UCI-ITS-WP-84-1

A Discrete Choice Model for Ordered Alternatives

UCI-ITS-WP-84-1

Kenneth A. Small

School of Social Sciences and
Institute of Transportation Studies
University of California, Irvine

February 1984

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

ABSTRACT

A generalization of the multinomial logit (MNL) model is developed for cases in which discrete alternatives are ordered so as to induce stochastic correlation among alternatives in close proximity. The model belongs to the Generalized Extreme Value class introduced by McFadden, and is therefore consistent with random utility maximization. If the true model is nearly MNL, iterative estimation on an ordinary MNL computer package provides approximate parameter estimates and a test for the hypothesized failure of the MNL's "independence from irrelevant alternatives" assumption. A straightforward extension can handle cases where observations have been selected on the basis of a truncated choice set. The model's properties are investigated through a numerical example, and through two empirical applications whose rather unsatisfactory results are very briefly described.

The work in this paper was supported by Grant No. SES-8007010 from the National Science Foundation to Princeton University and by the Institute of Transportation Studies, University of California, Irvine. Of the many people providing helpful comments, I especially want to acknowledge David Brownstone, Gregory Chow, David Crawford, Dale Poirier, Richard Quandt, and George Tauchen. I thank Kenneth Boese and Avner Bar-Ilan for research assistance.

I. INTRODUCTION

The popularity of discrete choice models has led to their application in a number of situations where the alternatives can be ordered in some natural way. Examples include choice of occupation, number of automobiles owned, time of day of travel to work, biological effects of pesticides, and degree of labor force participation. Standard econometric models, however, impose stringent requirements on the degree of independence between the unobserved influences on various choices. The most well known is the "independence from irrelevant alternatives" (IIA) property of the multinomial logit (NML) model, which requires that the stochastic components of utility attached to the alternatives be independent (McFadden, 1973).

Yet in any of the examples above, it is plausible that unobserved traits will tend to affect the desirability of two or more alternatives similarly. A parent's disparaging view of education may make all higher-status occupations less likely to be chosen. A family's recreational pursuits may raise the desirability of owning two, three, or four cars. The need to transport a child to school may favor a certain group of possible schedules for a work trip. Formally, each of these implies correlation among the unobserved random utility components for alternatives which are close together on the natural ordering. I call this situation one of <u>proximate covariance</u>.

One way such correlations might arise is if the dependent variable is only a discrete representation of an underlying continuous variable.[1]

---

[1]I have encountered considerable resistance to the idea of using discrete models for such cases. Yet the literature contains many examples, and there is frequently no viable alternative. See, for example, Amemiya's (1981, p. 1518) discussion and grudging approval of Perloff and Wachter (1979), who for practical reasons used MNL on interval data even when the continuous data were available.

Ben-Akiva and Watanatada (1981) derive an MNL model from such a case. They carefully spell out the content of the IIA assumption, but do not discuss its validity. This paper questions its validity and proposes a way both to test and to correct for violation.

The model proposed in this paper permits a quite flexible covariance pattern, the main requirement being that for fixed i, the correlation between random utility components for alternatives i and j be a nonincreasing function of $|i-j|$. The model can, but need not, impose zero correlations between certain pairs of alternatives; in this feature it may be preferred to the nested logit (NL) model which <u>must</u> have some zero correlations. Like NL, the model proposed here is a member of the Generalized Extreme Value class, and therefore is consistent with random utility maximization and has MNL as a special case.

Two others models have been proposed for ordered alternatives, and discussed by Amemiya (1981) and Maddala (1983). One is the "ordered" or "ordered response" model, in which an ogive function is specified for the <u>sum</u> of the first (or last) k choice probabilities.[1] I show in the appendix that when this function is logistic (as in Deacon and Shapiro, 1975), the model is a special case of MNL. Thus while it may be

---

[1]See Crawford and Pollack (1982) for a more precise treatment of this and other concepts of ordering. Beggs et al. (1979) use the name "ordered logit" for a model which predicts not just the most preferred alternative but the entire preference ranking of alternatives. The problem addressed here should not be confused with correlation between successive <u>sample members</u> such as appears in panel data (Heckman, 1981; Cardell, 1977; Poirier & Ruud, 1982) and in spatial cross-sectional data (Fisher, 1971).

appealing for its parsimonious specification, it still suffers from a restrictive correlation structure.

The other is the "sequential" model, a special case of which is proposed independently by Sheffi (1979). This model is suitable when anyone choosing alternative n would always prefer $n-1$ to $n-2$, $n-2$ to $n-3$, and so forth; note that the direction of the ordering makes a difference here. Maddala warns that "these models are valid only if the random factors influencing responses at various stages are independent" (1983, p. 51), an unlikely assumption in most cases. Furthermore, it can be shown[1] that the sequential model is not a random utility model, but rather a peculiar limiting case of the nested logit model.

