*)| \cdot \gamma^i(x)\right), \tag{A.5}$$

and hence using the bounds (6) (iv):

$$\begin{aligned} \mathrm{E}u(y,x,\theta,d) &\leq d\left(1 + \sum_{i=1}^{m} \int \mathrm{d}p(x)\alpha^i(x)\gamma^i(x)\right) \equiv A_1 d, \\ \mathrm{E}u(y,x,\theta,d)^2 &\leq d^2\left(2A_1 + m\sum_{i=1}^{m} \mathrm{E}\left(y_i - f^i(x,\theta^*)\right)^2 \gamma^i(x)^2\right) \\ &\leq d^2\left(2A_1 + m\sum_{i=1}^{m} \int \mathrm{d}p(x)\alpha^i(x)\gamma^i(x)^2\right) \equiv A_2 d^2. \end{aligned} \tag{A.6}$$

These conditions imply Huber's assumptions $(N-1)$ to $(N-3)$. Define:

$$Z_T(\theta)=\frac{\left|\sum_{t=1}^{T}\left(\psi(y_t,x_t,\theta)-\psi(y_t,x_t,\theta^*)-\lambda(\theta)\right)\right|}{\sqrt{T}\left(1+\sqrt{T}\,|\theta-\theta^*|\right)}$$
$$=\frac{\left|\sum_{t=1}^{T}\sum_{i=1}^{m}\left(y_{it}-f^i(x_t,\theta^*)\right)\left[\frac{\partial\ln f^i(x,\theta)}{\partial\theta}-\frac{\partial\ln(f^i(x,\theta^*)}{\partial\theta}\right]\right|}{\sqrt{T}\left(1+\sqrt{T}\,|\theta-\theta^*|\right)}.\tag{A.7}$$

Then Huber's Lemma 3 states that:

$$\sup_{\Theta_0} Z_T(\theta)\to 0\tag{A.8}$$

in probability as $T\to\infty$. But

$$Z_T(\hat{\theta}_T)=B_T/\sqrt{T}\left(1+\sqrt{T}\,|\hat{\theta}_T-\theta^*|\right),\tag{A.9}$$

and Step 1 is complete.

Step 2. Since f^i is differentiable on Θ_1, the mean value theorem implies:

$$f^i(x_t,\theta^*)-f^i(x_t,\hat{\theta}_T)=-\left[\frac{\partial f^i(x_t,\tilde{\theta}_t)}{\partial\theta}\right]'(\hat{\theta}_T-\theta^*),\tag{A.10}$$

where $\tilde{\theta}_t$ is some interior point on the line segment connecting θ^* and $\hat{\theta}_T$. Substituting this expression in C_T yields $C_T=(F_T+G_T)\cdot(\hat{\theta}_T-\theta^*)$, with

$$F_T=-\sum_{t=1}^{T}\sum_{i=1}^{m}f^i(x_t,\theta^*)\left[\frac{\partial\ln f^i(x_t,\theta^*)}{\partial\theta}\right]\left[\frac{\partial\ln f^i(x_t,\theta^*)}{\partial\theta}\right]'+TJ(\theta^*)\tag{A.11}$$

and

$$G_T=\sum_{t=1}^{T}\sum_{i=1}^{m}\left\{\frac{\partial\ln f^i(x_t,\theta^*)}{\partial\theta}\frac{\partial f^i(x_t,\theta^*)}{\partial\theta'}-\frac{\partial\ln f^i(x_t,\hat{\theta}_T)}{\partial\theta}\frac{\partial f^i(x_t,\tilde{\theta}_t)}{\partial\theta'}\right\}.\tag{A.12}$$

Then $F_T/T \to 0$ in probability by the law of large numbers and

$$|G_t/T| \leqq |\hat{\theta}_T - \theta^*| \sum_{i=1}^{m} \int \mathrm{d}p(x)\alpha^i(x)\beta^i(x)\left[2\gamma^i(x)^2\right] \to 0 \tag{A.13}$$

in probability by Theorem 2. Since

$$\frac{C_T}{\sqrt{T}\left(1+\sqrt{T}\,|\hat{\theta}_T - \theta^*|\right)} = \left(\frac{F_T}{T} + \frac{G_T}{T}\right)\frac{\sqrt{T}\left(\hat{\theta}_T - \theta^*\right)}{1+\sqrt{T}\,|\hat{\theta}_T - \theta^*|}, \tag{A.14}$$

with the second term in the product stochastically bounded, this establishes Step 2.

Step 3. The first two steps establish that $(A_T - D_T)/\sqrt{T}(1+\sqrt{T}\,|\hat{\theta}_T - \theta^*|) \to 0$ in probability and hence $J(\theta^*)^{-1}(A_T - D_T)/\sqrt{T}(1+\sqrt{T}\,|\hat{\theta} - \theta^*|) \to 0$ in probability. Therefore given $\varepsilon > 0$, there exists T_ε such that for $T > T_\varepsilon$, the inequality

$$\left|J(\theta^*)^{-1}A_T/\sqrt{T} - \sqrt{T}\left(\hat{\theta}_T - \theta^*\right)\right| < \varepsilon\left(1+\sqrt{T}\,|\hat{\theta}_T - \theta^*|\right) \tag{A.15}$$

holds with probability at least $1-\varepsilon/2$. Chebyshev's inequality applied to $J(\theta^*)^{-1}A_T/\sqrt{T}$ implies [using assumptions (7) and (6) (iv)] that for some large constant K:

$$\left|J(\theta^*)^{-1}A_T/\sqrt{T}\right| < K \tag{A.16}$$

holds with probability at least $1-\varepsilon/2$. Then (A.15) and (A.16) imply

$$\sqrt{T}\,|\hat{\theta}_T - \theta^*| < (K+\varepsilon)/(1-\varepsilon), \tag{A.17}$$

and hence

$$\left|J(\theta^*)^{-1}A_T/\sqrt{T} - \sqrt{T}\left(\hat{\theta}_T - \theta^*\right)\right| \leq (K+1)\varepsilon/(1-\varepsilon) \tag{A.18}$$

with probability at least $1-\varepsilon$. Since ε can be made small, this establishes Step 3.

Step 4. The expression $J(\theta^*)^{-1}A_T/\sqrt{T}$ has mean zero and covariance matrix $J(\theta^*)^{-1}$, and satisfies the conditions of the Lindeberg–Levy central limit theorem. Therefore it converges in distribution to an asymptotically normal vector.

