

CHAPTER 2

MODELS WITHOUT IIA:

SEQUENTIAL LOGIT, GENERALIZED LOGIT, AND PROBIT MODELS

Introduction

The empirical work in this volume employs the multinomial logit model and, occasionally, variants of this model--the maximum model and the sequential logit model. The preceding chapter has established empirically that the multinomial logit model can provide an adequate fit to observed data, and that the IIA property of the multinomial logit model cannot be rejected. Nevertheless, there is considerable interest in alternatives to or extensions of this model for situations where the IIA property is unpalatable, or where the MNL model itself is cumbersome to estimate or manipulate.

Ad Hoc Choice Models

Several *ad hoc* alternatives to the MNL model have been suggested in transportation literature. McFadden (1974) introduced the maximum model, which, for joint choice of, say, mode *m* and destination *d*, would take the form

$$(1) \quad P_d = e^{\max_m v_{md}} / \sum_c e^{\max_m v_{mc}} \quad ;$$

$$(2) \quad P_{m|d} = e^{v_{md}} / \sum_n e^{v_{nd}} \quad ,$$

where v_{md} is the "mean" utility of the alternative (*m,d*), $P_{m|d}$ is the conditional probability of choosing mode *m* given destination *d*, and P_d is the marginal probability of destination *d*.

This model can be interpreted as corresponding to a tree decision structure of the form depicted in Figure 10, with mode choice, conditioned on destination, obeying a conventional MNL model, and with destination choice of the MNL form with the utility--or "inclusive value"--of each branch given by the maximum of the mean utilities of the alternatives in the branch.

McFadden (1975) has shown that the maximum model is not derivable, in general, from a population of utility-maximizing consumers.¹

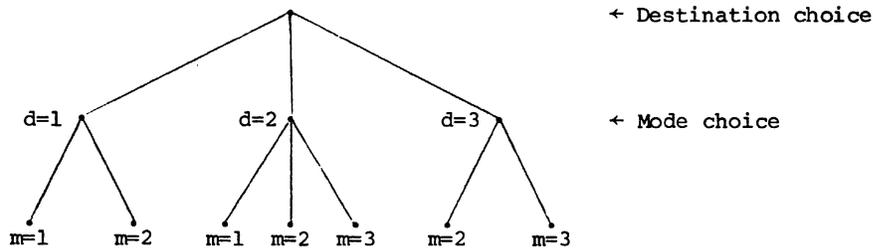


Figure 10

¹McFadden (1974) also suggested the cascade model, in which the desirability of a "branch" of alternatives is given by a probability-weighted sum of mean utilities of the alternatives. However, this model violates a basic axiom of choice theory, implied by utility maximization, that adding an alternative to a branch cannot lower the desirability of that branch--because the new alternative need be chosen only if it is better than the alternatives previously available. Hence, use of this model is not recommended.

A model that is useful for testing particular variable specifications, including the IIA property, is the universal, or "mother," logit form suggested by McFadden (1975). Consider an arbitrary choice model $P(i | C, z)$, where i indexes an alternative, C is the set of available alternatives, and z is a vector of variables describing the environment of the choice. Defining $v_i(z) = \log P(i | C, z)$ we have, trivially, the apparent MNL form:

$$(3) \quad P(i | C, z) = e^{v_i(z)} / \sum_{j \in C} e^{v_j(z)} ,$$

which differs from the MNL model considered in the introductory chapters only in that the "mean utility" $v_i(z)$ is a function of the attributes of all alternatives. The universal logit model is defined by taking $v_i(z)$ to be a linear-in-parameters expansion in known functions of z . This model is again generally inconsistent with utility maximization when the utility of an alternative is required to depend solely on the observed and unobserved attributes of the alternative.

McLynn (1973) has suggested a third *ad hoc* form, the fully competitive model, in which choice probabilities are given by a mapping of a vector of MNL probabilities into the unit simplex. This model is again inconsistent with utility maximization, and has the further drawback of retaining the essential structural restriction--called order-independence¹--that makes the IIA property unpalatable.

