into a system of equations to represent equilibrium conditions in a car market. A general structure of such equations is described by Manski (14). It is believed that the use of the dynamic models developed in this work, in the framework of equilibrium equations, can provide a useful system for the analysis of policies that affect the car market.

ACKNOWLEDGMENT

The research reported here was performed at the Transportation Research Institute, Technion--Israel Institute of Technology, Haifa. The study was sponsored by the Israeli Ministry of Transport, and we would like to thank Nurit Straus, the project monitor, for her continuing support. Partial support was also provided by a grant from the National Science Foundation to the Massachusetts Institute of Technology to study the demand for automobiles. We also greatly benefited from the enlightening discussions we had with Charles Manski.

REFERENCES

1. C. Lave and K. Train. A Behavioral Disaggregate Model of Auto Type Choice. Transportation Research, Vol. 13A, 1979, pp. 1-9.
2. Wharton Econometric Forecasting Associates. An Analysis of the Automobile Market: Modeling the Long Run Determinants of the Demand for Automobiles. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1977.
3. M. Ben-Akiva and S. Lerman. Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work. Transportation, Vol. 3, 1976, pp. 357-376.
4. S.R. Lerman and M. Ben-Akiva. Disaggregate Behavioral Model of Automobile Ownership. TRB, Transportation Research Record 569, 1976, pp. 34-55.
5. L.D. Burns, T.F. Golob, and G.C. Nicolaidis. Theory of Urban Household Automobile Ownership Decisions. TRB, Transportation Research Record 569, 1976, pp 56-75.
6. J. Kain and G. Fauth. Forecasting Auto Ownership and Mode Choice for U.S. Metropolitan Areas. Department of City and Regional Planning, Harvard Univ., Cambridge, Mass., Working Paper, 1977.
7. R.R. Johnson. Aggregation and the Demand for New and Used Automobiles. Review of Economic Studies, Vol. 45, 1978, pp. 311-327.
8. C.F. Manski and L. Sherman. An Empirical Analysis of Household Choice Among Motor Vehicles. Transportation Research, Vol. 14A, 1980, pp. 349-366.
9. D.A. Hensher and T. Manefield. A Structured Logit Model of Automobile Acquisition and Type Choice. Proc., Australian Transport Research Forum, Hobart, Tasmania, Australia, 1982.
10. N.S. Cardel and F. Dunbar. Measuring the Societal Impact of Automobile Downsizing. Transportation Research, Vol. 14A, 1980, pp. 423-434.
11. J.H. Boyd and R.E. Mellman. The Effect of Fuel Economy Standards on the U.S. Automobile Market: An Hedonic Demand Analysis. Transportation Research, Vol. 14A, 1980, pp. 367-378.
12. D. McFadden. Modelling the Choice of Residential Location. In Spatial Interaction Theory and Residential Location (A. Karlquist et al., eds.), North-Holland, Amsterdam, 1978.
13. M. Ben-Akiva and B. Francois. μ Homogeneous Generalized Extreme Value Model. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Working Paper, 1981.
14. C.F. Manski. Short Run Equilibrium in the Automobile Market. The Maurice Falk Institute for Economic Research in Israel, Jerusalem, Discussion Paper 8018, 1980.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Experiments with Optimal Sampling for Multinomial Logit Models

YOSEF SHEFFI AND ZVI TAREM

In this paper a recently published method for optimizing the sample used in estimating discrete-choice models is tested. The work is intended to identify and explore the elements that influence the effectiveness of this methodology in designing sampling procedures for estimating logit models. The investigation includes both analytical and numerical tests. The results indicate that the sample optimization method can improve the accuracy of the resulting estimates, as compared with random sample.

Data collection is, in many cases, the major cost item in studies that involve the estimation of econometric models. Techniques for sample design have therefore been developed for many econometric and statistical models (1). In this paper discrete choice models, which are extensively used in travel-demand analysis, are examined, and, in particular, the multinomial logit (MNL) model is discussed. The focus here is on a method for optimizing the sample used to estimate discrete-choice models. The applicability of this sample optimization approach to the collection of the sample points (the data) used to estimate MNL models is examined. Also examined is the appropriate amount of effort that should be invested in the sample optimization process.