References

Adler, T. and M. Ben-Akiva (1975) "A Joint Frequency, Destination and Mode Choice Model for Shopping Trips", *Transportation Research Record*, 569, 136–150.

Aitchinson, J. and J. Bennett (1970) "Polychotomous Quantal Response by Maximum Indicant", *Biometrika*, 57, 253–262.

Aitchinson, J. and S. Silvey (1957) "The Generalization of Probit Analysis to the Case of Multiple Responses", *Biometrika*, 44, 131–140.
Amemiya, T. (1973) "Regression Analysis When the Dependent Variable is Truncated Normal", *Econometrica*, 41, 997–1016.
Amemiya, T. (1974a) "Bivariate Probit Analysis: Minimum Chi-Square Methods", *Journal of the American Statistical Association*, 69, 940–944.
Amemiya, T. (1974b) "Multivariate Regression of Simultaneous Equation Models When the Dependent Variables Are Truncated Normal", *Econometrica*, 42, 999–1012.
Amemiya, T. (1974c) "A Note on the Fair and Jaffee Model", *Econometrica*, 42, 759–762.
Amemiya, T. (1975) "Qualitative Response Models", *Annals of Economic and Social Measurement*, 4, 363–372.
Amemiya, T. (1976) "The Maximum Likelihood, the Minimum Chi-Square, and the Non-Linear Weighted Least Squares Estimator in the General Qualitative Response Model", *Journal of the American Statistical Association*, 71, 347–351.
Amemiya, T. (1978a) "The Estimation of a Simultaneous Equation Generalized Probit Model", *Econometrica*, 46, 1193–1205.
Amemiya, T. (1978b) "A Note on the Estimation of a Time Dependent Markov Chain Model", Department of Economics, Stanford University.
Amemiya, T. (1978c) "On a Two-Step Estimation of a Multivariate Logit Model", *Journal of Econometrics*, 8, 13–21.
Amemiya, T. (1979) "The Estimation of a Simultaneous Equation Tobit Model", *International Economic Review*, 20, 169–181.
Amemiya, T. (1980) "The n^2-Order Mean Squared Errors of the Maximum Likelihood and the Minimum Logit Chi-Square Estimates", *The Annals of Statistics*, 8, 488–505.
Amemiya, T. (1981) "Qualitative Response Models: A Survey", *Journal of Economic Literature*, 19, 1483–1536.
Amemiya, T. and F. Nold (1975) "A Modified Logit Model", *Review of Economics and Statistics*, 57, 255–257.
Amemiya, T. and J. Powell (1980) "A Comparison of the Logit Model and Normal Discriminant Analysis When the Independent Variables are Binary", Tech. Report 320, IMSSS, Stanford Univ.
Anas, A. (1981) "Discrete Choice Theory, Information Theory, and the Multinomial Logit and Gravity Models", manuscript, Northwestern Univ.
Anderson, T. (1958) *Introduction to Multivariate Statistical Analysis*. New York: Wiley.
Anscombe, E. J. (1956) "On Estimating Binomial Response Relations", *Biometrika*, 43, 461–464.
Antle, C., L. Klimko, and W. Harkness (1970) "Confidence Intervals for the Parameters of the Logistic Distribution", *Biometrika*, 57, 397–402.
Arabmazar, A. and P. Schmidt (1982) "An Investigation of the Robustness of the Tobit Estimator to Non-Normality", *Econometrica*, 50, 1055–1064.
Ashford, J. R. and R. R. Sowden (1970) "Multivariate Probit Analysis", *Biometrics*, 26, 535–546.
Ashton, W. (1972) *The Logit Transformation*. New York: Hafner.
Atkinson, A. (1972) "A Test of the Linear Logistic and Bradley-Terry Models", *Biometrika*, 59, 37–42.
Barlett, M. S. (1935) "Contingency Table Interactions", *Journal of the Royal Statistical Society* (Supplement) 2, 248–252.
Beggs, S., S. Cardell and J. Hausman (1981) "Assessing the Potential Demand for Electric Cars", *Journal of Econometrics*, 16, 1–19.
Ben-Akiva, M. (1973) "Structure of Passenger Travel Demand Models", *Transportation Research Board Record*, No. 526, Washington, D.C.
Ben-Akiva, M. (1974) "Multi-Dimensional Choice Models: Alternative Structures of Travel Demand Models", *Transportation Research Board Special Report 149*, Washington, D.C.
Ben-Akiva, M. (1977) "Choice Models with Simple Choice Set Generating Processes", Working Paper, MIT.
Ben-Akiva, M. and S. Lerman (1974) "Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work", *Transportation*, 3, 357–376.
Ben-Akiva, M. and S. Lerman (1979) "Disaggregate Travel and Mobility Choice Modes and Measures of Accessibility". In *Behavior Travel Modelling* edited by D. Hensher and P. Stopher, pp. 654–679. London: Coom Helm.

