UC Berkeley
Earlier Faculty Research

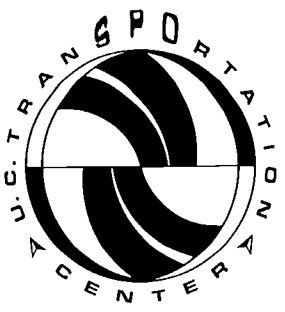
Title
Probit Model Estimation Revisited: Trinomial Models of Household Car Ownership

Permalink
https://escholarship.org/uc/item/2hr8d4bs

Authors
Bunch, David S.
Kitamura, Ryuichi

Publication Date
1991-09-01

eScholarship.org Powered by the California Digital Library
University of California



Probit Model Estimation Revisited: Trinomial Models of Household Car Ownership

David S. Bunch
Ryuichi Kitamura

September 1991
Working Paper, No. 70

The University of California
Transportation Center

University of California
Berkeley, CA 94720

The University of California Transportation Center

The University of California Transportation Center (UCTC) is one of ten regional units mandated by Congress and established in Fall 1988 to support research, education, and training in surface transportation. The UC Center serves federal Region IX and is supported by matching grants from the U.S. Department of Transportation, the California State Department of Transportation (Caltrans), and the University.

Based on the Berkeley Campus, UCTC draws upon existing capabilities and resources of the Institutes of Transportation Studies at Berkeley, Davis, and Irvine; the Institute of Urban and Regional Development at Berkeley; the Graduate School of Architecture and Urban Planning at Los Angeles; and several academic departments at the Berkeley, Davis, Irvine, and Los Angeles campuses. Faculty and students on other University of California campuses may participate in Center activities. Researchers at other universities within the region also have opportunities to collaborate on selected studies. Currently faculty at California State University, Long Beach, and at Arizona State University, Tempe, are active participants.

UCTC's educational and research programs are focused on strategic planning for improving metropolitan accessibility, with emphasis on the special conditions in Region IX. Particular attention is directed to strategies for using transportation as an instrument of economic development, while also accommodating to the region's persistent expansion and while maintaining and enhancing the quality of life there.

The Center distributes reports on its research in working papers, monographs, and in reprints of published articles. For a list of publications in print, write to the address below.

University of California Transportation Center

108 Naval Architecture Building
Berkeley, California 94720
Tel: 415/643-7378
FAX: 415/643-5456

Authors of papers reporting on UCTC-sponsored research are solely responsible for their content. This research was supported by the U.S. Department of Transportation and the California State Department of Transportation, neither of which assumes liability for its content or use.

Probit Model Estimation Revisited: Trinomial Models of Household Car Ownership

David S. Bunch
Ryuichi Kitamura

University of California at Davis

Working Paper, No. 70

September 1991

The University of California Transportation Center
University of California at Berkeley

Abstract

In this paper we revisit various important issues relating to practical estimation of the multinomial probit model, using an empirical analysis of car ownership as a test case. To provide context, a brief literature review of empirical probit studies is included. Estimates are obtained for a full range of model specifications, including models with random (uncorrelated and correlated) taste variation and/or a general random error structure, and issues of estimability, specification testing, and alternative normalizations for probit models are addressed. Three model trust region algorithms for finding maximum likelihood estimates are compared, and the superiority of a structured quasi-Newton method employing "model switching" over more traditional approaches (Broyden-Fletcher-Goldfarb-Shanno secant update, Berndt-Hall-Hall-Hausman) is demonstrated. The trust region algorithms have reliable convergence properties and provide useful diagnostic information. Finally, a comparison of some probit integral approximation schemes (Clark, and two variants of Mendell-Elston) versus numerical integration is included. There is additional evidence against using Clark's approximation, but a variant of the Mendell-Elston approach appears promising. Numerical problems with variable-ordering techniques (such as separated-split) are demonstrated.

1. Introduction

Beginning with McFadden (1974) and the subsequent popularization of the multinomial logit model in the social sciences, a substantial amount of research activity has occurred in the area of discrete choice analysis, with much of the work centering on travel demand issues. The 1970's and early 1980's witnessed numerous theoretical advances in choice modeling, with suggestions for more flexible (and complex) classes of models which could in principle capture more realistic patterns of choice behavior than the logit model. These include elimination-by-aspects (EBA), hierarchical elimination-by-aspects (HEBA), elimination-by-strategy (EBS), generalized extreme value (GEV), tree extreme value (TEV)/nested multinomial logit (NMNL), and multinomial probit (MNP) models--see McFadden (1981). The most theoretically appealing of these is arguably multinomial probit, which assumes that decision makers may be modeled as coming from a population of random utility maximizers, where the error components in the (unobserved) utilities arise from a multivariate normal distribution.

Despite the obvious flexibilities and advantages of the multinomial probit model, relatively few empirical results using this class of models have appeared in the published literature. For example, consider household car ownership, which provides the empirical test case for this paper. Studies of household car ownership have so far used multinomial logit models (e.g., Ben-Akiva and Lerman 1974), occasionally combined with linear utilization models (Train 1984) or ordered-response probit models (Kitamura 1987), but to our knowledge, no study has adopted multinomial probit models.

The limited appearance of MNP empirical applications is undoubtedly related to the various computational difficulties associated with obtaining parameter estimates. First, maximum likelihood estimation of complex nonlinear models is still problematic for many practitioners, and the MNP model is more complex than most. Second, the more useful specifications require estimation of covariance parameters, and the properties of the likelihood function are virtually unknown in these circumstances. Third, evaluating the

MNP choice probabilities requires integration of the multivariate normal probability density. This is in general quite difficult: for standard integration approaches the work required increases exponentially with the number of alternatives in the choice set.

Most attention in the literature has focused on this last issue. For a time the primary hopes for probit were pinned on the use of Clark's approximation (Clark, 1961) for evaluating choice probabilities--see Daganzo (1979). However, despite early encouraging results, evidence casting serious doubts on this approach eventually emerged--see Horowitz, et al. (1982). Alternative approaches that have been proposed are Monte' Carlo simulation (Lerman and Manksi, 1981), tabulation (Sparmann, et al., 1983), and "separated-split" (Langdon, 1984). More recently, Kamakura (1989) considered the Mendell-Elston approximation, and McFadden (1989) has suggested a variety of "choice simulators." However, none of these approaches has yet led to any published empirical results that we are aware of. In addition, numerical integration for three- and four-alternative probit models has been cited as "practical" in the literature, yet there has been only one of these of which we are aware.

The issue of the unknown concavity properties of the MNP log-likelihood function is often raised as a disadvantage, yet this in and of itself should not preclude applications: the problem of determining whether a local optimum is also a global optimum often arises in practice, and is essentially a computational issue that can be resolved by a variety of techniques, e.g., trying numerous starting points. But like the concavity issue, there are other potential obstacles caused by MNP's inherent complexity. The presence of covariance parameters, while providing flexibility, could cause the model to collapse under its own weight in the absence of huge datasets. Another problem which has not been particularly well-addressed is the difficulty of verifying if parameters in complex MNP model specifications are actually identified, leading to the possibility of overparameterized

models. This was highlighted by Dansie (1985), who pointed out that a relatively simple and seemingly innocuous model specification was, in fact, unidentified. Poorly-behaved log-likelihood functions would arise under such circumstances, perhaps causing numerical difficulties in parameter estimation that could quickly overcome the most patient analyst. In any case, the parameter estimates, unbeknownst to the user, would not have valid interpretations.

In this paper we revisit practical issues of MNP estimation, including model identification, algorithms for maximizing likelihood functions, and approximation methods for evaluating probit probabilities. Our goal is to carefully address all these issues simultaneously, establishing new norms for the state-of-the-art. We use as a test case an empirical study of household car ownership, and consider a full range of model specifications based on a linear-in-the-parameters MNP framework similar to that of Hausman and Wise (1978) and Albright, Lerman, and Manski (1977). Parameter identification issues are addressed using results from Bunch (1990). This approach to model estimation avoids numerical problems associated with overparameterization, and clarifies the interpretation of parameter estimates.

Due to the heavy computational requirements for MNP estimation, efficient optimization is an important consideration, irrespective of the method used to approximate the probit integrals. We compare three different algorithms for maximum likelihood estimation, and demonstrate that a recently developed approach is much more efficient, reliable, and robust than those previously used in applications of MNP. We also compare different numerical methods for multivariate normal integration. Accurate estimates for the car ownership models are first obtained through a numerical integration approach, and are then used in a comparison of three different integral approximation schemes, i.e., the

familiar Clark approach, and two methods based on the work of Mendell and Elston (1974).

The paper is organized as follows. Useful notation and definition of model specifications are developed in the next section, and a review of empirical MNP results in the literature is presented in section 3. Section 4 describes the empirical data and model formulations of household car ownership that form the basis for our analysis, and presents model parameter estimates obtained using an accurate numerical integration method. Issues of estimability, specification testing, and interpretation of alternative normalizations for the error covariance matrix are discussed. Section 5 addresses optimization issues related to maximum likelihood estimation, and presents a numerical comparison of three algorithms. Section 6 compares the three integral approximation schemes, and a summary and conclusions appear in section 7.

2. Linear-in-Parameters MNP Framework

Consider a sample of N households, indexed by n = 1, ..., N, with each choosing one alternative from a set of J alternatives. Explanatory data are collected for household n such that each of the J alternatives is characterized by K (generic) attributes, which are stored in the K x J matrix X_n. In what follows, we assume the following linear-in-parameters random utility framework:

$$U_n = X_n^T(\theta + \delta_n) + \mu + \varepsilon_n, \quad n = 1, ..., N, \tag{1}$$

where $U_n, \mu, \varepsilon_n \in \Re^J$, and $\theta, \delta_n \in \Re^K$. The vector U_n contains household n's (unobserved) true utilities for the J alternatives, and the observed choice will be the one with the largest utility. The K-vector $(\theta + \delta_n)$ contains the attribute taste weights for household n, where θ is the mean taste weight for the population and δ_n is the (unobserved) random deviation from the mean for household n, where $E(\delta_n) = 0$. The sum

$(\mu + \varepsilon_n)$ represents the effect of (unobserved) random errors on utility, where μ is the alternative-specific mean and ε_n is random error with $E(\varepsilon_n) = 0$. These terms may be regarded as including the effect of unobserved attributes of the individual and/or the alternatives, as well as other sources of error. In the sequel, the term "error" will refer to any effects associated with the ε_n term.