The original development of the sample optimization method, which is the subject of this paper, is from Daganzo (2). Daganzo's method is a stratified sampling technique. It assumes that the population to be sampled from can be partitioned into separate groups (or strata) and that observations can be sampled independently from each group. The objective of the sampling method is to determine how many observations should be drawn from each group so that the total estimation error is minimized. The estimation error is a composite measure of the error in all the model parameters. Naturally, this minimization is subject to a budget constraint. This sampling method attempts to determine the best allocation of the sampling budget. (The companion problem, that of determining the minimum budget required to achieve a certain accuracy, is somewhat more difficult. Its solution, however, can be inferred from the solution of the problem under consideration.)

Three main points are discussed in this paper. The first point is the applicability of the approach in terms of potential. The question examined in this context is the sensitivity of the sampling error to different sample designs. The second point is that the solution of the sample optimization (SO) problem requires prior estimates (or guesses) of the values of the parameters of the model to be estimated. The applicability of the whole concept depends, naturally, on the required accuracy of these prior estimates. The tests described in this paper explore this point in some detail. The third point is related to the first point. It has to do with the question of the amount of effort that should be invested in obtaining these prior estimates. Such an effort should be judged in comparison to the level of effort of the entire study, which means that the relevant question is the allocation of effort between obtaining the prior estimates and the estimation itself.

This paper is organized as follows. First, Daganzo's SO method is outlined. Then the application of this method to the MNL model is reviewed. Next, the question of the applicability of the SO method is explored by looking at a simple one-parameter model and a two-parameter model. Then the aforementioned issue of resource allocation in the framework of a small case study is discussed, and finally conclusions are given.

It should be noted that the conclusions of this paper are based on numerical experiments, which means that not all the results can be generalized in all circumstances. The experiments are described in further detail by Sheffi and Tarem (3).

SAMPLE OPTIMIZATION PROGRAM

Daganzo's SO method attempts to minimize the error associated with the estimation of the parameters of a discrete-choice model. The optimization problem is formulated as a mathematical minimization program, where a composite measure of the estimation error serves as the objective function and the sample group sizes are the decision variables. This approach assumes that the model under consideration is estimated by using the maximum likelihood (ML) method. It also assumes that the distribution of explanatory variables in each group is known. [This information may not be available, in which case the methods discussed by Lerman and Manski (4) may be used.]

The objective function of the SO program relates the sampling error to the sample group sizes. This expression can be derived from the Kramer-Rao lower bound on the covariance matrix of ML estimators. Letting $\underline{x}$ be a vector of explanatory variables, y be the dependent variable, and $\underline{\theta}$ be the vector of parameters for some model, this bound ($\Sigma_{\hat{\theta}}$) is given by

$$\Sigma_{\hat{\theta}} = \left\{-E[\nabla_\theta^2 L(\underline{\theta}\,|\,y,\underline{x})]\right\}^{-1} \tag{1}$$

where $L(\cdot|\cdot,\cdot)$ is the log-likelihood of the sample $(y,\underline{x})$ evaluated at $\underline{\theta}$, $\nabla_\theta^2 L(\cdot)$ is the θ-Hessian of $L(\cdot)$, and $E[\cdot]$ denotes the expectation operator that, in Equation 1, is carried out with respect to both y and $\underline{x}$.

For stratified sampling, where all observations are independent, the sample log-likelihood is given by the sum

$$L(\underline{\theta}\,|\,y,\underline{x}) = \sum_{k=1}^{K}\sum_{n=1}^{N_k} L(\underline{\theta}\,|\,y_n^k,\underline{x}_n^k) \tag{2}$$

where

$L(\underline{\theta}|y_n^k,\underline{x}_n^k)$ = log-likelihood of sample point n from group k,
$(y_n^k,\underline{x}_n^k)$ = observed values at this point,
N_k = number of observations in group k, and
K = number of groups in the sample.

The Hessian of this function is

$$\nabla_\theta^2 L(\underline{\theta}\,|\,y,\underline{x}) = \sum_{k=1}^{K}\sum_{n=1}^{N_k} \nabla_\theta^2 L(\underline{\theta}\,|\,y_n^k,\underline{x}_n^k) \tag{3}$$