In terms of using behavioral axioms to restrict the structure of choice models in reasonable ways, it is desirable to derive choice models directly from utility maximization rather than use *ad hoc* models. This route in the past has been encumbered by problems of computational and analytic intractability. However, recent developments have made several relatively general utility-maximization models practical for empirical application. We outline the structure of these models. However, we have not applied these models empirically in the current demand study.

¹McFadden (1975) discusses the property of order-independence and its consequences for choice models.

The Multinomial Logit Model for Joint Choice

We begin a discussion of alternatives to the MNL model with a reformulation of this model in a way that suggests generalizations. Joint choice of mode, destination, and auto can be interpreted (without loss of generality) as occurring in a tree decision structure of the form depicted in Figure 11. An MNL choice model for this joint choice will typically have the form

$$(4) \quad P_{mda} = e^{v_{mda}} / \sum_{n,c,b} e^{v_{ncb}} \quad ,$$

where $m = \text{mode}$; $d = \text{destination}$; $a = \text{auto availability}$, and $v_{mda} = \alpha x_{mda} + \beta y_{da} + \gamma z_a = \text{utility}$. Letting $P_{m|da}$ denote a conditional choice probability and P_m denote a marginal choice probability, one derives from (1) the formulae:

$$(5) \quad P_{m|da} = e^{v_{mda}} / \sum_n e^{v_{nda}} = e^{\alpha x_{mda}} / \sum_n e^{\alpha x_{nda}} \quad ;$$

$$(6) \quad P_{d|a} = \sum_n e^{v_{nda}} / \sum_{n,c} e^{v_{nca}} = \sum_n e^{\alpha x_{nda} + \beta y_{da}} / \sum_{n,c} e^{\alpha x_{nca} + \beta y_{ca}} \quad ;$$

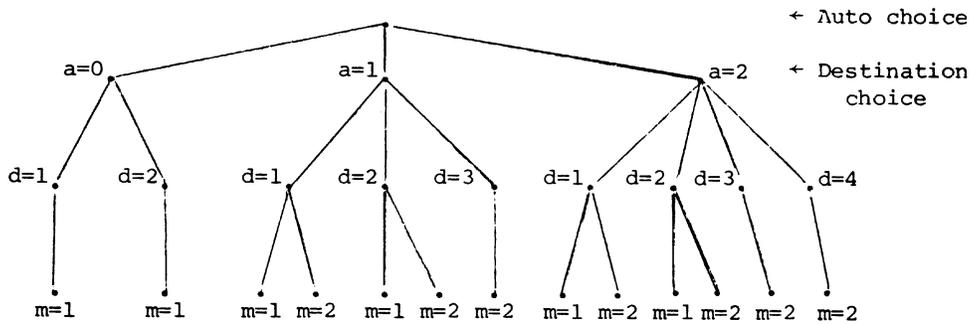


Figure 11

$$(7) \quad P_a = \sum_{n,c} e^{v_{nca}} / \sum_{n,c,b} e^{v_{ncb}} = \sum_{n,c} e^{\alpha x_{nca} + \beta y_{ca} + \gamma z_a} / \sum_{n,c,b} e^{\alpha x_{ncb} + \beta y_{cb} + \gamma z_b} \quad ;$$

$$(8) \quad I_{da} = \log \sum_n e^{\alpha x_{nda}} \quad ;$$

$$(9) \quad J_a = \log \sum_{n,c} e^{\alpha x_{nca} + \beta y_{ca}} = \log \sum_c e^{I_{ca} + \beta y_{ca}} \quad .$$

Then the choice probabilities can be written

$$(10) \quad P_{mda} = P_{m|da} P_{d|a} P_a \quad ;$$

$$(11) \quad P_{m|da} = e^{\alpha x_{mda}} / e^{I_{da}} \quad ;$$

$$(12) \quad P_{d|a} = e^{I_{da} + \beta y_{da}} / \sum_c e^{I_{ca} + \beta y_{ca}} = e^{I_{da} + \beta y_{da}} / e^{J_a} \quad ;$$

$$(13) \quad P_a = e^{J_a + \gamma z_a} / \sum_b e^{J_b + \gamma z_b} \quad .$$

Discussion of the historical development of the MNL model can be found in McFadden (1976b); the properties of the model, including its derivation from the theory of individual utility maximization, are given in McFadden (1973).

