RP-90-1

Paper No. 890348

PREPRINT

Title: Multinomial Probit Model Estimation Revisited: Testing of New Algorithms and Evaluation of Alternative Model Specifications for Trinomial Models of Household Car Ownership

Authors: David S. Bunch
Ryuichi Kitamura

University of California, Davis

Transportation Research Board
69th Annual Meeting
January 7-11, 1989
Washington, D.C.

Transportation Research Group
University of California at Davis

MULTINOMIAL PROBIT MODEL ESTIMATION REVISITED: TESTING OF NEW ALGORITHMS AND EVALUATION OF ALTERNATIVE MODEL SPECIFICATIONS FOR TRINOMIAL MODELS OF HOUSEHOLD CAR OWNERSHIP

David S. Bunch, and Ryuichi Kitamura

Research Report
UCD-TRG-RR-89-4

May 1989

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 for a full range of models, including specifications with random (uncorrelated and correlated) taste variation and/or a general covariance structure for alternative-specific errors, are obtained using a recently developed maximum likelihood estimation routine for choice models. The results provide useful insights into specification and testing of probit models, and the MLE algorithm demonstrates many desirable features, including useful convergence diagnostics. Finally, a numerical comparison of Clark's approximation versus numerical integration provides additional evidence against the use of Clark's approximation in probit estimation.

1. Introduction

Beginning with McFadden (1973) 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 a group of 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 the case of household car ownership, which is the subject of an empirical analysis presented later in this paper. Empirical 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. The more useful specifications require estimation of covariance parameters, and the properties of the likelihood function are virtually unknown in these circumstances. In addition, evaluating the MNP choice probabilities requires integration of the multivariate normal probability density, which is in general quite difficult, getting exponentially worse as the number of alternatives in the choice set increases.

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 has emerged--see Horowitz, et. al. (1982). Alternative approaches have been proposed by Sparmann, et. al. (1983), Langdon (1984), and McFadden (1986); however, none of these approaches has yet led to any published empirical results that we are aware of. On the other hand, 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.

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. In particular, a difficulty which has not been particularly well-addressed concerns whether one can verify if the 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 under these circumstances.

In this paper we revisit many of the above issues by taking advantage of a recently developed maximum likelihood estimation routine which uses state-of-the-art numerical optimization techniques, and is, we believe, much more efficient, reliable, and robust than those used in previous applications of MNP. We examine a full range of model specifications based on a linear-in-the-parameters MNP framework similar to those used by Hausman and Wise (1978) and Albright, Lerman, and Manski (1977), and estimate models using a recently collected data set on car ownership.

The utility model for car ownership is derived from basic economic assumptions, and gives rise to a set of explanatory variables which, although theoretically appealing, contains an obviously high degree of collinearity. The specifications vary from simple to extremely complex, beginning with IID probit, which makes extremely strong behavioral assumptions regarding independence among alternatives, all the way to a very general model involving fully correlated taste variation and non-IID alternative-specific errors. The demands of these analyses provide an excellent platform from which to address practical issues of probit estimation, including the testing of behavioral assumptions, examination of alternative but equivalent model normalizations, the accuracy of Clark's approximation, and the reliability and utility of numerical optimization algorithms.

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 briefly describes the maximum likelihood algorithm, and addresses convergence issues. In section 5 we derive the utility model for car ownership which is used to obtain the empirical results of section 6. We compare various model specifications, and discuss the accuracy of Clark's approximation. Comments on practical issues related to the numerical computation of probit estimates are included in section 7, followed by discussion and conclusions in section 8.

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:

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

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) alternative-specific random errors on utility, where μ is the mean and ε_n is random error with $E(\varepsilon_n) = 0$. This may be interpreted as the effect of unobserved attributes of the individual and/or the alternatives. In the sequel, the term "error" will refer to the alternative-specific errors associated with the ε_n term.

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

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

Note that μ in (1) could alternatively be represented via a formulation which allows dummy variables in X_n. However, by using (1) and (2) we choose to assume that the δ and ε

terms are always independent. Now, the probability that individual n selects alternative j is given by:

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

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

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 $\phi(x|\ m, S)$ 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 considerations since all J(J+1)/2 parameters are not identified.

A complete development of identification and normalization results is beyond the scope of this paper, but they may be found in Bunch (1989). Due to the structure of the MNP model, only the covariance matrix for the *differences* of the error terms is estimable, and in addition, one parameter may be fixed for scaling purposes. The maximum number of identifiable parameters in Σ_ε is therefore J(J-1)/2 - 1. For trinomial probit, one can consider finding estimates for the two free parameters in the matrix C given by

(4) $$C = \begin{bmatrix} 1 & C_{21} \\ C_{21} & C_{22} \end{bmatrix} \equiv \begin{bmatrix} \sigma_{11}-2\sigma_{31}+\sigma_{33} & \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} \\ \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} & \sigma_{22}-2\sigma_{32}+\sigma_{33} \end{bmatrix}.$$

where the σ_{ij} are elements of Σ_ε. In this study, model specifications containing a "general Σ_ε matrix" will mean that the specification includes the maximum number of identifiable error covariance parameters, and consequently estimation can be performed using the matrix C defined in equation (4). As discussed in section 6 below, such estimates may correspond to many different possible normalizations of Σ_ε.

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

In this section, available empirical results using multinomial probit models are reviewed. We are concerned with studies having more than two alternatives, with a focus on factors which relate to practical estimation issues: 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. The intent is to establish the domain which encompasses empirical results in the literature.

Hausman and Wise (1978)

This study, which we denote "HW78," estimates trinomial probit models of travel mode choice using a subsample of a data set from the Washington, D.C. area. 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 $\Sigma_U = X^TDX+I$, (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., $\Sigma_U = X^TDX+\Sigma_\varepsilon$. (They refer to these models as independent probit, covariance probit, saturated independent probit, and saturated covariance probit, respectively. They specify Σ_ε via non-identical diagonal terms and zero off-diagonal terms. This will be discussed in more detail later.)

All models are formulated with the same set of three explanatory variables and estimated with 100 observations from the data set. The optimization method used is based on the approach of Berndt, Hall, Hall, and Hausman (1974), which we will denote "BHHH." The choice probabilities are evaluated numerically using Owen's method (Owen, 1956; Johnson and Kotz, 1972). The gradient vectors for the search are calculated using the analytic expressions given in the paper.

The IID probit model results were found to be quite similar to those from the logit model, both in in terms of the estimated coefficients and log-likelihood values. This is consistent with the analysis of Horowitz (1980). In addition, they found that including uncorrelated taste variation on the generic attributes significantly improved their model's fit to the data. The introduction of alternative-specific dummies and non-IID errors, on the other hand, led to only marginal improvements. This was especially true for the $\mu \equiv 0 / \Sigma_U = X^TDX+I$ model versus the $\Sigma_U = X^TDX+\Sigma_\varepsilon$ model: the introduction of μ and Σ_ε only produced minute improvement. The latter model was the most general one considered, and contained the most parameters (10). Correlated taste variation was not examined in the HW78 study.