II. THE ORDERED GEV MODEL

McFadden (1978) has proposed a class of random utility models known as generalized extreme value (GEV) which have some of the computational properties of the MNL. Both MNL and NL are special cases. Letting $j = 1,\ldots,J$ index the set of alternatives, a GEV model is derived from a function $G(y_1,\ldots,y_J)$ defined on the orthant $y_j \geq 0$ which is nonnegative, homogeneous of degree one, tending toward $+\infty$ when any of its arguments tend toward $+\infty$, and whose n-th partial derivatives (with respect to distinct arguments) are nonnegative for odd n and

[1]Appendix available from the author.

nonpositive for even n. Any such function defines a cumulative distribution function (cdf)

$$(1) \qquad F(\varepsilon_1,\ldots,\varepsilon_J) = \exp\{-G(e^{-\varepsilon_1},\ldots,e^{-\varepsilon_J})\}$$

whose marginal distribution with respect to each variate is the extreme value distribution:

$$(2) \qquad F^k(\varepsilon_k) \equiv \lim_{\substack{\varepsilon_j\to\infty,\\ j\neq k}} F(\{e_j\}) = \exp[-c_k\exp(-\varepsilon_k)] \quad ,$$

where $c_k = G(\delta_{k1},\ldots,\delta_{kJ})$ and where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. If random utilities $\{\varepsilon_j, j=1,\ldots,J\}$ follow distribution (1), and V_j are observable utility components (sometimes called "strict utility" in the decision theory literature), then total utility

$$(3) \qquad U_j = V_j + \varepsilon_j \ , \quad j = 1,\ldots,J$$

is maximized at alternative k with probability

$$(4) \qquad P_k = e^{V_k} G_k(e^{V_1},\ldots,e^{V_J})/G(e^{V_1},\ldots,e^{V_J})$$

where G_k denotes the k-th partial derivative of G.

The MNL model is derived from the function

$$(5) \qquad G = \sum_{j=1}^{J} y_j \quad .$$

Its cdf (1) is a product of univariate extreme value distribution functions each of the form

(6) $$H_{MNL}(\varepsilon_j) = \exp[-\exp(-\varepsilon_j)] ,$$

and (4) yields the familiar form for choice probabilities:

(7) $$P_k = \exp(V_k) \Big/ \sum_{j=1}^{J} \exp(V_j) \quad .$$

The two-level NL model results from the function

(8) $$G = \sum_{r=1}^{R} \left(\sum_{j \in B_r} y_j^{1/\rho_r} \right)^{\rho_r}$$

where $B_r \subset \{1,\ldots,J\}$ is one of an exhaustive and mutually exclusive collection of R subsets of alternatives which define the "tree structure." Its cdf (1) is a product of R multivariate extreme value cdf's each of the form

(9) $$H_{NL}^{\rho_r}(\{\varepsilon_j | j \quad B_r\}) = \exp\left\{ -\left[\sum_{j \in B_r} \exp(-\varepsilon_j/\rho_r) \right]^{\rho_r} \right\} .$$

The NL choice probability derived from (4) can be written as the conditional probability of choosing an alternative from within a group times the probability of choosing that group:

(10) $$P_k = \frac{\exp(V_k/\rho_s)}{\sum_{j \in B_s} \exp(V_j/\rho_s)} \cdot \frac{\exp(\rho_s I_s)}{\sum_{r=1}^{R} \exp(\rho_r I_r)}$$

where B_s is the subset containing k and where

$$(11) \qquad I_r = \log \sum_{j \in B_r} \exp(V_j/\rho_r)$$

defines the <u>inclusive value</u> of subset B_r. Note that each factor in (10) has the MNL form.

It can be shown that only alternatives within a group B_r in the NL model have stochastic terms which are correlated with each other, and this correlation is inversely related to ρ_r. We wish to describe a case where only alternatives which lie close to each other along a natural ordering are correlated. To accomplish this, let the alternative labels j increase along this natural ordering,[1] and define a GEV model as follows:

DEFINITION 1: The Ordered Generalized Extreme Value (OGEV) model of discrete choice is the GEV model resulting from the function

$$(12) \qquad G(y_1,\ldots,y_J) = \sum_{r=1}^{J+M} \left(\sum_{j \in B_r} w_{r-j} y_j^{1/\rho_r} \right)^{\rho_r}$$

where M is a positive integer, ρ_r and w_m are constants satisfying

$$(13) \qquad 0 < \rho_r \leq 1 \quad , \qquad r = 1,\ldots,J+M$$

$$(14) \qquad w_m \geq 0 \quad , \qquad m = 0,\ldots,M$$

---

[1]Dale Poirier has pointed out to me that the model could easily be modified to deal with a cyclical ordering such as seasonality. This would require replacing the first summation in (12) by summation from $r = 1$ to J, and the second by summation over $j \in B_r$ when $r > M$ and $j \in B_r \cup B_{J+r}$ when $r \leq M$. In this way there are J subsets B_r, each with $(M+1)$ members.

$$\sum_{m=0}^{M} w_m = 1 \quad , \tag{15}$$

and where

$$B_r = \{j \quad \{1,\ldots,J\} | r-M \leq j \leq r\} \quad . \tag{16}$$

The main difference between this definition and the NL model defined by (8) is that the subsets B_r overlap. Each of the $(J+M)$ subsets contains up to $(M+1)$ contiguous alternatives; for example with $M=2$, $B_1 = \{1\}$, $B_2 = \{1,2\}$, $B_3 = \{1,2,3\}$, $B_4 = \{2,3,4\}$, $B_J = \{J-2,J-1,J\}$, $B_{J+1} = \{J-1,J\}$, and $B_{J+M} = \{J\}$. By including subsets with less than $(M+1)$ alternatives when those alternatives lie at one end of the ordering, we ensure that each alternative belongs to exactly $(M+1)$ different subsets. Since each subset can have its own parameter ρ_r, this provides considerable flexibility to the correlation patterns. The desired property of proximate covariance comes from the fact that the covariance between any two alternatives receives a contribution from each subset to which they belong in common; the closer they are together the more common subsets they belong to. Any two alternatives separated by fewer than M alternatives on the ordering belong to at least one common subset B_r, hence have a nonzero covariance (provided $\rho_r < 1$).

PROPOSITION 1: The OGEV model is of the GEV class and reduces to MNL when $\rho_r = 1$ for all r.