In stratified sampling it is assumed that all observations from a given group (k) are realizations of some underlying distribution $f_{y,\underline{x}}^{(k)}(y,\underline{x})$ that characterizes the group. Thus all these observations have the same expectation. The expectation of Equation 3 is therefore

$$E[\nabla_\theta^2 L(\underline{\theta}\,|\,y,\underline{x})] = \sum_{k=1}^{K} N_k E^{(k)}[\nabla_\theta^2 L(\underline{\theta}\,|\,y,\underline{x})] \tag{4}$$

where $E^{(k)}[\cdot]$ denotes the expectation taken over the distribution $f_{y,\underline{x}}^{(k)}(y,\underline{x})$, and the designations n and k are omitted from the notation of the likelihood function in order to clarify the presentation. The final expression for the bound on the parameter covariance matrix is obtained by combining Equation 4 with Equation 1, i.e.,

$$\Sigma_{\hat{\theta}}(\underline{N}) = \left\{-\sum_{k=1}^{K} N_k E^{(k)}[\nabla_\theta^2 L(\underline{\theta}\,|\,y,\underline{x})]\right\}^{-1} \tag{5}$$

To minimize the estimation error, a scalar measure of the size of the parameter covariance matrix has to be defined. A family of such measures can be defined by using a quadratic form of the covariance matrix with a (column) vector of constants, $\underline{z}$, i.e.,

$$F = \underline{z}^T \Sigma_\theta \underline{z} \tag{6}$$

where F is the estimation error, Σ_θ is the true parameter covariance matrix, and the superscript T denotes the transposition operation. Because the true covariance matrix is not known, the approximation in Equation 5, which holds asymptotically for maximum likelihood estimators, is used instead. Thus $F(\underline{N}) = \underline{z}^T\Sigma_{\hat{\theta}}(\underline{N})\underline{z}$, where $\underline{N} = (\ldots,N_k,\ldots)$. The form of the error measure used in this paper uses a vector $\underline{z} = (1,1,\ldots,1)$, i.e., $F(\underline{N})$ is the sum of the elements of the parameter covariance matrix.

The optimal sample composition is derived by minimizing $F(\underline{N})$ with respect to the N_k's. The unconstrained solution to the minimization is, obviously, to sample an infinitely large number of observations from each group. The estimation error then approaches zero. The sample size, however, is bounded by the budget available for sampling, and possibly by some other physical size constraints. The total budget constraint may be expressed by

$$\sum_{k=1}^{K} c_k N_k \leqslant B \tag{7}$$

where c_k is the cost of sampling one unit from group k, and B is the total budget available. Physical group size constraints may be expressed as

$$N_k \leq N_k^{max} \quad \text{for some groups k} \tag{8}$$

In addition, the constraint set should always include nonnegativity of the group sizes.

The SO program can be summarized as follows:

$$\underset{N_k}{\text{Min}} \quad F = \underline{z}^T \left\{ - \sum_{k=1}^{K} N_k E^{(k)} [\nabla_\theta^2 L(\underline{\theta} | y, \underline{x})] \right\}^{-1} \underline{z} \tag{9a}$$

Subject to

$$\sum_{k=1}^{K} c_k N_k \leq B \tag{9b}$$

$$0 \leq N_k \leq N_k^{max} \quad \text{for all k} \tag{9c}$$

Daganzo (5) indicates that this program has a unique local minimum for any constant vector $\underline{z}$ and any form of the log-likelihood function $L(\underline{\theta}|y,\underline{x})$. This means that the problem can be solved by using standard nonlinear, constrained optimization methods. The algorithm used in this work is based on the gradient projection (6) method.

The exact form of the objective function depends on the specific model for which the sample is designed. Sheffi and Tarem (3) formulate and solve this program for several model forms. In the next section the derivation of this expression for MNL models is reviewed. The remainder of the paper is aimed at evaluating the usefulness and applicability of the approach.

SAMPLE OPTIMIZATION FOR LOGIT MODELS

The logit formula is the most widely used discrete-choice model because of the simplicity of its form. A detailed description of the model can be found in Domeneich and McFadden (7).

The logit model can be used to quantify some aspects of individuals' choice among a set of alternatives. The model can be interpreted in the framework of random utility maximization by assuming that each decision maker attaches a measure of utility to each alternative and chooses the one with the largest utility. The utility of alternative j to an individual randomly drawn from the population (u_j) is modeled as the sum of a systematic utility term (v_j) and an error term that is assumed to be randomly distributed across the population. The systematic utility captures the model specification in terms of the relationships between the utility and the explanatory variables; thus $v_j = v_j(\underline{\theta},\underline{x})$. The specification of the random part determines the family of models to be used. If these random terms are assumed identically and independently Gumbel distributed, the resulting model is the MNL mode. The MNL model gives the probability that each available alternative is chosen (i.e., it has the highest utility)--$P_j(\underline{\theta},\underline{x})$--as

$$P_j(\underline{\theta}, \underline{x}) = \exp [v_j(\underline{\theta}, \underline{x})] / \sum_{i \in I} \exp [v_i(\underline{\theta}, \underline{x})] \tag{10}$$

where I is the index set of the available alternatives. In most cases the systematic utility is assumed to be linear in the parameters, and thus $v_j(\underline{\theta},\underline{x}) = \underline{\theta}^T\underline{x}$.