Albright, Lerman and Manski (1977)

This study ("ALM77") appeared as a technical report on the development of a MNP estimation program for the Federal Highway Administration, and includes an analysis of trinomial travel mode choice using the same data source as HW78. There are numerous differences between the two studies: ALM77 (i) uses many more observations than HW78 (557 versus 100), (ii) allows choice set size to vary where appropriate (J=2 or J=3), (iii) uses a different set of attributes, and (iv) estimates a much more general model specification.

The ALM77 model specification framework differs from equation (1): it suppresses μ and assumes that alternative-specific errors may be incorporated by including dummy variables in the matrix X, thus using "random taste variation" for the dummies to represent alternative-specific errors. The empirical model estimated by ALM uses seven "attributes" (including two alternative-specific dummies) and assumes fully correlated taste variation. It is therefore similar to the most general specification in Table 1, except that the δ and ε terms are not assumed to be independent. Constraints are necessary for identification purposes, so that the model contains 34 parameters (versus 10 for the most general HW78 specification).

The optimization algorithm used is a simple gradient search technique, and the choice probabilities are evaluated via Clark's approximation. The gradient vectors are calculated via finite differences. The use of Clark's approximation is justified in the report by a study involving test examples with three- and five-alternative choice sets; this study also appears in Lerman and Manski (1981). Choice probabilities evaluated via both Clark's approximation and a simulation method are compared, and ALM conclude based on these examples that the Clark approximation is accurate. They also indicate that their gradient search approach may be used with simulated probabilities, although they do not give an empirical example, and such an approach seems impractical.

ALM compare their 34-parameter probit estimates to the 7-parameter logit estimates and conclude that the improvement in log-likelihood is much too small to justify use of MNP for this dataset.

Currim (1982)

This is yet another travel mode choice study, but in contrast to HW78 and ALM77 it is based on marketing survey data. The analysis uses attributes based on the respondents "perceptions" of nine different mode characteristics. In addition to the standard time and cost measures, measures were developed for such "perceptual" attributes as availability, cleanliness, and personal safety. The analysis includes five alternatives, and eschews the use of alternative-specific constants.

A variety of model specifications are estimated, including logit, IID probit, and the HW78 model $\Sigma_U = X^T DX+I$, which this author calls the "perceptual interdependence" model. In addition, three model specifications which assume non-IID errors (but no taste variation) are estimated: unfortunately, none of these three appear to be identified, which casts serious doubts on the interpretability of the results.

A sample of 490 households was used in the study, where 369 were used for estimation purposes. The remaining 121 were used as a validation sample for purposes of evaluating the competing model specifications. The maximum number of parameters in any model appears to be 24. All the models were estimated using CHOMP--see Daganzo and Schoenfeld (1978)--in which probit choice probabilities are evaluated via Clark's approximation. The optimization procedure uses a line search with a least-change secant update to Davidon (1959) and Fletcher and Powell (1963), the so-called "DFP method." See section 4. Gradients are evaluated by finite differences. There appears to have been serious difficulties with obtaining the parameter estimates in this study. In addition to the identifiability problems, we note that Currim reports log-likelihoods for the more general model specifications which are *smaller* than the IID probit log-likelihood. Since IID probit is a nested special case in each instance, the validity of the results is highly questionable.

Johnson and Hensher (1982)

The MNP model is used here in a two-period panel analysis of travel mode choice. There were two modes ("car" and "train"), and five model specifications were estimated. Although four of the models are essentially binomial probit, the fifth incorporates intertemporal correlation of the alternative-specific error terms and can be written as a trinomial probit model, thus making the estimation problem comparable to HW78 and ALM77.

Each model includes an alternative-specific dummy and three explanatory variables, but none includes taste variation. Alternative cross-sectional formulations include lagged indicators of past choice or utilities of past choice. The covariance matrix for the two-period probit model is normalized by fixing the diagonal elements to one and leaving free the covariance term, which represents serial correlation of the random error term. We note that the error covariance matrices for the four binary probit models appear to be unidentified, including an inestimable covariance parameter--see Bunch (1989). This study uses observations on 163 commuters, and parameters estimates are obtained using CHOMP.

Miller and Lerman (1981)

This study examines decisions by retail firms regarding location and store-size. There are seven possible locations and two possible sizes, giving a total of 14 alternatives per choice set, the largest number of all the studies reviewed here. Seven attributes related to revenues were used, but in addition all alternatives were characterized by combinations of dummy variables for location and size. Three probit specifications were estimated: (i)

IID probit, (ii) a model including taste variation on the cost attribute only, and (iii) a model including taste variation on the cost, along with a variance parameter for each location, and a correlation parameter between stores of different sizes. The sample included observations on 181 stores.

Parameter estimates are obtained using CHOMP. 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)

This is a study from the marketing literature which develops a new multinomial probit model and compares it to the IID probit and HW78 X^TDX+I models on the basis of goodness-of-fit and predictive validity. The authors refer to these models as proposed multinomial probit (PMNP), IPROBIT, and random coefficients probit (RCP), respectively. They emphasize the importance of estimating models in terms of attributes only, so that the parameter estimates may be extended to predictions involving new products.

The PMNP model specifies a flexible covariance matrix in terms of weighted Euclidean distances between the pairs of attribute profiles, where the importance weights are the same as the taste weights in the utility function. The covariance matrix is a parsimonious function of these distances, using three free parameters to capture the degree and type of interactions between alternatives and the size of the random component in the random utility formulation. From the discussion it appears that all models have been correctly specified.

In addition to simulation experiments involving five alternatives per choice set, empirical data involving selection from a set of hypothetical wagers is used to estimate and test the models. A wager consists of a vector of equally probable payoffs, and may be conveniently characterized by two attributes (expected value and standard deviation). Each choice set contained three wagers (J=3). Choices by 100 student subjects were observed, where each subject made choices on 52 subsets. Estimations are performed using various subsets of the pooled data, with the number of observations ranging from 2000 to 2600. Predictive validity was measured by calculating log-likelihoods for subsets of the data not included in the estimation ("holdout samples"). The choice probabilities in this study are calculated using Clark's approximation, and the optimizations are 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.