PROOF: G is nonnegative because of the condition (14). Its homogeneity of degree one can be checked directly. Since the weights are nonnegative and at least one is strictly positive, (13) implies that G tends to infinity when any of its arguments does. The partial derivatives alternate in sign because of (13). (15) guarantees that when $\rho_r = 1$ for all r, G reduces to (5). Q.E.D.

The next proposition states that the OGEV model has the desired property of proximate covariance. The bivariate cdf of any two stochastic elements is a mixture of multinomial logit cdf's and components of nested logit cdf's, with fewer of the latter as the alternatives get farther apart.

PROPOSITION 2: Each of the stochastic utility elements ε_k in the OGEV model has a univariate marginal distribution which is extreme value according to (2) with parameter $c_k = \sum_{m=0}^{M} w_m^{\rho_{k+m}}$. Any two stochastic utility elements ε_j and ε_k are independent if $|j-k| > M$; if $0 < k-j \leq M$ their bivariate marginal distribution has a cdf which, provided all w_m are strictly positive, is given by

$$H(\varepsilon_j,\varepsilon_k) = H_{MNL}(\varepsilon_j - \log W_{jk}^A) \cdot H_{MNL}(\varepsilon_k - \log W_{jk}^B) \tag{17}$$

$$\cdot \prod_{r=k}^{j+M} H_{NL}^{\rho_r}(\varepsilon_j - \rho_r \log w_{r-j}\ ,\ \varepsilon_k - \rho_r \log w_{r-k})$$

where

$$W_{jk}^A = \sum_{m=0}^{k-j-1} w_m^{\rho_{j+m}} \quad \text{and} \quad W_{jk}^B = \sum_{m=M-(k-j-1)}^{M} w_m^{\rho_{k+m}} .$$

PROOF: The proof follows directly by substituting (12) into (1) and letting the appropriate arguments tend to infinity. Q.E.D.

The OGEV choice probabilities are obtained directly from (4):

$$P_k = \sum_{r=k}^{k+M} q(k|B_r)Q(B_r) \tag{18a}$$

where

$$q(k|B_r) = \frac{w_{r-k}\exp(V_k/\rho_r)}{\exp(I_r)} \tag{18b}$$

$$Q(B_r) = \frac{\exp(\rho_r I_r)}{\sum_{s=1}^{J+M} \exp(\rho_s I_s)} \tag{18c}$$

$$I_r = \log \sum_{j\in B_r} w_{r-j}\exp(V_j/\rho_r) \quad . \tag{18d}$$

Note that P_k is a sum of $(M+1)$ terms, each resembling the choice probability (10) for a nested logit model. It looks like an expansion of P_k into conditional probabilities of the MNL form, but it is not because the sets B_r are not mutually exclusive. We therefore cannot perform the analogue of the sequential estimation of the NL model (McFadden, 1981) since we cannot observe which B_r was "chosen" by a given individual. (More formally, it is because the log-likelihood function does not separate into two terms each of the MNL form). We can, however, use maximum likelihood. In practice, it usually will be necessary to reduce the number of free parameters by imposing equality restrictions on w_m and/or on ρ_r, and by fixing M in advance. For example, given integer M, the following has just one parameter in addition to those describing the strict utilities:

DEFINITION 2: The Standard OGEV model is the OGEV model for which $w_m = 1/(M+1)$ for all m, and $\rho_r = \rho$ for all r.

For the simplest case, $M=2$, the Standard OGEV model is generated by

$$G(y_1,\ldots,y_J) = \sum_{r=1}^{J+1} \left(\frac{1}{2}\, y_{r-1}^{1/\rho} + \frac{1}{2}\, y_r^{1/\rho}\right)^{\rho}$$

and has choice probabilities

$$P_k = \frac{e^{V_k/\rho}\left[\left(e^{V_{k-1}/\rho} + e^{V_k/\rho}\right)^{\rho-1} + \left(e^{V_k/\rho} + e^{V_{k+1}/\rho}\right)^{\rho-1}\right]}{\sum_{r=1}^{J+1} \left(e^{V_r/\rho} + e^{V_{r-1}/\rho}\right)^{\rho}}$$

with the convention that $e^{V_r/\rho} = 0$ for $r < 1$ or $r > J$.

Intuitively, the OGEV model causes P_k to be diminished if there is an attractive alternative ℓ nearby, because the latter will increase the denominator of (18b) for each B_r containing ℓ. This effect is greater the more sets B_r contain both ℓ and k, and the smaller are the corresponding ρ_r. At the extreme $\rho_r \to 0$ for all r, $\rho_r I_r$ tends to $\max\{V_j | j \in B_r\}$ and P_k tends to $n_k \exp(V_k) / \sum_{j=1}^{J} n_j \exp(V_j)$, where n_k is the number of sets B_r within which alternative k has the largest[1] strict utility. In this limit the probability of choosing an alternative dominated by all its M neighbors on each side

---

[1]An N-way tie involving alternative k within a set B_r contributes $1/N$ to n_k.

is 0. This limit is in fact a valid random utility model whose cdf is continuous but whose density function is discontinuous at points $\varepsilon_j = \varepsilon_k$ whenever $|j-k| \leq M$.

Proposition 2 and the intuitive argument above make it plausible to expect ε_j and ε_k to be more closely correlated the smaller is $|j-k|$. However, the correlation computed from the cdf (17) cannot be written in closed form.[1] Table 1 reports the results of some numerical integrations for the Standard OGEV Model, which confirm the expected pattern. It appears the model is able to generate correlations up to about 0.7. The correlations increase only slightly as ρ is decreased from 0.5, particularly for large M, which could lead to relatively flat likelihood functions in this range.