Berkovec, J. and J. Rust (1982) "A Nested Logit Model of Automobile Holdings for One-Vehicle Households", Working Paper, Department of Economics, MIT.
Berkson, J. (1949) "Application of the Logistic Function to Bioassay", *Journal of the American Statistical Association*, 39, 357–365.
Berkson, J. (1951) "Why I Prefer Logits to Probits", *Biometrika*, 7, 327–339.
Berkson, J. (1953) "A Statistically Precise and Relatively Simple Method of Estimating the Bio-Assay with Quantal Response, Based on the Logistic Function", *Journal of the American Statistical Association*, 48, 565–599.
Berkson, J. (1955a) "Estimation of the Integrated Normal Curve by Minimum Normit Chi-Square with Particular Reference to Bio-Assay", *Journal of the American Statistical Association*, 50, 529–549.
Berkson, J. (1955b) "Maximum Likelihood and Minimum Chi-Square Estimations of the Logistic Function", *Journal of the American Statistical Association*, 50, 130–161.
Berndt, E., B. Hall, R. Hall, and J. Hausman (1974) "Estimation and Inference in Non-Linear Structural Models", *Annals of Economic and Social Measurement*, 3, 653–666.
Bishop, T., S. Fienberg, and P. Holland (1975) *Discrete Multivariate Analysis*. Cambridge: MIT Press.
Block, H. and J. Marschak (1960) "Random Orderings and Stochastic Theories of Response". In *Contributions to Probability and Statistics*, edited by I. Olkin, 97–132. Stanford University Press.
Bock, R. D. and L. Jones (1969) *The Measurement and Prediction of Judgement and Choice*. San Francisco: Holden-Day.
Boersch-Supan, A. and J. Pitkin (1982) "Multinomial Logit Models of Housing Choices". Working Paper, No. 79, Joint Center for Urban Studies, MIT and Harvard.
Boskin, M. (1974) "A Conditional Logit Model of Occupational Choice", *Journal of Political Economy*, 82, 389–398.
Boskin, M. (1975) "A Markov Model of Turnover in Aid to Families with Dependent Children", *Journal of Human Resources*, 10, 467–481.
Brownstone, D. (1978) "An Econometric Model of Consumer Durable Choice and Utilization Rate". Paper IP-258, Center for Research in Management Science, University of California, Berkeley.
Cardell, S. (1977) "Multinomial Logit with Correlated Stochastic Terms". Working Paper, Charles River Associates.
Carlton, D. (1979) "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables". Working Paper, Univ. of Chicago.
Cavanagh, C. (1982) *Hypothesis Testing in Models with Discrete Dependent Variables*. Ph.D. Thesis, Dept. of Economics, Univ. of California, Berkeley.
Chambers, E. A. and D. R. Cox (1967) "Discrimination Between Alternative Binary Response Models", *Biometrika*, 54, 573–578.
Chung, C. and A. Goldberger (1982) "Proportional Projections in Limited Dependent Variable Models". Working Paper, Univ. of Wisconsin.
Clark, C. (1961) "The Greatest of a Finite Set of Random Variables", *Operations Research*, 9, 145–162.
Cook, P. (1975) "The Correctional Carrot: Better for Parolees", *Policy Analysis*, 1, 11–54.
Cosslett, S. (1978) "Efficient Estimation of Discrete-Choice Models from Choice-Based Samples". Ph.D. dissertation, Department of Economics, University of California, Berkeley.
Cosslett, S. (1980) "Estimation of Random Utility Models with Unknown Probability Distribution". Working Paper, Univ. of Florida.
Cosslett, S. (1981) "Efficient Estimators of Discrete Choice Models". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, 51–111. Cambridge: MIT Press.
Cosslett, S. (1981a) "Maximum Likelihood Estimator for Choice-Based Samples", *Econometrica*, 9, 1289–1316.
Cox, D. R. (1958) "The Regression Analysis of Binary Sequences", *Journal of the Royal Statistical Society* (Series B), 820, 215–242.
Cox, D. R. (1966) "Some Procedures Connected with the Logistic Response Curve". In *Research Papers in Statistics*, edited by F. David, 55–71. New York: Wiley.
Cox, D. R. (1970) *Analysis of Binary Data*. London: Methuen.
Cox, D. R. (1972) "The Analysis of Multivariate Binary Data", *Applied Statistics*, 21, 113–120.
Cox, D. R. and E. Snell (1968) "A General Definition of Residuals", *Journal of the Royal Statistical Society* (Series B), 30, 248–265.

Cox, D. R. and E. Snell (1971) "On Test Statistics Calculated from Residuals", *Biometrika*, 58, 589–594.

Cragg, J. G. (1971) "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", *Econometrica*, 39, 829–844.

Cragg, J. G. and R. Uhler (1970) "The Demand for Automobiles", *Canadian Journal of Economics*, 3, 386–406.

Crawford, D. and R. Pollak (1982) "Order and Inference in Qualitative Response Models". Discussion Paper 82-4, Rutgers Univ.

Cripps, T. F. and R. J. Tarling (1974) "An Analysis of the Duration of Male Unemployment in Great Britain 1932–1973", *Economic Journal*, 84, 289–316.

Daganzo, C. (1980) *Multinomial Probit*. New York: Academic.

Daganzo, C., F. Bouthelier, and Y. Sheffi (1977) "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm", *Transportation Science*, 11, 338–358.

Dagenais, M. G. (1975) "Application of a Threshold Regression Model to Household Purchases of Automobiles", *Review of Economics and Statistics*, 57, 275–285.

Daly, A. and S. Zachary (1979) "Improved Multiple Choice Models". In *Identifying and Measuring the Determinants of Model Choice*, edited by D. Hensher and O. Dalvi, pp. 187–201.

Debreu, G. (1960) "Review of R. D. Luce Individual Choice Behavior", *American Economic Review*, 50, 186–188.

Diewert, E. (1978) "Duality Approaches to Microeconomic Theory". Technical Report 281, Institute of Mathematical Studies in the Social Sciences, Stanford University.

Domencich, T. and D. McFadden (1975) *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland.

Dubin, J. (1982) *Economic Theory and Estimation of the Demand for Consumer Durables and their Utilization*. Ph.D. Dissertation, MIT.

Dubin, J. and D. McFadden (1980) "An Econometric Analysis of Residential Electrical Appliance Holdings and Usage". Working Paper, Department of Economics, Massachusetts Institute of Technology.

Duncan, G. (1980a) "Formulation and Statistical Analysis of the Mixed Continuous/Discrete Model in Classical Production Theory", *Econometrica*, 48, 839–852.

Duncan, G. (1980b) "Mixed Continuous Discrete Choice Models in the Presence of Hedonic or Exogenous Price Functions". Working Paper, Department of Economics, Washington State University.

Durling, F. (1969) "Bivariate Probit, Logit, and Burrit Analysis". Tech. Report 41, Southern Methodist Univ.

Efron, B. (1975) "The Efficiency of Logistic Regression Compared to Normal Discrimination Analysis", *Journal of the American Statistical Association*, 70, 892–898.

Fair, R. C. and D. M. Jaffee (1972) "Methods of Estimation for Markets in Disequilibrium", *Econometrica*, 40, 497–514.

Farber, H. (1978) "Bargaining Theory, Wage Outcomes, and the Occurrence of Strikes", *American Economic Review*, 68, 262–271.

Farber, H. and D. Saks, (1981) "Why Workers Want Unions", *Journal of Public Economics*, to appear.

Fienberg, S. (1977) *The Analysis of Cross-Classified Data*. Cambridge: MIT Press.

Finney, D. (1964) *Statistical Method in Bio-Assay*. London: Griffin.

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

Fischer, G. and D. Nagin (1981) "Random versus Fixed Coefficient Quantal Choice Models" in C. Manski and D. McFadden, ed., *Structural Analysis of Discrete Data*. Cambridge: MIT.

Flinn, C. and J. Heckman (1982) "Models for the Analysis of Labor Force Dynamics", *Advances in Econometrics*, 1, 35–95.

Flinn, C. and J. Heckman (1982) "New Methods for Analyzing Structural Models of Labor Force Dynamics", *Journal of Econometrics*, 18, 115–168.

