

ORDERED LOGIT:
A DISCRETE CHOICE MODEL WITH
PROXIMATE COVARIANCE AMONG ALTERNATIVES

Kenneth A. Small

Econometric Research Program
Research Memorandum No. 292

December 1981

The work in this paper was supported by Grant No. SES-8007010
from the National Science Foundation.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

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 between alternatives in close proximity. The model belongs to the Generalized Extreme Value class introduced by McFadden, and is therefore consistent with random utility maximization. Iterative estimation on ordinary MNL computer packages is possible if the true model is "nearly" MNL, and the first such iteration serves as 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 estimation of a model explaining commuters' work-trip scheduling choices.

I. INTRODUCTION

One of the most popular econometric models for choice among discrete alternatives is the multinomial logit (MNL). Its theoretical basis as a random utility model is well established (McFadden, 1973). It is convenient both because maximum likelihood estimates can be computed quickly and because in many instances a choice among many alternatives can be broken computationally into a sequence of simpler choices (Ben-Akiva, 1973; Daly and Zachary, 1978; Williams, 1977; McFadden, 1978).

A widely discussed limitation of the MNL model is its property of "independence from irrelevant alternatives" (IIA), which unfortunately holds whether or not the alternatives in question are "relevant." This property follows from the assumption that the unobserved preferences for the various alternatives are stochastically independent; it implies that the ratio of choice probabilities for any two alternatives is independent of the properties of all other alternatives.

This paper investigates a generalization of the MNL model which allows a particular kind of departure from the IIA assumption. An example is the number of automobiles owned by a household.¹ The household is assumed to choose among discrete alternatives representing different numbers of automobiles. The stochastic utility component for alternative j represents unobserved traits that alter the desirability of owning j automobiles. Common sense suggests that a household with an idiosyncratic

¹See Train (1977) for a critique of many such models.

preference for owning, say, 4 automobiles is also likely to have a similar preference for owning 3 or 5, since it probably has some unobserved trait (such as several teenage children) leading it to want a lot of cars. The empirical implication is that if the option of owning 4 or more cars were for some reason foreclosed, this household would be more likely to choose 3 than would some other household with identical observed characteristics. Yet the IIA property prohibits this.

Another example is provided by the choice of time-of-day for work trips as modelled by McFadden et al. (1977), Small (1982), and Abkowitz (1980). Because of practical difficulties with treating the choice as a continuous one (especially the tendency of respondents to round off replies to the nearest five minutes), all of these authors estimated an MNL model of choice among 12 discrete alternatives each representing arrival at work within a particular 5-minute interval. The choice set consisted of intervals covering the range from 42 1/2 minutes before to 17 1/2 minutes after the "official" work start time. Given the natural ordering of these alternatives, it is likely that certain unobserved traits (such as employee independence or existence of a comfortable place to wait until starting time) are correlated among nearby alternatives.

These examples share the property that the alternatives can be arrayed along some obvious dimension, hence have a natural ordering which can be determined a priori. Variables reflecting this ordering may be included in the model--for example, Train's auto ownership model includes a cost variable which is simply proportional to the number of autos owned. If such variables completely

captured the determinants of individual behavior, there would be no remaining stochastic term related to this ordering and no need for the model proposed here. It is precisely because observable traits are usually not sufficient to describe these preferences that the problems addressed in this paper arise.

The problem is analogous to, but really quite different from, autocorrelation in time series (or its generalization in cross-sectional data from two-dimensional space: see Fisher, 1971). In the latter cases the problem is that successive observations are arrayed along one or more dimensions, so that error correlation depends inversely on some measure of distance between observations. With the advent of panel data with discrete dependent variables, this phenomenon has received some attention in the discrete choice literature as well (Cardell, ; Heckman, 1981). In contrast, the type of error discussed in this paper involves only correlation among the unobserved properties of the choices themselves, and has nothing to do with resemblance among different members of the sample. In this sense it is more akin to the existence, rather than correlation pattern, of an error term in a continuous model. The postulate is simply that the dependent variable itself is only a discrete representation of an underlying continuous variable with an unobserved stochastic component. This view of discrete choice models in appropriate circumstances has been made explicit by Ben-Akiva and Watanatada (1981), who provide a formal derivation of a logit model from an underlying continuous process. Although they spell out carefully the content of the IIA assumption in such a context, they do not discuss its validity. This paper questions its validity and proposes a way both to test and to correct for violation.

McFadden (1978; 1981) has provided a framework for defining generalizations of the MNL model which are consistent with stochastic utility maximization, but which need not have the IIA property. The most thoroughly

explored is the nested logit (NL) model, also known as structured logit (Ben-Akiva, 1973; Amemiya, 1978; Train, 1980; Brownstone, 1980). This model permits the alternatives to be divided into groups within which the stochastic terms are correlated. Parameters ρ_r (one for each group) determine the within-group relative to the between-group correlations. Special cases include the MNL ($\rho_r = 1$) and the maximal NL ($\rho_r = 0$). In the latter, originally devised to solve the "red bus - blue bus" paradox (McFadden, 1973), the within-group stochastic utilities are perfectly correlated, indicating that any unobserved traits distinguishing members of this group from each other are completely ignored in choices among groups.

For example, the time-of-day model could be dividing into two groups of alternatives--"not late" and "late"--on the assumption that some unobserved traits would similarly affect all the alternatives within a group. Somewhat more arbitrarily, the auto ownership possibilities could be divided into several groups, such as "0" and "1 or more." Sheffi (1979) has proposed a model for cases such as auto ownership where it might plausibly be asserted that anyone choosing n would always prefer $n-1$ to $n-2$, $n-2$ to $n-3$, etc. I show in the Appendix that Sheffi's model, while not itself a random utility model, is the limit of a multi-level NL model (with as many levels as alternatives) as the ρ 's at each level approach 0 in a special way.

All these models require some a priori division of the choice set into subsets. It is apparent from the above examples, however, that a more symmetric treatment of the choice set is appropriate in some cases. Unobserved traits may affect a group of two or more adjacent alternatives simultaneously, but this group cannot be specified in advance and, in fact, varies within the sample. I call this situation one of proximate covariance among the stochastic component of utility. When the alternatives are arrayed along some

dimension which provides a natural ordering, the stochastic terms for alternatives lying close to each other are more closely correlated than those lying at a greater distance.

The random utility model proposed in this paper, which I call "ordered logit" (OL),¹ allows for proximate covariance among the stochastic utility components. Like the nested logit model, it belongs to the class characterized by McFadden (1978) as "generalized extreme value" (GEV) models. Like the nested logit model, it contains parameters ρ_r determining the pattern of covariances, it reduces to MNL when $\rho_r = 1$ for all r , and it has a special "maximal" structure when $\rho_r = 0$. In fact, one special case has the same mathematical form as the nested logit except that the subsets of alternatives are not mutually exclusive.

Unlike NL, the ordered logit model cannot in general be estimated as a sequence of MNL steps. A more general technique such as maximum likelihood is required.² However, in the special case where the deviations from MNL are small (ρ_r 's all near 1), the model can be estimated using iterations of an ordinary MNL computational routine. The first step is an MNL estimation. The second step is an example of a diagnostic test for IIA departures using the "universal logit method" of McFadden, Train and Tye (1977), in which an MNL

¹This model should not be confused with models in the biostatistics literature in which alternatives are ordered, and a normal or logistic function is fitted to the cumulative probability distribution. Such models are termed "ordered response" by Amemiya (1975), with "ordered normal" and "ordered logistic" special cases; they are more akin to the model for "nested alternatives" proposed by Sheffi (1979) than to that proposed here.

²This may not be such a serious drawback as it first appears, since current research by the author with David Brownstone suggests that sequential estimation of the NL model is often greatly inferior to maximum likelihood estimation.

algorithm is used including variables which are not allowed in a true MNL model because they mix characteristics of different alternatives.

The paper is organized as follows. In Section II, a fairly intuitive special case of the model is presented and shown to have the desired properties. Section III derives a first-order approximation when the departure from IIA is small, and shows that it can be estimated iteratively using an ordinary MNL estimation routine. It proposes this procedure both as a diagnostic test for the seriousness of the IIA violation, and as a practical tool for assessing the biases in the MNL coefficient estimates. Section IV then constructs an example in which the nature of the IIA violation is obvious, and compares predictions of the ordered logit model with those of the MNL and NL models (including Sheffi's special NL). Section V presents the general ordered logit model, and Section VI generalizes it still further for cases where the true choice set is truncated prior to observation. Section VII then describes an empirical implementation to the time-of-day choice, building on the models presented in Small (1982). Some concluding comments follow.

II. A GEV MODEL WITH CORRELATION AMONG ADJACENT ALTERNATIVES

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 $j = 1, \dots, J$ index the set of alternatives, a GEV model is derived from a function $G(y_1, \dots, y_J)$ defined on the orthant $y_j \geq 0$ which is nonnegative, homogenous of degree one, tending toward $+\infty$ when any of its arguments tends toward $+\infty$, and whose n -th partial derivatives (with respect to distinct arguments) are nonnegative for odd n and nonpositive for even n . Any such function defines a cumulative distribution function (cdf)

$$(2.1) \quad F(\varepsilon_1, \dots, \varepsilon_J) = \exp\{-G(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_J})\}$$

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

$$(2.2) \quad F^k(\varepsilon_k) \equiv \lim_{\substack{\varepsilon_j \rightarrow \infty, \\ j \neq k}} F(\{\varepsilon_j\}) = \exp\{-c_k e^{-\varepsilon_k}\}$$

where $c_k = G(\delta_{k1}, \dots, \delta_{kJ})$ and where $\delta_{ij} = 1$ if $i=j$ and 0 otherwise. For a decision-maker maximizing utility

$$(2.3) \quad U_j = V_j + \varepsilon_j$$

the probability (given V_j) of choosing a particular alternative k is

$$(2.4) \quad P_k = e^{V_k G_k} (e^{V_1}, \dots, e^{V_J}) / G(e^{V_1}, \dots, e^{V_J})$$

where G_k denotes the partial derivative of G with respect to its k -th argument.