The authors conclude that PMNP performs slightly better than RCP based on log-likelihood measures of fit, but they also perform calculations which demonstrate that for this set of data the primary source of randomness is taste variation.

van Lierop(1986)

In a study of residential location analysis various multinomial probit models are applied to such issues as decision to move-or-stay, and choice of dwelling type. In the latter case, seven dwelling types (detached single-family housing, semi-detached single-family housing, etc.) are grouped into three classes to formulate trinomial probit models. No taste variation is assumed in this study. The analyses, however, appear to suffer from overparameterization of the covariance matrix, which casts doubts on the interpretability and validity of the study's results. The overall study had a sample of 1107 households, with varying numbers used for estimation, depending upon the particular model. As in many of the above studies, CHOMP was used for obtaining the estimates.

Summary

Although the theoretical advantages of the the multinomial probit model have often been recognized, relatively few empirical results have appeared. Seven empirical studies were found in the published literature, and relevant features of these have been summarized in Table 2. There may be more that we have somehow overlooked, as well as unpublished studies, but it is obvious that MNP has not become a commonplace technique for discrete choice analysis.

It is surprising to us that only one study uses numerical integration for evaluating choice probabilities, despite 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. Four of the studies used CHOMP, so software availability would also appear 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. Note, however, that some of these results could have had problems due to a unidentified model specification--see Dansie (1985). 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 HW78 and ALM77 studies use the same data source and fit models for the same mode choice problem, but unfortunately raise more questions than they answer. HW78 conclude that their probit model is a significant improvement over logit, while in contrast ALM77 conclude that their (more general) probit model is *not* an improvement over logit.

ALM77 use steepest descent and Clark's approximation, but more data points and a more explanatory variables. HW78 use BHHH and numerical integration, but fewer data points and fewer explanatory variables. Thus their basic conclusions are not robust to these features of the analysis, and a more controlled comparison would have been illuminating.

Particularly disturbing is that a relatively large proportion of the studies we reviewed (three of seven) contain major conceptual flaws, i.e., they use 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.

4. A New Algorithm for MNP Estimation

As noted above, 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 study we limit ourselves to three-alternative probit, and thus it will be practical to compute choice probabilities via numerical integration, which also allows a comparison of numerical integration versus Clark's approximation. The optimization algorithm used here is based on the work of Bunch (1987, 1988), and Dennis, Gay, and Welsch (1981a, b). It is specifically designed to find MLE's for probabilistic choice models, where any particular choice model is coded as a separate module. The following sections include a brief description of the algorithm, but a more detailed development is available in Bunch (1987). In Bunch (1988), the numerical performance of the algorithm is compared to other popular methods (BHHH, DFP, and BFGS) for the MNL model and a non-IIA model due to Batsell and Polking (1985).

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 work until now has 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 next four subsections discuss global and local strategies, our algorithm, and stopping rules, respectively. 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 convention in the optimization literature is to find local minimizers. Therefore the following discussion assumes that the maximum likelihood estimation problem has been restated as a minimization problem by taking the negative of the log-likelihood function.

4.1 Global Strategies

Let $L(x)$ denote the objective function, where x is the vector of parameters. As noted above, most traditional iterative methods employ line searches. At the k^{th} iteration a search direction $d_k = - H_k^{-1} \nabla L(x_k)$ is determined, where H_k is an approximation to the Hessian of $L(x_k)$. A step is taken along this direction, and the new iterate is given by $x_{k+1} = x_k +$

αd_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 algorithm used here employs 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 x_{k+1} is obtained by minimizing a quadratic model of the objective function $L(x)$ using information available at the current iterate x_k. Consider the quadratic model $m(s)$ given by

$$m_k(s) = L(x_k) + \nabla L(x_k)^T s + \frac{1}{2} s^T H_k s, \tag{5}$$

where $s \equiv x\text{-}x_k$ 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 δ_k about the current iterate x_k, then we can find the next iterate by solving the following constrained optimization problem:

$$\min m_k(s) \text{ subject to } \|s\|_2 \leq \delta_k, \tag{6}$$

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.)

This approach, combined with appropriate rules for selecting δ_k, leads to algorithms which are globally convergent. Close to the solution, both the line search and trust region methods should take full quasi-Newton steps, so as to preserve the convergence properties of the local strategy.

4.2 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., $H_k = \nabla^2 L(x_k)$, which gives the very fast local convergence properties ("q-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 ("q-linear"). These include steepest descent (used by ALM77), and BHHH (used by HW78). Actually, BHHH can be fast under certain circumstances, but in general this cannot be guaranteed, and numerical tests show 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. 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 ("q-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.

The special structure of the MLE problem for probabilistic choice models produces an expression for the Hessian which may be represented by

(7) $$\nabla^2 L(x) = C(x) + A(x),$$

where C(x) may be readily computed using information already available from the gradient calculation--see Bunch (1987). Computing A(x) involves a matrix of second derivatives for each observation, and is extremely expensive. BHHH approximates $\nabla^2 L(x)$ by ignoring A(x). The least-change secant update methods (DFP and BFGS) discard C(x) and approximate $\nabla^2 L(x)$ directly. Our algorithm calculates C(x) at each iteration, but builds up an approximation to A(x). There are therefore two quadratic models available at each iteration, defined by $H_k{}^1 = C(x_k)$, and $H_k{}^2 = C(x_k) + A_k$, and the algorithm switches between the two when one is performing better than the other ("model switching"), where "performance" is evaluated via heuristics based on comparing the amount function decrease predicted by the model versus the actual observed function decrease. So, the algorithm will always perform at least as well as the BHHH method, and for difficult problems it retains local q-superlinear convergence.

4.3 The Algorithm

The algorithm combines the model trust region with the model switching scheme described above. The implementation is as described in Bunch (1987), where appropriate modules have been added for the MNP model. The data arrays and some of the modules are structured similar to those of Daganzo (1979). A SPEC module implements the model specifications from Table 1, and various Cholesky factorizations are used to ensure positive definiteness of the estimated covariance parameters.

The numerical integration routine was modified from one obtained from Mark Schervish--see Schervish (1984). Schervish's routine makes use of a bivariate normal routine due to Donnelly (1973), and since this study only involves three alternatives, the relevant integration is actually performed by this routine. The Clark's approximation routine was a coded version of the description in Daganzo (1979). Finite difference gradients were used, although future plans include developing modules for "direct" evaluation of gradients using various analytical expressions.

4.4 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 implementation used in this paper is that it naturally lends itself to a suite of stopping rules which has convenient diagnostic interpretations when performing statistical parameter estimation. These were developed by David M. Gay, and for a complete discussion see Gay (1982) and Dennis, Gay, and Welsch (1981a, b).

There are three "favorable" stopping rules: x-convergence, relative function-convergence, and absolute function-convergence. The first two hold when the relative step size or the relative change in $L(x)$, respectively, fall below their user-defined tolerances. The third is included for the rare case in which both $L(x^*) = 0$ and $x^* = 0$, where x^* is the solution. Now, these rules are only good for cases in which the quadratic model is deemed to be "trustworthy," and the Hessian approximation appears to be positive definite. If one 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 $L(x)$ is small, but the Hessian of $L(x)$ 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 could indicate that 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. Examples of the utility of these stopping rules appear in section 7.