III. APPROXIMATION WHEN THE MODEL IS ALMOST MULTINOMIAL LOGIT

As already noted, the OGEV model cannot in general be estimated using computer software designed for MNL. However, if the departure from MNL is small in the sense that ρ_r are near one, there is a procedure involving iteration of MNL-type estimations which provides a good approximation. It uses "pseudo-variables" whose values for a given

---

[1] In the case $M=1$, there is a theoretical upper limit of $1/\sqrt{2} = .707$ on $r_{i,i+1}$, where r_{ij} is the correlation between ε_i and ε_j. This limit is imposed by the requirements that $r_{i,i+2} = 0$, and that the covariance matrix of $\varepsilon_i, \varepsilon_{i+1}, \varepsilon_{i+2}$ be positive definite. A similar approach for $M=2$ yields $r_{i,i+1} \leq \sqrt{3}/2 = .866$.

Table 1

Correlation Between ε_j and ε_k: Standard OGEV

M	k-j	ρ 0.9	ρ 0.5	ρ 0.1
1	1	.094	.354	.427
3	1	.142	.546	.643
3	2	.094	.354	.397
3	3	.047	.172	.192
5	1	.158	.612	.673
5	2	.126	.481	.532
5	3	.094	.354	.393
5	4	.062	.232	.259
5	5	.031	.114	.133
10	1	.172	.674	.686

alternative[1] involve the utilities (estimated in the previous iteration) of nearby alternatives. Such a variable would not be allowed in a true MNL model based on random utility maximization: In fact, McFadden, Train, and Tye (1977) suggest including pseudo-variables as a test for departures from MNL, though they do not suggest how to construct them. The procedure described below constructs a pseudovariable suitable for testing the particular type of departure from MNL embodied in the OGEV model, and also provides approximate estimates of the parameters ρ_r.

Let $\sigma_r \equiv 1 - \rho_r << 1$ for all r; $\sigma \equiv (\sigma_1,\ldots,\sigma_{J+M})'$; and write V_j as a linear function of observable variable vector z_j with coefficient vector β. The idea is to find pseudovariables N^r, which may depend on β, such that the OGEV choice probability (18) can be approximated by the MNL-like formula:

$$(19) \qquad P_k \cong \exp(\beta' z_k + \sigma' N_k) / \sum_{j=1}^{J} \exp(\beta' z_j + \sigma' N_j)$$

where $N_j \equiv (N_j^1,\ldots,N_j^{J+M})'$ is a column vector whose r-th component is is the value of pseudo-variable N^r at alternative j. Expanding (18) and (19) in Taylor series, we find they are identical up to first order in σ if

---

[1]As is evident from (19), I place the alternative subscript on the variable rather than the coefficient. A single variable thus takes on values for each alternative and each sample member. In the terminology of Maddala (1983, p. 42), this is the "conditional logit" rather than the "multinomial logit" formulation. As Maddala shows, the two are equivalent, so I have followed his lead in disregarding the semantic distinction and referring to both as MNL.

$$(20) \qquad N_j^r = \begin{cases} w_{r-j}(V_j - I_r^o) = -w_{r-j} \log \sum_{\ell \in B_r} w_{\ell-r}(P_\ell^o/P_j^o) \ , & j \in B_r \\ 0 \ , & j \notin B_r \ , \end{cases}$$

where P_j^o is given by (7), and I_r^o by (18d) with ρ_r set to one.

To estimate, let each iteration be as follows. Starting with some initial value $\hat{\beta}$ ($\hat{\beta} = 0$ may be used for the initial iteration), compute $\{N_j^r\}$ from (20) and treat them as fixed in (19); an ordinary MNL step then provides an estimate $\hat{\sigma}$ and a new $\hat{\beta}$. Keep iterating until $\hat{\beta}$ changes negligibly. While this procedure is not guaranteed to converge, it has done so fairly quickly in a number of applications, so long as $|\sigma^r|$ do not exceed about 0.75. An equality constraint $\sigma_r = \sigma$, $r \in B_m$, is easily dealt with by replacing the corresponding $\{N^r\}$ by the single variable $N^B = \sum_{r \in B_m} N^r$.

The estimate $\hat{\sigma}$ also provides a simple way to test whether the model does depart from MNL. It can be shown[1] that under the null hypothesis $\sigma = 0$, the usual MNL covariance matrix provided at the last iteration is a consistent estimate of the true covariance matrix of the estimated parameters $\hat{\beta}$ and $\hat{\sigma}$. This may seem surprising in light of numerous examples in econometrics of two-stage estimators where standard errors are underestimated by the output of a standard algorithm at the second stage. The reason the "naive" procedure works in this case is that β and σ are estimated simultaneously in each MNL step.

---

[1] Appendix available from the author.

The intuition behind the procedure can best be illustrated by considering the standard OGEV model with $M=1$. There is just one pseudovariable, with values

$$\text{(21)} \qquad N_j \equiv \sum_{r=1}^{J+1} N_j^r = \left[\log \tfrac{1}{2} + \log\left(1 + e^{V_{j-1}-V_j}\right) + \log\left(1 + e^{V_{j+1}-V_j}\right)\right]$$

with the convention $V_0 = V_{J+1} = -\infty$. Aside from the constant (which has no effect), N_j takes on large positive values for those alternatives adjacent to unpopular alternatives. To the extent that the true model departs from MNL toward OGEV, being adjacent to an alternative with low utility makes a given alternative more likely to be chosen, since any stochastic effect favoring both will be magnified. For example, in choice among six ordered occupations, suppose numbers 3 and 4 denote middle-status occupations and 3 is seldom chosen by females; then a (stochastic) preference for middle-status occupations will have a greater effect on P_4 for a female than it would for a male, because for the male its effect would be spread more evenly between P_3 and P_4. If the data are generated by a process like this, the pseudo-variable N will show explanatory power, and its coefficient σ will tend to be positive.