Friedman, P. (1973) "Suggestions for the Analysis of Qualitative Dependent Variables", *Public Finance Quarterly*, 1, 345–355.

Fuller, W., Manski, C. and D. Wise (1980) "New Evidence on the Economic Determinants of Post-Secondary Schooling Choices". J.F.K. School of Government Discussion Paper 9UD.

Gart, J. and J. Zweifel (1967) "On the Bias of Various Estimators of the Logit and its Variance", *Biometrika*, 54, 181–187.
Gilbert, C. (1979) "Econometric Models for Discrete Economic Processes", Working Paper, Oxford Univ.
Gillen, D. W. (1977) "Estimation and Specification of the Effects of Parking Costs on Urban Transport Model Choice", *Journal of Urban Economics*, 4, 186–199.
Goldberger, A. S. (1964) *Econometric Theory*. New York: Wiley.
Goldberger, A. S. (1971) "Econometrics and Psychometrics: A Survey of Communalities", *Psychometrika*, 36, 83–107.
Goldberger, A. S. (1973) "Correlations Between Binary Outcomes and Probabilistic Predictions", *Journal of the American Statistical Association*, 68, 84.
Goldfeld, S. M. and R. E. Quandt (1972) *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland.
Goldfeld, S. and R. E. Quandt (1973) "The Estimation of Structural Shifts by Switching Regressions", *Annals of Economic and Social Measurement*, 2, 475–486.
Goldfeld, S. and R. E. Quandt (1976) "Techniques for Estimating Switching Regressions". In *Studies in Non-Linear Estimation*, edited by S. Goldfeld and R. E. Quandt, 3–37. Cambridge: Ballinger.
Goodman, L. A. (1970) "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications", *Journal of the American Statistical Association*, 65, 226–256.
Goodman, L. A. (1971) "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications", *Technometrics*, 13, 33–61.
Goodman, L. A. (1972a) "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables", *American Sociological Review*, 37, 28–46.
Goodman, L. A. (1972b) "A General Model for the Analysis of Surveys", *American Journal of Sociology*, 77, 1035–1086.
Goodman, L. A. (1973) "Causal Analysis of Panel Study Data and Other Kinds of Survey Data", *American Journal of Sociology*, 78, 1135–1191.
Goodman, L. A. and W. H. Kruskal (1954) "Measures of Association for Cross Classifications", *Journal of the American Statistical Association*, 49, 732–764.
Goodman, L. A. and W. H. Kruskal (1954a) "Measures of Association for Cross Classification II, Further Discussion and References", *Journal of the American Statistical Association*, 54, 123–163.
Gourieroux, C. and A. Monfort (1981) "Asymptotic Properties of the Maximum Likelihood Estimator in Dichotomous Logit Models", *Journal of Econometrics*, 17, 83–97.
Gourieroux, C., J. Laffont, and A. Monfort (1980) "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes", *Econometrica*, 48, 675–695.
Gourieroux, C., A. Holly, and A. Monfort (1980) "Kuhn–Tucker, Likelihood Ratio, and Wald Tests for Non-Linear Models with Inequality Constraints on Parameters". Discussion Paper 770, Harvard Univ.
Greene, W. (1981) "Estimation of Some Limited Dependent Variable Models Using Least Squares and The Method of Moments". Working Paper 245, Cornell Univ.
Griliches, Z., B. Hall, and J. Hausman (1978) "Missing Data and Self-Selection in Large Panels", *Annals de l'Insee* 30–31, 137–176.
Griliches, Z., B. Hall, and J. Hausman (1981) "Econometric Models for Count Data with an Application to the Patents R&D Relationship", *Econometrica*, forthcoming.
Grizzle, J. (1962) "Asymptotic Power of Tests of Linear Hypotheses Using the Probit and Logit Transformations", *Journal of the American Statistical Association*, 57, 877–894.
Grizzle, J. (1971) "Multivariate Logit Analysis", *Biometrics*, 27, 1057–1062.
Gronau, R. (1973) "The Effect of Children on the Housewife's Value of Time", *Journal of Political Economy*, 81, 168–199.
Gronau, R. (1974) "Wage Comparisons—A Selectivity Bias", *Journal of Political Economy*, 82, 1119–1143.
Gurland, J., I. Lee, and P. Dahm (1960) "Polychotomous Quantal Response in Biological Assay", *Biometrics*, 16, 382–398.
Haberman, S. (1974) *The Analysis of Frequency Data*. University of Chicago Press.
Haldane, J. (1955) "The Estimation and Significance of the Logarithm of a Ratio of Frequencies",

Annals of Human Genetics, 20, 309–311.
Hall, R. (1970) "Turnover in the Labor Force", *Brookings Papers on Economic Activity*, 3, 709–756.
Hansen, L. (1982) "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029–1054.
Harter, J. and A. Moore (1967) "Maximum Likelihood Estimation, from Censored Samples, of the Parameters of a Logistic Distribution", *Journal of the American Statistical Association*, 62, 675–683.
Hausman, J. A. (1978) "Specification Tests in Econometrics", *Econometrica*, 46, 1251–1271.
Hausman, J. A. (1979) "Individual Discount Rates and the Purchase and Utilization of Energy Using Durables", *Bell Journal of Economics*, 10, 33–54.
Hausman, J. A. (1983) "Specification and Estimation of Simultaneous Equations Models", *Handbook of Econometrics*.
Hausman, J. A. and D. McFadden (1984) "A Specification Test for the Multinomial Logit Model", *Econometrica*, forthcoming.
Hausman, J. A. and D. A. Wise (1976) "The Evaluation of Results from Truncated Samples: The New Jersey Negative Income Tax Experiment", *Annals of Economic and Social Measurement*, 5, 421–445.
Hausman, J. A. and D. A. Wise (1977) "Social Experimentation, Truncated Distribution and Efficient Estimation", *Econometrica*, 45, 319–339.
Hausman, J. A. and D. A. Wise (1978) "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences", *Econometrica*, 46, 403–426.
Hausman, J. A. and D. Wise (1981) "Stratification on Endogenous Variables and Estimation: The Gary Experiment". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, 365–391. Cambridge: MIT Press.
Hay, J. (1979) "An Analysis of Occupational Choice and Income". Ph.D. dissertation, Yale University.
Heckman, J. (1974) "Shadow Prices, Market Wages, and Labor Supply", *Econometrica*, 42, 679–694.
Heckman, J. (1976a) "Simultaneous Equations Model with Continuous and Discrete Endogenous Variables and Structural Shifts". In *Studies in Non-Linear Estimation*, edited by S. Goldfeld and R. Quandt, 235–272. Cambridge: Ballinger.
Heckman, J. (1978a) "Dummy Exogenous Variables in a Simultaneous Equation System", *Econometrica*, 46, 931–959.
Heckman, J. (1978b) "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of the True State Dependence Against the Hypothesis of Spurious State Dependence", *Annals de l'Insee*, 30–31, 227–269.
Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153–161.
Heckman, J. (1981a) "Heterogeneity and State Dependence in Dynamic Models of Labor Supply". In *Conference on Low Income Labor Markets*, edited by S. Rosen. University of Chicago Press.
Heckman, J. (1981b) "Statistical Models for the Analysis of Discrete Panel Data". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, pp. 114–178. Cambridge: MIT Press.
Heckman, J. (1981c) "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Stochastic Process and Some Monte Carlo Evidence on Their Practical Importance". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, pp. 179–185. Cambridge: MIT Press.
Heckman, J. and R. Willis (1975) "Estimation of a Stochastic Model of Reproduction: An Econometric Approach". In *Household Production and Consumption*, edited by N. Terleckyj, pp. 99–138. New York: National Bureau of Economic Research.
Heckman, J. and R. Willis (1977) "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Married Women", *Journal of Political Economy*, 85, 27–58.
Heckman, J. and T. McCurdy (1980) "A Dynamic Model of Female Labor Supply", *Review of Economic Studies*, 47, 47–74.
Heckman, J. and B. Singer (1980) *Longitudinal Labor Market Studies: Theory, Methods and Empirical Results*. Social Science Research Council Monograph. New York: Academic.
Heckman, J. and T. McCurdy (1982) "New Methods for Estimating Labor Supply Functions: A Survey", *Research in Labor Economics*, 4, 65–102.
Heckman, J. and B. Singer (1982) "The Identification Problem in Econometric Models for Duration