5. Formulation of Household Car Ownership Model

The model we use is based on the assumption that the household is the decision making unit which maximizes its utility by choosing the number of cars to own. Let the household utility be

$$U(y, t, N_{nw}) = \pi y^{\alpha} t^{\beta} N_{nw}^{\tau}, \tag{8}$$

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. Using this, formulate household behavior as

$$\text{Maximize } U(y, t, N_{nw}) = \pi y^{\alpha} t^{\beta} N_{nw}^{\tau} \tag{9}$$

Subject to:

$$\begin{aligned} C(n) + y + S(h) + q_{nw}N_{nw} + q_wN_w &= Y \\ &= WT_w \\ T_w + t + t_{nw}N_{nw} + t_wN_w &= aT \end{aligned}$$

where

n	=	number of cars,
h	=	number of household members,
a	=	number of adults,
Y	=	total household income,
W	=	wage rate,
T_w	=	total time spent at work,
T	=	total time available,
N_w	=	number of work trips,
q_k	=	average cost per trip for purpose k, where k = w (work), or nw(non-work),
t_k	=	average trip time for purpose k,
$C(n)$	=	cost of owning n cars, and
$S(h)$	=	subsistence cost of a household of size h,

where all time, cost and trip variables are defined per unit time (a year, say). This formulation is an extension of that in Beckmann, Gustafson, and Golob (1973) to include the time constraint as well as the monetary budget constraint. Combining these two constraints as in Becker (1965), solving the optimization problem and substituting the optimal solution into the utility function, the logarithm of the indirect utility, V, can be obtained as

$$\ln V = a_0 + \tau \ln[WaT - C(n) - S(h) - S_w N_w] - \beta \ln W - \tau \ln S_{nw} \tag{10}$$

where

$$a_0 = \ln \pi + \alpha \ln \alpha + \beta \ln \beta + \tau \ln \tau - (\alpha + \beta + t) \ln(\alpha + \beta + \tau)$$
$$S_k = q_k + W t_k$$

Note that S_k is the average generalized cost of a trip for purpose k.

Several assumptions and approximations need to be applied to this indirect utility function to obtain an estimable functional form. The average cost of car ownership and the subsistence expenditure are assumed to be linear functions of the number of cars and the number of household members, respectively:

where N is the number of destination zones. Then,

$$(16)\qquad \Omega_i = \ln N + A_i - a_1 q_i - a_2 t_i,$$

therefore,

$$(17)\qquad q_i + \frac{a_2}{a_1} t_i = \frac{\ln N + A - \Omega_i}{a_1} = r - s\Omega_i$$

where

$$r \equiv \frac{\ln N + A}{a_1}, \text{ and}$$
$$s \equiv \frac{1}{a_1}.$$

Note that N and A are common to all observations. Using this, generalized cost of travel can be approximated as

$$(18a)\qquad S_w{}^m = r_w - s_w \Omega_{i,w}{}^m$$

$$(18b)\qquad S_{nw}{}^m = r_{nw} - s_{nw} \Omega_{i,nw}{}^m$$

where superscript m refers to the mode of travel. Using the approximation

$$(19)\qquad \ln(a + b) = \ln a + \frac{b}{a},$$

which offers accurate results for a wide range of b when a is moderately large (5, say), we rewrite the indirect utility for a household in zone i as

$$(20)\qquad \ln V = a_0 + \tau \ln[WaT - C(n) - S(h) - S_w{}^* N_w] - \beta \ln W - \tau \ln S_{nw}{}^*$$
$$= a_0 + \tau \ln I_R - r_w \frac{N_w}{I_R} + s_w \frac{N_w \Omega_{iw}{}^*}{I_R} - \beta \ln W$$
$$- \tau \ln r_{nw} + \frac{\tau\, s_{nw}}{r_{nw}} \Omega_{i,nw}{}^*$$
$$= \beta_0 + \beta_1 \ln I_R + \beta_2 \frac{N_w}{I_R} + \beta_3 \frac{N_w \Omega_{iw}{}^*}{I_R} + \beta_4 \ln W + \beta_5 \Omega_{i,nw}{}^*$$

where superscript * refers to the mode used by the household, and

(21) $$I_R = WaT - C(n) - S(h)$$

The coefficients of this model are now estimable using commonly available explanatory data. The only exception is β_4, which is not estimable since lnW is common and invariant across the alternatives. These coefficients are related to the parameters defined earlier as follows:

(22)
$$\begin{aligned}
\beta_0 &= a_0 - \tau \ln r_{nw}, \\
\beta_1 &= \tau, \\
\beta_2 &= -r_w, \\
\beta_3 &= s_w, \\
\beta_4 &= -\beta, \text{ and} \\
\beta_5 &= \frac{\tau\, s_{nw}}{r_{nw}}.
\end{aligned}$$

The accessibility measure may be expressed in terms of measurable variables based on some assumption of car allocation. For example, let $\mu_{i,k}$ be a measure of car availability for purpose k, and $0 \le \mu_{i,k} \le 1$. It may then be assumed that $\Omega_{i,k}{}^*$ is a function of μ_k,

(23) $$\Omega_{i,k}{}^* = \mu_{i,k}\Omega_{i,k}{}^c + (1 - \mu_{i,k})\Omega_{i,k}{}^t$$

where superscript c refers to car and t to transit. A simple measure of car availability may be formulated as

(24a) $$\mu_{i,w} = \min\{1, \frac{\text{No. of cars}}{\text{No. of workers}}\}$$

$$= \min\{1, \frac{n}{w}\}, \text{ where w is the number of workers, and}$$

(24b) $$\mu_{i,nw} = \min\{1, \frac{\text{No. of cars}}{\text{No. of adults}}\}$$

$$= \min\{1, \frac{n}{a}\}.$$

6. Empirical Results

Trinomial probit models of household car ownership are formulated using the indirect utility derived above. The three-alternative choice set for all households is {no car, one car, two or more cars}. An important feature of this model estimation exercise is the fact that the utility specification is theoretically deduced prior to analysis, and not formulated through statistical search for an optimal empirically-derived model. One of the consequences is highly collinear data, creating inclement conditions for model estimation. This problem thus offers a useful test ground for the algorithm described in section 4.

The following analysis uses the model specifications in Table 1, and one of our objectives is to examine the various possible nested model specifications, thereby testing the behavioral hypotheses embodied in them. We first describe the data set, and then discuss the car ownership model results obtained using numerical choice probabilities. We then consider the accuracy of Clark's approximation.

6.1 Car Ownership Data Set

A subsample from the ongoing 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. 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 the model variables produced a 945-household sample for model estimation.