To get a multinomial probit model, one adds the theoretically appealing assumption that both random terms have multivariate normal distributions:

$$(2) \qquad \delta_n \sim MVN(0, \Sigma_\delta),\ \Sigma_\delta \in \Re^{K \times K} \text{ and } \varepsilon_n \sim MVN(0, \Sigma_\varepsilon),\ \Sigma_\varepsilon \in \Re^{J \times J}$$

Note that μ in equation (1) could be represented via an alternative framework which allows dummy variables in X_n; the (alternative-specific) random errors would then be represented by the δ terms corresponding to the dummies, and no ε terms would be required. (See, for example, Albright, Lerman, and Manski, 1977.) By using (1) and (2) instead, we choose a slightly different framework that assumes the δ and ε terms are always independent, simplifying analysis of model specifications. Now, the probability that individual n selects alternative j is given by:

$$(3) \qquad P_j(V_U(\theta, \mu, X_n), \Sigma_U(\Sigma_\delta, \Sigma_\varepsilon, X_n)) = \text{Prob}[\, U_{nj} > U_{ni} \text{ for all } i \neq j]$$

$$= \int_{u_j=-\infty}^{\infty} \cdot \int_{u_1=-\infty}^{u_j} \cdots \int_{u_{j-1}=-\infty}^{u_j} \cdot \int_{u_{j+1}=-\infty}^{u_j} \cdots \int_{u_J=-\infty}^{u_j} \phi(u \mid V_U, \Sigma_U)\, du_1 \cdots du_J$$

where $V_U(\theta, \mu, X_n) = X_n^T\theta + \mu$, $\Sigma_U(\Sigma_\delta, \Sigma_\varepsilon, X_n) = X_n^T\Sigma_\delta X_n + \Sigma_\varepsilon$, and $\phi(x \mid m, S)$ is the multivariate normal density function with mean m and covariance S.

Various model specifications may be formulated by placing restrictions on the random taste covariance matrix Σ_δ and the error covariance matrix Σ_ε in (2). Three types of taste variation are considered: (i) fixed (non-random) tastes with $\Sigma_\delta \equiv 0$, (ii) uncorrelated

random tastes with $\Sigma_\delta \equiv D$ (a diagonal matrix), and (iii) correlated random tastes with a general covariance matrix Σ_δ. Two types of errors are considered: (i) IID errors with $\Sigma_\varepsilon = kI$, where k is a scale constant, and (ii) non-IID errors with a general Σ_ε. Scaling in the probit model is arbitrary, and we will assume that k=1 for the IID error specifications. However, the specification of a "general" Σ_ε requires special consideration, since all J(J+1)/2 parameters are not estimable.

For example, consider the trinomial case and suppose we wish to calculate the probability of choosing the third alternative. The estimability problem arises from the fact that the three-dimensional integral in equation (3) may be reduced to two dimensions by transformation to the space of utility *differences*:

$$(4) \quad P_3(V_U, \Sigma_U) = \Pr[U_3 \geq U_1 \text{ and } U_3 \geq U_2] = \Pr[U_1 - U_3 \leq 0 \text{ and } U_2 - U_3 \leq 0]$$

$$= \Pr[Z_1^{(3)} \leq 0 \text{ and } Z_2^{(3)} \leq 0] = \Pr[Z \leq 0]$$

$$= \Phi(0 | V_Z, \Sigma_Z)$$

where Φ is the bivariate normal cumulative distribution function,

$$Z_j^{(3)} = U_j - U_3, \; j = 1 \text{ and } 2,$$

$$Z = (Z_1^{(3)}, Z_2^{(3)}) = \Delta_3 U,$$

$$V_Z = \Delta_3 V_U,$$

$$\Sigma_Z = \Delta_3 \Sigma_U \Delta_3^T = \Delta_3 (X_n^T \Sigma_\delta X_n) \Delta_3^T + \Delta_3 \Sigma_\varepsilon \Delta_3^T,$$

and Δ_3 is the transformation matrix given by

$$(5) \quad \Delta_3 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}.$$

Next consider the term in the Σ_Z matrix given by

$$(6) \qquad \Delta_3\Sigma_\varepsilon\Delta_3^T = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} \\ \sigma_{21} & \sigma_{22} & \sigma_{32} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11}-2\sigma_{31}+\sigma_{33} & \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} \\ \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} & \sigma_{22}-2\sigma_{32}+\sigma_{33} \end{bmatrix}$$

$$\equiv \begin{bmatrix} M_{11} & M_{21} \\ M_{21} & M_{22} \end{bmatrix}$$

Only the covariance matrix M for the *differences* of the error terms is estimable; in addition, one parameter may be fixed for scaling purposes. The maximum number of identifiable parameters in Σ_ε when J=3 is therefore two, and in general is J(J-1)/2 - 1. In this study we estimate the two free parameters M_{21} and M_{22} in the matrix

$$(6) \qquad M = \begin{bmatrix} 1 & M_{21} \\ M_{21} & M_{22} \end{bmatrix}$$

for model specifications containing a "general Σ_ε matrix," i.e., we estimate the maximum number of identified covariance parameters. As discussed in section 4 below, such estimates may correspond to many different possible normalizations of Σ_ε. For a more complete development of estimability and normalization results, see Bunch (1990).

All possible combinations of the various covariance features yield the six MNP model specifications used in our empirical analysis: these are summarized in Table 1, and it will be convenient to refer to a specific model specification according to the functional form assumed by Σ_U. In addition, the specification with $\Sigma_\varepsilon = I$ and $\Sigma_\delta = 0$ shall be termed an IID probit model. Similar specifications using this framework have been estimated by other authors, as described in the next section.

3. Empirical Multinomial Probit Results in the Literature

Although the theoretical advantages of the multinomial probit model have often been recognized, we found relatively few empirical studies in the published literature which apply the model to problems with more than two alternatives. Table 2 lists the seven studies we found, and summarizes features which relate to practical estimation issues, e.g., the number of alternatives, type of optimization algorithm, method of choice probability evaluation, types of model specifications (e. g., random versus fixed taste weights, IID versus non-IID errors), and number of exogenous variables. There may be more studies that were somehow overlooked, but it is obvious that MNP has not become a commonplace technique for discrete choice analysis. The following briefly identifies the domain of empirical results in the literature; for a more detailed review see our technical report (Bunch and Kitamura 1989).

Hausman and Wise (1978)

Topic: travel mode choice. The optimization method approximates the Hessian by the method of Berndt, Hall, Hall, and Hausman (1974) ["BHHH"], and uses analytic expressions for the gradients. Probit integrals are evaluated using the same accurate numerical integration approach used here, i.e., Owen's method--see section 6. Four probit specifications are used in the study. Using the notation of the previous section they are: (i) IID probit with generic attributes only, i.e., $\mu \equiv 0$ and $\Sigma_U \equiv I$, (ii) uncorrelated random tastes for generic attributes with zero-mean IID error terms, i.e., $\mu \equiv 0$ and $\Sigma_U = X^T D X + I$, (iii) IID probit with alternative specific dummy variables, i.e., μ is estimated and $\Sigma_U \equiv I$, and (iv) uncorrelated random tastes for generic attributes with alternative specific dummy variables and non-IID error terms, i.e., $\Sigma_U = X^T D X + \Sigma_\varepsilon$.

Albright, Lerman and Manski (1977)

Topic: travel mode choice. This study appeared as a technical report on the development of a MNP estimation program for the Federal Highway Administration. The optimization method is a simple gradient search ("steepest descent method") using finite difference gradients. Probit integrals are obtained using Clark's approximation--see section 6. The model specification is quite general, and includes correlated taste variation.

This report justifies use of Clark's approximation by giving simulation results for examples with three- and five-alternative choice sets; these results also appear in Lerman and Manski (1981). The report claims that the gradient search approach may be used with simulated probabilities, although no examples are given, and such an approach seems impractical.

Currim (1982)

Topic: travel mode choice. Models were estimated using CHOMP--see Daganzo and Schoenfeld (1978). CHOMP uses Clark's approximation for the integrals and performs optimization via a least-change secant update due to Davidon (1959) and Fletcher and Powell (1963), the so-called "DFP method." Many model specifications, including one with uncorrelated random taste variation are estimated; unfortunately, many of the models are unidentified, casting doubt on the validity of the results.

Johnson and Hensher (1982)

Topic: two-period panel analysis of travel mode choice. Parameters are obtained using CHOMP. The four binary probit models include an inestimable covariance parameter, i.e., they are unidentified.

Miller and Lerman (1981)

Topic: choices of location and store size by retail firms. This study is notable in that it includes 14 alternatives per choice set. Parameters are estimated via CHOMP. Two of the three model specifications include (uncorrelated) random taste variation. The authors indicate that there were difficulties in obtaining the parameter estimates, and that the final results were extremely sensitive to the starting point used.

Kamakura and Srivastava (1984)

Topic: choices among wagers. The authors develop a new multinomial probit model and compare it to the IID probit and X^TDX+I models. From the discussion, it appears that all models have been correctly specified. Probit integrals are calculated using Clark's approximation, and optimization is performed using the general constrained nonlinear programming algorithm GRG2 of Lasdon, Warren, and Ratner (1982). This code uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS) secant update to approximate the Hessian--see section 4.

van Lierop(1986)

Topic: choice of residential location. CHOMP is used to obtain parameter estimates. All analyses appear to be overparameterized, casting doubts on the interpretability and validity of the study's results.

Remarks

It is perhaps surprising that we found only one study which uses numerical integration for evaluating choice probabilities, given claims that three- or four-alternative studies are "practical." Six of the seven studies use Clark's approximation. For the two studies involving more than three alternatives this was a logical approach due to practical considerations; however, five of the studies could in principle have used numerical

integration. Since four of the studies used CHOMP, software availability seems to be an obvious factor.

Concerns about the accuracy of Clark's approximation could explain the small number of published MNP studies. Only two of the studies were published after the appearance of Horowitz, et. al. (1982), which presents simulations demonstrating that Clark's approximation can be poor under some circumstances. It would appear that the MNP estimation problems described in the introduction have probably not been adequately addressed by the approaches attempted thus far, and potential users of MNP are perhaps waiting for advances in computing to catch up with the theory.

The general disarray surrounding the topic of probit estimation is illustrated by comparing the Hausman and Wise (HW78) and Albright, Lerman, and Manski (ALM77) studies. They both use the same data source and fit models for the same mode choice problem, but unfortunately they reach different conclusions. In HW78 the probit model is a significant improvement over logit, but ALM77 conclude that their (more general) probit model is *not* an improvement over logit. ALM77 use steepest descent and Clark's approximation. HW78 use BHHH and numerical integration, but use fewer data points and fewer explanatory variables. Thus the basic conclusions for mode-choice models are not robust to these features of the analyses, and a more controlled comparison would have been illuminating.

Particularly disturbing is the relatively large proportion of the studies (three of seven) containing major conceptual flaws, i.e., unidentified model specifications. This indicates that MNP is not generally well-understood. Furthermore, such misspecifications give rise to poorly-behaved log-likelihood functions and may have resulted in unreported abortive attempts at probit estimation, further complicating the advancement of the model in practical applications. These issues are addressed in the next section.

4. Test Case: Models of Car Ownership

Our "test problem" is to successfully estimate and evaluate several alternative models for predicting the number of cars owned by a household. The three-alternative choice set for all households is {no car, one car, two-or-more cars}. Twelve trinomial probit models were constructed as follows: (i) two different formulations ("theoretical" and "empirical") were identified for the utility function in equation (1), and (ii) the six model specifications in Table 1 were implemented for both formulations. The two formulations and the data set are described in more detail below.