IV. NUMERICAL EXAMPLE

In this section a numerical example is constructed to illustrate the kind of situation for which the ordered logit model is designed. Properties of several models are compared, both for initial fit and for ability to predict results of exogenous changes.

Suppose a household can own zero, one, two, or three autos; denote these discrete alternatives by $j = 1,2,3$, and 4, respectively. There are three types of households: "small" ones which prefer smaller numbers of cars, "large" ones which prefer larger numbers, and "medium-sized" households which are indifferent among numbers of autos. Most households (90%) are medium sized, and this group is split evenly among the available ownership alternatives; 5% of households are small and choose the smallest possible j, while 5% are large and choose the largest possible j. Table 2 shows a set of utilities for each group compatible with the above description.

Now, suppose only alternatives 1-3 are available to the members of an observed sample, so that the observed choice frequencies are as shown in the next to the last row of the table. The problem is to fit a choice model to these observed frequencies when household size is not observed. A reasonable specification might be $U_j = V_j + \varepsilon_j$ with

$$V_j = j \cdot \alpha \tag{22}$$

and with the distribution of ε_j depending on the model. Note that the true model may be characterized[1] by $\alpha = 0$ and $\varepsilon_j \equiv U_j$ distributed as in Table 2, from which we calculate correlations $\text{corr}(\varepsilon_1,\varepsilon_2) = \text{corr}(\varepsilon_2,\varepsilon_3) = .688$, $\text{corr}(\varepsilon_1,\varepsilon_3) = -.053$.

This is just the kind of error structure for which the simple ordered logit model is appropriate (though not exact, since these ε_j do not

---

[1]This characterization is not unique, because of both the arbitrary constant in EU_j (which is 0.1 in Table 2) and, more importantly, the arbitrary constant which could be added to U_j for any single household type (the latter would affect the correlations).

Table 2

Parameters for Numerical Example

Household Type	Proportion of Total	Utility				Choice Frequency			
		U_1	U_2	U_3	U_4	f_1	f_2	f_3	f_4
Small	.05	2	1	0	-1	.05	.00	.00	--
Medium	.90	0	0	0	0	.30	.30	.30	--
Large	.05	0	1	2	3	.00	.00	.05	--
Aggregate	1.00					.35	.30	.35	--

have an extreme value distribution). The choice frequencies "pile up" at the extreme alternatives ($j = 1$ and $j = 3$) because of unobserved preferences which are systematic in the index j. It turns out that the standard OGEV model with $M = 1$ correctly recognizes this and uses the extra parameter ρ to help account for the larger observed frequencies of these two alternatives. This enables it to predict more accurately the results of removing or adding alternatives.

Since household characteristics are not observed, only the chosen alternative distinguishes one observation from another. Hence the log-likelihood function, given the observed frequencies f_j, is simply

$$L(\theta) = \sum_{j=1}^{3} f_j \log P_j(\theta) \tag{23}$$

where parameter vector θ consists of α and, where relevant, ρ.

Table 3 gives the parameter values maximizing L, as well as predictions for three scenarios: (A) removal of alternative 3 from the choice set, for example by prohibiting on-street parking in a neighborhood of two-car garages; (B) removal of alternative 1 from the choice set, for example by making all non-automotive forms of travel infeasible; (C) addition of a new alternative 4 (representing three cars) to the choice set, for example by removing a prohibitive tax on owning three cars. Predictions are made using the appropriate choice probability formula[1] from Section II. V_j is calculated from (22), replacing α by its estimated value $\hat{\alpha}$, except that V_3 is set

[1]The formula for sequential logit choice probability is

$$P_k = (1 - P_{(k+1)|k}) \prod_{j=1}^{k} P_{(j+1)|j} \quad \text{where} \quad P_{(j+1)|j} = [1 + \exp(V_j - V_{j+1})]^{-1}.$$

to $-\infty$ for scenario A, and V_1 is set to $-\infty$ for scenario B. It is convenient for scenarios A and B to let P_2, the predicted share of alternative 2, serve as a basis for comparing models; note that the true result is $P_2^A = P_2^B = .5$, since alternative 2 will be chosen by half the medium-sized households and all the large (scenario A) or small (scenario B) households. For scenario C, the table shows P_4^C, the predicted share of the newly added alternative, whose true value is $P_4^C = .275$ (.90/4 from the medium-sized and .05 from the large households. Scenario C is not considered for the NL model because the new tree structure is ambiguous.

In order to give MNL a fairer comparison with OGEV and NL, each of which has two parameters, a two-parameter MNL model is also estimated. It is specified by $V_1 = \alpha_1$, $V_2 = 0$, $V_3 = \alpha_3$. It should be noted that two parameters are sufficient to fit these data exactly in terms of aggregate shares, so our performance test must rely on predictions.

We see from Table 3 that the OGEV model gives the best predictions. Although the 1-parameter MNL model estimates α correctly, it does not fit the initial shares exactly; this qualifies its success in correctly predicting P_2 in scenarios A and B, since the <u>change</u> in P_2 is underpredicted (an analyst using one-parameter MNL would predict a 50% increase in number of people owning one car, rather than the actual 66.7%). One-parameter MNL also underpredicts the shift to newly added alternative 4 in Scenario C. Two-parameter MNL under-predicts the change in P_2 from either scenario A or B, and provides no basis for prediction under scenario C (since we would not know what to assume for V_4). Using NL, the fitted parameters achieve a compromise such that the shift toward alternative 2 is overpredicted in one scenario

Table 3

Results of Numerical Example

	Estimated Parameters				Fitted Shares		Predicted Shares		
	$\hat{\alpha}$	$\hat{\alpha}_1$	$\hat{\alpha}_3$	$\hat{\rho}$	P_1	P_2	P_2^A	P_2^B	P_4^C
TRUE VALUES					.35	.30	.50	.50	.275
FITTED MODELS:									
MNL: 1-parameter	.000	---	---	---	.33	.33	.50	.50	.250
MNL: 2-parameter	---	.154	.154	---	.35	.30	.46	.46	---
OGEV[a]	.000	---	---	.5850	.35	.30	.50	.50	.269
NL (1;2,3)[b]	.103	---	---	.6675	.35	.30	.53	.46	---
NL (1,2,;3)[c]	-.103	---	---	.6675	.35	.30	.46	.53	---
Sequential Logit	.431	---	---	---	.39	.24	.61	.39	.223

[a]Standard OGEV with $M = 1$.