The explanatory variables for the car ownership model are I_R, N_w/I_R, $N_w\Omega_w^*/I_R$, and Ω_{nw}^*, defined in section 5 above, where the zone subscripts have been suppressed for notational convenience. The functions C(n) and S(h) defined by equations (11) and (12) use K_1 = 1200 Dfl per person and K_2 = 1830 Dfl per car, respectively. In equation (21) for I_R, the quantity WaT was replaced by the household's reported total personal income. In equation (23), the quantities $\Omega_{i,k}{}^m$ for k = w, nw and m = c, t were represented by the accessibility measures discussed above; "non-work" accessibilities were assumed to be accessibilities to shopping destinations.

6.2 Estimation Results

Maximum likelihood parameter estimates for the logit model and the six MNP specifications in Table 1 were obtained using the algorithm described in section 4. In this section we examine results obtained when choice probabilities are evaluated via numerical integration. Estimated coefficients for the four explanatory variables variables defined above and two alternative-specific dummies, as well as log-likelihoods and goodness-of-fit measures, are summarized in Table 3. The coefficients generally have significant t-scores, as well as the theoretically anticipated signs. For example, the coefficient of N_w/I_R represents $-r_w$, and thus the negative signs shown in the table are expected.

We conclude that our empirical results generally support the economic model formulation presented in section 5. Note that care should be taken in comparing the absolute values of the coefficients from various models, since the scaling will depend upon the particular specification used. However, the coefficients of the three models with IID error terms may be directly compared, and in addition relative coefficients may be computed for comparison purposes, if necessary.

The estimated model coefficients do imply some minor deviations from the theoretically constructed model. The fourth variable, Ω_{nw}^*, has insignificant coefficients across the models, suggesting that non-work travel is not a major determinant of household utility. Secondly, the generally significant alternative-specific constant terms suggest that some important explanatory variables which contribute to a household's utility for car ownership have been omitted from the analysis. The coefficient estimates, t-statistics, and log-likelihood of the IID probit model are very close to those obtained for the multinomial logit model. This is consistent with previously reported results indicating that for practical purposes these two models offer the same statistical results.

For the models including random taste variation, the coefficient estimates in Table 3 represent the expected coefficient values (or average taste weights) for the sample population. Consider the models containing IID errors: the expected tastes for the random taste formulations tend to be larger in absolute value than the corresponding coefficient estimates obtained from a fixed-taste model. (This also appears to be the case for the non-IID error models, but as noted above, care should be taken with such comparisons.) We first consider hypothesis tests of the various nested specifications, and then discuss other interpretations of the numerical results.

Significance of Random Tastes and Non-IID Errors

The log-likelihoods presented in Table 3 can be used to test the significance of: (i) the assumption of randomness in the model coefficients, and (ii) the presence non-IID alternative-specific errors. Table 4 summarizes the results of likelihood-ratio (LR) tests that may be performed on the various nested models.

These LR test statistics offer evidence that *tastes are random* and *error terms are not IID*. Statistics for comparing various taste variation hypotheses are presented in part (a) of Table 4; results for the IID error models and the non-IID error models appear separately. For the IID error models, the LR statistic for the comparison of fixed tastes versus uncorrelated random tastes has a highly significant value of 15.14 with four degrees of freedom (4 dfs). For fixed tastes versus correlated random tastes LR = 18.28 (10 dfs), and for uncorrelated random tastes versus correlated random tastes LR = 3.1 (6 dfs). The results for models with IID errors thus reject the hypothesis of fixed tastes, although there is no significant difference between the correlated and uncorrelated taste variation models.

On the other hand, the hypotheses of fixed tastes and uncorrelated tastes versus correlated tastes are *both* rejected for models with non-IID errors. In any case, the importance of allowing for random taste variation is evident.

The LR test statistic comparing the simplest model (IID probit) versus the most complex model (correlated random tastes and non-IID errors) is 45.4 with 12 dfs, and thus the IID probit model is clearly inadequate. Interestingly, the LR for IID probit versus the fixed taste/non-IID error model is not significant, while in contrast, the difference between the correlated taste/IID error model and the correlated taste/non-IID error model is highly significant (see part (b) of Table 4). This result demonstrates the danger of testing model specifications without examining all possibilities; the hypothesis of IID errors could have been accepted if models with fixed tastes alone had been studied.

Random Taste Covariance Matrix (Σ_δ)

Having established that correlated random tastes are important, further interpretation of the results is naturally of interest. The estimated taste covariance matrix Σ_δ and the corresponding correlation matrix for the $X^T\Sigma_\delta X + \Sigma_\varepsilon$ model are summarized in Table 5.

This matrix includes some very large correlations. The correlation between the random coefficients for $-N_w/I_R$ and $N_w\Omega_w^*/I_R$ is large and positive, while the correlation between Ω_{nw}^* and both $-N_w/I_R$ and $N_w\Omega_w^*/I_R$ are large and negative. This would indicate that households placing large weights on work trip-related costs tend also to place large weights on workplace accessibility. And, those households placing large weight on shopping accessibility will tend to place smaller weights on the work trip-related attributes.

Recall that a complicating factor with these attributes is the high degree of multicollinearity; this could be a factor in the high correlations. We estimated an alternative three-attribute specification with $-N_w/I_R$ omitted: a LR test rejected the hypothesis that the three- and four-attribute models were the not significantly different.

Alternative-Specific Error Covariance Matrix (Σ_ε)

The statistical analysis presented above rejected the assumption of IID error terms. As discussed previously, a matrix C is obtained by our estimation routines for models involving a "general Σ_ε." Recall that the C matrix is defined by equation (4), which we reproduce here for convenience:

$$C = \begin{bmatrix} 1 & C_{21} \\ C_{21} & C_{22} \end{bmatrix} \equiv \begin{bmatrix} \sigma_{11}-2\sigma_{31}+\sigma_{33} & \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} \\ \sigma_{21}-\sigma_{31}-\sigma_{32}+\sigma_{33} & \sigma_{22}-2\sigma_{32}+\sigma_{33} \end{bmatrix}. \tag{25}$$

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

From (25) 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 C. We note, however, that one may not arbitrarily choose any two-parameter expression for Σ_ε. Three allowable choices (see Bunch, 1989) are:

$$(26) \qquad \Sigma_\varepsilon^1 = \begin{bmatrix} 1 & \sigma_{21} & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$(27) \qquad \Sigma_\varepsilon^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix},$$

and

$$(28) \qquad \Sigma_\varepsilon^3 = \begin{bmatrix} \sigma_{11} & \sigma_{21} & 0 \\ \sigma_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

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 errors for the various alternatives are completely uncorrelated but with 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 $\sigma_{21} = C_{21}$ and $\sigma_{22} = C_{22}$.