The MNL model is derived from the function

$$(2.5) \quad G = \sum_{j=1}^J y_j .$$

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

$$(2.6) \quad H_{\text{MNL}}(\varepsilon_j) = \exp\{-e^{-\varepsilon_j}\}$$

and it has the familiar form for choice probabilities:

$$(2.7) \quad P_k = \frac{e^{V_k}}{\sum_{j=1}^J e^{V_j}} .$$

The two-level nested logit model results from the function

$$(2.8) \quad G = \sum_{r=1}^R \left(\sum_{j \in B_r} y_j^{1/\rho_r} \right)^{\rho_r}$$

where $B_r \subset \{1, \dots, J\}$ is one of an exhaustive and mutually exclusive collection of R subsets of alternatives. Its cdf is a product of R multivariate extreme value cdf's each of the form

$$(2.9) \quad H_{\text{NL}}^{\rho_r}(\{\varepsilon_j | j \in B_r\}) = \exp\left\{-\left(\sum_{j \in B_r} e^{-\varepsilon_j/\rho_r}\right)^{\rho_r}\right\} .$$

The NL choice probabilities can be written as products of the probability of choosing a group and the conditional probability of choosing an alternative from within that group:

$$(2.10) \quad P_k = \frac{e^{V_k/\rho_s}}{\sum_{j \in B_s} e^{V_j/\rho_s}} \cdot \frac{e^{\rho_s I_s}}{\sum_{r=1}^R e^{\rho_r I_r}}$$

where B_s is the subset containing k and where

$$(2.11) \quad I_r = \log \sum_{j \in B_r} e^{V_j / \rho_r}$$

defines the inclusive value of subset B_r . Note that each factor in (2.10) has the MNL form.

Alternatives within a group B_r in the NL model have stochastic terms which are correlated, with a correlation coefficient inversely related to ρ_r . We wish to exploit this property for the case where any two alternatives are correlated only if they are adjacent along a natural ordering. To accomplish this using the simplest possible model, let the alternative labels j increase along this natural ordering, and define a GEV model as follows:

DEFINITION 2.1: The Simple Ordered Logit (SOL) model of discrete choice is the GEV model resulting from the function

$$(2.12) \quad G(y_1, \dots, y_J) = \sum_{r=1}^{J+1} \left(\frac{1}{2} y_r^{1/\rho} + \frac{1}{2} y_{r-1}^{1/\rho} \right)^\rho$$

where ρ is a constant satisfying $0 < \rho \leq 1$, and where by convention $y_0 \equiv y_{J+1} \equiv 0$.

It is easy to verify:

PROPOSITION 2.1: The function defined in (2.12) satisfies the conditions in Theorem 1 of McFadden (1978) for a GEV model.

PROOF: These conditions, stated at the beginning of this section, can be checked directly. The oscillating sign of the partial derivatives follows from the fact that the n -th partial derivative is a positive number multiplied by $\rho(\rho-1)\dots(\rho-n)$. QED

The distribution function for the model is easily found from (2.1). It has, from (2.2), a univariate marginal distribution $F^k(\epsilon_k)$ which is extreme value with parameter $c_k = 2^{1-\rho}$.

PROPOSITION 2.2: Given the SOL model defined on choice set $\{1, \dots, J\}$, any two distinct stochastic utility elements ε_j and ε_k are independent unless $|j-k| = 1$, in which case they have a bivariate marginal cdf which is a product of two univariate cdf's of the MNL form (2.6) and one bivariate cdf of the NL form (2.9) all to the power $2^{-\rho}$.

PROOF: Both statements follow by letting all but two arguments in F tend to $+\infty$. For $k = j+1$, the result is

$$(2.13) \quad \begin{aligned} H_{\text{SOL}}(\varepsilon_j, \varepsilon_{j+1}) &= H_{\text{MNL}}(\varepsilon'_j) \cdot H_{\text{MNL}}(\varepsilon'_{j+1}) \cdot H_{\text{NL}}^\rho(\varepsilon'_j, \varepsilon'_{j+1}) \\ &= [H_{\text{MNL}}(\varepsilon_j) \cdot H_{\text{MNL}}(\varepsilon_{j+1}) \cdot H_{\text{NL}}^\rho(\varepsilon_j, \varepsilon_{j+1})]^{2^{-\rho}} \end{aligned}$$

where $\varepsilon'_i = \varepsilon_i + \rho \cdot \log(2)$. QED

The choice probabilities for the SOL model can be written from (2.4) as

$$(2.14) \quad P_k = \frac{e^{v_k/\rho} [(e^{v_{k-1}/\rho} + e^{v_k/\rho})^{\rho-1} + (e^{v_k/\rho} + e^{v_{k+1}/\rho})^{\rho-1}]}{\sum_{r=1}^{J+1} (e^{v_r/\rho} + e^{v_{r-1}/\rho})^\rho}$$

with the convention that $e^{v_r/\rho} = 0$ for $r < 1$ or $r > J$. These may also be written in the suggestive form

$$(2.15) \quad P_k = q(k; B_k) Q(B_k) + q(k; B_{k+1}) Q(B_{k+1})$$

where

$$(2.16) \quad q(k; B_s) = (e^{v_k/\rho}) / e^{I_s}$$

$$(2.17) \quad Q(B_s) = (e^{\rho I_s}) / \sum_{r=1}^{J+1} e^{\rho I_r}$$

$$(2.18) \quad I_r = \log(e^{v_r/\rho} + e^{v_{r-1}/\rho})$$

This form looks like an expansion of P_k into conditional probabilities of the MNL form, but it is not because the sets $B_k = \{k, k-1\}$ are not mutually exclusive. We cannot perform the analogue of the sequential estimation of the NL model since we cannot observe which B_k 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 or any other general technique to obtain estimates of ρ and of the unknown parameters in the V_j .

Intuitively, the SOL model causes the choice probability of a given alternative to be diminished (as compared to the MNL model) if it is adjacent to a very attractive alternative, for then one or both of the q 's in (2.15) will be small. One way to see this more clearly is to consider two extremes. At the extreme $\rho = 1$ it is easy to see that the model is just the MNL model. At the opposite extreme, as $\rho \rightarrow 0$, each of the quantities $(e^{V_r/\rho} + e^{V_{r-1}/\rho})^{-\rho}$ tends to $\exp(\text{Max}\{V_r, V_{r-1}\})$; ρI_r tends to $\text{Max}\{V_r, V_{r-1}\}$; and

$$(2.19) \quad P_k \xrightarrow{\rho \rightarrow 0} n_k e^{V_k} / \sum_{j=1}^J n_j e^{V_j}$$

where n_j is the number of pairs of adjacent alternatives within which alternative j has the largest strict utility V_j . A tie between V_j and V_{j+1} contributes $1/2$ to both n_j and n_{j+1} ; while the pairs $(0,1)$ and $(J, J+1)$ contribute 1 to n_1 and n_J , respectively. In this limit the probability of choosing any alternative dominated by both its neighbors is 0 . This limit is in fact a valid random utility model whose cdf is continuous but whose density function is discontinuous at points $\epsilon_j = \epsilon_{j+1}$; its bivariate marginal for two adjacent alternatives is

$$H(\epsilon_j, \epsilon_{j+1}) = H_{\text{MNL}}(\epsilon_j) \cdot H_{\text{MNL}}(\epsilon_{j+1}) \cdot H_M(\epsilon_j, \epsilon_{j+1})$$

where H_M is the bivariate cdf for two variates grouped together in the maximal model, obtained from (2.9) by letting ρ_r tend to 0:

$$H_M(\varepsilon_j, \varepsilon_{j+1}) = \exp(-e^{-\text{Min}\{\varepsilon_j, \varepsilon_{j+1}\}}).$$

The limiting case (2.19) is more useful for understanding the behavior of the model than for actual estimation, because it is rare that a data set could fit a model which predicts a zero probability for one or more alternatives. More likely, the estimated coefficients would imply $(V_r - V_{r-1}) \rightarrow 0$ as $\rho_r \rightarrow 0$ in such a way that $(V_r - V_{r-1})/\rho_r$ remains finite. The example in the next section has this property. In this way an estimated SOL may approach a limit as $\rho_r \rightarrow 0$ quite different from (2.19). The process is analogous to the manner demonstrated in the Appendix in which a particular (J-1)-level nested logit model approaches the model proposed by Sheffi.

One feature of the SOL is that each of the end-point alternatives, 1 and J, are treated as though they are adjacent to an alternative with utility $-\infty$. This makes sense if the choice set truly represents all available choices, for then it could be argued that any unobserved preferences pushing an individual toward the low or high end of the scale would be likely to land him at the end point. In practice, however, there may be no logical end to the ordering and the end points may be chosen as a matter of convenience to include all or most of the observations. In such a situation, the SOL model would introduce a selectivity bias due to truncation of the dependent variable, a bias absent in the MNL model due to its IIA property.¹ The proper estimation of the SOL (or the more general OL) model in such a case is explained in Section VI.