The Sequential MNL Model

We next define the sequential, or nested, MNL model. A typical sequential model differs from the joint MNL model solely in that the coefficients of inclusive values are not constrained to equal one. Hence, the joint MNL model is a linear restriction on any of the sequential models. Specifically, a sequential model is defined by

$$(14) \quad P_{mda} = P_{m|da} P_{d|a} P_a \quad ;$$

$$(15) \quad P_{m|da} = e^{\alpha x_{mda}} / \sum_n e^{\alpha x_{nda}} \quad ;$$

$$(16) \quad I_{da} = \log \sum_n e^{\alpha x_{nda}} \quad ;$$

$$(17) \quad P_{d|a} = e^{\theta I_{da} + \beta y_{da}} / \sum_c e^{\theta I_{ca} + \beta y_{ca}} \quad ;$$

$$(18) \quad J_a = \log \sum_c e^{\theta I_{ca} + \beta y_{ca}} \quad ;$$

$$(19) \quad P_a = e^{\lambda J_a + \gamma z_a} / \sum_b e^{\lambda J_b + \gamma z_b} \quad .$$

When $\theta = \lambda = 1$, this model is identical to the joint MNL model. More generally, when $\theta \neq 1$, equations (17) and (18) differ in the two models, and when $\lambda \neq 1$, equation (19) differs in the two models.

The sequential model was introduced by Domencich and McFadden (1975), and studied by Ben-Akiva (1973). We shall demonstrate that, under specified restrictions on the coefficients of inclusive values, the sequential model is consistent with utility maximization. To do this we introduce a family of choice models derived from utility maximization that are of interest in their own right and that contain the sequential models as a special case.

The Generalized Extreme Value Model

McFadden (1977b) has recently proposed a family of generalized extreme value (GEV) choice models that allow a general pattern of dependence among alternatives and yield a closed form for the choice probabilities. The following result characterizes the family:

Theorem: Suppose $G(y_1, \dots, y_J)$ is a non-negative, homogeneous-of-degree-one function of $(y_1, \dots, y_J) \geq 0$. Suppose $\lim_{y_i \rightarrow +\infty} G(y_1, \dots, y_J) = +\infty$ for $i = 1, \dots, J$.

Suppose for any distinct (i_1, \dots, i_k) from $\{1, \dots, J\}$, $\partial^k G / \partial y_{i_1} \dots \partial y_{i_k}$ is nonnegative if k is even and non-positive if k is odd. Then,

$$(20) \quad P_i = e^{v_i} G_i(e^{v_1}, \dots, e^{v_J}) / G(e^{v_1}, \dots, e^{v_J})$$

defines a choice model that is consistent with utility maximization.

Proof: Consider the function

$$(21) \quad F(\varepsilon_1, \dots, \varepsilon_J) = \exp \left\{ -G(e^{-\varepsilon_1}, \dots, e^{\varepsilon_J}) \right\} .$$

If $\varepsilon_i \rightarrow -\infty$, then $G \rightarrow +\infty$, implying $F \rightarrow 0$. If $(\varepsilon_1, \dots, \varepsilon_J) \rightarrow +\infty$, then $G \rightarrow 0$, implying $F \rightarrow 1$. Define, recursively, $Q_1 = G_1$ and $Q_k = Q_{k-1} G_k - \partial Q_{k-1} / \partial y_k$. Then Q_k is a sum of signed terms, with each term a product of cross derivatives of G of various orders. Suppose each signed term in Q_{k-1} is nonnegative. Then $Q_{k-1} G_k$ is nonnegative. Further, each term in $\partial Q_{k-1} / \partial y_k$ is non-positive, because one of the derivatives within each term has increased in order, changing from even to odd or vice versa, with a hypothesized change in sign. Hence, each term in Q_k is nonnegative. By induction, Q_k is nonnegative for $k = 1, \dots, J$.