[b]Tree structure: $B_1 = \{1\}$; $B_2 = \{2,3\}$.

[c]Tree structure: $B_1 = \{1,2\}$; $B_2 = \{3\}$.

and underpredicted in the other. Sequential logit gives worse fit and worse predictions than any of the other models.

Estimating OGEV by the first-order approximation procedure of Section III yields an initial $\hat{\alpha}$ of 0, and the pseudovariable

$$N_j = \begin{cases} 0 & j = 1,3 \\ -\frac{1}{2}\log(2) & j = 2 \,. \end{cases}$$

Further iterations give $\hat{\alpha} = 0$ and $\hat{\rho} = .5552$. With this $\hat{\alpha}$ and $\hat{\rho}$, predictions for the three scenarios are identical to those shown in the table for the exact OGEV model. Note that N captures the positional "advantage" of whatever alternatives are at the extreme ends of the choice set, so that (19) correctly attributes the observed higher choice frequencies for these alternatives to the correlation structure; because changes in the choice set cause appropriate changes in N through equation (20), this positional advantage is correctly transferred to the new choice set inherent in each of the three scenarios, thereby resulting in the good predictions. Though it is risky to generalize from such a simple example, it appears the first-order approximation may do very well at predicting even when ρ differs substantially from one.

V. DEPENDENT VARIABLE FROM A TRUNCATED CHOICE SET

In general, GEV models cannot be estimated consistently from observations on only a subset of the available alternatives. MNL is the only known exception. The Ordered GEV model, however, provides a way of incorporating unobserved portions of the choice set into the error structure. This section shows the form taken by the choice probabilities conditional on a choice from a contiguous subset $j = 1,\ldots,J$ when the true model is Ordered GEV on a larger choice set.

DEFINITION 3: The Extended Ordered GEV (EOGEV) model of discrete choice is the GEV model resulting from the function (12) with w_{r-j} replaced by

(24a) $$w'_{r,(r-j)} = w_{r-j}\exp(-\sigma_r a_r)$$

where a_r are parameters and

(24b) $$\sigma_r = 1 - \rho_r \quad .$$

Making the weights depend on r in this way maintains the assumptions necessary for the model to be a GEV random-utility model, and ensures that it still has MNL as a special case. Equations (18) are still valid with w_{r-j} replaced by (24a) and I_r by:

(24c) $$I'_r = I_r - \sigma_r a_r \quad ,$$

in which case they become:

(25) $$P_k = \frac{\sum_{r=k}^{k+m} w_{r-k}\exp[V_k/\rho_r - \sigma_r(I'_r + a_r)]}{\sum_{r=1}^{J+M} \exp(\rho_r I'_r)} \quad .$$

The approximation for ρ_r near one is

(26) $$P_k \cong \frac{\exp[V_k + \sum_r \sigma_r N_k^r + \sum_r (\sigma_r a_r) A_k^r]}{\sum_{j=1}^{J} \exp[V_j + \sum_r \sigma_r N_j^r + \sum_r (\sigma_r a_r) A_j^r]}$$

where N_j^r is given by (20), and where $A_j^r = -w_{r-j}$ if $j \in B_r$ and $A_j^r = 0$ otherwise. Note that pseudo-variables A^1 and N^1 are collinear, as are A^{J+M} and N^{J+M}; hence some equality restrictions are necessary to identify σ^1, a^1, σ^{J+M}, and a^{J+M}.

PROPOSITION 3: Suppose the true choice model is Ordered GEV with the alternative set $\tilde{B} = \{-J_1+1,\ldots,0,1,\ldots,J+J_2\}$ and with strict utilities $\tilde{V}_j$, where J_1, J_2 are nonnegative integers. Let $B = \{1,\ldots,J\}$, $\tilde{B}_r = \{j \in \tilde{B} | r \leq j \leq j+M\}$ as before, and $B_r = \{j \in B | r \leq j \leq j+M\}$. Then for $k \in B$, the choice probability conditional on some alternative from the subset B being chosen is given by an EOGEV model with strict utilities $V_j = \tilde{V}_j$, $j \in B$, and with parameters

$$a_r = (\tilde{I}_r - I_r)/\rho_r \quad , \tag{27}$$

where

$$\tilde{I}_r = \log \sum_{j \in \tilde{B}_r} w_{r-j} \exp(\tilde{V}_j/\rho_r) \quad . \tag{28}$$

Note that $\tilde{B}_r \supseteq B_r$ so that $a_r \geq 0$; also that for $M+1 \leq r \leq J$, $\tilde{B}_r = B_r$ hence $a_r = 0$.