¹In general, MNL estimates on choice-based samples are not consistent: see Manski and Lerman (1977); Manski and McFadden (1981). However, when certain alternatives are sampled with probability zero and are removed from the hypothesized choice set, the resulting MNL estimates are consistent as proved in McFadden (1978), pp. 87-91.

III. APPROXIMATION WHEN THE MODEL IS ALMOST MULTINOMIAL LOGIT

As already noted, the SOL model of the previous section cannot in general be estimated using computer software designed for MNL. However, it can when the departure from MNL is small in the sense that ρ is near one. As shown below, this can be done by first estimating the MNL model; then reestimating with the addition of a variable whose value for alternative j involves the utilities (estimated in the first step) of alternatives $j-1$ and $j+1$; and reiterating. A variable which mixes traits of different alternatives would not be allowed in a true MNL model based on random utility maximization, and in fact can serve as a diagnostic test for departures from MNL (McFadden, Train, and Tye, 1977). Thus the particular variable described below can be added as a test for the particular type of departure from MNL embodied in the SOL model, a test which requires only one iteration; its coefficient after further iterations provides an estimate of the SOL parameter $1-\rho$.

The intuition behind the procedure is not hard to grasp. The SOL model with ρ slightly less than one differs from the MNL model in that individuals have unobserved preferences for groups of adjacent alternatives as well as for specific alternatives. Compared to MNL, the SOL choice probability for a given alternative will be slightly diminished if it is adjacent to an alternative with a high utility, since the latter is likely to absorb a disproportionate share of an unobserved preference common to both alternatives. Thus a pseudo-variable whose value varies inversely with the estimated utilities of adjacent utilities should receive a positive coefficient, which is larger the more ρ differs from one. By using a first-order Taylor approximation to the SOL choice probabilities, valid for small $\sigma \equiv 1-\rho$, we can choose this variable so that its coefficient is an estimate of σ itself.

It is desired to approximate the choice probability (2.14) by a form which looks like the MNL choice probability (2.7) with an additional term σN , where

$N = (N_1, \dots, N_J)$ is the new pseudo-variable:

$$(3.1) \quad P_k \approx \frac{e^{V_k + \sigma N_k}}{\sum_{j=1}^J e^{V_j + \sigma N_j}}.$$

This can be accomplished by approximating both (2.14) and (3.1) by the first two terms in a Taylor Series expansion about the point $\sigma = 0$:

$$(3.2) \quad P_k \approx P_k^0 [1 + \sigma (\partial \log P_k / \partial \sigma)^0]$$

where the superscript o indicates evaluation at $\sigma = 0$. Since (2.14) and (3.1) both reduce to the same thing when $\sigma = 0$, their approximations will be identical if N is chosen to make the terms $(\partial \log P_k / \partial \sigma)^0$ identical. Computing this derivative from (3.1) and (2.14) respectively and equating, we thus require that

$$N_k - \sum_j P_j^0 N_j = (V_k - \bar{I}_k^0) - \sum_{j=1}^J P_j^0 (V_j - \bar{I}_j^0)$$

where

$$(3.3) \quad \bar{I}_j^0 \equiv \frac{1}{2} (I_j^0 + I_{j+1}^0)$$

and I_r^0 is obtained from (2.18) by setting $\rho = 1$ (recall the convention $e^{V_0} = e^{V_{J+1}} = 0$). Thus, the desired pseudo-variable N has values

$$(3.4) \quad N_j = V_j - \bar{I}_j^0$$

$$(3.5) \quad = -\frac{1}{2} [\log(\frac{1}{2}) + \log(1 + e^{V_{j-1} - V_j}) + \log(1 + e^{V_{j+1} - V_j})]$$

$$(3.6) \quad = -\frac{1}{2} [\log(\frac{1}{2}) + \log(1 + P_{j-1}^0 / P_j^0) + \log(1 + P_{j+1}^0 / P_j^0)]$$

with the convention $P_0 = P_{J+1} = 0$. The form (3.5) shows that N has the

interpretation described above.

The variable N is not observed, but depends on the unknown parameters in the "strict utilities" V_j . Let

$$(3.7) \quad V_j = \beta' z_j$$

where β is the vector of these unknown parameters, β' is its transpose, and z_j is a vector of observable traits of alternative j (possibly depending on observable traits of the sample member as well). The log-likelihood function for (3.1) is a sum of terms, one for each member of the sample, of the form

$$(3.8) \quad L(\beta, \sigma) = \beta' z_k + \sigma N_k(\beta) - \log \sum_{j=1}^J \exp[\beta' z_j + \sigma N_j(\beta)]$$

where k indexes the alternative chosen by that sample member. We can maximize this likelihood function on an ordinary MNL computer package in a stepwise fashion. First, use MNL to obtain an initial estimate of β . Second, use this estimate to construct an estimate of N , and use the MNL algorithm with this pseudo-variable N to obtain an estimate of σ and a new estimate of β . Continue iterating the second step until the estimates are stable. This should converge to the maximum-likelihood estimate of (3.1) if the model is reasonably well behaved and σ is small, though as far as I know such convergence is not guaranteed. As shown in the Appendix, the standard error computed by the MNL package for σ is correct, but those for β should be corrected as shown.

To test for the validity of the MNL model against the more general SOL, only the first iteration need be carried out. Given the null hypothesis $\sigma = 0$, the initial estimate for β is consistent, thus so are the subsequent estimates of N and σ . The asymptotic t -statistic for σ then provides a test of the null hypothesis; this test is an example of the "universal logit" technique for detecting departures from the MNL model (McFadden, Train, and Tye, 1977).

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 degree of fit and for ability to predict results of exogenous changes.

Suppose a household can own zero, one, or two autos; denote these discrete alternatives by $j = 1, 2, \text{ and } 3$, 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 three ownership alternatives; 5% of households are small and choose $j=1$, while 5% are large and choose $j=3$. Table 4.1 shows the observed choice frequencies in a sample of N households, along with a set of utilities for each group compatible with the above description.

The problem is to fit a choice model to these observed frequencies when household size is not observed. Whatever "strict utilities" V_j are fit by the model, there will be an unobserved utility component $\varepsilon_j = U_j - V_j$ which varies within the sample with a substantial correlation (.688) between ε_1 and ε_2 and between ε_2 and ε_3 ; but only a small correlation (-.053) between ε_1 and ε_3 . This is just the kind of error structure for which the simple ordered logit model is appropriate (though not exact). Note that 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 . The SOL model of Section II 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 better predict the results of removing or adding alternatives. The MNL model, in contrast, falsely assigns larger utilities to alternatives 1 and 3, thereby tending to predict too small a

TABLE 4.1

Parameters for Numerical Example

	<u>Alternative (j)</u>			<u>Total</u>
	1	2	3	
Utility (U_j)				
Small households	2	1	0	
Medium households	0	0	0	
Large households	0	1	2	
Choice Frequency (f_j)				
Small households	.05N	0	0	.05N
Medium households	.30N	.30N	.30N	.90N
Large households	0	0	.05N	.05N
Total (observed)	.35N	.30N	.35N	N

share for alternative 2 when conditions are changed. The NL model falsely introduces an asymmetry to the utility structure; it predicts well at one extreme but poorly at the other.

Each of the models will be considered in two forms. In the non-generic (alternative-specific) form, two free parameters V_1 and V_3 are allowed, with V_2 normalized to 0. In the generic form, the V_j are constrained by the one-parameter function

$$(4.1) \quad V_j = \alpha(j-2)$$

which postulates that strict utility varies linearly with number of cars owned (again, normalized so that $V_2=0$). Since household characteristics are not observed, the log-likelihood function given the observed frequencies f_j is simply

$$(4.2) \quad L = N \sum_{j=1}^3 f_j \log(P_j)$$

where P_j are the predicted choice probabilities given values for the free parameters. Since there are only two independent observations, f_1/f_2 and f_3/f_2 , there can be at most two free parameters; when there are two, L is maximized by setting at $\hat{P}_j = f_j$. Thus in the non-generic NL and SOL models the additional parameter ρ is treated as exogenous, whereas in the generic forms it is estimated along with α .

To compare predictive ability, two scenarios are used throughout: (A) removal of alternative 3 from the choice set, for example by prohibiting on-street parking in a neighborhood of two-car garages; and (B) removal of alternative 1 from the choice set, for example by making all non-automotive forms of travel infeasible. I shall use the predicted share of alternative 2, denoted P_2^A and P_2^B respectively, as the basis for comparison. The true result of either of these situations 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. Thus either scenario results in an increase of .20 in the number of households choosing alternative 2. For the generic models, a third scenario is also used: (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. Extending the pattern of Table 4.1, small households are assumed to regard owning three cars as even worse than owning two; large households regard it as better; and medium-sized ones are indifferent. Thus the true prediction of the share garnered by this new alternative is $P_4^C = .275$ (.225 from the medium-sized and .050 from the large households).

The results of these experiments are summarized in Table 4.2 and discussed in the rest of this section.

Non-generic Models

The MNL likelihood function is maximized at $V_1 = V_3 = .154$. As is well known, when an alternative is removed from the choice set the model predicts that all other choice probabilities change proportionally. Thus when either of the extreme alternatives is removed, the remaining probabilities are both predicted to rise by a factor of $1/(1-.35)$. This understates the true rise in P_2 by nearly .04 or 20 percent.