The maximum likelihood estimation results in this section were obtained using the Model Switching algorithm described in section 5. Choice probabilities were calculated using an accurate numerical integration method--see section 6. The focus of this section is to illustrate the practical evaluation of alternative models, specifications, and normalizations using results from Bunch (1990). Although some observations regarding model validity will be made in the course of the discussion, a detailed treatment of car ownership is beyond the scope of this paper, and is the subject of other work in progress.

4.1 Data Set

A subsample from the Dutch National Mobility Panel data set is used in this study. The Dutch panel survey is a general purpose transportation panel survey of approximately 1,500 households in each wave, that are scattered across 20 municipalities and intended to represent the Dutch population (van Wissen and Meurs, 1989). The data set contains general demographic and socioeconomic information, and trip information obtained from seven-day travel diaries filled out by those household members who are at least 12 years old. The data file used in this study also contains accessibility measures obtained from a nationwide network model and a set of destination choice models (Geinzer and Daly, 1981).

The use of this panel data set was motivated by the possibility of extending the cross-sectional analysis presented here to a longitudinal analysis at a later time. For this reason, households that participated in both the first and third waves of the survey, i.e., those participating both in April 1984 and in April 1985, are included in the sample for this study. Eliminating households with missing values for important model variables produced a 945-household sample for model estimation. Summary measures for some key variables appear in Table 3.

4.2 Utility Formulations

The "theoretical" formulation postulates that each household is a decision making unit which maximizes its utility by choosing the number of cars to own. Household utility is given by

$$U(y, t, N_{nw}) = \pi\, y^{\alpha}\, t^{\beta}\, N_{nw}^{\tau}, \tag{7}$$

where

y = remaining (discretionary) income,
t = discretionary time,
N_{nw} = number of non-work trips,
$\pi > 0$, and
$0 < \alpha, \beta, \tau < 1$.

The number of non-work trips, N_{nw}, is included in this formulation on the grounds that, for a given amount of monetary and time resources, household utility can be optimized by consuming them at optimal locations, which naturally induces traveling. Travel behavior is determined by assuming that the household maximizes utility, subject to constraints on both monetary and time resources. (This is an extension of Beckmann, Gustafson, and Golob, 1973.)

An indirect utility function may be obtained by solving the optimization problem and substituting the optimal solution into the utility function, as in Becker (1965). Several

assumptions and approximations may then be used to obtain an estimable, linear-in-parameters functional form. We omit the details of this derivation, which is beyond the scope of this paper and will be discussed elsewhere; the final formulation is defined by the four explanatory variables given in Table 4, part b.

The "empirical" formulation is defined by a set of four variables chosen from considering previous studies on car ownership, e.g., Ben-Akiva and Atherton (1977). We include measures for accessibility to non-work destinations, car availability to adults in the household, and car availability to workers in the household--see Table 4, part b. In both formulations the non-work accessibilities are taken to be accessibly to shopping destinations.

4.3 Estimation Results

Estimated coefficients for the explanatory variables and two alternative-specific dummies, as well as log-likelihoods and goodness-of-fit statistics, are reported in Table 5 for models based on the theoretical formulation. Corresponding results for models based on the empirical formulation are reported in Table 6. Results in both tables were obtained using accurate numerical integrals, as noted above. Coefficients have been standardized relative to the estimate of the income variable coefficient to facilitate model comparisons.

The t-scores for the alternative-specific dummies are generally significant for both formulations. This suggests that some important explanatory variables contributing to a household's utility for car ownership have been omitted from both analyses. For the theoretical formulation models, the t-scores for $WtdAcc_{nw}$ are always insignificant, and t-scores are insignificant for N_w/RI in half the cases. This calls into question the usefulness of the theoretical formulation, which has severe problems with multicollinearity. In contrast, the t-scores are significant for the coefficients in all empirical formulation models.

A model-by-model comparison of log-likelihoods for theoretical versus empirical formulations appears in Table 7. Models based on the empirical formulation clearly dominate those from the theoretical formulation in terms of fitting the data. In fact, the log-likelihood for the most simple empirical formulation model (IID probit, number of parameters, NPar = 6) is roughly comparable to the log-likelihood for the most complex theoretical formulation model ($X^T\Sigma_\delta X+\Sigma_\epsilon$, NPar = 18). Because of these findings, as well as other features of the estimations discussed below and in section 5, we will often focus on the empirical formulation models in the sequel. We complete our evaluation of the estimation results by first considering hypothesis tests of the various nested model specifications, and then discussing other interpretations of the numerical results.

Significance of Random Tastes and Non-IID Errors

The log-likelihoods presented in Table 7 can be used to test the significance of: (i) the assumption of IID versus non-IID random errors, and (ii) the assumption of random taste variation in the population. Tables 8 and 9 summarize the results of some likelihood-ratio tests that may be performed on various nested model specifications for the two formulations.

For empirical formulation results, see Table 8. Statistics for comparing various taste variation hypotheses are presented in part (a); results for the IID and non-IID error models appear in separate columns. For the IID error models, the likelihood-ratio (LR) statistic for the comparison of fixed tastes versus uncorrelated random tastes has a value of 8.77 with four degrees of freedom (4 dfs), which is not significant at the 5% level. For fixed tastes versus correlated random tastes LR = 16.57 (10 dfs), and for uncorrelated random tastes versus correlated random tastes LR = 7.80 (6 dfs): neither is significant at the 5% level. The results for models with IID errors thus do not reject the hypothesis of fixed tastes.

On the other hand, the hypotheses of fixed tastes versus uncorrelated tastes (LR = 16.26, 4 dfs) is rejected at the 1% level, and fixed versus correlated tastes (LR = 20.79, 10 dfs) is rejected for non-IID error models at the 5% level. However, the hypothesis of uncorrelated versus correlated random tastes (LR = 9.08) is *not* rejected at the 5% level. Thus, conclusions regarding random tastes depend upon the assumption of IID versus non-IID errors.

The LR statistics for the IID probit model versus the two most complex models (uncorrelated/non-IID and correlated/non-IID) are 20.97 with 6 dfs and 29.87 with 12 dfs, respectively (these do not appear in Table 8). The hypothesis of IID probit is rejected versus these alternatives at the 1% level. We conclude that the most appropriate specification for this data is a non-IID error model with uncorrelated random tastes.

Consider Table 8, part (b) for LR tests of IID versus non-IID errors. The assumption of IID errors is strongly rejected for the two random-taste specifications ($p < 0.005$) but is not rejected for the fixed taste specification at the 5% level. This result demonstrates the risk of testing model specifications without examining all the possibilities; an analysis which excludes models with random taste variation would accept the IID probit model as the correct specification.

A similar analysis for the theoretical formulation models (see Table 9) leads to a strong rejection of the fixed tastes hypothesis and an acceptance of the correlated random taste/non-IID error model, in contrast to the empirical formulation results. Recall that the theoretical formulation models do not fit the data as well as the empirical formulation models, indicating that the theoretical formulation's attributes do not contain as much useful explanatory information. This is consistent with the outcome of the LR test analyses: the preferred model specification using the theoretical formulation is the one that relies most

heavily on complex patterns of random taste variation to explain the variability in the observed choices.

Alternative-Specific Error Covariance Matrix (Σ_ϵ)

The statistical analysis presented above rejects the assumption of IID error terms. As discussed previously, a matrix M is obtained by our estimation routines for models involving a "general Σ_ϵ." Recall that the M matrix is defined by equation (6), which we reproduce here for convenience:

$$(7) \qquad M = \begin{bmatrix} 1 & M_{21} \\ M_{21} & M_{22} \end{bmatrix} \equiv \begin{bmatrix} \sigma_{11}-2\sigma_{31}+\sigma_{33} & \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} \\ \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} & \sigma_{22}-2\sigma_{32}+\sigma_{33} \end{bmatrix}.$$

So, M_{21} and M_{22} are the estimated quantities, expressed in terms of the σ_{ij}'s which are elements of Σ_ϵ.

From (7) it is obvious that only two of the σ_{ij} terms are estimable and that the remaining terms are arbitrary. Thus, many possible normalizations for Σ_ϵ may be consistent with the estimated matrix M. We note, however, that one may not arbitrarily choose any two-parameter expression for Σ_ϵ. Three allowable choices (see Bunch, 1990) are:

$$(8) \qquad \Sigma_\epsilon^1 = \begin{bmatrix} 1 & \sigma_{21} & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$(9) \qquad \Sigma_\epsilon^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix},$$

and

$$(10) \qquad \Sigma_\varepsilon^3 = \begin{bmatrix} \sigma_{11} & \sigma_{21} & 0 \\ \sigma_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Unfortunately, these three normalizations correspond to different behavioral interpretations despite the fact that they are not *a priori* distinguishable. For example, Σ_ε^2 assumes that the unobserved error terms for the various alternatives are completely uncorrelated but have different variances. This could be the result of a large number of unobservable factors that affect the utilities for the three alternatives in a totally random way. In contrast, Σ_ε^1 or Σ_ε^3 might imply the presence of omitted variables that commonly occur in both alternatives 1 and 2 but not in alternative 3. Note that Σ_ε^1 is a convenient normalization to consider, since $\sigma_{21} = M_{21}$ and $\sigma_{22} = M_{22}$.

Assume that the indices {1, 2, 3} in (8) through (10) are assigned to the choice alternatives {one car, two or more cars, no car}. It is then straightforward to calculate each Σ_ε matrix using the estimated values for M. This is done in Table 10 using empirical formulation estimates for the three models with non-IID errors. The correlation between alternatives 1 and 2 (i.e., $\rho = \sigma_{21}/\sqrt{\sigma_{11}\sigma_{22}} = \sigma_{21}/\sqrt{\sigma_{22}}$) has also been computed and reported where appropriate.

First consider the Σ_ε^1 results. The model with no taste variation has a small-but-negative ρ, in contrast to the two models which include taste variation. The model including uncorrelated tastes produces $\rho = 0.766$, while $\rho = 0.398$ for the correlated taste model. Recall that model $X^TDX+\Sigma_\varepsilon$ was preferred based on LR tests: this normalization is consistent with the interpretation that the one-car and two-or-more-cars alternatives share unobserved attributes which are highly correlated, but do not share attributes with the no-car alternative. The $X^T\Sigma_\delta X+\Sigma_\varepsilon$ model is more complex with regard to the variation in random tastes, and produces a smaller ρ.

The interpretations of the Σ_ε^3 normalizations are very similar to Σ_ε^1 for all three specifications; the only differences are due to the assumption of a non-zero variance for the errors associated with the no-car alternative. In contrast, the Σ_ε^2 normalization for the (preferred) $X^T DX+\Sigma_\varepsilon$ model produces a negative variance estimate, and thus this normalization is behaviorally inconsistent with the estimates obtained from the data. This is not unexpected, since it seems a bit unrealistic that the one- and two-or-more-cars alternatives would have zero correlation.

This example also illustrates a practical advantage to using the M matrix to get estimates, followed by an analysis of various normalizations. If the maximum likelihood estimation routine were attempting to directly estimate a MNP model with normalization Σ_ε^2, then it would produce either a boundary solution or encounter a singularity for this particular case. Corresponding results for the theoretical formulation models appear in Table 11. Many of the normalizations do not have reasonable behavioral interpretations, either because they have negative variances or because $|\rho| > 1$. Note that the "preferred" model specification $X^T\Sigma_\delta X+\Sigma_\varepsilon$ exhibits this behavior, creating additional concerns regarding the theoretical formulation.