Data". Discussion Paper 82-6, N.O.R.C.
Henderson, J. and Y. Ioannides (1982) "Tenure Choice and the Demand for Housing". Working Paper, Dept. of Economics, Boston Univ.
Hockerman, I., Prashker, J. and M. Ben-Akiva (1982) "Estimation and Use of Dynamic Transaction Models of Automobile Ownership". Working Paper, Dept. of Civil Engineering, MIT.
Horowitz, J. (1979a) "Identification and Diagnosis of Specification Errors in the Multinomial Logit Model". Mimeograph, Environmental Protection Agency.
Horowitz, J. (1979b) "A Note on the Accuracy of the Clark Approximation for the Multinomial Probit Model". Mimeograph, Department of Transportation, Massachusetts Institute of Technology.
Horowitz, J. (1980) "The Accuracy of the Multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Demand", *Transportation Research* (Part B).
Horowitz, J. (1981) "Sampling, Specification and Data Errors in Probabilistic Discrete Choice Models". In *Applied Discrete Choice Modeling*, edited by D. Hensher and L. Johnson, 417–435. London: Croom Helm.
Horowitz, J. (1981a) "Statistical Comparison of Non-Nested, Discrete-Choice, Random-Utility Models". Working Paper, Environmental Protection Agency.
Horowitz, J. (1981b) "Testing the Multinomial Logit Model Against the Multinomial Probit Model Without Estimating the Probit Parameters", *Transportation Science*, 15, 153–163.
Horowitz, J., J. Sparmonn, and C. Daganzo (1981) "An Investigation of the Accuracy of the Clark Approximation for the Multinomial Probit Model". Working Paper, EPA.
Huber, P. (1965) "The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions". *Fifth Berkeley Symposium on Statistics and Probability*, 1, 221–233.
Hulett, J. R. (1973) "On the Use of Regression Analysis with a Qualitative Dependent Variable". *Public Finance Quarterly*, 1, 339–344.
Ito, T. (1980) "Methods of Estimation for Multi-Market Disequilibrium Models", *Econometrica*, 48, 97–126.
Johnson, L. and D. Hensker (1981) "Application of Multinomial Probit to a Two-Period Panel Data Set". Working Paper, Macquarie Univ.
Joreskog, K. and A. Goldberger (1975) "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable Model", *Journal of the American Statistical Association*, 70, 631–639.
Judge, G., W. Griffiths, R. Hill, and T. Lee (1980) *The Theory and Practice of Econometrics*. New York: Wiley, p. 601ff.
Kiefer, N. (1978) "Discrete Parametre Variation: Efficient Estimation of a Switching Regression Model", *Econometrica*, 46, 427–434.
Kiefer, N. (1980) "A Note on Switching Regressions and Discrimination", *Econometrica*, 48, 1065–1070.
Kiefer, N. M. (1979) "On the Value of Sample Separation Information", *Econometrica*, 47, 997–1003.
Kiefer, N. and G. Neumann (1979) "An Empirical Job Search Model with a Test of the Constant Reservation Wage Hypothesis", *Journal of Political Economy*, 87, 89–107.
King, M. (1980) "An Econometric Model of Tenure Choice and Demand for Housing as a Joint Decision". Working Paper, Department of Economics, University of Birmingham.
Klecka, W. (1980) *Discriminant Analysis*. Beverly Hills: Sage.
Knoke, D. and P. Burke (1980) *Log-Linear Models*. Beverly Hills: Sage.
Kohn, M., C. Manski, and D. Mundel (1976) "An Empirical Investigation of Factors Influencing College Going Behavior", *Annals of Economic and Social Measurement*, 5, 391–419.
Ladd, G. (1966) "Linear Probability Functions and Discriminant Functions", *Econometrica*, 34, 873–885.
Lave, C. (1970) "The Demand for Urban Mass Transit", *Review of Economics and Statistics*, 52, 320–323.
Lee, L. F. (1978) "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables", *International Economic Review*, 19, 415–433.
Lee, L. F. (1979) "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables", *Econometrica*, 47, 977–996.
Lee, L. F. (1980) "Fully Recursive Probability Models and Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data", Discussion Paper 83-32, Univ. of Mass.