Assume that the indices {1, 2, 3} in (26) through (28) are assigned to the choice alternatives {one car, two or more cars, no car}. It is straightforward to calculate each Σ_ε matrix using our estimated values for C: these have been summarized in Table 6 for the three models with non-IID errors, along with their correlation coefficients (where appropriate).

First consider the Σ_ε^1 results. The model with no taste variation has a negative ρ in contrast to the two models which include taste variation. The model including fully correlated tastes produces a ρ equal to one, in contrast to $\rho = 0.329$ for the uncorrelated taste model. Recall that model $X^T\Sigma_\delta X+\Sigma_\varepsilon$ was preferred based on LR tests, and 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. While the very large ρ is interpretable, it also raises troubling questions regarding the completeness of the model specification and/or the problems with multicollinearity in the data. It might also raise concerns regarding the quality of the maximum likelihood estimation results.

For both the $\Sigma_\varepsilon{}^2$ and $\Sigma_\varepsilon{}^3$ normalizations, the uncorrelated taste model is the only one producing proper covariance matrices. The remaining two models both produce negative variances in $\Sigma_\varepsilon{}^2$, and $\Sigma_\varepsilon{}^3$'s for which $|\rho| > 1$. In particular, the correlated taste model results are behaviorally inconsistent with these normalizations. This is not unexpected, since it seems unrealistic that the one- and two-or-more-cars alternatives would have zero correlation. Also, one might wonder about the assumption in (28) that the variances for the two-or-more alternative and the no-car alternative are equal, but unequal to the one-car variance. Based on these considerations, one might perhaps like to consider the following candidate normalizations:

(29) $$\Sigma_\varepsilon{}^4 = \begin{bmatrix} 1 & \sigma_{21} & 0 \\ \sigma_{21} & 1 & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix},$$

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

Unfortunately, these normalizations are not identified even though they each appear to have two free parameters--see Bunch (1989). This 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.

6.3 Accuracy of Clark's Approximation

The data set, model specifications, and numerical algorithms compiled for this study provide an opportunity for comparing estimation results obtained using numerical integration versus Clark's approximation. Note that previous comparisons in the literature (e.g., Horowitz, et. al.1982, Lerman and Manski 1981) rely on simulated data.

Comparison results were obtained by taking the numerical integration (NI) parameter estimates as the starting point for each model specification, and re-running the estimation procedure using Clark approximation (CA) choice probabilities. Some useful measures are summarized in Tables 7a and 7b, for the IID error and non-IID error models, respectively. Both parts of Table 7 include: (i) the log-likelihood of the NI estimates, evaluated using NI probabilities, (ii) the log-likelihood of the NI estimates, evaluated using CA probabilities, (iii) the log-likelihood of the CA estimates, evaluated using CA probabilities, (iv) the relative difference between the NI and CA parameter estimates, i.e., coefficients plus Cholesky parameters, and (v) the relative difference between the NI and CA coefficient estimates only. The relative difference between two vectors **a** and **b** is defined by

$$\text{(31)} \qquad \mathrm{RDIFF}(a, b) = \frac{\sqrt{\sum_{k=1}^{p} (a_k - b_k)^2}}{\sqrt{\sum_{k=1}^{p} b_k^2}}$$

In addition, part (b) of Table 7 includes the C matrix estimates and correlation coefficients for comparison purposes.

Examination of Table 7a reveals that the NI and Clark estimation results are very similar for the IID error models. This is perhaps to be expected for IID probit, since

Clark's approximation should be reasonably stable in this simple case. On the other hand, both X^TDX+I and $X^T\Sigma_\delta X+I$ potentially allow a much more flexible pattern of covariances than IID probit or logit, and on the face of it could produce more difficult choice model integrals to evaluate, thus possibly causing difficulties for the use of Clark's approximation. In fact, the correlation matrices for estimated Σ_δ parameters are:

$$R_\delta^{NI} = \begin{bmatrix} 1 & & & \\ 0.218 & 1 & & \\ 0.288 & -0.872 & 1 & \\ 0.837 & 0.716 & -0.292 & 1 \end{bmatrix}, \tag{32}$$

and,

$$R_\delta^{CA} = \begin{bmatrix} 1 & & & \\ 0.218 & 1 & & \\ 0.260 & -0.885 & 1 & \\ 0.870 & 0.672 & -0.250 & 1 \end{bmatrix}, \tag{33}$$

which exhibit substantial correlation of taste variation. However, the largest RDIFF for the IID error models is only 6.7%, and these results seem to support the use of Clark's approximation for MNP models which include taste variation but not alternative specific errors.

In contrast, Table 7b reveals that substantial differences occur between NI and CA estimates for the non-IID error models. The log-likelihood values are extremely different for all three models, and the relative differences in the estimates are quite large. In the case of the $X^T\Omega X+\Sigma_\varepsilon$ model the relative difference between coefficient estimates exceeds 300%, and many estimates (both coefficients and taste covariances) were substantially different, with some exhibiting large swings combined with sign changes. For the Σ_ε and $X^TDX+\Sigma_\varepsilon$ models the relative differences for the parameter estimates are 26.6% and 50%, respectively, and the relative differences for the coefficients exceed 35% for both models. In addition, the Clark estimates for the C matrix are extremely different from the more "reasonable" NI estimates, and are extremely ill-conditioned. It is worth noting that these examples probably represent an extremely stringent test of Clark's approximation: in addition to the collinear data, the "true" (NI-estimated) C matrices are highly correlated, and thus the integrals to be evaluated could be extremely difficult.

Another interesting feature of these results is that the Clark estimate for the most general model $X^T\Sigma_\delta X+\Sigma_\varepsilon$ has a *smaller* log-likelihood than either of the two more restricted non-IID models. This is clearly a poor result, and in our experiments this represented the only numerical evidence for multiple local optima. In an alternative run, we used the $X^TDX+\Sigma_\varepsilon$ Clark estimates as the basis for a starting point and obtained a more reasonable solution, i.e., a solution with a log-likelihood of -663.63.

The Clark log-likelihoods have been used to create the LR test statistics in Table 8, which is the analog to Table 4 above. Note that an analysis based on Table 8 leads to a different set of conclusions than those from Table 4. The hypothesis of IID errors is always rejected, but the hypothesis of uncorrelated tastes is never rejected. The preferred model would be $X^TDX+\Sigma_\varepsilon$, rather than $X^T\Sigma_\delta X+\Sigma_\varepsilon$.

7. Comments on Numerical Estimation and Convergence

In addition to analyzing car ownership models, this study was regarded to be a useful test of a new method for probit estimation. As noted above, the models estimated here provide an excellent basis for testing. They are reasonably complex, cover an interesting range of possibilities, and use observational empirical data which are likely to be collinear. Although a detailed tabulation of numerical experiments is beyond the scope of this paper, we make a few general comments about the experience gained in estimating these models, and give some examples of the useful features of the algorithm described in section 4.