Equations (8) through (10) are not the only possibilities for two-parameter normalizations. Consider the class of normalizations where the (unrescaled) Σ_ε matrix has either $\sigma_{ij} = \sigma_{kl}$ or $\sigma_{ij} = 0$ (see Bunch 1990). Then it is possible to exhaustively write down all the possibilities, and one might, for example, like to consider the following candidate normalizations:

$$\Sigma_\varepsilon^4 = \begin{bmatrix} 1 & \sigma_{21} & 0 \\ \sigma_{21} & 1 & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}, \tag{11}$$

$$(12) \qquad \Sigma_{\varepsilon}^{5} = \begin{bmatrix} 1 & \sigma_{21} & \sigma_{31} \\ \sigma_{21} & 1 & \sigma_{31} \\ \sigma_{31} & \sigma_{31} & 1 \end{bmatrix}.$$

Unfortunately, these normalizations are *not identified* even though they each appear to have two free parameters--see Bunch (1990). This further illustrates that, even though the MNP model exhibits a considerable degree of flexibility, there are still obstacles in formulation and estimation of intuitively appealing model specifications which can take into account unobserved but correlated attributes. Fortunately, the ML algorithms discussed in the next section can numerically detect difficulties of this type.

5. Comparison of Maximum Likelihood Estimation Algorithms

As noted in the introduction, there are two major computational concerns in probit estimation: (i) evaluation of the choice probabilities and (ii) searching for a (local) maximizer of the log-likelihood function. The former is important since evaluation of the multivariate normal CDF is difficult and expensive. This makes the latter even more crucial, since the search must be extremely efficient, requiring as few function evaluations as possible. In this section, we extend previous work (Bunch, 1987, 1988) by comparing the performance of three MLE algorithms on the difficult problem of MNP estimation. A complete discussion of optimization algorithms is well beyond the scope of this paper. However, we provide a brief overview of relevant optimization issues, and refer the reader to more detailed references.

There are two important questions that must be considered in formulating an iterative search algorithm: (i) does it reliably converge to a local optimum from an arbitrary starting point, and (ii) does it do so as quickly and as cheaply as possible? Algorithms satisfying (i) are called globally convergent, and employ a *global strategy*. (Note: "global convergence" should not be confused with "finding a global minimizer.") Two global

strategies are the well-known *line search* approach, and the more recently developed *model trust region* approach. All empirical probit studies until now (including those in section 3) have employed line searches; we use the model trust region approach, described below. The speed with which the algorithm converges to a local minimizer is generally dependent on the method used to approximate the Hessian matrix, i.e., the *local strategy*. A search algorithm may be characterized by particular combinations of global and local strategies.

Another important issue is deciding when to stop the algorithm; the model trust region codes used here provide a useful suite of stopping rules which also yield useful diagnostic information. The next four subsections describe global and local strategies, the implementation of three algorithms used in our comparison, and stopping rules, respectively. The final subsection provides a numerical comparison of the three algorithms, using the car ownership estimation problems as test cases. The descriptions given below are necessarily brief, and the reader is referred to Dennis and Schnabel (1983) for details of nonlinear optimization techniques. In addition, note that the standard convention in the optimization literature is to find local minimizers. The maximum likelihood estimation problem can be conveniently restated as a minimization problem by taking the negative of the log-likelihood function.

5.1 Global Strategies

Let $L(\beta)$ denote the objective function, where β is the vector of parameters. As noted above, most traditional iterative methods employ line searches. At the k^{th} iteration a search direction $d_k = - H_k^{-1} \nabla L(\beta_k)$ is determined, where H_k is an approximation to the Hessian of $L(\beta_k)$. A step is taken along this direction, and the new iterate is given by $\beta_{k+1} = \beta_k + \alpha d_k$, where α is the step length. One or more values of α may be attempted until a new point is identified which exhibits "sufficient decrease" in the objective function (negative log-likelihood). Setting appropriate conditions on this process produces

algorithms which have good global convergence properties, but with step lengths of one near the solution. When $\alpha = 1$ the step is just d_k, the so-called "quasi-Newton step."

The algorithms used here employ the more recently developed model trust region approach. The idea behind iterative techniques based on Newton's method is that, close to the solution, the new iterate β_{k+1} is obtained by minimizing a quadratic model of the objective function $L(\beta)$ using information available at the current iterate β_k. Consider the quadratic model m(s) given by

$$m_k(s) = L(\beta_k) + \nabla L(\beta_k)^T s + \frac{1}{2} s^T H_k s, \tag{13}$$

where $s \equiv \beta - \beta_k$ and H_k is assumed to be positive definite. Finding the minimizer of $m_k(s)$ gives the quasi-Newton step d_k defined above.

Suppose, though, that taking the quasi-Newton step produces an *increase* in the objective function we are trying to minimize. Then m_k is a poor model for L in the region containing the full quasi-Newton step. However, there is some smaller *region* in which we can *trust* m_k to *model* L. If we characterize this region by a sphere of radius r_k about the current iterate β_k, then we can find the next iterate by solving the following constrained optimization problem:

$$\min m_k(s) \text{ subject to } \|s\|_2 \leq r_k, \tag{14}$$

where $\|\bullet\|_2$ denotes the l_2-(or Euclidean) norm. (While more complex norms could be used, we have no need of such generality here.)

Given a current iterate and quadratic model, a simple version of the model trust region approach is:

1. Solve (14) to get a candidate step s_c for the next iterate.

2. Evaluate $L(\beta_c)$, where $\beta_c = \beta_k + s_c$. Is β_c acceptable?. If NO, shrink r_k and go to step 1.

3. Convergence? If NO, then:
   (i) let $\beta_{k+1} = \beta_c$,
   (ii) evaluate $\nabla L(\beta_{k+1})$,
   (iii) get a new quadratic model m_{k+1},
   iv) determine r_{k+1}, and
   (v) go to step 1.

Acceptability of β_c in step 2 is determined by conditions similar to those used in linesearch routines; in practice this is little more restrictive than requiring $L(\beta_c)$ to be smaller than $L(\beta_k)$. The trust region radius r_k is adjusted by comparing the actual function decrease $L(\beta_{k+1}) - L(\beta_k)$ to the function decrease predicted by the quadratic model, i.e., $m_k(\beta_{k+1}) - L(\beta_k)$. If the agreement is good, then the trust region can be expanded. If the agreement is poor, then the trust region is shrunk. A more complete discussion of the details of this approach, such as how (14) is solved and rules for selecting r_k to ensure global convergence, is well beyond the scope of this paper--we refer the interested reader to Dennis and Schnabel (1983, Chapter 6). Close to the solution, both the line search and trust region methods take full quasi-Newton steps, so as to preserve the convergence properties of the local strategy.

5.2 Local Strategies

Choosing a "local strategy" is equivalent to choosing a method for determining H_k at each iteration. The ideal choice is the true Hessian, i.e., $H_k = \nabla^2 L(\beta_k)$, which gives the very fast local convergence properties ("quadratic convergence") of Newton's method. Unfortunately, calculating the true Hessian is extremely expensive, both computationally and from the standpoint of writing computer programs.

Other choices for H_k are convenient, but slow ("linear"). These include steepest descent ($H = I$), and BHHH (defined below). Actually, BHHH can be fast under certain circumstances, but in general this cannot be guaranteed, and previous experience with discrete choice models shows that it can often be quite slow--see Bunch (1988). We note that BHHH is an analog to the Gauss-Newton method for nonlinear least squares, which can also be quite slow. Finally, the least-change secant update methods (or "variable metric methods") such as DFP (used in CHOMP) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) exhibit fast local convergence properties ("superlinear")--see Dennis and Schnabel (1983). These methods use the gradient information obtained at each iteration to update a stored approximation to the Hessian. There are two drawbacks to these methods: (i) they can take quite a number of iterations to build up a reasonable approximation, and (ii) they ignore useful structural information about the problem being solved.

To see this special structure, consider the maximum likelihood estimation problem

$$(15) \qquad \min_{\beta \in R^p} L(\beta) = - \sum_{n=1}^{Nobs} \ln P_{c(n)}(\beta),$$

where c(n) denotes the choice by household n and other notation has been suppressed. The full expression for the Hessian is given by

$$(16) \qquad \nabla^2 L(\beta) = \sum_{n=1}^{Nobs} \frac{1}{P_{c(n)}^2} \nabla P_{c(n)} \nabla P_{c(n)}^T - \sum_{n=1}^{Nobs} \frac{1}{P_{c(n)}} \nabla^2 P_{c(n)}$$

$$\equiv C(\beta) + A(\beta).$$

where $C(\beta)$ may be readily computed using information already available from the gradient calculation. However, computing $A(\beta)$ involves a matrix of second derivatives for each observation, and is extremely expensive. BHHH approximates $\nabla^2 L(\beta)$ by ignoring $A(\beta)$. The least-change secant update methods (DFP and BFGS) discard $C(\beta)$ and approximate the matrix $\nabla^2 L(\beta)$ directly.

Equation (16) motivates a special-purpose algorithm which calculates $C(\beta)$ at each iteration, but which also builds up a secant approximation to $A(\beta)$. There are therefore two quadratic models available at each iteration, given by the following Hessian approximations:

(17a) $$H_k^1 = C(\beta_k), \text{ and}$$

(17b) $$H_k^2 = C(\beta_k) + A_k.$$

where A_k is the secant approximation to $A(\beta_k)$. For a discussion on computing A_k, see Bunch (1987).

Each model in (17) can be used to calculate the predicted function decrease in going from β_k to β_{k+1}. The actual function decrease versus the predicted function decrease for each quadratic model is used to determine which is performing better. The algorithm can then *switch* between the two for purposes of determining successive iterates. For more details on the model switching approach, see Dennis, Gay, and Welsch (1981a, b). Since H_k^1 alone gives the BHHH approach, the *Model Switching* algorithm will always perform at least as well as BHHH, and for more difficult problems it retains local q-superlinear convergence due to the availability of the the model given by equation (17b).

5.3 Three MLE Algorithms

The three algorithms compared here are produced by combining BFGS, BHHH, and Model Switching (local strategies) with the model trust region (global strategy). The

trust region subroutines used in this study are extensions of the codes described in Gay (1983) and Dennis, Gay, and Welsch (1981a, b). For a more detailed discussion of model switching algorithms for choice models, see Bunch (1987). The BHHH method was implemented by using the same code as for Model Switching, but with the second quadratic model always suppressed.

In addition to the trust region subroutines, other modules are required. Bunch (1987) describes the general modules required for probabilistic choice models; appropriate modules for MNP were produced using data arrays and subroutines similar to those described in Daganzo (1979). One module implements the model specifications from Table 1, and various Cholesky factorizations are used to ensure positive definiteness of the estimated covariance parameters. Another module to compute probit integrals is also required. Comparison results for this section were obtained using the fixed-order Mendell-Elston choice probabilities described in section 6. (This method is accurate enough for purposes of comparing MLE algorithms, but is much faster than numerical integration.) Finite difference gradients were used, although future plans include developing modules for "direct" evaluation of gradients using various analytical expressions.