Lee, L. F. (1980a) "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias". Working Paper, Department of Economics, University of Minnesota.
Lee, L. F. (1980b) "Statistical Analysis of Econometric Models of Discrete Panel Data". Working Paper, Department of Economics, University of Minnesota.
Lee, L. F. (1981a) "Estimation of Some Non-Normal Limited Dependent Variable Models". Discussion Paper, No. 43, Center for Econometrics and Decision Sciences, Univ. of Florida.
Lee, L. F. (1981b) "Simultaneous Equations Models with Discrete and Censored Variables". In *Structural Analysis of Discrete Data*, edited C. Manski and D. McFadden, 346–364. Cambridge MIT Press.
Lee, L. F. (1981c) "A Specification Test for Normality Assumption for the Truncated and Censored Tobit Models". Discussion Paper No. 44, Center for Econometrics and Decision Sciences", Univ. of Florida.
Lee, L. F., G. S. Maddala, and R. P. Trost (1979) "Testing for Structural Change by *D*-Methods in Switching Simultaneous Equation Models". *Proceedings of the American Statistical Association*, forthcoming.
Lee, L. F. and R. P. Trost (1978) "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand", *Journal of Econometrics*, 8, 357–382.
Lee, L. F., G. Maddala, and R. Trost (1980) "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity", *Econometrica*, 48, 491–504.
Lerman, S. R. (1977) "Location, Housing, Automobile Ownership and Model to Work: A Joint Choice Model". *Transportation Research Board Record*, No. 610, Washington, D.C.
Lerman, S. and C. Manski (1980) "Information Diffusion in Discrete Choice Contexts". Presented to CEME Conference on Discrete Econometrics.
Lerman, S. and C. Manski (1981) "On the Use of Simulated Frequencies to Approximate Choice Probabilities". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, 305–319. Cambridge: MIT Press.
Li, M. (1977) "A Logit Model of Home Ownership", *Econometrica*, 45, 1081–1097.
Little, R. E. (1968) "A Note on Estimation for Quantal Response Data", *Biometrika* 55, 578–579.
Loikkanen, H. (1982) "A Logit Model of Intra-Urban Mobility". Discussion Paper 22, Labour Institute for Economic Research, Helsinki.
Luce, R. D. (1959) *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
Luce, R. D. (1977) "The Choice Axiom After Twenty Years", *Journal of Mathematical Psychology*, 15, 215–233.
Luce, R. D. and P. Suppes (1965) "Preference, Utility, and Subjective Probability". In *Handbook of Psychology III*, edited by R. Luce, R. Bush, and E. Galanter, 249–410. New York: Wiley.
Maddala, G. S. (1977a) "Self-Selectivity Problem in Econometric Models". In *Applications of Statistics*, edited by K. Krishniah. Amsterdam: North-Holland.
Maddala, G. S. (1977b) "Identification and Estimation Problems in Limited Dependent Variable Models". In *Natural Resources, Uncertainty and General Equilibrium Systems: Essays in Memory of Rafael Lusky*, edited by A. S. Blinder and P. Friedman, 423–450. New York: Academic.
Maddala, G. S. (1978) "Selectivity Problems in Longitudinal Data", *Annals de l'Insee*, 30-1, 423–450.
Maddala, G. S. (1982) *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge Univ. Press.
Maddala, G. S. and L. F. Lee (1976) "Recursive Models with Qualitative Endogenous Variables", *Annals of Economic and Social Measurement*, 5, 525–545.
Maddala, G. and F. Nelson (1974) "Maximum Likelihood Methods for Markets in Disequilibrium", *Econometrica*, 42, 1013–1030.
Maddala, G. S. and F. D. Nelson (1975) "Switching Regression Models with Exogenous and Endogenous Switching". *Proceedings of the American Statistical Association* (Business and Economics Section), 423–426.
Maddala, G. S. and R. Trost (1978) "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand", *Journal of Econometrics*, 8, 357–382.
Maddala, G. S. and R. Trost (1980) "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity", *Econometrica*, 48, 491–503.
Malhotra, N. (1982) "A Comparison of the Predictive Validity of Procedures for Analyzing Binary

Data". Working Paper, Georgia Institute of Technology.
Manski, C. (1975) "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205–228.
Manski, C. (1977) "The Structure of Random Utility Models", *Theory and Decision*, 8, 229–254.
Manski, C. (1981) "Closest Empirical Distribution Estimation". Working Paper, Hebrew Univ.
Manski, C. (1981a) "Structural Models for Discrete Data". *Sociological Methodology*, pp. 58–109.
Manski, C. (1982) "Analysis of Equilibrium Automobile Holdings in Israel with Aggregate Discrete Choice Models". Discussion Paper 82.06, Falk Institute.
Manski, C. and S. Lerman (1977) "The Estimation of Choice Probabilities from Choice-Based Samples", *Econometrica*, 45, 1977–1988.
Manski, C. and D. McFadden (1981) "Alterntative Estimates and Sample Designs for Discrete Choice Analysis". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, pp. 2–50. Cambridge: MIT Press.
Mardia, V. (1970) *Families of Bivariate Distributions*. Connecticut: Hafner.
Marschak, J. (1960) "Binary-Choice Constraints and Random Utility Indicators". In *Mathematical Methods in the Social Sciences*, edited by K. Arrow, S. Karlin, and P. Suppes, pp. 312–329. Stanford University Press.
McFadden, D. (1973) "Conditional Logit Analysis of Qualitative Choice Behavior". In *Frontiers in Econometrics*, edited by P. Zarembka, pp. 105–142. New York: Academic.
McFadden, D. (1974) "The Measurement of Urban Travel Demand", *Journal of Public Economics*, 3, 303–328.
McFadden, D. (1976a) "A Comment on Discriminant Analysis 'Versus' Logit Analysis", *Annals of Economics and Social Measurement*, 5, 511–523.
McFadden, D. (1976b) "Quantal Choice Analysis: A Survey", *Annals of Economic and Social Measurement*, 5, 363–390.
McFadden, D. (1976c) "The Revealed Preferences of a Public Bureaucracy", *Bell Journal*, 7, 55–72.
McFadden, D. (1978a) "Cost, Revenue, and Profit Functions, and the Generalized Linear Profit Function". In *Production of Economics*, edited by M. Fuss and D. McFadden, 3–110. Amsterdam: North-Holland.
McFadden, D. (1978b) "Modelling the Choice of Residential Location". In *Spatial Interaction Theory and Residential Location*, edited by A. Karlquist et al., 75–96. Amsterdam: North-Holland.
McFadden, D. (1979a) "Econometric Net Supply Systems for Firms with Continuous and Discrete Commodities". Working Paper, Department of Economics, Massachusetts Institute of Technology.
McFadden, D. (1979b) "Quantitative Methods for Analysing Travel Behavior of Individuals". *Behavioral Travel Modeling*, edited by D. Hensher and P. Stopher, 279–318. London: Croom Helm.
McFadden, D. (1979c) "Econometric Analysis of Discrete Data". Fisher–Schultz Lecture, Econometric Society, Athens.
McFadden, (1981) "Econometric Models of Probabilistic Choice". In *Structural Analysis of Discrete Data*, edited by C. Manski and D. McFadden, pp. 198–272. Cambridge: MIT Press.
McFadden, D., C. Puig, and D. Kirschner (1977) "Determinants of the Long-Run Demand for Electricity". *Proceedings of the American Statistical Association*, Business and Economics Section, Vol. 1.
McFadden, D. and F. Reid (1975) "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models". *Transportation Research Board Record*, 534, 24–37, Washington, D.C.
McFadden, D., A. Talvitie et al. (1977) "Demand Model Estimation and Validation". Final Report Series, Vol. 5. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.
McFadden, C., W. Tye, and K. Train (1976) "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model". *Transportation Research Board Record* No. 637, pp. 39–45, Washington, D.C.
McKelvey, R. and W. Zovoina (1975) "A Statistical Model for the Analysis of Ordinal Level Dependent Variables", *Journal of Mathematical Sociology*, 4, 103–120.
Miller, L. and R. Radner (1970) "Demand and Supply in U.S. Higher Education", *American Economic Review*, 60, 326–334.
Mitchell, O. and G. Fields (1982) "The Effects of Changing the Budget Constraint: Parametric and Non-Parametric Approaches to Retirement Decisions". Working Paper, Cornell Univ.