Numerous pathways were followed in generating the final set of numerical results for this paper. Generally, the more simple models were estimated first, and the results of the more simple models were appropriately modified for use as starting points in estimating the more complex models. Preliminary estimates were obtained for models without alternative-specific dummies, and these results were used as starting points for the models with dummies. Unfortunately, the dummies were always highly significant, as noted above. Many of our preliminary runs used Clark's approximation due to speed considerations, and the output of these runs were often used as starting points for the numerical integration runs. Sometimes the Clark results provided poor starting points, i.e., points for which the numerical integrals could not be calculated, and other pathways were taken.

We make the following observation: *during the course of extensive numerical testing we saw almost no evidence of multiple "strong" local optima.* The one exception was the Clark run discussed above. On the other hand, situations often arose in which the "singular convergence" condition was raised, indicating that the Hessian of the log-likelihood was singular, or nearly so. Specifically, "singular convergence" was always obtained for the following two models: $\Sigma_U = X^T D X + I$, and $\Sigma_U = X^T \Sigma_\delta X + I$, i.e., models with random tastes and IID errors. Recall, though, that our LR tests *rejected* the assumption of IID errors. Thus these results are consistent with a scenario in which the specifications are (i) constrained to lie in an incorrect subspace, and (ii) are overparameterized relative to the highly-collinear dataset when forced to lie in this incorrect subspace. Because we are dealing with nonlinear models, the more general "correct" specification may not be overparameterized relative to the dataset, even if the constrained specification is.

Generally, the non-IID error models (and the IID probit model) converged with the relative function-convergence condition. Convergence was slowest for the most complex model, $\Sigma_U = X^T \Sigma_\delta X + \Sigma_\varepsilon$, with many small steps taken before the algorithm stopped. Furthermore, for some starting points, the "singular convergence" condition was observed, and it is likely that this model--which has the most parameters of all the models tested (18)--has a relatively flat log-likelihood function.

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 that 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:

(34) $$\Sigma_1 = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

(35) $$\Sigma_3 = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & 1 & 0 \\ 0 & 0 & \theta_3 \end{bmatrix},$$

(36) $$\Sigma_2 = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & \theta_3 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

(37) $$\Sigma_4 = \begin{bmatrix} 1 & \theta_2 & \theta_4 \\ \theta_2 & \theta_3 & \theta_5 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix}.$$

Now, (34) and (36) are identified, while (35) and (37) are not, based on earlier discussions. Furthermore, (34) is a special case of (35) and (36) is a special case of (37). For a summary of numerical results using these examples, see Table 9. 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 in previous discussions. 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.

8. Conclusions

In this paper we have revisited various issues relating to practical estimation of the multinomial probit model by applying state-of-the-art numerical methods to an analysis of highly collinear empirical data and a wide range of model specifications. To set the context, we included a brief review of empirical probit results in the literature.

The MLE algorithm--which uses model trust regions and model switching--performed very well on these problems, reliably and efficiently producing parameter estimates. Bunch (1988) previously compared this algorithm to alternative methods and found it to have advantages; although an analogous comparison of numerical methods using probit examples is beyond the scope of this paper, we found the performance of the algorithm on these problems to be encouraging.

In addition, the algorithm is 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. In our empirical study, models with IID error terms produced the singular convergence condition with both numerical integration and Clark approximation choice probabilities, and later tests indicated that the assumption of IID errors was incorrect. An obvious next step is to test the algorithm on cases involving more than three alternatives: this work is now underway and should be the subject of a forthcoming paper.

Our comparison of Clark's approximation versus numerical integration 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. Perhaps most disturbing is that the likelihood ratio tests using the Clark results produce different conclusions regarding the correct model specification. This finding is somewhat discouraging, highlighting what is still a major difficulty in estimating MNP models: evaluation of choice probabilities for choice sets with large numbers of alternatives. On the other hand, recent advances in computer technology, including parallel and distributed processing techniques in conjunction with improved algorithms like the one used here, may in the short run push back the barriers to using numerical integration a bit more--see Schervish (1988).

In addition to the more obvious numerical issues, the discussion in section 6 illustrates some important points regarding the selection and testing of model specifications. First, we note that our study obtains estimates for a full range of specifications, including those with correlated taste variation and non-IID errors. This turned out to be important: we could not reject the most general model in our framework using a likelihood ratio test. Thus, these empirical results are the first of which we are aware in which both correlated taste variation and non-IID errors are found to be significant.

Second, we feel that identification issues for models with non-IID errors have received insufficient attention in the literature. Although Albright, Lerman, and Manski (1977) provide some useful discussion, there is a general lack of understanding among practitioners that has resulted in a disturbing number of published papers containing specification errors, as noted in section 3. This insidious problem, which is an outgrowth of the complexity of MNP, could prove to be as much of an obstacle to practical probit applications as the more obvious computational issues.

Bunch (1989) provides the background material on identification which is used in section 6 for obtaining the various possible behavioral interpretations of our non-IID error covariance estimates. Specifically, the problem is that the estimates do not lead to a unique normalization but to many possible equivalent normalizations, and it is therefore up to the modeler to choose one. Fortunately, as our results illustrate, improper covariance matrices can sometimes arise allowing some normalizations to be eliminated. The main point, though, is that it is important for modelers to consider these interpretations as a necessary part of their analyses.

Acknowledgements

This study was supported in part by the USDOT 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.

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.

Batsell, R. R., and J. Polking (1985) A generalized model of market share. *Marketing Science*, 4, 177-198.

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 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. (1989) On the estimability of the multinomial probit model. Transportation Research Group Working Paper, 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.

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.

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.

Johnson, N. L. and S. Kotz (1972) *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons, New York, NY.

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.

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

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.

McFadden, D. (1973) 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. (1986) A method of simulated moments for estimation of multinomial probits without numerical integration. Department of Economics Working Paper, Massachusetts Institute of Technology, Cambridge, MA.

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.

Schervish, M. J. (1984) Algorithm AS 195: Multivariate Normal Probabilities with Error Bound. *Applied Statistics*, 33, pp. 81-93. Correction, 34, pp.103-104.

Schervish, M. J. (1988) Applications of parallel computation to statistical inference. *Journal of the American Statistical Association*, 83, 976-983.

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.