Differentiating F , $\partial F/\partial \epsilon_1 = e^{-\epsilon_1} Q_1 F$. Suppose

$$\partial^{k-1} F/\partial \epsilon_1, \dots, \partial \epsilon_{k-1} = e^{-\epsilon_1} \dots e^{-\epsilon_{k-1}} Q_{k-1} F \quad .$$

Then,

$$\begin{aligned} \partial^k F/\partial \epsilon_1, \dots, \partial \epsilon_k &= e^{-\epsilon_1} \dots e^{-\epsilon_k} \{Q_{k-1} G_k F - F \partial Q_{k-1} / \partial y_k\} \\ &= e^{-\epsilon_1} \dots e^{-\epsilon_k} Q_k F \quad . \end{aligned}$$

By induction,

$$\partial^J F/\partial \epsilon_1, \dots, \partial \epsilon_J = e^{-\epsilon_1} \dots e^{-\epsilon_J} Q_J F \geq 0 \quad .$$

Hence, F is a cumulative distribution function. When $\epsilon_j = +\infty$ for $j \neq i$, $F = \exp \{-a_i e^{-\epsilon_i}\}$, where $a_i = G(0, \dots, 0, 1, 0, \dots, 0)$. [i^{th} place] This is the extreme value (Weibull, Gumbel) distribution. Hence, F is a multivariate extreme value distribution.

Suppose a population has utilities $u_i = V_i + \epsilon_i$, where $(\epsilon_1, \dots, \epsilon_J)$ is distributed F . Then, the probability that the first alternative is selected satisfies

$$\begin{aligned} (22) \quad P_1 &= \int_{\epsilon=-\infty}^{+\infty} F_1(\epsilon, V_1 - V_2 + \epsilon, \dots, V_1 - V_J + \epsilon) d\epsilon \\ &= \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon} G_1(e^{-\epsilon}, e^{-\epsilon-V_1+V_2}, \dots, e^{-\epsilon-V_1+V_J}) \exp \left\{ -G(e^{-\epsilon}, e^{-\epsilon-V_1+V_2}, \dots, e^{-\epsilon-V_1+V_J}) \right\} d\epsilon \\ &= \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon} G_1(e^{V_1}, e^{V_2}, \dots, e^{V_J}) \exp \left\{ -e^{-\epsilon} e^{-V_1} G(e^{V_1}, e^{V_2}, \dots, e^{V_J}) \right\} d\epsilon \\ &= e^{V_1} G_1(e^{V_1}, e^{V_2}, \dots, e^{V_J}) / G(e^{V_1}, \dots, e^{V_J}) \quad , \end{aligned}$$

where the third equality uses the homogeneity-of-degree-one of G , and consequent homogeneity-of-degree-zero of G_1 . Because this argument can be applied to any alternative, the theorem is proved. Q.E.D.

The special case $G(y_1, \dots, y_J) = \sum_{j=1}^J y_j$ yields the MNL model. An example of a more general G function satisfying the hypotheses of the theorem is

$$(23) \quad G(y) = \sum_{m=1}^M a_m \left(\sum_{i \in B_m} y_i \frac{1}{1-\sigma_m} \right)^{1-\sigma_m}$$

where $B_m \subseteq \{1, \dots, J\}$, $\bigcup_{m=1}^M B_m = \{1, \dots, J\}$, $a_m > 0$, and $0 \leq \sigma_m < 1$. The parameter σ_m is an index of the similarity of the unobserved attributes of alternatives in B_m . The choice probabilities for this function satisfy

$$(24) \quad P_i = \sum_{m: i \in B_m} e^{\frac{V_i}{1-\sigma_m}} a_m \left(\sum_{j \in B_m} e^{\frac{V_j}{1-\sigma_m}} \right)^{-\sigma_m} \sum_{n=1}^M a_n \left(\sum_{k \in B_n} e^{\frac{V_k}{1-\sigma_n}} \right)^{1-\sigma_n}.$$