Moore, D. H. (1973) "Evaluation of Five Discrimination Procedures for Binary Variables", *Journal of the American Statistical Association*, 68, 399–404.
Morimune, K. (1970) "Comparisons of Normal and Logistic Models in the Bivariate Dichotomous Analysis", *Econometrica*, 47, 957–975.
Moses, L., R. Beals, and M. Levy (1967) "Rationality and Migration in Ghana", *Review of Economics and Statistics*, 49, 480–486.
Mosteller, F. (1968) "Association and Estimation in Contingency Tables", *Journal of the American Statistical Association*, 63, 1–28.
Nelson, F. (1977) "Censored Regression Models with Unobserved Stochastic Censoring Thresholds", *Econometrica*, 6, 309–327.
Nelson, F. and L. Olsen (1978) "Specification and Estimation of a Simultaneous Equation Model with Limited Variables", *International Economic Review*, 19, 685–710.
Nerlove, M. (1978) "Econometric Analysis of Longitudinal Data: Approaches, Problems and Prospects, The Econometrics of Panel Data", *Annals de l'Insee*, 30:1, 7–22.
Nerlove, M. and J. Press (1973) "Univariable and Multivariable Log-Linear and Logistic Models". RAND report No. R-1306-EDA/NIH.
Nerlove, M. and J. Press (1976) "Multivariate and Log Linear Probability Models for the Analysis of Qualitative Data". Discussion Paper, Department of Economics, Northwestern University.
Newey, W. (1982) "Generalized Method of Moments Specification Testing". Working Paper, Dept. of Economics, MIT.
Nickell, S. (1979) "Estimating the Probability of Leaving Employment", *Econometrica*, 47, 1249–1266.
Oliveira, J. T., de (1958) "External Distributions". *Revista de Faculdada du Ciencia*, *Lisboa* (Serie A) 7, 215–227.
Olsen, R. (1978a) "Comment on 'The Effect of Unions of Earnings and Earnings on Unions: A Mixed Logit Approach'", *International Economic Review*, 19, 259–261.
Olsen, R. (1978b) "Tests for the Presence of Selectivity Bias and Their Relation to Specifications of Functional Form and Error Distribution". Working Paper No. 812, revised, Yale University.
Olsen, R. (1980a) "A Least Squares Correction for Selectivity Bias", *Econometrica*, 48, 1815–1820.
Olsen, R. (1980b) "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator". Working Paper, Institute for Social and Policy Studies, Yale University.
Olsen, R. (1980c) "Estimating the Effect of Child Mortality on the Number of Births". Economic Growth Center, Yale University.
Palepu, (1982) *A Stochastic Model of Acquisitions*. Ph.D. thesis, Sloan School of Management, MIT.
Plackett, R. L. (1974) *The Analysis of Categorical Data*. London: Charles Griffin.
Poirier, D. J. (1977) "The Determinants of Home Buying" in *The New Jersey Income-Maintenance Experiment*". In Vol. II, *Expenditures, Health and Social Behavior, and the Quality of the Evidence*, edited by H. W. Watts and A. Rees, pp. 73–91. New York: Academic.
Poirier, D. J. (1978) "A Switching Simultaneous Equation Model of Physician Behavior in Ontario". In *Structural Analysis of Discrete Data: with Econometric Applications*, edited by D. McFadden and W. Manski, pp. 392–421. Cambridge: MIT Press.
Poirier, D. J. and P. Ruud (1980) "On the Appropriateness of Endogenous Switching", *J. Econometrics*, forthcoming.
Poirier, D. J. and P. Ruud (1981) "Conditional Minimum Distance Estimation and Autocorrelation in Limited Dependent Variable Models". Working paper, Univ. of Toronto.
Pollakowski, H. (1974) "The Effects of Local Public Service on Residential Location Decision: An Empirical Study of the San Francisco Bay Area". Ph.D. Dissertation, Department of Economics, University of California, Berkeley.
Pollakowski, H. (1980) *Residential Location and Urban Housing Markets*. Lexington, Mass.: D.C. Heath.
Poterba, J. and L. Summers (1982) "Unemployment Benefits, Labor Market Transitions, and Spurious Flows: A Multinomial Logit Model with Errors in Classification", Working Paper, MIT.
Powell, J. (1982) "Least Absolute Deviations Estimation for Censored and Truncated Regression Models". Ph.D. thesis, Department of Economics, Stanford Univ.
Pratt, J. (1977) "Concavity of the Log Likelihood of a Model of Ordinal Categorical Regression". Working Paper, Harvard Univ.
Press, J. and S. Wilson (1978) "Choosing Between Logistic Regression and Discriminant Analysis", *JASA*, 73, 699–705.