PROOF: The true choice probabilities $\tilde{P}_k$ are given by (18) with B replaced by $\tilde{B}$, V by $\tilde{V}$, and I by $\tilde{I}$. For $k \in B$, the conditional choice probability is

$$P_k = \tilde{P}_k / \sum_{j=1}^{J} \tilde{P}_j = \frac{\sum_{r=k}^{k+m} w_{r-k} \exp[(V_k/\rho_r) - \sigma_r \tilde{I}_r]}{\sum_{j=1}^{J} \sum_{r=j}^{j+M} w_{r-k} \exp[(V_j/\rho_r) - \sigma_r \tilde{I}_r]} \tag{29}$$

Using equations (24c) and (27) to write $\tilde{I}_r = I_r' + a_r$, we see the numerators of (29) and (25) are identical. To show that the denominators are also identical, reverse the order of summation in the denominator of (29) and use (18d), (24b), and (24c). Q.E.D.

VI. EMPIRICAL APPLICATIONS

The models discussed in this paper have been implemented on two data sets. Although computational feasibility has been established, the results do not give strong support for the model's applicability to these examples. For this reason, the examples and results are described only briefly; more detailed information can be obtained from the author.

The first data set is one used previously with the MNL model (Small, 1982) to study scheduling by automobile commuters facing congestion. The problem is to explain how far (in 5-minute intervals) a commuter's planned time of arrival deviates (in either direction) from the official work start time. Explanatory variables include several employer and employee characteristics, and a detailed description of how much travel time would be encountered for any of twelve alternative choices of arrival time. The most successful MNL model involved 9 variables; it was estimated here on a sample of 527 commuters each with 12 alternatives specified, ranging from 40 minutes early to 15 minutes late.[1] In the earlier paper I presented a heuristic test which, as one would expect a priori, suggested departures from MNL of the type dealt with here.

Table 4 presents some selected results for five extensions of MNL. They are: (a) nested logit[2] with the 9 "earliest" alternatives and the 3

---

[1]The model is "Model 4" of Small (1982), p. 473. The sample of 453 used in the earlier paper was expanded for the present study by reconstructing some previously missing data on carpooling.

[2]See Small and Brownstone (1982) for further results on NL models using these data.

Table 4

Empirical Results: Trip Timing

	Model				
	NL[a]	OGEV[b]	EOGEV[c]	OGEV[d]	EOGEV[e]
$\hat{\rho}$	0.76 (0.13)	0.75 (0.21)	0.37 (0.15)	0.76 (0.28)	0.38 (0.33)
$\widehat{(1-\rho)a_1}$	--	--	∞	-- --	5.91 (3.01)
χ^2 statistic[f]	2.90	0.25	4.94	0.74	5.09

Estimated standard errors are in parentheses.

[a]Nested logit: $B_1 = \{1,\ldots,9\}$ and $B_2 = \{10,11,12\}$; $\rho_1 = \rho_2 \equiv \rho$.

[b]Standard Ordered GEV: $M = 5$

[c]Extended Standard OGEV: $M = 5$, $a_r = 0$ for $r \geq 2$.

[d]Same as b, approximate estimator, one iteration starting with MNL values for β.

[e]Same as c, approximate estimator, one iteration starting with MNL values for β.

[f]Twice the difference between the log likelihood achieved by the model and that achieved by MNL.

"latest" alternatives forming groups B_1 and B_2, respectively, and with $\rho_1 = \rho_2 = \rho$; (b) Standard Ordered GEV with $M = 5$; (c) Extended Ordered GEV as in (b) but with provision for a single a_1 representing effects of eliminating from the sample commuters who chose to arrive more than 40 minutes early for work; (d) and (e) incomplete estimation of the first-order approximations to (b) and (c), respectively, stopping after just one iteration. The first three extensions were estimated by full information maximum likelihood (FIML),[1] the other two using a standard MNL package.[2]

These results show that the OGEV model, though performing plausibly, does not perform as well as NL (as judged by likelihood achieved), nor is it better than MNL at conventional significance levels (the one-sided test of $\rho < 1$ against $\rho = 1$ rejects the null at a 12% level). EOGEV, in contrast, rejects MNL rather too vigorously: The algorithm tried to make a_1 infinite. I conclude that EOGEV is misinterpreting

[1]The likelihood function and its first derivatives were programmed as FORTRAN functions; the second derivatives were approximated by the expected value of the cross-product of the first derivatives (as in the algorithm by Berndt et al., 1974); and the modified quadratic hill-climbing method of Goldfeld and Quandt (1972) was applied, as embodied in their numerical optimization package GQOPT.

For models giving parameter estimates in the proper range, this procedure proved quite reasonable in cost. The OGEV model converged in 3 iterations (starting from the MNL parameter estimates and $\rho = 1$), requiring 7 function evaluations and using 36 seconds of central processing time on the IBM 3033 computer at Princeton University. The NL model (starting from 1 for ρ and 0 for all other parameters) required 5 iterations, 12 function evaluations, and 26 seconds. However, reasonable starting values sometimes resulted in the algorithm getting "stuck" for unknown reasons, so a certain amount of trial and error must be anticipated. Also, trials at various starting values are advisable to guard against non-uniqueness of local maxima.

[2]QUAIL, developed by D. McFadden and others, required 8 seconds of cpu time to estimate MNL, and an additional 14 seconds to estimate one iteration of the approximation for EOGEV.

some spurious effect, perhaps a misspecification of the determination of very early arrivals. On the brighter side, the rather simple programming required to estimate a single iteration of Section III's procedure provides a good approximation to the FIML estimates of ρ, and gives ample warning of an excessively large a_1 in the extended model.

The other application attempted was to reestimate the model of automobile ownership in Train (1980). In order to keep the problem manageable, I ignored the simultaneity between this decision and choice of travel mode to work (a major focus of Train's paper). Thus I attempted to model auto ownership alone, first as independent of and then as conditional upon modal choice. Although this sufficed to obtain plausible MNL models, all attempts to generalize to ordered GEV resulted in violations of the conditions $\rho_r \leq 1$. Again, the first-order approximation gave warning that FIML estimation would not produce results compatible with the model.