Functions of the form in (23) can also be nested to yield a wider class satisfying the theorem hypotheses. For example, the function

$$(25) \quad G = \sum_{q=1}^Q a_q \left(\sum_{m \in D_q} \left(\sum_{j \in B_m} y_j \frac{1}{1-\sigma_m} \right)^{\frac{1-\sigma_m}{1-\delta_q}} \right)^{1-\delta_q}$$

where B_m is defined as in (20) and $D_q \subseteq \{1, \dots, M\}$, satisfies the hypotheses provided $1 > \sigma_m \geq \delta_q \geq 0$.

The reader's understanding of (24) may be aided by a simple example. Suppose an individual must choose among three modes, auto ($m = 1$), bus ($m = 2$), and rail ($m = 3$), and suppose that the unobserved attributes of bus and rail are correlated. Then, G may have the form

$$(26) \quad G(y) = a_1 y_1 + a_2 \left(y_2^{\frac{1}{1-\sigma}} + y_3^{\frac{1}{1-\sigma}} \right)^{1-\sigma},$$

where, from the proof of the theorem, one can see that the size of σ is related to the correlation of the unobserved attributes of alternatives 2 and 3. From (24), the choice probabilities from $C = \{1,2,3\}$ are

$$(27) \quad P(1 | C) = a_1 e^{V_1} \left\{ a_1 e^{V_1} + a_2 \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}} \right)^{1-\sigma} \right\}$$

and

$$(28) \quad P(2 | C) = a_2 e^{\frac{V_2}{1-\sigma}} \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}} \right)^{-\sigma} \left\{ a_1 e^{V_1} + a_2 \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}} \right)^{1-\sigma} \right\}.$$

When alternative 1 is unavailable and the choice set is $D = \{2,3\}$, the choice probability is

$$(29) \quad P(2 | D) = e^{\frac{V_2}{1-\sigma}} \left\{ e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}} \right\}^{-1}.$$

As the correlation between the unobserved components of modes 2 and 3 becomes large, and σ approaches one,

$$(30) \quad \lim_{\sigma \rightarrow 1} \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}} \right)^{1-\sigma} = e^{\max(V_2, V_3)}.$$

Then, the choice probabilities have the following limits:

$$(31) \quad P(1 | C) \rightarrow a_1 e^{V_1} / \left\{ a_1 e^{V_1} + a_2 e^{\max(V_2, V_3)} \right\};$$

$$(32) \quad P(2 | C) = \begin{cases} 1 - P(1|C) & \text{if } V_2 > V_3 \\ 1/2 (1 - P(1|C)) & \text{if } V_2 = V_3 \\ 0 & \text{if } V_2 < V_3 \end{cases} ;$$

$$(33) \quad P(2 | D) = \begin{cases} 1 & \text{if } V_2 > V_3 \\ 1/2 & \text{if } V_2 = V_3 \\ 0 & \text{if } V_2 < V_3 \end{cases} .$$

The limiting probability (31) has the form associated with the *ad hoc* maximum model marginal decision among branches. However, the conditional choice probability for alternatives within a branch does not match the form for this choice postulated in the maximum model.

Relation of Sequential MNL and GEV Model

The choice probabilities corresponding to (25) can be specialized to the sequential MNL model described in (14) - (19), as we shall now show. This result establishes that sequential MNL models are consistent with individual utility maximization for appropriate parameter values, and that the coefficients of inclusive prices can be used to obtain estimates of the similarity parameters σ and δ . It is hence possible to estimate some GEV choice models using sequential MNL models and inclusive prices. Further, the GEV class provides a generalization containing alternative sequential MNL models, and could be estimated directly to test the presence of a sequential or tree structure.