To fit an SOL model to the observed choice frequencies, note that the symmetry of the problem requires $V_1 = V_3$. This leads to a single equation in V_1 which, given any value of ρ between 0 and 1, can be solved numerically. Given the resulting estimate of $V_1 = V_3$, prediction of scenario A or B is made by letting V_1 or V_3 , respectively, go to $-\infty$. The resulting predictions for P_2^A and P_2^B range from the MNL value of .46 (at $\rho=1$) to .53 (as $\rho \rightarrow 0$). The true value of .50 is predicted when $\rho = .5850$; this is the value at which all three strict

TABLE 4.2

NUMERICAL EXAMPLE

	Estimated Parameters		Fitted Shares		Predicted Shares		
	\hat{V}_1	\hat{V}_3	P_1	P_2	P_2^A	P_2^B	P_4^C
<u>TRUE VALUES</u>			.35	.30	.50	.50	.275
<u>NON-GENERIC MODELS:</u>							
MNL	.154	.154	.35	.30	.46	.46	
SOL							
$\rho = .7$.034	.034			.49	.49	
$\rho = .5850$.000	.000			.50	.50	
$\rho = .3$	-.046	-.046			.52	.52	
$\rho \rightarrow 0$.000	.000 ^a			.53	.53	
NL grouping (2,3)							
$\rho = .8006$.000	.123			.50	.46	
$\rho = .5$	-.232	.077			.56	.46	
$\rho \rightarrow 0$ (Sheffi)	-.619	.000 ^b			.65	.46	
NL grouping (1,3)							
$\rho = .5$.501	.501			.38	.38	
$\rho \rightarrow 0$.847	.847			.30	.30	
<u>GENERIC MODELS:</u>							
	$\hat{\rho}$	$\hat{\alpha}$					
MNL	-	0	.33	.33	.50	.50	.250
SOL	.5850	0	.35	.30	.50	.50	.269
SOL, first-order approx.	.5552	0			.50	.50	.269
NL grouping (2,3)	.6675	.103			.53	.46	
Sheffi	-	.431	.39	.24	.61	.39	.223

$${}^a \lim_{\rho \rightarrow 0} (V_3/\rho) = \lim_{\rho \rightarrow 0} (V_1/\rho) = -.405.$$

$${}^b \lim_{\rho \rightarrow 0} (V_3/\rho) = .154; \text{ this is the estimated parameter } W_3 \text{ in the Sheffi model.}$$

utilities are equal, in which case the differences among the observed frequencies are correctly explained as artifacts of the error distribution.

Fitting a nested logit model is accomplished similarly, leading to the estimates of V_1 and V_3 shown for the NL with $B_1 = \{1\}$ and $B_2 = \{2,3\}$. Predictions of the share P_2^A after deleting alternative 3 range from .46 at $\rho=1$ (the MNL value) to .65 at $\rho \rightarrow 0$. This model also has a value of ρ , namely .8006, for which the correct prediction for P_2^A is obtained. However, for scenario B the model predicts the same as MNL: the odds P_3/P_2 remain unchanged. Conversely, the NL model $B_1 = \{1,2\}$ and $B_2 = \{3\}$ can correctly predict scenario B but not A. Thus it is impossible for either model to predict both scenarios correctly. As $\rho \rightarrow 0$, the model shown approaches Sheffi's model with parameter $W_3 = \lim_{\rho \rightarrow 0} (V_3/\rho) = .154$ (see Appendix). By assuming that all households which would otherwise have chosen alternative 3 prefer 2 to 1, Sheffi's model considerably overpredicts the shift to alternative 2 when alternative 3 is deleted; like the MNL model, it underpredicts the shift when alternative 1 is deleted.

The other possible NL tree structure, which groups alternatives 1 and 3, has little plausibility and its predictions are worse than the MNL model.

Generic Models

With non-generic models, we saw that there is some value of ρ for which the SOL model is superior in predictive ability. However, nothing could be said about whether a reasonable value of ρ would be obtained by fitting to the original data. To remedy this, we now consider models with one fewer free parameter, so that ρ can be estimated.

The MNL and SOL models must, by symmetry, achieve a maximum likelihood at $\hat{\alpha} = 0$ (it is easy to check that this is a maximum, not a minimum). Fortuitously, the MNL model with this estimate of α correctly predicts the new shares P_2^A and P_2^B ; however, since it incorrectly predicts them in the original state, it is still not a perfect tool for stimulating these scenarios (it underpredicts the increase in number of households owning one car by 17 percent). The SOL model, having an additional parameter, achieves a higher log likelihood; in fact it can fit the observed shares exactly. What is of more interest is that it does so at precisely that value of ρ for which it correctly predicts both P_2^A and P_2^B .

Also shown in the table is the result of estimating the SOL model by the first-order approximation procedure of Section III. In this procedure, the MNL is first estimated, with the result $\hat{\alpha} = 0$. This value is then used to compute the variable N from equation (3.5), resulting in

$$(4.3) \quad N_j = \begin{cases} 0 & j=1,3 \\ -\frac{1}{2} \log(2) & j=2. \end{cases}$$

This variable assigns a larger value to alternatives 1 and 3, reflecting their positions immediately adjacent to a "neighbor" with implicit utility of negative infinity. The likelihood function formed from (3.1) and (4.2) is maximized at $\hat{\alpha} = 0$ and $\hat{\rho} = 1 - \hat{\alpha} = .5552$. This is slightly less than the value of ρ obtained from the unapproximated estimate. Yet when the predictions are carried out using the same approximation, by setting V_1 or V_3 equal to $-\infty$ and recalculating (3.5) and (3.1), the predictions are exactly the same as those of the exact SOL model. 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.

The nested logit model (with alternatives 2 and 3 grouped) also gives an exact fit to the observed frequencies, with $\rho = .6675$ and a small positive value for α . However, it does not predict scenarios A or B as well as the SOL. The Sheffi model was also fitted, with W_j replacing V_j in equation (4.1); its predictions are worse than any of the other generic models.

A final comparison, involving scenario C, is considered to further illustrate the improvement of the SOL over the MNL for this example. Because it does not recognize the "piling up" of frequency at the extreme alternatives caused by the unobserved systematic tastes, the MNL model underpredicts the shift into a newly-allowed alternative $j=4$ by about 9 percent. SOL does better, underpredicting by 2 percent. Such a prediction cannot be made in general for nested logit models, because there is no obvious way to assign a new tree structure. The exception is the Sheffi model, which specifies that the new mode be grouped with mode 3 as a new level in the tree; however, it does worse than the MNL in this prediction test.

In conclusion, the SOL seems to perform better than any of the other models considered here for this example. The reason is the existence of an omitted variable systematically related to the alternative number, making the MNL assumptions invalid; and the fact that this variable is symmetric about the middle alternative, which makes the NL models unsuitable. The appropriateness of the SOL in a real situation depends on how closely it mirrors traits such as these.

V. A GENERAL MODEL WITH PROXIMATE COVARIANCE: ORDERED LOGIT

The model presented thus far can be generalized in four ways. First, the stochastic term associated with an alternative may be correlated with nearby but not immediately adjacent alternatives. This is dealt with by including $M+1$ terms rather than just two within each of the parentheses of equations (2.12) and (2.14). Second, the correlation between nearby alternatives j and k may decline slowly or rapidly with $|k-j|$, or even show an uneven pattern (for example an especially high correlation between j and $j+2$ might result from roundoff effects). This is dealt with by multiplying the M terms just mentioned by a set of weights. Third, the correlation between nearby alternatives may not be the same at one end of the scale as at the other end. This is accounted for by allowing ρ to depend on r in the summations. The model incorporating these three generalizations is what I call the Ordered Logit model, and is presented in this section. The fourth generalization is described in the next section.

DEFINITION 5.1: The Orderet Logit (OL) model of discrete choice is the GEV model resulting from the function

$$(5.1) \quad G(y_1, \dots, y_J) = \sum_{r=1}^{J+M} \left(\sum_{m=0}^M w_m y_{r-m} \right)^{1/\rho_r} \rho_r$$

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

$$(5.2) \quad 0 < \rho_r \leq 1, \quad r = 1, \dots, J+M$$

$$(5.3) \quad w_m \geq 0, \quad m = 0, \dots, M$$

$$(5.4) \quad \sum_{m=0}^M w_m = 1,$$

and where by convention $y_j \equiv 0$ for $j \notin \{1, \dots, J\}$. I will also adopt the convention that

$$(5.5) \quad w_m = 0, \quad m \notin \{0, 1, \dots, M\}.$$

PROPOSITION 5.1: The OL model is a model of the GEV class which reduces to the MNL when $\rho_r = 1$ for all r , and to the SOL when $M = 1$, $\rho_r = \rho$ for all r , and $w_0 = w_1 = 1/2$.

PROOF: G is non-negative because of the condition (5.3). Its homogeneity of degree one can be checked directly. Since (5.3) and (5.4) ensure that at least one of the weights is positive, G tends to infinity when any of its arguments does. Its partial derivatives alternate in sign just as in the SOL model. (5.4) guarantees that when $\rho_r = 1$ for all r , G reduces to (2.5); it can be checked directly that it reduces to (2.12) when the conditions in the last statement are met. QED

The next proposition states that 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 5.2: Each of the stochastic utility elements ε_k in the OL model has a univariate marginal distribution which is extreme value according to (2.2) with parameter $c_k = \sum_{m=0}^M w_m \rho_{k+m}$. Any two distinct stochastic utility elements ε_j and ε_k are independent unless $|j-k| \leq M$. The bivariate marginal distribution of ε_j and ε_k for $0 < k-j \leq M$, provided that $a_m > 0$ for all $m \leq M$, has a cdf given by

$$(5.6) \quad H_{OL}(\varepsilon_j, \varepsilon_k) = H_{MNL}(\varepsilon_j - \log w_{jk}^A) \cdot H_{MNL}(\varepsilon_k - \log w_{jk}^B) \\ \cdot \prod_{r=k}^{j+M} H_{NL}^{r}(\varepsilon_j - \rho_r \log w_{r-j}, \varepsilon_k - \rho_r \log w_{r-k})$$

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

PROOF: The proof follows directly by calculating (2.1) and letting the appropriate arguments tend to infinity. QED.