These limited results suggest that the model presented in this paper is sensitive to misspecification. Thus the first priority of the researcher should be to provide an adequate set of explanatory variables, for which MNL is probably satisfactory as a tool for exploratory work. The extra complication involved in the Ordered GEV model may or may not pay off in additional quantitative precision, depending on whether or not the data are sufficiently numerous and accurate to measure the rather subtle effects that such ordering produces.

APPENDIX

The class of "ordered" or "ordered-response" models defined by Amemiya (1981, pp. 1513-1516) and Maddala (1983, pp. 46-49) are of the general form $P_1 + \ldots + P_k = F(W_k)$, $k = 1,\ldots,J-1$, where F is a cumulative distribution function, W_k depends on data vector x and on unknown parameters, and $W_1 < W_2 < \ldots < W_{J-1}$. (In Maddala, this holds after relabelling the alternatives in reverse order.) If F is logistic, we have an ordered logit model. Now define

$$V_1 = -\log[1 + \exp(-W_1)]$$

$$V_k = \log\left[-\sum_{j=1}^{k-1} \exp(V_j) + \{1 + \exp(-W_k)\}^{-1}\right], \quad k = 2,\ldots,J-1$$

$$V_J = \log\left[-\sum_{j=1}^{J-1} \exp(V_j) + 1\right] \quad .$$

Substitution into equation (7) yields $P_1 + \ldots + P_k = F(W_k)$, with F logistic. That is, the ordered logit model is identical to an MNL defined with these V_j. Furthermore, in Madalla and in all but one of Amemiya's examples, the restriction $W_k = \gamma_k - x'\beta$ is imposed, making ordered logit a rather strongly restricted MNL model.

BIBLIOGRAPHY

Ameyiya, Takeshi: "Qualitative Response Models: A Survey," Journal of Economic Literature, 19 (1981), 1483-1536.

Ben-Akiva, Moshe, and Thawat Watanatada: "Applications of a Continuous Spatial Choice Logit Model," in C. Manski and D. McFadden (Eds.), Structural Analysis of Discrete Data with Econometric Applications (Cambridge, Massachusetts: M.I.T. Press, 1981).

Beggs, S., S. Cardell, and J. A. Hausman: "Assessing the Potential Demand for Electric Cars," Working paper, Charles River Associates, Boston, 1979.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman: "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, 3 (1974), 653-665.

Cardell, Scott: "Multinomial Logit with Correlated Stochatic Terms," working paper, Charles River Associates, Boston, 1977.

Crawford, David L., and Robert A. Pollak: "Order and Inference in Qualitative Response Models," Discussion Paper 82-4, Bureau of Economic Research, Rutgers University (1982).

Daly, A., and S. Zachary: "Improved Multiple Choice Models," in D. A. Hensher and M. Q. Dalvi (Eds.), Determinants of Travel Choice (Farnborough, England: Saxon House, 1978).

Deacon, Robert, and Perry Shapiro: "Private Preference for Collective Goods Revealed Through Voting on Referenda," American Economic Review, 65 (1975), 943-955.

Fisher, Walter D.: "Econometric Estimation with Spatial Dependence," Regional and Urban Economics, 1 (1971), 19-40.

Goldfeld, S. M., and R. E. Quandt: Nonlinear Methods in Econometrics (Amsterdam: North-Holland, 1972).

Heckman, James: "Models for Testing the Hypothesis of True State Dependence against that of Spurious State Dependence," in C. Manski and D. McFadden (Eds.), Structural Analysis of Discrete Data with Econometric Applications (Cambridge, Massachusetts: M.I.T. Press, 1981).

McFadden, Daniel: "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (Ed.), Frontiers in Econometrics (New York: Academic Press, 1973), 105-142.

McFadden, Daniel: "Modelling the Choice of Residential Location," in Anders Karlqvist et al. (Eds.), Spatial Interaction Theory and Planning Models (Amsterdam: North-Holland Publishing Company, 1978), 75-96.

McFadden, Daniel: "Econometric Models of Probabilistic Choice," in C. F. Manski and D. McFadden (Eds.), Structural Analysis of Discrete Data with Econometric Applications (Cambridge, Massachusetts: M.I.T. Press, 1981).

McFadden, Daniel, Kenneth Train, and William B. Tye: "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," Transportation Research Record, 637 (1977), 39-45.

Maddala, G. S: Limited-Dependent and Qualitative Variables in Econometrics (Cambrige: Cambridge University Press, 1983).

Perloff, J. M., and M. L. Wachter: "The New Jobs Tax Credit: An Evaluation of the 1977-78 Wage Subsidy Program," American Economic Review, Papers and Proceedings 69 (1979), 173-179.

Poirier, Dale and Paul Ruud: "Conditional Minimum Distance Estimation and Autocorrelation in Limited Dependent Variable Models." Paper presented to the Econometric Society annual meeting, New York, 1981.

Sheffi, Yosef: "Estimating Choice Probabilities Among Nested Alternatives," Transportation Rsearch, 13B (1979), 189-205.

Small, Kenneth A.: "The Scheduling of Consumer Activities: Work Trips," American Economic Review, 72 (1982), 467-479.

Small, K. A., and D. Brownstone: "Efficient Estimation of Nested Logit Models: An Application to Trip Timing," Research Memorandum No. 296, Econometric Reseach Program, Princeton University, 1982.

Train, Kenneth: "A Structured Logit Model of Auto Ownership and Mode Choice," Review of Economic Studies, 47 (1980), 357-370.