To obtain the sequential model (14) - (19) from (25), index alternatives by mda for mode m, destination d, and auto availability a, and specialize (25) to the form

$$(34) \quad G = \sum_a \left(\sum_d \left(\sum_m y_{mda}^{\frac{1}{\theta\lambda}} \right)^\theta \right)^\lambda, \quad (0 < \theta \leq \lambda \leq 1) \quad .$$

Assume $V_{mda} = \theta\lambda\alpha'x_{mda} + \lambda\beta'y_{da} + \gamma'z_a$. Then (25) yields

$$(35) \quad P_{mda} = \frac{e^{V_{mda}/\theta\lambda}}{\sum_n e^{V_{nda}/\theta\lambda}} \frac{\left(\sum_n e^{V_{nda}/\theta\lambda} \right)^\theta}{\sum_c \left(\sum_n e^{V_{nca}/\theta\lambda} \right)^\theta} \frac{\left(\sum_c \left(\sum_n e^{V_{nca}/\theta\lambda} \right)^\theta \right)^\lambda}{\sum_b \left(\sum_c \left(\sum_n e^{V_{ncb}/\theta\lambda} \right)^\theta \right)^\lambda}$$

$$= \frac{e^{\alpha'x_{mda}}}{\sum_n e^{\alpha'x_{nda}}} \frac{e^{\beta'y_{da} + \theta I_{da}}}{\sum_c e^{\beta'y_{ca} + \theta I_{ca}}} \frac{e^{\gamma'z_a + \lambda J_a}}{\sum_b e^{\gamma'z_b + \lambda J_b}},$$

where

$$I_{da} = \log \sum_n e^{\alpha' x_{nda}} ;$$

$$J_a = \log \sum_c e^{\beta' y_{ca} + \theta I_{ca}} .$$

This is precisely the sequential model (14) - (19) . Hence, we have established that a sufficient condition for a sequential model to be consistent with individual utility maximization is that the coefficient of each inclusive price not exceed one, and that the coefficients of inclusive prices not decline as one moves up the tree to more inclusive nodes; i.e., $0 < \theta \leq \lambda \leq 1$.¹ McFadden (1976e) has shown that when the inclusive price coefficient in the sequential model exceeds two, the model is inconsistent with individual utility maximization. Hence, some limits on θ and λ are also necessary. When the sufficient condition $0 < \theta \leq \lambda \leq 1$ is satisfied, $\sigma = 1 - \theta$ is an index of the similarity of alternative modes, while $\delta = 1 - \lambda$ is an index of the similarity of alternative destinations.

¹The preceding demnstration for three-level trees is readily generalized to trees of any depth. The simplest proof is by induction.

The Multinomial Probit Model

The multinomial probit model (MNP) is an alternative to the logit model, which has the advantage of considerable flexibility in the description of the error structure, permitting deviations from the IIA or order-independence restrictions when they are unwarranted. Descriptions of the model formulation from utility maximization theory are given in McFadden (1974) and Domencich and McFadden (1975). The primary drawback of this model has been computational intractability when the number of alternatives exceeds three or four. Recent developments in the approximation of MNP probabilities have reduced these computational barriers, making the model potentially practical.

Consider again the joint choice of auto availability a , destination d , and mode m . A choice model can be obtained by assuming that each alternative has a utility function $u_{mda} = v_{mda} + \lambda_{mda} + \eta_{da} + v_a$, where λ_{mda} , η_{da} , and v_a are unobserved effects summarizing the influence of unobserved attributes and taste variations. Assume λ_{mda} , η_{da} , v_a to be jointly distributed normally across alternatives. If each individual maximizes utility, the proportion of the population choosing mda is

$$(36) \quad P_{mda} = \int_{\epsilon_{mda}=-\infty}^{+\infty} \dots \int_{\epsilon_{ncb}=-\infty}^{V_{mda}-V_{ncb}+\epsilon_{mda}} \dots \int_{\epsilon_{MDA}=-\infty}^{V_{mda}-V_{MDA}+\epsilon_{mda}} n(\epsilon;0,\Omega) d\epsilon \quad ,$$

where the number of integrals equals the number of alternatives, $n(\epsilon;0,\Omega)$ is the multivariate normal density with mean vector 0 and covariance matrix Ω , and $\epsilon_{mda} = \lambda_{mda} + \eta_{da} + v_a$, with the joint normal distribution of λ , η , and v determining Ω .