The choice probabilities are obtained directly from (2.4):

$$(5.7) \quad P_k = \frac{\sum_{r=k}^{k+M} w_{r-k} e^{V_k/\rho_r - (1-\rho_r)I_r}}{\sum_{r=1}^{J+M} e^{\rho_r I_r}}$$

where

$$(5.8) \quad I_r = \log \sum_{m=0}^M w_m e^{r-m/\rho_r}.$$

P_k is a sum of $M+1$ terms, each resembling the choice probability for a nested logit model. It has the limiting case (2.19) except that n_j is now the number of subsets of $M+1$ adjacent alternatives within which j has the largest strict utility ($0 \leq n_j \leq M$).

It is evident that many new parameters have been introduced in generalizing the Simple Ordered Logit model. In practice, it is difficult to think of a situation in which one would want more free parameters ρ_r than alternatives, and in most applications some strong conditions should be imposed on the w_m and ρ_r . One obvious such condition is a set of equality restrictions on ρ_r , though perhaps less stringent than forcing them all to be equal as in the SOL. Another is to specify constant or geometrically declining weights w_m . Both of these approaches are taken in the empirical work reported in Section VII.

As with the SOL model, the model proposed here has a convenient approximate form when the deviation from MNL (as measured by $1-\rho$) is small. The derivation parallels that in Section III, and the result is identical to (3.1) if σ and N

are now interpreted as vectors of parameters σ_r and variables N^r , the latter taking the form (for alternative j):

$$(5.9) \quad N_j^r = w_{r-j} (V_j - I_r^0)$$

$$(5.10) \quad = -w_{r-j} \log \sum_{m=0}^M w_m (P_{r-m}^0 / P_j^0).$$

Restrictions on $\{\sigma_r\}$ can be dealt with by replacing the complete set $\{N^r\}$ with a smaller number of combinations. For example, the equality restriction $\sigma_r = \sigma$ for $r \in B$ is imposed simply by including one variable $N^B = \sum_{r \in B} N^r$ instead of the separate variables N^r . Provided the weights are specified in advance, this approximate model can be estimated iteratively using an MNL estimation routine, and the null hypothesis $\sigma_r = 1$ for all r can be tested using the asymptotic standard errors computed by such a routine. Section VII provides an example.

VI. DEPENDENT VARIABLE FROM A TRUNCATED CHOICE SET

As noted earlier, GEV models cannot in general be consistently estimated from observations on only a subset of the available alternatives. The Ordered Logit model, however, does provide a way of incorporating unobserved portions of the choice set. The model can then be estimated provided some structure can be specified for the strict utilities of the unobserved alternatives. This structure might follow naturally from the form of the independent variables if they are generic; or, at the cost of some additional parameters, a structure can be specified. It turns out that if the true model is OL, the truncated data for J alternatives can be treated as OL with the following extension.

DEFINITION 6.1: The Extended Ordered Logit (EOL) model of discrete choice is the GEV model resulting from the function (5.1) with w_m replaced by

$$(6.1) \quad w_{rm} = w_m e^{-\sigma_r a_r}$$

where a_r are constants and $\sigma_r = 1 - \rho_r$.

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 (5.6) and (5.7) are replaced by straightforward generalizations, the latter taking the form

$$(6.2) \quad P_k = \frac{\sum_{r=k}^{k+M} w_{r-k} e^{-\sigma_r a_r V_k / \rho_r e^{-(1-\rho_r) I'_r}}}{\sum_{r=1}^{J+M} \rho_r I'_r}$$

with

$$(6.3) \quad I'_r = I_r - \sigma_r a_r .$$

The approximation for ρ_r near one is similar to that for OL but involves additional pseudo-variables A^r :

$$(6.4) \quad P_k \cong \frac{\exp[V_k + \sum_r \sigma_r N_{r-k}^r + \sum_r (\sigma_r a_r) A_{r-k}^r]}{\sum_{j=1}^J \exp[V_j + \sum_r \sigma_r N_{r-j}^r + \sum_r (\sigma_r a_r) A_{r-j}^r]}$$

where N_j^r is given by (5.10), and where

$$(6.5) \quad A_j^r = -w_{r-j} .$$

It should be noted that the pseudo-variables A^1 and N^1 are collinear, as are A^{J+M} and N^{J+M} .

PROPOSITION 6.1: Suppose the true choice model is Ordered Logit with the alternative set $\tilde{B} = \{-J_1+1, \dots, 0, 1, \dots, J+J_2\}$ and with strict utilities \tilde{V}_j , where J_1, J_2 are positive integers. Let $B = \{1, \dots, J\}$. Then for $k \in B$, the choice probability P_k conditional on some alternative from the subset B being chosen is given by an EOL model with strict utilities

$$(6.6) \quad V_j = \begin{cases} \tilde{V}_j & 1 \leq j \leq J \\ -\infty & \text{otherwise} \end{cases}$$

and with parameters

$$(6.7) \quad a_r = \begin{cases} 0 & , \quad M+1 \leq r \leq J \\ (1/\rho_r) \log \left[1 + \frac{\sum_{j \in B} w_{r-j} e^{\tilde{v}_j / \rho_r}}{\sum_{j \in B} w_{r-j} e^{\tilde{v}_j / \rho_r}} \right] & , \quad \text{otherwise} \end{cases}$$

PROOF: The true choice probabilities, given by the OL probabilities (5.7) with utilities \tilde{v}_j and extended choice set \tilde{B} , are

$$(6.8) \quad \tilde{p}_k = \frac{\sum_{r=k}^{k+M} w_{r-k} e^{\tilde{v}_k / \rho_r} e^{-(1-\rho_r) \tilde{I}_r}}{\sum_{r=-J_1+1}^{J+J+M} w_r e^{\rho_r \tilde{I}_r}}$$

where

$$(6.9) \quad \tilde{I}_r = \log \sum_{m=0}^M w_m e^{\tilde{v}_{r-m} / \rho_r}$$

$$(6.10) \quad = I_r + \rho_r a_r$$

with I_r given by (5.8) and a_r by (6.7). For $k \in B$, $V_k = \tilde{v}_k$ and the conditional probability is

$$(6.11) \quad p_k = \frac{\tilde{p}_k}{\sum_{j=1}^J \tilde{p}_j} = \frac{\sum_{r=k}^{k+M} w_{r-k} e^{\tilde{v}_k / \rho_r} e^{-(1-\rho_r) \tilde{I}_r}}{\sum_{j=1}^J \sum_{r=j}^{j+M} w_{r-j} e^{\tilde{v}_j / \rho_r} e^{-(1-\rho_r) \tilde{I}_r}}$$

The second exponential in the numerator can be written as

$$e^{-\sigma_r a_r} e^{-(1-\rho_r) \tilde{I}_r}$$

which makes the numerator identical to that of (6.2). Reverse the order of summation in the denominator of (6.11), then change the second summation index to $m = r-j$. Terms with $(r-m)$ outside the allowable range of j are automatically excluded by the convention (6.6), so the denominator becomes

$$\sum_{r=1}^{J+M} \left(\sum_{m=0}^M w_m^v e^{r-m/\rho_r} \right) e^{-(1-\rho_r)\tilde{I}_r} = \sum_{r=1}^{J+M} e^r e^{-(1-\rho_r)\tilde{I}_r}$$

which is easily put in the same form as the denominator of (6.2). Thus, (6.11) is equivalent to (6.2). QED

There are several reasons one might want to use this result rather than estimate (6.8) directly. It may be difficult to specify the strict utilities of unobserved alternatives. It may simply be more convenient to use the EOL model because of an existing computer program, or because of the approximation (6.4). In any case, if the EOL is to be used with this interpretation, (6.7) shows that those a_r not constrained to zero should be positive, and should decline toward zero as r approaches $M+1$ from below or J from above (recall convention (5.5)). In many cases it may be desirable to constrain a_r to be some simple function of r with these properties, so as to reduce the number of new parameters introduced; this approach is taken in the empirical work described in the next section.

VII. EMPIRICAL APPLICATION: TRIP TIMING

The first-order approximations to the models developed in this paper have been applied to the choice of time-of-day of work trips by commuters in the San Francisco Bay area. Data were collected on the actual work arrival times, the "official" work start times, and other characteristics of 527 individuals who commuted by auto in 1972 (McFadden, Talvitie, & Associates, 1977). These were supplemented with engineering calculations of the travel times each would have faced at each of 12 alternative arrival times, spanning a range of 40 minutes before to 15 minutes after the official starting time for his or her job.

In a previous paper (Small, 1982) I have described the rationale for treating these choices as discrete, the main reasons being rounding of the observed choices and analytical difficulties with a continuous model. In that paper, I estimated a series of 12-alternative MNL models. While noting that some departure from the MNL's assumed error structure was likely, a relatively crude test was unable to reject the MNL model.

The most well-behaved specification found in the earlier paper is shown in Table 7.1.¹ This model demonstrates the all-important trade-off between the desire to avoid congestion on the one hand, and the desire to avoid arriving too early or late on the other. The estimated marginal rates of substitution imply that the average non-carpooler would incur .53 minute of travel time to avoid arriving an extra minute early; 1.24 minute to avoid arriving an extra minute late; and an additional 1.53 minute to avoid arriving an extra minute beyond the reported flexibility range.