5.4 Stopping Rules

One of the more tricky issues in parameter estimation concerns when to stop an iterative search, and since probit estimation is a difficult problem this takes on added importance. A feature of the model trust region approach used in this paper is that it quite naturally leads to stopping rules which provide useful convergence diagnostics when solving parameter estimation problems. An integrated set of five stopping rules developed by David M. Gay is used by all three algorithms described above. One (or more) of the rules will be satisfied at convergence. For a complete discussion see Gay (1982) and Dennis, Gay, and Welsch (1981a, b).

Three of the rules are "favorable," only occurring when the quadratic model at the solution is deemed to be "trustworthy" and the Hessian approximation appears to be positive definite. They are called x-convergence, relative function-convergence, and absolute function-convergence, and more than one may occur at the solution. X-convergence holds when the relative step size for the iterate falls below a user-defined tolerance, and relative function-convergence holds when the relative change in $L(\beta)$ falls below another user-defined tolerance. Absolute function-convergence is included for the rare case in which both $L(\beta^*) = 0$ and $\beta^* = 0$, where β^* is the solution. If one (or more) of these favorable stopping rules does not hold, Gay (1982) shows that under reasonable assumptions one of two remaining conditions, "singular convergence" or "false convergence," *must* hold.

"Singular convergence" occurs when the relative change in $L(\beta)$ is small, but the Hessian of $L(\beta)$ appears to be singular (or nearly singular). This is a useful diagnostic for probit estimation, since two different problems seem possible: (i) the model specification is *a priori* unidentified, or (ii) multicollinearity in the data produces a MLE problem which is effectively overparameterized. "False convergence" occurs when the step sizes are getting small, but the iterates appear to be converging to a non-critical point, i.e., a point for which the gradient is nonzero. This condition could occur if the probit model is overparameterized to the degree that the log-likelihood is unbounded. It could also occur if there are errors in computing the probit model, or if the method used for computing choice probabilities is non-smooth (see section 6).

To further illustrate the utility of this algorithm's features, we include some small numerical examples which draw on the discussion of identification issues from the previous section. The examples use the trinomial probit data set from Daganzo (1979, page 17), in which each alternative is characterized by a measurement on one generic attribute.

Assume there are no alternative-specific dummies, no taste variation, and the mean of the utilities is given by $V = [\theta_1 a_1, \theta_1 a_2, \theta_1 a_3]^T$ for all specifications. The four specifications are defined in terms of the following error covariance matrices:

$$\Sigma_1 = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{18}$$

$$\Sigma_2 = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & 1 & 0 \\ 0 & 0 & \theta_3 \end{bmatrix}, \tag{19}$$

$$\Sigma_3 = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & \theta_3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{20}$$

$$\Sigma_4 = \begin{bmatrix} 1 & \theta_2 & \theta_4 \\ \theta_2 & \theta_3 & \theta_5 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix}. \tag{21}$$

Now, (18) and (20) are identified whereas (19) and (21) are not. Furthermore, (18) is a special case of (19) and (20) is a special case of (21). For a summary of numerical results using these examples, see Table 12. Note that the log-likelihoods for the overparameterized specifications are always identical to the those for the identified specifications. In every case but one, the algorithm gives an indication of the specification problem. (In the remaining case, the t-scores are insignificant for all the parameter estimates.) Singular convergence may be interpreted as discussed previously. False convergence occurs when the step sizes get very small, but the quadratic model is inadequate. This could happen if the search direction is very steep, as might occur on the side of a ridge with a flat bottom.

5.5 Numerical Comparison of MLE Algorithms

A numerical comparison of the three MLE algorithms defined above was made using the six probit models with the empirical formulation presented in section 4. Probit integrals were computed using the fixed-order Mendell-Elston approximation described in section 6. For each computer run the three algorithms were initialized using the same starting point. Starting points were obtained in a "realistic" fashion, i.e., they were obtained in a sequential manner by estimating the less complex models first, and then using these results to get starting points for the more complex models. The nesting of the specifications was exploited in the obvious way; for example, results for IID probit were used to get a starting point for the Σ_ε model.

Results are given in Table 13. We report iterations, function evaluations, and convergence criteria.[1] The number of iterations for these codes also corresponds to the number of gradient evaluations: at each iteration the estimate of H_k is updated, but multiple function evaluations using the same H_k may be attempted during each iteration. With the noted exceptions, the algorithms converged using the relative-function criterion. A user-defined limit of 200 function evaluation was imposed. The behavior of the algorithms was reliable and consistent, in the sense that the three algorithms always converged to the same solution. In most cases the log-likelihoods agreed to five decimal places (consistent with the relative-function stopping criterion), and the parameter estimates to two or more decimal places.

These results demonstrate many of the theoretically expected features of the algorithms, and corroborate previous work (Bunch 1988). The BHHH method showed satisfactory performance in only two cases (one being IID probit). These results illustrate the slow (linear) convergence of BHHH; in one case BHHH did not converge after 200 function evaluations. The inadequacy of the BHHH approximation to the Hessian for these models is supported by the appearance of the "singular convergence" stopping criterion in

half the cases. BHHH often took two to three times the number of function evaluations required by Model Switching, and in the worst case would have taken more than an order of magnitude more work.

The BFGS method was also outperformed by the Model Switching algorithm, behaving in a manner described earlier. BFGS, although a locally q-superlinear method, does not exploit the special structure of the problem as does Model Switching, and expends substantial effort in building up a Hessian approximation. It typically required 1.5 to 3.5 times as much work as the Model Switching method.

6. Comparison of Probit Integral Approximation Methods

The car ownership models considered here are trinomial, and it was therefore practical to use an accurate numerical integration method to obtain the estimates in section 4. These "accurate estimates" also provide a useful standard for comparing some alternative methods for calculating probit integrals. The methods considered here are Clark's approximation, and two variants of the Mendell-Elston approximation. The "separated-split" method of Langdon (1984) is similar in nature to Mendell-Elston, but is more cumbersome to code. For these and other reasons discussed below, we did not extend the study to include separated-split. The idea of "choice simulators" due to McFadden (1989) has recently been receiving attention, and may perhaps be the subject of a future study.

The results presented here add numerical evidence to the previous studies of Clark's approximation (Horowitz, et al., 1982) and Mendell-Elston (Kamakura, 1989); a notable difference is that our study uses empirical data rather than simulated data. Following these two studies, we compare the methods on the basis of "estimation accuracy" using a variety of measures. We briefly describe the methods and then give the comparison.

6.1 Algorithms for Calculating Probit Integrals

Starting from equation (3), the probit integral we wish to calculate is conveniently transformed as follows:

$$(22) \qquad P(j|\, V_U(\theta, \mu, X), \Sigma_U(\Sigma_\delta, \Sigma_\varepsilon, X)) = \text{Prob}[\, U_j > U_i \text{ for all } i \neq j]$$

$$= \text{Prob}[U_i - U_j < 0, i \neq j]$$

$$= \text{Prob}\,[Z_i^{(j)} < 0, i \neq j]$$

$$= \Phi(0|\, V, \Sigma)$$

where $Z_i^{(j)} = U_i - U_j$ and V and Σ are the mean and covariance matrix, respectively, for the J-1 vector of Z's. For this study J=3, and thus we require evaluation of the bivariate normal CDF. The first method is the accurate procedure due to Owen (1956). The specific algorithm used to calculate the bivariate normal integral is due to Donnelly (1973).

The second method is Clark's approximation, as described in Daganzo (1979). The method rewrites equation (22) as follows:

$$(23) \qquad \text{Prob}[\, U_j > U_i \text{ for all } i \neq j] = \text{Prob}[\, U_j > \max\,(U_i) \text{ for all } i \neq j]$$

$$= \text{Prob}[\max\,(Z_i^{(j)}) < 0, i \neq j]$$

and approximates the maximum of the Z's by a normally distributed random variable with mean and variance given by Clark's formulas. The algorithm used here was a modified routine from CHOMP.

For Mendell-Elston, it is convenient to limit discussion to the case J=3 and label the alternatives A, B, and C. The probability of choosing alternative A may be written as

$$(24) \qquad P_A = \text{Prob}[U_A > U_B \text{ and } U_A > U_C]$$

$$= \text{Prob}[U_A > U_B] \bullet \text{Prob}[U_A > U_C \mid U_A > U_B]$$

Each term can be approximated by standard normal distribution, after a suitable transformation and determination of the appropriate mean and variance expressions--see Kamakura (1989).

Note that the ordering of the decomposition in equation (24) is arbitrary: a different ordering is possible and the approximation would yield a different numerical result. The method used for determining the ordering is therefore an issue in defining an approximation. One simple option is to use a fixed lexicographic ordering to perform the calculation: we refer to this as the "fixed-order" Mendell-Elston approximation.

By arguments similar to Langdon (1984) for separated-split, the term Prob[$U_A > U_C$ | $U_A > U_B$] is apparently best approximated by a normal distribution when Var[$U_A - U_B$] < Var[$U_A - U_C$]. If these variances are used to determine the ordering of the decomposition, then the approximation will depend upon the actual values of the MNP model parameters. Since the ordering could *change* during the course of an iterative maximum likelihood search, we refer to this approach as the "variable-order" Mendell-Elston approximation. The fixed- and variable-order Mendell-Elston ("ME") approximations were coded in Fortran, and they use the routine ALNORM (Hill 1973) for evaluating the standard univariate normal CDF.

6.2 Numerical Comparison

Parameter estimates were obtained for the six empirical formulation MNP models using the Clark, fixed-order ME, and variable-order ME approximations, respectively. The "accurate estimates" using numerical integration (NI) from section 4 were used as starting values to save time and to facilitate comparison of the results. A summary of the runs appears in Table 14. The first column reports log-likelihoods evaluated using the method-estimated parameter values and method-estimated probabilities. The second column reports the percentage error of the method-log-likelihood versus the NI-log-likelihood. With the exception of IID Probit, the Clark method is dominated by the two ME methods by factors ranging from 3 to 15. The two ME methods are more comparable. For IID Probit the two methods are the same due to the fixed covariance matrix, but fixed-order ME is more

accurate than variable-order ME in all but one of the remaining cases. This is generally supported by examining the mean absolute value percentage errors (MAPE) for the method-coefficient estimates versus the NI-coefficient estimates, given by

$$\text{MAPE}(\beta^{\text{Method}}, \beta^{\text{NI}}) = \sum_{k=1}^{p} \frac{| \beta_k^{\text{Method}} - \beta_k^{\text{NI}} |}{| \beta_k^{\text{NI}} |}. \tag{25}$$

However, for two models (Σ_ε and $X'\Sigma_\delta X+\Sigma_\varepsilon$) the Clark method is comparable to ME(FO) and ME(VO) is less accurate.

Another interesting indicator of "performance" is the work required for finding the estimates. *Ceteris paribus*, a more accurate approximation method should requires less work to find estimates from the "accurate estimate" starting point. The Clark method generally requires more iterations that the ME methods.