The MNP model generalizes a classic model of Thurstone (1927) for binary choice. Bock and Jones (1969) applied the model to the three-alternative case. The model was suggested for transportation analysis by Domencich and McFadden (1975), and first applied to transportation data by Hausman and Wise (1976). The MNP model is conceptually appealing because it allows consideration of stochastic components for tastes and unobserved attributes within an alternative, and provides a way of specifying the structure of dependence between alternatives. However, MNP choice probabilities can be expressed exactly only as a multivariate or iterated integral of dimension $J - 1$, where J is the number of alternatives. Exact calculation by numerical integration is very fast for $J = 2$ or 3 , moderately costly for $J = 4$, and impractical on a large scale for J

≥ 5 . One of the more effective direct numerical integration methods, adapted for transportation applications is due to Hausman and Wise (1976).

Two recent contributions have provided techniques for approximating MNP choice probabilities at moderate cost. This has made MNP a practical alternative for many transportation applications. The first method, due to Manski (1976), applies a Monte Carlo procedure directly to the utilities of alternatives. Suppose $J + 1$ alternatives, with utilities $U_i = V_i + \epsilon_i$, where $(\epsilon_1, \dots, \epsilon_{J+1})$ is multivariate normal with zero means and covariance matrix $\Sigma = (\sigma_{ij})$. For given values of V_i , vectors $(\epsilon_1, \dots, \epsilon_{J+1})$ from the multivariate distribution can be drawn, and the frequency with which utility is maximized at alternative i recorded. These frequencies approximate the exact MNP probabilities when the number of Monte Carlo repetitions is large. Because this method involves repetitive simple calculations, it can be programmed in computer assembly language to operate quite efficiently. The approach is appealing in its generality--any joint distribution of the unobserved effects can be assumed. In practice, the method is most effective when a relatively good initial approximation to the frequencies is available.

The second approximation method, due to Daganzo, Bouthelier, and Sheffi (1976), uses a procedure suggested by Clark (1961) to approximate the maximum of bivariate normal variables by a normal variable. When the correlation of the variables is nonnegative, this approximation is accurate within a few percent. Suppose $J + 1$ alternatives, with utilities $U_i = V_i + \epsilon_i$ and $(\epsilon_1, \dots, \epsilon_{J+1})$ distributed multivariate normal, zero means and covariance matrix Σ . The probability that the first alternative is chosen is then

$$\begin{aligned}
 (37) \quad P_1 &= \text{Prob} [V_1 + \epsilon_1 > V_j + \epsilon_j \text{ for } j = 2, \dots, J + 1] \\
 &= \text{Prob} [V_1 - V_{J+1} + \epsilon_1 - \epsilon_{J+1} > V_j - V_{J+1} + \epsilon_j - \epsilon_{J+1} \\
 &\quad \text{for } j = 2, \dots, J \text{ and } V_1 - V_{J+1} + \epsilon_1 - \epsilon_{J+1} > 0] \ .
 \end{aligned}$$

Define $v_j = V_j - V_{J+1}$ and $y_i = \epsilon_i - \epsilon_{J+1}$. Then, (y_1, \dots, y_J) is multivariate normal with mean zero and covariance matrix $\Omega = (\omega_{ij})$, where $\omega_{ij} = \sigma_{ij} + \sigma_{J+1, J+1} - \sigma_{i, J+1} - \sigma_{j, J+1}$. Hence,