Quandt, R. E. (1956) "Probabilistic Theory of Consumer Behavior", *Quarterly Journal of Economics*, 70, 507–536.
Quandt, R. E. (1968) "Estimation of Model Splits", *Transportation Research*, 2, 41–50.
Quandt, R. E. (1970) *The Demand for Travel*. London: D.C. Heath.
Quandt, R. E. (1972) "A New Approach to Estimating Switching Regressions", *Journal of the American Statistical Association*, 67, 306–310.
Quandt, R. E. (1978) "Test of the Equilibrium vs Disequilibrium Hypothesis", *International Economic Review*, 19, 435–452.
Quandt, R. E. (1981) "Econometric Disequilibrium Models". Working Paper, Princeton Univ.
Quandt, R. E. (1982) "A Bibliography of Quantity Rationing and Disequilibrium Models", Working Paper, Princeton Univ.
Quandt, R. E. and W. Baumol (1966) "The Demand for Abstract Travel Modes: Theory and Measurement", *Journal of Regional Science*, 6, 13–26.
Quandt, R. E. and J. Ramsey (1978) "Estimating Mixtures of Normal Distributions and Switching Regressions", *Journal of the American Statistical Association*, 71, 730–752.
Quigley, J. (1976) "Housing Demand in the Short-Run: An Analysis of Polytomous Choice", *Explorations in Economic Research*, 3, 76–102.
Radner, R. and L. Miller (1975) *Demand and Supply in U.S. Higher Education*. New York: McGraw-Hill.
Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*. New York: Wiley.
Richards, M. G. and M. Ben-Akiva (1974) "A Simultaneous Destination and Mode Choice Model for Shopping Trips", *Transportation*, 3, 343–356.
Robinson, P. (1982) "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables", *Econometrica*, 50, 27–41.
Rosen, H. and K. Small (1979) "Applied Welfare Economics with Discrete Choice Models". Working Paper 319, National Bureau of Economic Research.
Rothenberg, T. (1982) "Approximating the Distributions of Econometric Estimators and Test Statistics", manuscript, Univ. of California, Berkeley.
Ruud, P. (1982) "A Score Test of Consistency". Working Paper, Dept. of Economics, Univ. of California, Berkeley.
Sattath, S. and A. Tversky (1977) "Additive Similarity Trees", *Psychometrika*, 42, 319–345.
Schultz, T. P. (1975) "The Determinants of Internal Migration in Venezuela: An Application of the Polytomous Logistic Model". Presented at the Third World Congress of the Econometric Society, Toronto, Canada.
Sickles, R. C. and P. Schmidt (1978) "Simultaneous Equation Models with Truncated Dependent Variables: A Simultaneous Tobit Model", *Journal of Economics and Business*, 31, 11–21.
Sjoberg, L. (1977) "Choice Frequency and Similarity", *Scandinavian Journal of Psychology*, 18, 103–115.
Small, K. (1981) "Ordered Logit: A Discrete Choice Model with Proximate Covariance Among Alternatives". Working Paper, Department of Economics, Princeton Univ.
Smith, K., Savin, N. and J. Robertson (1982) "A Monte Carlo Study of Maximum Likelihood and Minimum Chi-Square Estimators in Logit Analysis". Working Paper, USDA, Forest Service, Pacific Southwest Forest and Range Experiment Station.
Spilerman, S. (1972) "Extensions of the Mover-Stayer Model", *American Journal of Sociology*, 78, 599–626.
Stewman, S. (1976) "Markov Models of Occupational Mobility: Theoretical Development and Empirical Support", *Journal of Mathematical Sociology*, 4, 201–245.
Swartz, C. (1976) "Screening in the Labor Market: A Case Study". Ph.D. dissertation, University of Wisconsin, Madison.
Tachibanaki, T. (1981) Education, Occupation, Hierarchy, and Earnings: A Recursive Logit Approach for Japan". Research paper, Stanford Univ.
Talvities, A. (1972) "Comparison of Probabilistic Modal-Choice Models; Estimation Methods and System Inputs". Highway Research Board Record 392, pp. 111–120.
Theil, H. (1969) "A Multinomial Extension of the Linear Logit Model", *International Economic Review*, 10, 251–259.

Theil, H. (1970) "On the Estimation of Relationships Involving Qualitative Variables", *American Journal of Sociology*, 76, 103–154.
Thurstone, L. (1927) "A Law of Comparative Judgement", *Psychological Review*, 34, 273–286.
Tobin, J. (1958) "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24–36.
Train, K. (1978) "A Validation Test of a Disaggregate Mode Choice Model", *Transportation Research*, 12, 167–174.
Train, K., and D. McFadden (1978) "The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models", *Transportation Research*, 12, 349–353.
Train, K. and M. Lohrer (1982) "Vehicle Ownership and Usage: An Integrated System of Disaggregate Demand Models". Final Report, Cambridge Systematics.
Trost, R. and L. Lee (1982) "Technical Training and Earnings: A Polychotomous Choice Model with Selectivity". Working Paper, Dept. of Economics, Univ. of Florida.
Tversky, A. (1972a) "Choice by Elimination", *Journal of Mathematical Psychology*, 9, 341–367.
Tversky, A. (1972b) "Elimination by Aspects: A Theory of Choice", *Psychological Review*, 79, 281–299.
Tversky, A. and S. Sattath (1979) "Preference Trees", *Psychology Review*, 86, 542–573.
Vuong, Q. (1982) "Probability Feedback in a Recursive System of Logit Models: Estimation". Working Paper, California Institute of Technology.
Walker, S. and D. Duncan (1967) "Estimation of the Probability of an Event as a Function of Several Independent Variables", *Biometrika*, 54, 167–179.
Warner, S. (1962) *Stochastic Choice of Mode in Urban Travel*. Evanston: North-Western University Press.
Warner, S. L. (1963) "Multivariate Regression of Dummy Variates under Normality Assumptions", *Journal of the American Statistical Association*, 58, 1054–1063.
Westin, R. B. (1974) "Predictions from Binary Choice Models", *Journal of Econometrics*, 2, 1–16.
Westin, R. B. and D. W. Gillen (1978) "Parking Location and Transit Demand: A Case Study of Endogenous Attributes in Disaggregate Mode Choice Functions", *Journal of Econometrics*, 8, 75–101.
Williams, H. (1977) "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", *Environment Planning*, A9, 285–344.
Willis, R. and S. Rosen (1979) "Education and Self-Selection", *Journal of Political Economy*, 87, 507–536.
Winston, C. (1981) "A Disaggregate Model of the Demand for Intercity Freight Transport", *Econometrica*, 49, 981–1006.
Yatchew, A. (1980) *Design of Econometric Choice Models*. Ph.D. Thesis, Harvard Univ.
Yellot, J. (1977) "The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgement, and the Double Exponential Distribution", *Journal of Mathematical Psychology*, 15, 109–144.
Zellner, A. and T. Lee (1965) "Joint Estimation of Relationships Involving Discrete Random Variables", *Econometrica*, 33, 382–394.