An interesting point from Table 14 is that the ME(VO) method appears to get "hung up," expending large numbers of function evaluations and then stopping using the "false convergence" criterion. This (unanticipated) behavior occurs as a consequence of embedding the variable-ordering algorithm within the maximum likelihood estimation routine. The variable ordering feature results in non-smooth probabilities and hence a non-smooth likelihood function. A related feature is the possibility that different orderings may provide additional "degrees of freedom" for optimization purposes, yielding a distorted log-likelihood function. It is generally well-known that quasi-Newton methods perform poorly on such non-smooth problems. From a theoretical perspective, the fast local convergence properties of quasi-Newton methods strongly depend upon the smoothness assumptions for the objective function (see Dennis and Schnabel 1983).

The separated-split method (not considered here) includes more complex requirements for order-determination than does Mendell-Elston, but will produce similar

problems with non-smoothness and distortion of the log-likelihood function. The potential for this difficulty was, in fact, anticipated by Langdon (1984), and one very simple answer to the problem is to use the fixed-order method included here. Other approaches for improving approximations based on equation (24) await more research.

Another comparison based on the method-estimated probabilities evaluated using method-estimates appears in Table 15, which is patterned after Table II of Horowitz, et al. (1982). The table shows the maximum and root-means-square of (RMS) differences between the "accurate" and method-estimated probabilities; both absolute and relative differences are reported. These results are consistent with those of the previous table: the two ME methods are much more accurate than the Clark method, and the fixed-order ME method essentially dominates the variable-order ME method.

The approximations for the IID error models appear to be much better than for the non-IID error models. The identity matrix appears to dominate the covariance structure for the IID error models, producing conditions under which the approximations are known to perform reasonably well, i.e., the variances are not greatly different for the various alternatives, and there are no negative correlations. The absolute differences between the "accurate" choice probabilities and the Clark-estimated choice probabilities for the IID error models are ≤ 0.046 (maximum) and 0.012 (RMS). For fixed-order ME the differences are ≤ 0.010 (maximum) and 0.002 (RMS), a sizeable improvement over Clark.

For the non-IID error models the Clark approximation exhibits extremely poor behavior in terms of relative differences: 13.9 (maximum) and 0.748 (RMS) for $P \geq 0.05$. The worst percentage difference for fixed-order ME with $P \geq 0.05$, on the other hand, is 0.148 (maximum) and 0.029 (RMS). On the whole, the fixed-order ME method appears to perform well.

Finally, we include a comparison of the non-IID model error-covariance-matrix parameter estimates for the three methods in Table 16. Estimates are reported for the parameters of the M matrix in equation (6), as well as the computed value for the correlation coefficient ρ for the normalization in equation (8). The Clark method produces correlations of either -1 or +1, and in the case of the Σ_ϵ model actually produces the wrong sign. The variable-order ME method performs better than Clark, but not as well as fixed-order ME.

7. Summary and Conclusions

In this paper we have revisited various issues relating to practical estimation of the multinomial probit model by applying state-of-the-art methods to an analysis of two empirical data sets over a wide range of model specifications. To set the context, we included a brief review of empirical probit results in the literature. Use of estimability conditions, tests of alternative specifications, and alternative normalizations and their interpretations were illustrated.

A recently developed maximum likelihood algorithm for choice models--which uses model trust regions and model switching--performed very well on these MNP problems, reliably and efficiently producing parameter estimates. The algorithm was shown to be much more efficient than alternative algorithms used in previous probit estimation work. In addition, the model trust region algorithms were shown to produce useful diagnostic information through the various convergence conditions. For example, unidentified model specifications or overparameterized models can sometimes be detected through the singular or false convergence conditions, as demonstrated through some numerical examples.

Our comparison of alternative methods for calculating probit integrals provides additional evidence against the use of Clark's approximation. The Clark results for the non-IID specifications are substantially different from the numerical integration results, and

the Clark non-IID covariance estimates are extremely ill-conditioned. On the other hand, the fixed-order Mendell-Elston approximation performed well and appears to be a promising technique.

Since this method is much less expensive to use than numerical integration, it could open the door to applications of MNP involving more than three alternatives per choice set. Langdon (1984) claims that the separated-split method should be practical for choice sets up to about 15 alternatives. The fixed-order ME method is less expensive, and would therefore exceed this number. On the other hand, the accuracy for all these approximations remains to be demonstrated for larger choice sets, and is an area for future testing. Finally, the problems of imbedding a variable-order approximation technique were demonstrated, calling into question the feasibility of using either the variable-order Mendell-Elston method or the separated split method in maximum likelihood search routines.

Acknowledgements

The authors are grateful to Frank Koppelman and two referees for their helpful comments on an earlier version of this paper. This study was supported in part by the U. S. Department of Transportation through the Region Nine Transportation Center at University of California, Berkeley. The data file used in the study was prepared while the second author (RK) was at Bureau Goudappel Coffeng, Develnter, the Netherlands, with the assistance of Co Holzaphell. The accessibility measures used in the analysis were provided by Hugh Gunn of Hague Consulting Group, the Hague, the Netherlands. The Dutch National Mobility Panel survey is sponsored by the Projectbureau for Integrated Traffic and Transportation Studies, and the Directorate General of Transport of the Netherlands Ministry of Transport and Public Works.

Footnotes

1. We follow the numerical analysis convention of reporting function and gradient evaluations. Accurate CPU times are difficult to obtain (especially on time-shared machines) and are machine dependent. Function and gradient evaluations adequately reflect the amount of work required for solving expensive problems, since the amount work for algorithm overhead is relatively small in this case. The work required for a gradient evaluation requires roughly Npar times as much work as a function evaluation (where Npar = no. of parameters) if finite differences are used (as in this paper), but can be less for more efficient implementations. The default rule for the relative-function convergence tolerance for this family of codes is $\max\{10^{-10}, \text{machep}^{2/3}\}$, where "machep" is machine epsilon. The corresponding rule for the relative x-convergence tolerance is $\text{machep}^{1/2}$. The numerical results were obtained on a DEC MicroVAX II in double precision. To save work, convergence tolerances were set as though the runs were being obtained in single precision; the relative-function and relative-x tolerances were set to 2.4×10^{-6} and 6.1×10^{-5}, respectively.

References

Albright, R. L., S. R. Lerman and C. F. Manski (1977) *Report on the Development of an Estimation Program for the Multinomial Probit Model.* Preliminary report prepared for the Federal Highway Administration, October, 54p.

Becker, G. S. (1965) A theory of the allocation of time. *The Economic Journal*, 75, 493-517.

Beckmann, M. J., R. L. Gustafson, and T. F. Golob (1973) Locational factors in automobile ownership decisions. *Annals of Regional Science*, 7, 1-12.

Ben-Akiva, M. and T. J. Atherton (1977) Methodology for short-range travel demand predictions, *Journal of Transport Economics and Policy*, 11, 224-261.

Ben-Akiva, M. and S. R. Lerman (1974) Some estimation results of a simultaneous model of auto ownership and mode choice to work. *Transportation*, 3, 357-76.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974) Estimation and inference in non-linear structural models. *Annals of Economic and Social Measurement*, 3, 653-655.

Bunch, D. S. (1987) Maximum likelihood estimation of probabilistic choice models. *SIAM Journal on Scientific and Statistical Computing*, 8, 56-70.

Bunch, D. S. (1988) A comparison of algorithms for maximum likelihood estimation of choice models. *Journal of Econometrics*, 38, 145-67.

Bunch, D. S. (1990) Estimability in the multinomial probit model. *Transportation Research B* (to appear).

Bunch, D. S. and R, Kitamura (1989) Multinomial Probit Model Estimation Revisited: Testing of New Algorithms and Evaluation of Alternative Model Specifications for Trinomial Models of Household Car Ownership. Transportation Research Group Research Report UCD-TRG-RR-4, University of California, Davis.

Clark, C. E. (1961) The greatest of a finite set of random variables. *Operations Research*, 9, 145-62.

Currim, I.S. (1982) Predictive testing of consumer choice models not subject to independence of irrelevant alternatives. *Journal of Marketing Research*, 19, 208-22.

Daganzo, C. (1979) *Multinomial Probit : The Theory and Its Application to Demand Forecasting.* Academic Press, New York.

Daganzo, C. and L. Schoenfeld (1978) *CHOMP User's Manual.* Berkeley, CA: University of California.

Dansie, B. R. (1985) Parameter estimability in the multinomial probit model. *Transportation Research B*, 19B, 526-28.

Davidon, W. C. (1959) Variable metric method for minimization. A. E. C. research and development report ANL5990, revised.

Dennis, J. E., Jr., D. M. Gay, and R. E. Welsch (1981a) An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7, 348-368.

Dennis, J. E., Jr., D. M. Gay, and R. E. Welsch (1981b) Algorithm 573 NL2SOL - An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7, 369-383.

Dennis, J. E., Jr. and R. B. Schnabel (1983) *Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ.

Donnelly, T. G. (1973) Algorithm 462: Bivariate normal distribution. *Communications of the ACM*, 16, 636.

Fletcher, R. and M. J. D. Powell (1963) A rapidly convergent descent method for minimization, *Computer Journal*, 6, 163-168.

Gay, D. M. (1982) On convergence testing in model/trust-region algorithms for unconstrained optimization, Computing Science Technical Report No. 104, Bell Laboratories.

Gay, D. M. (1983) Algorithm 611. Subroutines for unconstrained minimization using a model/trust region approach, *ACM Transactions on Mathematical Software*, 8, 503-524.

Geizner, J. and A. Daly (1981) *Zuidvleugel Study Report 9: Models of Car Ownership and License Holding.* Cambridge Systematics Europe, The Hague, the Netherlands.

Hausman, J. A. and D. A. Wise (1978) A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46, 403-426.

Hill, I. D. (1973) Algorithm AS 66. The normal integral, *Applied Statistics*, 22, 424-427.

Horowitz, J. (1980) The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand. *Transportation Research*, 14B, 4, 331-42.

Horowitz, J. L., J. M. Sparmann and C.F. Daganzo (1982) An investigation of the accuracy of the Clark approximation for the multinomial probit model. *Transportation Science*, 16, 3, 382-401.

Johnson, L. and D. Hensher (1982) Application of multinomial probit to a two-period panel data set. *Transportation Research A*, 16A, 5/6, 457-64.

Kamakura, W. A. (1989) The estimation of multinomial probit models: a new calibration algorithm, *Transportation Science*, 23 (4), 253-265.

Kamakura, W. A. and R. K. Srivastava (1984) Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research*, 21, 420-34.

Kitamura, R. (1987) A panel analysis of household car ownership and mobility. *Proceedings of the Japan Society of Civil Engineers*, No. 383, IV-7, 13-27.

Langdon, M.G. (1984) Improved algorithms for estimating choice probabilities in the multinomial probit model. *Transportation Science*, 18, 3, 267-99.

Lasdon, L. S., A. D. Warren, and M. W. Ratner (1982) *GRG2 User's Guide*. Austin TX: Department of General Business, University of Texas at Austin.