$$\begin{aligned}
 (38) \quad P_1 &= \text{Prob} [v_1 + y_1 > 0 \text{ and } v_1 + y_1 > v_j + y_j \text{ from } j = 2, \dots, J] \\
 &= \int_{y_1 = -v_1}^{\infty} n_1(y_1) N_{(1)}((v_1 - v_j + y_1) | y_1) dy_1 \quad ,
 \end{aligned}$$

where $n_{Y(X)}(y | x)$ denotes the normal density for the vector of variables indexed by Y , conditioned on the vector of variables indexed by X ; $N_{Y(X)}(y | x)$ denotes the corresponding cumulative distribution function,

$$N_{Y(X)}(y | x) = \int_{-\infty}^y n_{Y(X)}(y' | x) dy' \quad ; \text{ and } n_Y(y) \text{ is the marginal density of the}$$

variables indexed by Y .¹ The form (26), involving J integrals, is the basis for exact calculations of P_1 . Alternately, write

$$(39) \quad P_1 = \text{Prob} [v_1 + y_1 > 0 \text{ and } v_1 + y_1 > \max_{j=2, \dots, J} (v_j + y_j)] \quad .$$

The Clark method considers trivariate normal random variables (X_1, X_2, X_3) , and approximates the bivariate distribution of $(X_1, \max(X_2, X_3))$ by a bivariate normal distribution with the same first and second moments. The approximation rests on the fact that these moments for $(X_1, \max(X_2, X_3))$ can be calculated exactly in a straightforward manner. Applied recursively to the expression

$$(40) \quad y_0 = \max(v_2 + y_2, \max(v_3 + y_3, \dots, \max(v_{J-1} + y_{J-1}, v_J + y_J, \dots))) \quad ,$$

the method allows the distribution of (y_1, y_0) to be approximated by a bivariate normal distribution $n_1(y) n_{0(1)}(y_0 | y_1)$, so that (38) is approximated by the univariate integral

¹As a shorthand, the set of all indices, or the set of all indices excluding those on which a distribution is conditioned, are omitted. Thus $N_{(1)}$ means $N_{2, \dots, J(1)}$.

$$(41) \quad P_1 = \int_{y_1 = -v_1}^{\infty} n_1(y_1) N_{0(1)}(v_1 + y_1 | y_1) dy \quad ,$$

where

$$N_{0(1)}(y_0 | y_1) = \int_{-\infty}^{y_0} n_{0(1)}(y_0' | y_1) dy_0' \quad .$$

Thus, an MNP choice probability for $J + 1$ alternatives is approximated by a univariate integral involving a univariate normal density and univariate normal cumulative distribution function (which can be accurately approximated computationally by a series expansion). The approximation requires $J - 2$ applications of the Clark formula.

Manski (1976) has reported good results in maximum likelihood search methods using the approximation above, with search directions determined by numerical evaluation of derivatives. This suggests that the bias caused by the approximation is relatively stationary for evaluation of probabilities at neighboring points. This fortuitous conclusion suggests that it is probably unnecessary to obtain analytic derivatives of P_1 with respect to parameters in statistical routines. On the other hand, it is possible that the use of analytic derivatives could decrease computation time. An argument given in McFadden (1977) shows that the Clark procedure can be applied to yield quick approximations to analytic derivatives.

The key to the accuracy of the Daganzo-Bouthelie-Sheffi approximation is the accuracy of the Clark procedure. Because the true distribution of the maximum of two normal variates is skewed to the right, one would expect the procedure to tend to underestimate small probabilities. The approximation will be best when the variates are positively correlated, with widely differing means, and worse when they are negatively correlated, with similar means. It may be possible to adjust the Clark formulae empirically to improve their accuracy for computation of small probabilities. Alternately, it would be interesting to explore

the possibility of adapting the Clark methodology to other trivariate distributions. In particular, if the generalized extreme value distribution introduced in the section on MNL and GEV models were used, then the only point of approximation would be the initial fit to the multivariate normal density, because maxima of GEV-distributed variates are again GEV-distributed. This would limit approximation error as J increases, in contrast to the Clark procedure, which becomes less accurate with large numbers of alternatives.