¹Estimates in the earlier paper were based on a subsample of 453 cases because of missing data in the variable indicating whether the person drives in a car pool. These data were subsequently reconstructed from original sources thus permitting the larger sample used here.

Several OL and EOL models were estimated using this specification of the strict utilities, in order to detect proximate covariance in the stochastic term and to determine its effect on the properties of the estimated model. One might expect the SOL to inadequately capture the covariance pattern, for several reasons. First, nearby but not immediately adjacent alternatives are likely to be correlated: for example, individuals might avoid arriving at any time prior to the (unobserved) opening time of the building in which they work. This would require an OL model with $M > 1$. Second, the asymmetry between early and late arrival suggests that the σ_r might not be identical. This is tested below by allowing σ_r to take on one value σ_E for $r \leq 8$; another value σ_L for $r \geq 10+M$; with a linear transition between these values for intermediate r .¹ Third, the strength of the variable R15 accounting for rounding to the nearest 15 minutes could introduce an even more complicated error structure, in which stochastic terms for alternatives j and $j+3$ are closely correlated. This is tested by specifying non-uniform weights $w_0=w_3 > w_1=w_2=\dots=w_M$.

Finally, the choice set used here was in fact truncated. Alternatives representing arrival more than 40 minutes early or 15 minutes late were deleted prior to the construction of the travel-time variables, and the few individuals choosing such alternatives were deleted from the sample. This suggests the need for the EOL model, which is tested by postulating nonzero $\{a_r\}$ in equation (6.4)

¹Other patterns with even more degrees of freedom were found to lead to severe collinearity problems.

which are exponentially declining¹ in r for $1 \leq r \leq M$, and in $(J+M-r)$ for $J+1 \leq r \leq J+M$, with parameters γ_E and γ_L respectively.

The results of one iteration of the two-step approximate estimator are summarized in Table 7.2. As shown in the appendix, this is sufficient for testing the null hypothesis $\sigma = 0$ by using the asymptotic t-statistic. In contrast, the chi-squared tests described below are only approximate because the true maximum likelihood has not been determined. Furthermore, successive iterations in the few cases tried showed the parameter estimates for the first iteration to be off by as much as 40 percent.²

¹With constant σ_r , this is accomplished by using the two additional variables AE and AL defined in the footnotes to Table 7.1 in the second stage of the linear approximation. The rationale is that the utility \tilde{V}_j of deleted alternatives is smaller the further j is from the range included in the model. For example, suppose $V_1 - \tilde{V}_j = \alpha_E(1-j)$ for $j < 1$; $\rho_r \equiv \rho$; and $w_m \equiv w_0$. Then from (6.7), for $1 \leq r \leq M$,

$$a_r = \frac{1}{\rho} \log \left[1 + \frac{\sum_{j=r-M}^0 (e^{\alpha_E/\rho})^j}{\sum_{j=1}^r (e^{\alpha_E/\rho})^j} \right]$$

If α_E is large, this can be approximated by

$$a_r \approx \frac{1}{\rho} \log \left[1 + \frac{1}{r e^{\alpha_E/\rho}} \right] \approx \frac{1}{\rho} \gamma_E^r, \quad 1 \leq r \leq M$$

with $\gamma_E = e^{-\alpha_E/\rho} < 1$. Similarly, if $V_J - \tilde{V}_j = \alpha_L(j-J)$ for $j > J$, and α_L is large,

$$a_r \approx \frac{1}{\rho} \gamma_L^{(J+M+1-r)}, \quad J+1 \leq r \leq J+M$$

where $\gamma_L = e^{-\alpha_L/\rho}$. On the basis of the coefficients of the MNL model, I chose $\alpha_L = 4\alpha_E$, hence set $\gamma_L = (\gamma_E)^4$ and scanned over several values of γ_E .

²Because of a tendency of parameter estimates to oscillate, each iteration used the average of the initial values (used to construct N) and the final values (estimated) from the previous two-step iteration. Even so, seven iterations were required to achieve convergence to three significant digits in all parameters.

Several of the expectations noted above were borne out. Neither the SOL, the OL, nor the EOL with $M=1$ raised the log likelihood significantly over that achieved by the MNL. In contrast, EOL models with higher values of M worked quite well; the one labelled "best fit" had $M=5$ and achieved a log likelihood which exceeded that of the MNL model by 4.98. The corresponding chi-squared statistic (9.96) is significant at the 2 percent level if we count only the three extra degrees of freedom actually appearing as estimable parameters. More honestly, there are about 5 degrees of freedom because the model shown was the best obtained by scanning over several values of M and γ_E ; even so, the model is significantly different from MNL at a 10 percent level, and the estimated σ is greater than zero at a 2 percent level (1-tailed test). Non-uniform weights w_m were not necessary or even helpful. Neither the pattern described above nor an exponentially declining pattern gave as high a log likelihood as uniform weights. Nor is there much evidence of σ_r varying with r : the two-parameter specification described above yielded nearly identical estimates of σ_E and σ_L and a log likelihood only 0.13 greater than under the constraint $\sigma_E = \sigma_L$.

The only problem with the "best-fit" model is the unreasonably large estimate of a_{J+M} . As a crude approximation, suppose only the deleted alternatives immediately adjacent to the choice set have utility greater than $-\infty$. Denoting these by \tilde{V}_0 and \tilde{V}_{J+1} , (6.7) can be solved for $(\tilde{V}_0 - V_1)$ and $(\tilde{V}_{J+1} - V_J)$ in terms of the estimated parameters. The results are 0.41 and 0.93, respectively. The first of these is plausible: $j=0$ represents arriving at work 45 minutes late, which according to the estimated MNL coefficients would have a utility $(\beta_{R15} - \beta_{R10} + 5 \cdot \beta_{SDE}) = 0.33$ greater than that of V_1 (for non-carpoolers) if its travel time were the same. The result for $(\tilde{V}_{J+1} - V_J)$, however, is totally implausible; it should be substantially negative. For this reason, the model was reestimated forcing $a_{J+M} = 0$. The result is the "Preferred EOL" model; after iterating to convergence as shown in Table 7.1, it differs from the MNL model with a chi-

squared statistic of 5.82, significant at a 25 percent level if we consider there to be four degrees of freedom. This seems a conservative assessment, given that two of these degrees of freedom were used up in scanning over only a few discrete values of M and γ_E ; that the resulting model has parameter estimates of the right signs and of plausible magnitude (indeed, better fitting models were rejected on plausibility grounds); and that a one-tailed t-test of either $\sigma = 0$ or $a_1 = 0$ separately would reject the null hypothesis at a 2 percent significance level. It should also be noted that the estimated σ is closer to one than to zero, casting doubt on the accuracy of the first-order approximation for this case.

Table 7.1 shows several properties of the parameter estimates in order to assess the quantitative differences between the MNL and the Preferred EOL models. The marginal rates of substitution of travel time for scheduling inconvenience all tend to be overpredicted, by up to 12 percent, by the MNL model; that is, the EOL model would predict an even greater tendency for commuters to shift schedules in response to changes in congestion. Another way to see this is in the last row of the table, which shows how much the fraction of commuters arriving more than 10 minutes early would rise if the congestion encountered at later arrival times were increased by one minute. This predicted response is about 6 percent greater from the EOL than from the MNL results.

To summarize, there seems a good case that the Extended Ordered Logit model is appropriate for these data. The signs and magnitudes of the new parameters are plausibly estimated, and there were good reasons why the simpler SOL and OL models were not satisfactory. The first-order approximation to the EOL is probably crude, though the example in Section IV suggests that it still may predict well. Estimated by using the two-step procedure on an ordinary computer package, its relative ease of use recommends it for testing departures from the MNL model. But for obtaining accurate parameter estimates, the need for several iterations is a drawback. For models such as this one with many

alternatives, these iterations are expensive and cumbersome; it would probably be more fruitful to go directly to a maximum likelihood program for the full EOL model. For smaller applications, iterations on existing MNL routines are an attractive option. Finally, in this example the MNL results are qualitatively correct and quantitatively quite satisfactory, despite the large departure from MNL indicated by the estimated parameters of the more general model.

TABLE 7.1

ESTIMATED COEFFICIENTS: TRIP-TIMING MODELS^a(asymptotic t-statistics^b in parentheses)

First-order Approximation

VARIABLE ^c	MNL	PREFERRED EOL	
		Iter. No. 1	Iter. No. 7 (convergence)
Rounding error:			
R15	1.106 (10.97)	.363 (0.96)	.595 (2.37)
R10	.398 (3.92)	.126 (0.72)	.207 (1.49)
Travel time:			
TIM	-.141 (-2.67)	-.104 (-1.84)	-.106 (-1.91)
TIM.CP	.105 (1.39)	.076 (1.00)	.080 (1.05)
Schedule delay, early:			
SDE	-.075 (-12.24)	-.049 (-3.75)	-.051 (-4.37)
SDE.CP	.023 (2.55)	.017 (1.79)	.017 (1.87)
Schedule delay, late:			
SDL	-.175 (-5.99)	-.117 (-2.83)	-.129 (-3.59)
SDLX	-.216 (-2.67)	-.122 (-1.32)	-.143 (-1.65)
D2L	-1.057 (-6.21)	-.683 (-2.84)	-.763 (-3.65)
Pseudo-variables for first-order approximation:			
N	-	.674 (1.96)	.824 (2.11)
AE	-	5.72 (1.99)	6.16 (2.10)
Log likelihood	-994.90	-992.30	-991.99
Marginal rates of sub- stitution (non- carpoolers):			
$-(\partial \text{TIM} / \partial \text{SDE})_V$	0.53	0.47	0.48
$-(\partial \text{TIM} / \partial \text{SDL})_V$	1.24	1.12	1.22
$-(\partial \text{TIM} / \partial \text{SDLX})_V$	1.53	1.18	1.35
Pred. share for alts 1-6:			
\bar{P}_{1-6} (actual = .3776)	.3853	.3781	.3770
$\Delta \bar{P}_{1-6} / \Delta \text{TIM}_{7-12}$ ^d	.0248		.0264

^aDependent variable is the choice among 12 time-of-day alternatives, each denoting an arrival time within a 5-minute interval centered between 40 minutes before and 15 minutes after the time of work officially begins. Sample size = 527.