Lerman, S.R. and C.F. Manski (1981) On the use of simulated frequencies to approximate choice probabilities. In *Structural Analysis of Discrete Data with Econometric Applications* (C.F. Manski and D. McFadden, editors), pp. 305-19. MIT Press, Cambridge, MA.

Lierop, W. van (1986) *Spatial Interaction Modelling and Residential Choice Analysis*. Gower Publishing, Aldershot, England.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In *Frontiers of Econometrics* (P. Zarembka, editor). Academic Press, New York.

McFadden, D. (1981) Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data with Econometric Applications* (C.F. Manski and D. McFadden, editors) pp. 198-272. MIT Press, Cambridge, MA.

McFadden, D. (1989) A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica*, 57(5), 995-1026.

Mendell, N. and R. C. Elston (1974) Multifactorial qualitative traits: Genetic analysis and prediction of recurrence risks. *Biometrics*, 30 (March), 41-57.

Miller, E. J. and S. R. Lerman (1981) Disaggregate modelling and decisions of retail firms: a case study of clothing retailers. *Environment and Planning A*, 13, 729-746.

Owen D. B. (1956) Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics*, 27, 1075-1090.

Sparmann, J. M., C. F. Daganzo and M. Soheily (1983) Linear probit models: Statistical properties and improved estimation methods. *Transportation Research B*, 17B, 67-86.

Train, K. (1984) *Qualitative Choice Analysis: Theory, Economics, and an Application to Automobile Ownership*. MIT Press, Cambridge, MA.

Wissen, L. J. G. van and H. J. Meurs (1989) The Dutch Mobility Panel: Experiences and evaluation. *Transportation* (forthcoming).

Table 1

Model Specifications for Linear-in-Parameters MNP Framework

Tastes (δ)	Errors (ε_n)	Σ_U
Fixed	IID	I
Fixed	Non-IID	Σ_ε
Random Uncorrelated	IID	$X^T D X + I$
Random Uncorrelated	Non-IID	$X^T D X + \Sigma_\varepsilon$
Random Correlated	IID	$X^T \Sigma_\delta X + I$
Random Correlated	Non-IID	$X^T \Sigma_\delta X + \Sigma_\varepsilon$

Table 2
Summary of Multinomial Probit Empirical Studies

Reference	No. of Alts[a]	No. of Specs[b]	NOBS[c]	Max No. Attributes	Taste Variation?	Alt. Spec Constants	Prob/Opt[d]
Hausman and Wise (1978)	3	4	100	3	uncorr	Y	NI BHHH
Albright, Lerman, and Manski (1977)	2, 3	1	557	5	corr	Y	Clark St. Descent
Currim (1982)	5	5	369	9	uncorr	N	Clark DFP
Hensher and Johnson (1982)	2, 3	5	100	3	No	Y	Clark DFP
Miller and Lerman (1982)	14	3	181	7	1 variable	Y	Clark DFP
Kamakura and Srivistava (1984)	3	3	2000, 2600	2	uncorr	N	Clark BFGS
van Lierop (1986)	2, 3	9	≤1107	3	No	Y	Clark DFP

a Number of alternatives per choice set.

b Number of different model specifications estimated in the study.

c Total number of observed choices in the data set (s).

d Method for calculating choice probabilities (NI = "numerical integration," Clark = Clark's approximation), and optimization method used in obtaining maximum likelihood estimates--see text for abbreviations.

Table 3
Summary Measures for Subsample of Dutch Mobility Panel

(945 Households from 1984)

Summary Statistics:

Variable	Abbrev.	Ave. per household
No. of Workers	Workers	0.9672
Total personal income	TPInc	32,221 Guilders
No. of Cars	NCars	0.907
Household Size	HHSize	2.964
No. of Adults (≥17 yrs.)	Adults	2.298

Distribution of Cars:

NCARS	No. of Households	Percentage
0	206	(22%)
1	632	(67%)
2 or more	107	(11%)

Table 4
Explanatory Data Formulations for Household Utility Functions

a. Variables (see also Table 3):

Abbreviation	Description
N_w	No. of annual work-trips
CCost	Average annual cost of owning one car
SubCost	Average annual subsistence cost per person
RI	Remaining income = TPInc - NCars * CCost - HHSize * SubCost
N_w	No. of annual work-trips
Acc_k^m	Accessibility measure for trip destinations for mode m and trip-type k,k = w (work), nw (non-work), and m = c (car), t (transit)
CPAdult	Cars per adult = NCars/Adults
CPWorker	Cars per worker = NCars/Workers if Workers > 0, 0 otherwise.
$WtdAcc_w$	Weighted accessibility for work trips $= \min(1, CPWorker)*Acc_w^c + \min(1, CPWorker)*Acc_w^t$
$WtdAcc_{nw}$	Weighted accessibility for non-work trips $= \min(1, CPAdult)*Acc_{nw}^c + \min(1, CPAdult)*Acc_{nw}^t$

b. Formulations:

Theoretical	Empirical
Log(RI)	Log(RI)
N_w/RI	Acc_{nw}^m, m = c or t
$N_wWtdAcc_w/RI$	CPAdult
$WtdAcc_{nw}$	CPWorker

Table 5
Estimated Coefficients and t-Scores for Alternative Car Ownership Models
(Theoretical Formulation)

COEFFICIENT ESTIMATES							
		IID Errors			Non-IID Errors		
		Fixed	Random Taste		Fixed	Random Taste	
Explanatory Variable	Logit	Taste	Uncor.	Cor.	Taste	Uncor.	Cor.
One Car Dummy	0.13	0.14	0.13	0.12	0.13	0.13	0.14
	11.2	12.2	8.8	9.8	7.6	6	12.4
Two Car Dummy	0.10	0.10	0.10	0.10	-0.01	0.10	0.16
	5.4	5.5	5.1	5.3	-0.3	2	10.8
ln(RI)[a]	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10.8	13.2	6.3	6.4	7.2	6.0	7.7
N_w/RI (x 100)	-0.12	-0.13	-0.14	-0.18	-0.13	-0.22	-0.10
	-3.2	-3.2	-1.2	-1.2	-3.1	-1.3	-1.04
$N_w WtdAcc_w$/RI	0.19	0.21	0.21	0.22	0.24	0.26	0.19
(x 10)	5.3	5.4	3.2	2.9	5.3	3.0	3.7
$WtdAcc_{nw}$	-0.006	-0.006	-0.003	-0.002	-0.001	-0.002	-0.015
	-0.6	-0.6	-0.4	-0.2	-0.1	-0.2	-1.9
L(0)[b]	-1038.19	-1038.19	-1038.19	-1038.19	-1038.19	-1038.19	-1038.19
L(C)	-801.14	-801.14	-801.14	-801.14	-801.14	-801.14	-801.14
L(β)	-688.47	-689.26	-681.69	-680.12	-686.68	-677.81	-666.79
-2[L(0)-L(β)]	699.43	697.86	713.00	716.15	703.02	720.75	743.23
d.f. (= No. of parameters)	6	6	10	16	8	12	18
-2[L(C)-L(β)]	225.32	223.76	238.90	242.04	228.91	246.64	269.13
d.f.	4	4	8	14	6	10	16
1 - L(β)/L(0)	0.337	0.336	0.343	0.345	0.339	0.347	0.358
1 - L(β)/L(C)	0.141	0.140	0.149	0.151	0.143	0.154	0.168

[a] Coefficient estimates have been standardized relative to ln(RI) for comparison purposes. See Table 1 for model specifications, and Table 4 for definitions of explanatory variables.

[b] L(0) denotes log-likelihood for a naive model which assumes equi-probable outcomes, L(C) is the log-likelihood for a model with alternative-specific constants only, and IID covariance matrix; L(β) is the log-likelihood at the estimated solution for the model.

Table 6
Estimated Coefficients and t-Scores for Alternative Car Ownership Models (Empirical Formulation)

COEFFICIENT ESTIMATES							
		IID Errors			Non-IID Errors		
		Fixed	Random Taste		Fixed	Random Taste	
Explanatory Variable	Logit	Taste	Uncor.	Cor.	Taste	Uncor.	Cor.
One Car Dummy	0.16	0.18	0.15	0.15	0.20	0.16	0.09
	6.2	6.4	6.2	6.3	5.9	6.2	5.9
Two Car Dummy	0.14	0.14	0.13	0.14	0.14	0.17	0.08
	3.7	3.7	3.8	4.1	2.2	4.1	3.5
ln(RI)[a]	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10.6	9.4	5.1	7.3	5.8	5.2	7.7
Acc^{m}_{nw}	0.05	0.06	0.05	0.04	0.07	0.06	0.03
	4.3	4.4	2.1	4.4	4.5	3.8	4.3
CPAdult	-0.16	-0.19	-0.13	-0.13	-0.23	-0.15	-0.08
	-6.1	-6.4	-5.1	-5	-6.4	-5.1	-4.4
CPWorker	0.05	0.05	0.04	0.04	0.05	0.03	0.02
	4.0	4.3	3.1	3.7	3.1	3.1	2.8
L(0)[b]	-1038.19	-1038.19	-1038.19	-1038.19	-1038.19	-1038.19	-1038.19
L(C)	-801.14	-801.14	-801.14	-801.14	-801.14	-801.14	-801.14
L(β)	-668.33	-666.93	-662.54	-658.64	-664.66	-656.53	-651.91
-2[L(0)-L(β)]	739.72	742.52	751.30	759.10	747.06	763.32	772.56
d.f. (= No. of parameters)	6	6	10	16	8	12	18
-2[L(C)-L(β)]	265.62	268.42	277.20	285.00	272.96	289.22	298.46
d.f.	4	4	8	14	6	10	16
1 - L(β)/L(0)	0.356	0.358	0.362	0.366	0.360	0.368	0.372
1 - L(β)/L(C)	0.166	0.168	0.173	0.178	0.170	0.181	0.186

a Coefficient estimates have been standardized relative to ln(RI) for comparison purposes. See Table 1 for model specifications, and Table 4 for definitions of explanatory variables.

b L(0) denotes log-likelihood for a naive model which assumes equi-probable outcomes, L(C) is the log-likelihood for a model with alternative-specific constants only, and IID covariance matrix; L(β) is the log-likelihood at the estimated solution for the model.

Table 7
Comparison of Log-Likelihoods for Two Formulations

Model	No. of Params	Formulation	
		Theoretical	Empirical
IID Errors			
I	6	-689.26	-666.93
$X^T DX+I$	10	-681.69	-662.54
$X^T \Sigma_\delta X+I$	16	-680.12	-658.64
Non-IID Errors			
Σ_ε	8	-686.68	-664.66
$X^T DX+\Sigma_\varepsilon$	12	-677.81	-656.53
$X^T \Sigma_\delta X+\Sigma_\varepsilon$	18	-666.79	-651.99

Table 8

Likelihood Ratio Tests of the Significance of Random Tastes and Non-IID Errors (Empirical Formulation Models)

(a) Tests of Random Tastes

		IID Error Terms (Random Tastes)			Non-IID Error Terms (Random Tastes)	
		Uncor.	Cor.		Uncor.	Cor.
Fixed Tastes	LR	8.77	16.57	LR	16.26*	20.79*
	df	4	10	df	4	10
Uncor. Tastes	LR		7.80	LR		9.08
	df		6	df		6

(b) Tests of Non-IID Errors

		Non-IID Errors		
		Fixed Tastes	Random Tastes	
			Uncor.	Cor.
IID Errors	LR	4.53	12.02**	13.30**
	df	2	2	2

* Significant at the 5% level.
** Significant at the 0.5% level.