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^TDX+I
Random Uncorrelated	Non-IID	$X^TDX+\Sigma_\varepsilon$
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
Johnson and Hensher (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
Alternative Models of Household Car Ownership
(Wave 1 Dutch Panel Data Set)

		IID Error Terms			Non-IID Error Terms		
		Fixed	Random Taste		Fixed	Random Taste	
Explanatory Variable	Logit	Taste	Uncor.	Cor.	Taste	Uncor.	Cor.
COEFFICIENT ESTIMATES							
One Car Dummy	1.99	1.90	2.21	2.16	.90	1.10	1.80
	11.2	12.2	8.8	9.8	7.6	6.0	12.4
Two Car Dummy	1.55	1.36	1.73	1.71	-.09	.78	2.10
	5.4	5.5	5.1	5.3	-.3	2.0	10.8
$\ln(I_R)$	15.03	13.80	17.65	17.74	7.10	8.17	12.90
	10.8	13.2	6.3	6.4	7.2	6.0	7.7
N_w/I_R (x 100)	-1.74	-1.75	-2.54	-3.20	-.94	-1.77	-1.34
	-3.2	-3.2	-1.2	-1.2	-3.1	-1.3	-1.04
$N_w\Omega_w^*/I_R$ (x 10)	2.91	2.92	3.68	3.94	1.67	2.15	2.42
	5.3	5.4	3.2	2.9	5.3	3.0	3.7
Ω_{nw}^*	-.09	-.08	-.06	-.03	-.01	-.02	-0.19
	-.6	-.6	-.4	-.2	-.1	-.2	-1.9
L(0)	-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.	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

See text for definitions of explanatory variables. 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 4
Likelihood Ratio Tests of the Significance of Random Tastes and Non-IID Errors in Car Ownership Models (Numerical Integration Results)

(a) Tests of Random Tastes

		IID Error Terms (Random Tastes) Uncor.	IID Error Terms (Random Tastes) Cor.		Non-IID Error Terms (Random Tastes) Uncor.	Non-IID Error Terms (Random Tastes) 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	Non-IID Errors: Random Tastes Uncor.	Non-IID Errors: Random Tastes Cor.
IID Errors	LR	5.15	7.75	26.66
	df	2	2	2

TABLE 5

Estimated Taste Covariance Matrix (Σ_δ) and Correlations for $X^T\Sigma_\delta X + \Sigma_\varepsilon$ Model

Explanatory variable	$\ln I_R$	$-N_w/I_R$ (x 100)	$N_w\Omega_w^*/I_R$ (x 10)	Ω_{nw}^*	SE(δ)
$\ln I_R$	22.47	7.73	0.826	0.462	4.74
$-N_w/I_R$ (x 100)	0.279	34.17	20.38	-2.55	5.85
$N_w\Omega_w^*/I_R$ (x 10)	0.049	0.973	12.84	-1.54	3.58
Ω_{nw}^*	0.223	-0.998	-0.984	0.190	0.44

Note: Covariances are shown in the upper triangular cells and correlation coefficients are shown in the lower triangular cells.

Table 6
Derived Σ_ε's Using Non-IID Model Estimates

Normalization	Σ_ε (Fixed Tastes)	$X^T DX + \Sigma_\varepsilon$ (Uncorrelated Tastes)	$X^T \Sigma_\delta X + \Sigma_\varepsilon$ (Correlated Tastes)
	Model Specification		
Σ_ε^1 [equation (26)]	$\begin{bmatrix} 1 & -0.526 & 0 \\ -0.526 & 0.555 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.138 & 0 \\ 0.138 & 0.175 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1.142 & 0 \\ 1.142 & 1.303 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
ρ_1	-0.707	0.329	1
Σ_ε^2 [equation (27)]	$\begin{bmatrix} 1 & & \\ & -1.134 & \\ & & -8.04 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ & 0.043 & \\ & & 0.160 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ & 0.709 & \\ & & -0.345 \end{bmatrix}$
Σ_ε^3 [equation (28)]	$\begin{bmatrix} 2.602 & -2.897 & \\ -2.897 & 1 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 10.44 & 0.576 & \\ 0.576 & 1 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 0.535 & 0.753 & \\ 0.753 & 1 & \\ & & 1 \end{bmatrix}$
ρ_3	-1.80	0.178	1.03

Table 7
Comparison of Clark Approximation (CA) and Numerical Integration (NI) Results

(a) IID Error Models	I	$X^TDX + I$	$X^T\Sigma_\delta X + I$
$L(\beta^{NI})$ - NI probs	-689.26	-681.69	-680.12
$L(\beta^{NI})$ - CA probs	-690.41	-682.36	-679.37
$L(\beta^{CA})$ - CA probs	-690.34	-681.94	-679.13
RDIFF(β^{NI}, β^{CA})	0.015	0.078	0.0018
RDIFF($COEFF^{NI}$, $COEFF^{CA}$)	0.015	0.067	2.6×10^{-5}

(b) Non-IID Error Models	Σ_ε	$X^TDX + \Sigma_\varepsilon$	$X^T\Sigma_\delta X + \Sigma_\varepsilon$
$L(\beta^{NI})$ - NI probs	-686.68	-677.81	-666.79
$L(\beta^{NI})$ - CA probs	-704.32	-675.39	-702.79
$L(\beta^{CA})$ - CA probs	-672.48	-665.92	-676.76
RDIFF(β^{NI}, β^{CA})	0.266	0.497	1.59
RDIFF($COEFF^{NI}$, $COEFF^{CA}$)	0.3621	0.391	3.10
C_{21}^{NI}	-0.527	0.138	1.131
C_{21}^{CA}	-0.069	0.015	1.701
C_{22}^{NI}	0.555	0.175	1.279
C_{22}^{CA}	0.0047	0.0002	2.895
ρ^{NI}	-0.707	0.329	1
ρ^{CA}	-1	1	1

TABLE 8
Likelihood Ratio Tests of the Significance of Random Tastes and Non-IID Errors in Car Ownership Models (Clark Approximation Results)

(a) Tests of Random Tastes

		IID Error Terms (Random Tastes) Uncor.	IID Error Terms (Random Tastes) Cor.		Non-IID Error Terms (Random Tastes) Uncor.	Non-IID Error Terms (Random Tastes) Cor.
Fixed Tastes	LR df	16.78 4	22.42 10	LR df	13.12 4	17.70 10
Uncor. Tastes	LR df		5.64 6	LR df		4.58 6

(b) Tests of Non-IID Errors

		Non-IID Errors: Fixed Tastes	Non-IID Errors: Random Tastes Uncor.	Non-IID Errors: Random Tastes Cor.
IID Errors	LR df	35.72 2	32.06 2	31.00 2

Table 9
Convergence Examples

Spec	Starting Pt. $[\theta_1, \theta_2, \text{--}]$	Iterations	F evals[a]	$L(\theta^{MLE})$	Convergence[b]
Eq. (34)	[0, 0]	5	6	339.34	Rel-Function
Eq. (35)	[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
Eq. (36)	[0, 0, 1]	7	8	335.33	Rel-Function
Eq. (37)	$[\theta^{MLE}_{(\text{Eq. } 36)}, 0\text{--}]$[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 (36) were used to get the starting point for (37).