^bUnder null hypothesis that model is MNL; not corrected as in Appendix.

^cDefinition of independent variables:

SD = Schedule Delay: actual arrival time minus official work start time, in minutes, for a given alternative. Thus its value for alternative j is $SD_j = 5(j-9)$, $j = 1, \dots, 12$.

R15 = $\begin{cases} 1 & \text{if } SD = -30, -15, 0, 15 \\ 0 & \text{otherwise.} \end{cases}$

R10 = $\begin{cases} 1 & \text{if } SD = -40, -30, -20, -10, 0, 10 \\ 0 & \text{otherwise.} \end{cases}$

TIM = Travel Time in minutes

SDE = Max. $\{-SD, 0\}$.

SDL = Max. $\{SD, 0\}$.

FLEX = Answer to question: "How many minutes late can you arrive at work without it mattering very much?".

SDLX = Max. $\{SD-FLEX, 0\}$.

D2L = $\begin{cases} 1 & \text{if } SD \geq FLEX \\ 0 & \text{otherwise.} \end{cases}$

CP = Dummy for car pool.

$N = \sum_{r=1}^{J+M} N^r$ (N^r from equation 5.10)

$AE = \sum_{r=1}^M (\gamma_E^{(r-1)}) A^r$ (A^r from equation 6.5)

$AL = \sum_{r=J+1}^{J+M} (\gamma_L^{(J+M-r)}) A^r$ (A^r from equation 6.5)

^dPredicted change in \bar{P}_{1-6} from increasing TIM by one minute for alternatives 7-12.

TABLE 7.2
 PREFERRED EOL AND VARIANTS: TRIP-TIMING MODELS
 First-order Approximation, One Iteration

	SOL	OL	EOL			
			Variants			Preferred
			M = 1	Best fit	$\gamma \neq 0$	$a_{J+M} = 0$
Exogenous parameters ^a						
M	1	5	1	5	5	5
γ_E	-	-	.1	.1	$\neq 0$.1
γ_L	-	-	-	.0001	-	-
Estimated coefficients ^b						
σ	-.252 (-0.63)	+.240 (0.87)	.140 (0.26)	.760 (2.19)	+.620 (1.88)	+.674 (1.96)
σ_{a_1}	-	-	1.11 (1.06)	5.91 (2.05)	5.91 (3.01)	5.72 (1.99)
$\sigma_{a_{J+M}}$	-	-	-	12.34 (1.83)	-	-
Increase in log likelihood over MNL value	0.20	0.37	0.79	4.98	2.54	2.60

^a γ_L was not varied independently, but was set equal to $(\gamma_E)^4$ as explained in the footnote in the text; see Table 7.1, footnote a, for an explanation of how they are used in constructing variables AE and AL (whose coefficients are σ_{a_1} and $\sigma_{a_{J+M}}$, respectively). $\gamma_E \neq 0$ implies $AE = A^1$. Where γ_L is not given, AL was omitted from specification.

^bAfter only one iteration of two-step estimator. Asymptotic t-statistic under null hypothesis $\sigma_r = 0$ is given in parentheses.

VIII. CONCLUSION

This paper has presented the Ordered Logit (OL) model, a generalization of multinomial logit (MNL). It applies to cases in which discrete alternatives are arrayed along some dimension on which individuals have systematic unobserved preferences, thereby inducing stochastic correlation between alternatives in close proximity. The model belongs to the class of Generalized Extreme Value models defined by McFadden, and is therefore consistent with random utility maximization. When the departure from MNL is small, it can be estimated iteratively using an ordinary MNL routine; the first iteration serves as a diagnostic test for failure of the MNL assumptions. The model can be modified to eliminate selectivity bias from truncation of the dependent variable. The model's applicability has been demonstrated both with a numerical example and with real data on scheduling choices by commuters.

Several extensions to the work here can be suggested. There is need to further develop theoretically justified ways of specifying the many new parameters introduced by the OL (and its extension for truncated choice sets), as functions of a manageable number of estimable parameters. Monte Carlo tests comparing the OL with other discrete choice models such as MNL, multinomial probit, and nested logit would help define the conditions under which it is useful: for example, it may be that the OL could satisfactorily handle some cases where the true model is multinomial probit. Finally, the OL can substitute for the MNL portion of more complicated models such as nested logit; in fact, it is simple to write down a nested logit model in which the conditional probabilities within each subgroup are OL instead of MNL, and there seems little to deter one from estimating it.

APPENDIX

Acronyms and Symbols Frequently Used in this Paper

cdf	Cumulative distribution function
EOL	Extended Ordered Logit
GEV	Generalized Extreme Value
IIA	Independence from Irrelevant Alternatives
MNL	Multinomial Logit
NL	Nested Logit
OL	Ordered Logit
SOL	Simple Ordered Logit
j, k	Alternative label (runs from 1 through J)
m	Index for weights within subset of alternatives in OL model
r, s	Index for a subset of alternatives in NL or OL model
w	Weights
z	Vector of independent variables
A^r	New variable used in 2nd step of approximate EOL estimator
B	Subset of alternatives
F	Cumulative distribution function of one or more error terms
G	Function defining a GEV model
H	A particular cumulative distribution function
I	Inclusive value
J	Number of alternatives
L	Log-likelihood function
M	Number of adjacent alternatives grouped together in OL model minus one
N^r	New variable used in 2nd step of approximate OL estimator
N	Number of cases in sample
P	Choice probability
U	Utility (strict plus stochastic)
V	Strict (non-stochastic, observable) utility component
β	Vector of coefficients of independent variables
ϵ	Stochastic utility component (error term)
ρ	Parameter of NL and OL models measuring dissimilarity within a group ($0 < \rho \leq 1$)
σ	($1-\rho$)

Variance of the Approximate Estimator

Let $V_j = \beta'z_j$ be the strict utility of mode j for a member of the sample, where z_j is a vector of observable characteristics of alternatives and β a vector of unknown coefficients. Let $\sigma = (\sigma^1, \dots, \sigma^{J+M})'$ and $N_j(\beta) = (N_j^1, \dots, N_j^{J+M})'$,

where N_j^r is defined by (5.10). Note that

$$(A.1) \quad \partial N_j^r / \partial \beta = w_{r-j} (z_j - \bar{z}^r)$$

where

$$(A.2) \quad \bar{z}^r = \frac{(\sum_m z_{r-m} w_m P_m^0)}{(\sum_m w_m P_m^0)} .$$

If k is the mode chosen, the log-likelihood function for the model given by (3.1) receives a contribution from this sample member of

$$(A.3) \quad L(\beta, \sigma) = \beta' z_k + \sigma' N_k(\beta) - \log \sum_j \exp[\beta' z_j + \sigma' N_j(\beta)]$$

Then

$$\begin{aligned} \frac{\partial L(\beta, \sigma)}{\partial \beta} &= z_k + \sum_r w_{r-k} (z_k - \bar{z}^r) \\ &\quad - \frac{\sum_j [z_j + \sum_r w_{r-j} (z_j - \bar{z}^r)] \exp(\beta' z_j + \sigma' N_j)}{\sum_\ell \exp(\beta' z_\ell + \sigma' N_\ell)} \end{aligned}$$

$$(A.4) \quad = (z_k - \bar{z}) + (y_k - \bar{y})$$

and

$$\frac{\partial L(\beta, \sigma)}{\partial \sigma} = N_k - \frac{\sum_j N_j \exp(\beta' z_j + \sigma' N_j)}{\sum_\ell \exp(\beta' z_\ell + \sigma' N_\ell)}$$

$$(A.5) \quad = N_k - \bar{N}$$

where

$$y_j = \sum_r w_{r-j} (z_j - \bar{z}^r)$$

$$\bar{y} = \sum_j P_j y_j$$

$$\bar{z} = \sum_j P_j z_j$$

$$\bar{N} = \sum_j P_j N_j$$

and where P_j is given by (3.1). The asymptotic variance-covariance matrix of the maximum-likelihood estimate (β, σ) is \mathcal{M}^{-1} , where \mathcal{M} is the expectation of the cross product of first derivatives of the log-likelihood function:

$$(A.6) \quad \mathcal{M} = \sum_{\text{sample}} \sum_k P_k \begin{pmatrix} \partial L / \partial \beta \\ \partial L / \partial \sigma \end{pmatrix} \begin{pmatrix} \partial L / \partial \beta \\ \partial L / \partial \sigma \end{pmatrix}'$$

$$(A.7) \quad = \mathcal{M}^A + \mathcal{M}^B$$

where

$$(A.8) \quad \mathcal{M}^A = \sum_{\text{sample}} \sum_j P_j \begin{pmatrix} z_j - \bar{z} \\ N_j - \bar{N} \end{pmatrix} \begin{pmatrix} z_j - \bar{z} \\ N_j - \bar{N} \end{pmatrix}'$$

$$(A.9) \quad \mathcal{M}^B = \sum_{\text{sample}} \sum_j P_j \begin{pmatrix} m_{11} & m'_{12} \\ m_{12} & 0 \end{pmatrix}$$

$$(A.10) \quad m_{11} = (y_j - \bar{y})(z_j - \bar{z})' + (z_j - \bar{z})(y_j - \bar{y})' + (y_j - \bar{y})(y_j - \bar{y})'$$

$$(A.11) \quad m_{12} = (N_j - \bar{N})(y_j - \bar{y})'$$

\mathcal{M}^A is the inverse of the variance-covariance matrix of (β, σ) if N were treated as an ordinary variable in the last MNL step of the iterative procedure described in Section V; \mathcal{M}^A is sometimes called the moment matrix and is computed by most MNL computer packages. Thus, the procedure is to retrieve \mathcal{M}^A at the last step of the estimation, compute \mathcal{M}^B and add it to \mathcal{M}^A , then invert the resulting matrix to obtain the estimated variance-covariance matrix of (β, σ) .