Table 9

Likelihood Ratio Tests of the Significance of Random Tastes and Non-IID Errors (Theoretical Formulation Models)

(a) Tests of Random Tastes

		IID Error Terms (Random Tastes)			Non-IID Error Terms (Random Tastes)	
		Uncor.	Cor.		Uncor.	Cor.
Fixed Tastes	LR	15.14**	18.28*	LR	17.73**	39.78**
	df	4	10	df	4	10
Uncor. Tastes	LR		3.14	LR		22.04**
	df		6	df		6

(b) Tests of Non-IID Errors

		Non-IID Errors		
		Fixed Tastes	Random Tastes	
			Uncor.	Cor.
IID Errors	LR	5.15	7.75*	26.66**
	df	2	2	2

* Significant at the 5% level.

** Significant at the 0.5% level.

Table 10
Derived Σ_ε's Using Non-IID Model Estimates (Empirical Formulation)

Normalization	Model Specification: Σ_ε (Fixed Tastes)	Model Specification: $X^TDX + \Sigma_\varepsilon$ (Uncorrelated Tastes)	Model Specification: $X^T\Sigma_\delta X + \Sigma_\varepsilon$ (Correlated Tastes)
Σ_ε^1 (Equation 8)	$\begin{bmatrix} 1 & -0.080 & 0 \\ -0.080 & 0.436 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.351 & 0 \\ 0.351 & 0.210 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.171 & 0 \\ 0.171 & 0.186 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
ρ_1	- 0.121	0.766	0.398
Σ_ε^2 (Equation 9)	$\begin{bmatrix} 1 & & \\ & 0.388 & \\ & & 0.087 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ & -0.217 & \\ & & 0.541 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ & 0.017 & \\ & & 0.207 \end{bmatrix}$
Σ_ε^3 (Equation 10)	$\begin{bmatrix} 4.400 & -0.634 & \\ -0.634 & 1 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 8.528 & 2.344 & \\ 2.344 & 1 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 9.782 & 0.847 & \\ 0.847 & 1 & \\ & & 1 \end{bmatrix}$
ρ_3	- 0.302	0.803	0.271

Table 11

Derived Σ_ε's Using Non-IID Model Estimates (Theoretical Formulation)

Normalization	Model Specification: Σ_ε (Fixed Tastes)	Model Specification: $X^TDX + \Sigma_\varepsilon$ (Uncorrelated Tastes)	Model Specification: $X^T\Sigma_\delta X + \Sigma_\varepsilon$ (Correlated Tastes)
Σ_ε^1 (Equation 8)	$\begin{bmatrix} 1 & -0.526 & 0 \\ -0.526 & 0.555 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.138 & 0 \\ 0.138 & 0.175 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1.142 & 0 \\ 1.142 & 1.303 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
ρ_1	- 0.707	0.329	1
Σ_ε^2 (Equation 9)	$\begin{bmatrix} 1 & & \\ & 0.709 & \\ & & -0.345 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ & 0.043 & \\ & & 0.160 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ & -1.134 & \\ & & -8.04 \end{bmatrix}$
Σ_ε^3 (Equation 10)	$\begin{bmatrix} 2.602 & -2.897 & \\ -2.897 & 1 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 10.44 & 0.576 & \\ 0.576 & 1 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 0.535 & 0.753 & \\ 0.753 & 1 & \\ & & 1 \end{bmatrix}$
ρ_3	- 1.80	0.178	1.03

Table 12
Convergence Examples

Σ	Starting Pt. $[\theta_1, \theta_2, --]$	Iterations	Func. Evals[a]	$L(\theta^{MLE})$	Convergence[b]
Equ. (18)	[0, 0]	5	6	339.34	Rel-Function
Equ. (19)	[0, 0, 1]	7	10	339.34	Rel-Function[c]
	[0.2, 0.5, 1]	12	47	339.34	False
	[0.1, 0.4, 1]	7	8	339.34	Singular
Equ. (20)	[0, 0, 1]	7	8	335.33	Rel-Function
Equ. (21)	$[\theta^{MLE}_{(Equ.\ 20)}, 0, --]$[d]	2	1	335.33	Singular

a Total number of function evaluations required.
b Type of convergence--see text.
c The t-scores for all estimates were insignificant.
d The estimates for (20) were used to get the starting point for (21).

Table 13
Comparison of MLE Algorithm Performance

Model	Source of Starting Point*	Algorithm		
		Model Switching	BHHH	BFGS
a. Iterations				
I	$\beta = 0$	11	11	23
Σ_ε	I	8	11	26
X^TDX+I	I	15	25[s]	26
$X^TDX+\Sigma_\varepsilon$	Σ_ε	22	47[s]	37
	X^TDX+I	12	18[s]	40
$X^T\Sigma_\delta X+I$	X^TDX+I	19	101[fl]	30
$X^T\Sigma_\delta X+\Sigma_\varepsilon$	$X^TDX+\Sigma_\varepsilon$	17	39[s]	51
	$X^T\Sigma_\delta X+I$	17	70[s]	52
Average:		**15.1**	**40.3**	**35.6**
b. Function Evaluations				
I	$\beta = 0$	15	14	34
Σ_ε	I	10	13	43
X^TDX+I	I	25	51[s]	33
$X^TDX+\Sigma_\varepsilon$	Σ_ε	35	91[s]	54
	X^TDX+I	20	37[s]	62
$X^T\Sigma_\delta X+I$	X^TDX+I	28	200[fl]	45
$X^T\Sigma_\delta X+\Sigma_\varepsilon$	$X^TDX+\Sigma_\varepsilon$	23	52[s]	72
	$X^T\Sigma_\delta X+I$	21	99[s]	79
Average:		**22.1**	**69.6**	**52.8**

* Nested model whose estimates were used to get a starting point.
"s" = Singular Convergence
"fl" = algorithm termintated when number of function evaluations reached the function limit of 200.

Table 14
Comparison of Integration Methods: Log-likelihoods, Mean of Absolute Percentage Error (MAPE) of Coefficients, and MLE Convergence Behavior (Empirical Formulation Models)

Model	Int. Method	Log-Like	Log-Lik % Error	MAPE of Coeff.	Iterations	Function Evals	Converg. Cond.
IID Errors							
I	NI	669.33					
	Clark	668.08	-0.19	1.50	10	17	Rel-F
	ME(VO)	667.04	-0.34	0.26	5	11	Rel-F
	ME(FO)	667.04	-0.34	0.26	4	7	Rel-F
X^TDX+I	NI	662.54					
	Clark	663.74	0.18	2.71	3	4	Rel-F
	ME(VO)	662.94	0.06	0.45	2	3	Rel-F
	ME(FO)	662.90	0.05	0.44	2	3	Rel-F
$X^T\Sigma_\delta X+I$	NI	658.64					
	Clark	660.95	0.35	2.77	13	19	Rel-F
	ME(VO)	658.24	-0.06	0.36	4	37	False
	ME(FO)	658.58	-0.01	0.22	4	8	Singular
Non-IID Errors							
Σ_ε	NI	664.66					
	Clark	654.74	-1.49	4.23	13	18	Rel-F
	ME(VO)	664.88	0.03	9.06	10	59	False
	ME(FO)	665.19	0.08	4.14	4	6	Rel-F
$X'DX+\Sigma_\varepsilon$	NI	656.53					
	Clark	649.18	-1.12	7.75	12	23	Rel-F
	ME(VO)	652.71	-0.58	6.51	27	97	False
	ME(FO)	658.14	0.25	6.94	9	17	Rel-F
$X^T\Sigma_\delta X+\Sigma_\varepsilon$	NI	651.99					
	Clark	645.36	-1.02	6.45	19	29	Rel-F
	ME(VO)	650.70	-0.20	9.71	12	55	False
	ME(FO)	652.70	0.11	6.36	13	22	Rel-F

Table 15
Comparison of Integration Methods: Accuracy of Approximated Choice Probabilities versus Numerical Integration[a]

Model	Integration Method	Absolute Difference		Relative Difference for P's ≥ 0.05		Relative Difference for P's < 0.05	
		Max	RMS[b]	Max	RMS	Max	RMS
IID Errors							
I							
	Clark	0.008	0.006	0.066	0.014	1.000	0.506
	ME(VO)	0.004	0.001	0.011	0.003	0.009	0.006
	ME(FO)	0.004	0.001	0.011	0.003	0.009	0.006
X^TDX+I							
	Clark	0.020	0.007	0.101	0.015	0.689	0.273
	ME(VO)	0.003	0.001	0.009	0.002	0.053	0.021
	ME(FO)	0.003	0.001	0.011	0.002	0.079	0.032
$X^T\Sigma_\delta X+I$							
	Clark	0.046	0.012	0.216	0.031	0.706	0.273
	ME(VO)	0.012	0.002	0.043	0.005	0.054	0.021
	ME(FO)	0.010	0.002	0.025	0.004	0.022	0.012
Non-IID Errors							
Σ_ε							
	Clark	0.778	0.078	12.840	0.748	3401.500	3401.500
	ME(VO)	0.033	0.011	0.204	0.026	0.203	0.143
	ME(FO)	0.029	0.009	0.098	0.020	0.099	0.070
$X'DX+\Sigma_\varepsilon$							
	Clark	0.735	0.048	13.850	0.475	17.620	7.224
	ME(VO)	0.084	0.021	0.348	0.058	0.615	0.333
	ME(FO)	0.028	0.010	0.107	0.022	0.171	0.086
$X^T\Sigma_\delta X+\Sigma_\varepsilon$							
	Clark	0.730	0.057	13.000	0.582	15.470	6.959
	ME(VO)	0.077	0.020	0.261	0.053	0.515	0.342
	ME(FO)	0.045	0.013	0.148	0.029	0.169	0.094

a Measures are based those used in Table II of Horowitz, et al. (1982). Results are based on 945 observations for empirical formulation models.

b Root-Mean-Square

Table 16
Comparison of Integration Methods: Covariance Estimates for Non-IID Errors in Empirical Formulation Models

Estimate	Method	Σ_ε	$X^TDX+\Sigma_\varepsilon$	$X^T\Sigma_\delta X+\Sigma_\varepsilon$
M_{12}	NI	0.080	0.351	0.171
	CA	-0.072	0.040	0.065
	ME(VO)	0.293	0.560	0.429
	ME(FO)	0.142	0.384	0.154
M_{22}	NI	0.656	-0.294	0.395
	CA	0.001	-0.000	-0.000
	ME(VO)	0.707	0.007	0.278
	ME(FO)	0.702	-0.319	0.422
ρ	NI	0.121	0.766	0.398
	CA	-1.000	1.000	1.000
	ME(VO)	0.383	1.000	0.839
	ME(FO)	0.198	0.769	0.341