Two things should be noted. First, the formula still holds if some of the variables N^F are replaced by combinations $\sum_{r \in B} N^F$ in order to constrain the

corresponding σ_r to be equal. Simply reduce the number of components of N_k accordingly.

More important, under the null hypothesis $\sigma = 0$, the variance-covariance matrix reduces to $(M^A)^{-1}$, i.e. to that computed as though N^r were exogenous. Thus statistical tests for deviation of the model from MNL can be carried out using the asymptotic standard errors for σ_r computed by the MNL estimating routine.

Equivalence of Sheffi and Nested Logit Models as ρ Approaches Zero

Sheffi (1979) proposes a choice model for cases "where the alternatives are associated with some ranking and the choice of any alternative implies that all lower-ranked alternatives have been chosen as well" [p. 189]. Citing automobile ownership and shopping-trip frequency as examples, he proposes two postulates:

Postulate A implies that the utilities [strict utility plus stochastic component] are a monotonically increasing function of their index set for all alternatives ranked lower than the chosen one [and] that the utility of any alternatives ranked higher than the one following the chosen alternative is lower than the utility of the alternative following the chosen one.

Postulate B [implies] that the random variables formed from pairwise subtraction of the utilities of the given alternatives are independent... [p. 192]

Sheffi then derives the following choice probabilities:¹

¹I have adapted his formula to apply to the choice set $\{1, \dots, J\}$ used throughout this paper.

$$(A.12) \quad P_k = (1 - P_{k+1|k}) \prod_{j=2}^k P_{j|j-1}$$

where $P_{j|j-1}$ is an independent bivariate choice depending only on the utilities of alternatives j and $j-1$. In his empirical application, these bivariate choices are modelled as logit:

$$(A.13) \quad P_{j|j-1} = \frac{1}{1 + e^{-(W_j - W_{j-1})}}$$

where I have used the symbol W instead of Sheffi's V to avoid confusion with the strict utility as defined in this paper.

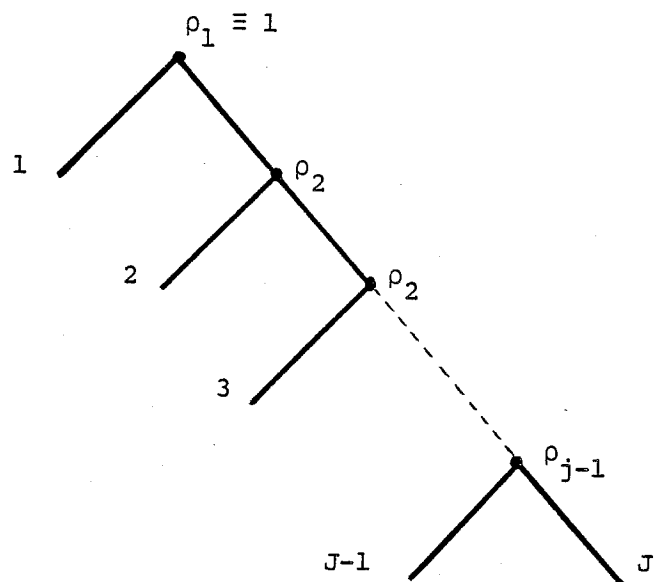
Although Sheffi describes his model as "based on the [standard] random utility theory" [p. 195], postulates A and B are actually inconsistent within such a theory. Let $U_j = V_j + \epsilon_j$ be the utility of alternative j , divided into its strict and stochastic components. Postulate A requires that $(\epsilon_j - \epsilon_{j-1}) > (V_{j-1} - V_j)$ whenever some alternative $k > j$ is chosen. But this means that the distribution of $(\epsilon_j - \epsilon_{j-1})$ must depend on whether or not $(\epsilon_k - \epsilon_{k-1})$ exceeds $(V_{k-1} - V_k)$ for some $k > j$, thereby contradicting Postulate B.

Sheffi has made a subtle error because his model is actually a limiting case of a nested logit model with $(J-1)$ levels, as each level approaches a maximal model. The W_j which determine (A.13) are not the strict utilities themselves but rather can be derived as a limit of a transformation of the true stochastic utilities, which I shall continue to denote as V_j . In this limit, both $(V_{k-1} - V_k)$ and the variance of $(\epsilon_k - \epsilon_{k-1})$ shrink to zero in such a way that level k can effectively be isolated from lower levels as though they were truly independent.

To see this, consider a nested logit model with the tree structure shown in Figure A.1. A parameter ρ_j is associated with the node from which alternative j branches off, so that the choice probabilities are given by (A.12) with

TABLE A.1

A Nested Logit Tree Structure



$$\begin{aligned}
 P_{j|j-1} &= \frac{e^{\rho_j I_j / \rho_{j-1}}}{e^{v_{j-1} / \rho_{j-1}} + e^{\rho_j I_j / \rho_{j-1}}} \\
 (A.14) \qquad &= \frac{1}{1 + \exp[(v_{j-1} - \rho_j I_j) / \rho_{j-1}]}
 \end{aligned}$$

$$I_J = v_J / \rho_J$$

$$I_j = \log(e^{v_j / \rho_j} + e^{\rho_{j+1} I_{j+1} / \rho_j}), \quad j = 1, \dots, J-1$$

and with $\rho_1 \equiv 1$ by convention. For the model to be a valid NL, the condition $0 < \rho_J < \rho_{j-1} < \dots < \rho_1 \equiv 1$ must be met. Suppose the ρ_j decrease geometrically

$$\rho_j = r^{j-1}$$

for some positive r . Consider the following limiting process: $r \rightarrow 0$, and $v_j \rightarrow v_2$ for $j > 2$ in such a way that

$$(A.15) \quad \lim_{r \rightarrow 0} (v_j - v_{j-1}) / r^{j-2} = \Delta_j$$

exists. Then

$$\lim_{r \rightarrow 0} (\rho_j I_j) = v_j$$

and

$$\lim_{r \rightarrow 0} P_{j|j-1} = \frac{1}{1 + e^{-\Delta_j}}$$

This is precisely the model (A.12) and (A.13) with $w_j - w_{j-1} = \Delta_j$.

Sheffi's model has considerable plausibility, is easy to estimate efficiently, and seems to perform well in appropriate circumstances. However, this

demonstration poses two dilemmas for applying and interpreting it. First, in specifying generic variables it is important to realize that one is not specifying the strict utilities themselves, but rather differences between them divided by a ratio of ρ 's. How one should think about a plausible utility specification in such a model is not obvious, and probably depends on the resolution to the second dilemma: namely, no one seems to have worked out the welfare implications of exogenous changes in such a model. To do so should be a straightforward matter of determining the expected maximum utility in the NL model of Figure A.1, and taking the appropriate limit.

BIBLIOGRAPHY

- Abkowitz, Mark D.: The Impact of Service Reliability on Work Travel Behavior, Ph.D. Dissertation, M.I.T., 1980.
- Amemiya, Takeshi: "Qualitative Response Models," Annals of Economic and Social Measurement, 4 (1975), 363-372.
- Amemiya, Takeshi: "On a Two-step Estimation of a Multivariate Logit Model," Journal of Econometrics, 8 (1978), 13-21.
- Ben-Akiva, Moshe: Structure of Passenger Travel Demand Models, Ph.D. Dissertation, M.I.T., 1973.
- 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).
- Brownstone, David: An Econometric Model of Consumer Durable Choice and Utilization Rate, Ph.D. Dissertation, University of California, Berkeley, 1980.
- Cardell, Scott:
- 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).
- Fisher, Walter D.: "Econometric Estimation with Spatial Dependence," Regional and Urban Economics, 1 (1971), 19-40.

- 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).
- Manski, Charles F., and Steven R. Lerman: "The Estimation of Choice Probabilities from Choice Based Samples," Econometrica, 45 (1977), 1977-1988.
- Manski, Charles F. and Daniel McFadden: "Alternative Estimators and Sample Designs for Discrete Choice Analysis," 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: "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, Antti P. Talvitie, and Associates: Demand Model Estimation and Validation, Vol. V of The Urban Travel Demand Forecasting Project Phase I Final Report Series, Institute of Transportation Studies, University of California, Berkeley, June 1977.

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.

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

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

Train, Kenneth: Auto Ownership and Mode Choice within Households, PH.D. Dissertation, University of California, Berkeley, 1977.

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

Williams, H. C. L.: "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," Environment and Planning - A, 9 (1977), 285-344.