To develop the SO objective function for the MNL model, the θ-Hessian of the log-likelihood function has to be derived for such models. The likelihood of a sample point n can be written as

$$L^*(\underline{\theta}|\underline{a}_n, \underline{x}_n) = \prod_{i \in I} P_i(\underline{\theta}, \underline{x}_n)^{a_{ni}} \tag{11}$$

where $\underline{a}_n$ is an indicator variable vector that contains the observed choice, i.e., $a_{nj} = 1$ if alternative j is chosen by the nth decision maker in the sample, and $a_{nj} = 0$ otherwise. The vector $\underline{x}_n$ includes the explanatory variables for the nth observation. The choice probabilities are given by Equation 10. The logarithm of Equation 11 is simply

$$L(\underline{\theta}|\underline{a}_n, \underline{x}_n) = \sum_{i \in I} a_{nj} \log P_{nj} \tag{12}$$

where $P_{nj} = P_j(\underline{\theta},\underline{x}_n)$ for ease of notation. The sample log-likelihood includes the sum over n of $L(\underline{\theta}|\underline{a}_n,\underline{x}_n)$, i.e.,

$$L(\underline{\theta}|\underline{a}, \underline{x}) = \sum_{n=1}^{N} \sum_{j \in I} a_{nj} \log P_{nj} \tag{13}$$

where N is the total sample size.

The derivation of the θ-Hessian of the sample log-likelihood function is simple but somewhat lengthy (3). The final result of applying the Hessian operator to the log-likelihood function is

$$\nabla^2 L(\underline{\theta}|\underline{a}, \underline{x}) = -W^T Q W \tag{14}$$

where W is the matrix of attribute differences for an individual randomly drawn from the population, i.e., row j of W is the difference $\underline{x}_j - \underline{x}_I$, where I is the index of the last alternative (any other alternative can be chosen as a base). Q is a square matrix with the elements,

$$[Q]_{ij} = P_i (\delta_{ij} - P_j) \quad \text{for } i, j = 1, 2, \ldots, I-1 \tag{15}$$

where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. After inserting Equation 14 into the objective function of the sample optimization program (Equation 9a), this function becomes

$$F = \underline{z}^T \left\{ - \sum_{k=1}^{K} N_k E^{(k)} [-W^T Q W] \right\}^{-1} \underline{z} \tag{16}$$

Computing the expectations of $E^{(k)}[\cdot]$ in Equation 16 requires prior knowledge of both the distribution of the attributes in all groups and the values of the unknown parameter vector ($\underline{\theta}$). The latter is required for computing the choice probabilities that appear in the elements of Q. As previously mentioned, it is assumed in this paper that the attribute distributions are known before sample optimization. The main concern of this paper is with the required accuracy of the initial parameter guesses.

Because the function under the expectation operator is complicated, a numerical Monte Carlo approach for computing these expectations was adopted. With this approach, M observations were drawn from the distribution of the attributes and the average, where

$$(1/M) \sum_{m=1}^{M} [-W_m^T Q_m W_m] \tag{17}$$

was used as an approximation of the true expectations.

INACCURACIES IN INITIAL GUESSES: ONE-PARAMETER MODEL

In this section two of the issues that determine the applicability of the SO approach are examined. These questions are addressed in the context of a simple logit model that includes only two alternatives and a single parameter.

The usefulness of the SO method depends on two separate questions. The first is whether SO actually improves the accuracy of the resulting parameter estimates. Although SO assures minimum error in estimation, the improvement relative to other sample designs may be insignificant. In this case the optimization process is not cost effective. The second question is the dependence of the optimization results on the accuracy of the initial parameter guesses used in the optimization. If the optimization process requires accurate parameter values to yield satisfactory sample composition, its usefulness will be limited because having such accurate parameter values obviates the need for the estimation process.

Thus, for the SO method to be useful, it is necessary that the estimation error will decrease when an optimal sample is used, but also that this optimal composition may be obtained without an accurate initial parameter guess.

The tests described in the following sections are designed to determine if and when these conditions can be met for a simple logit model, where the issue can be addressed analytically. The simple logit model chosen for this analysis includes two alternatives and one parameter. The systematic utilities of these alternatives are $x_1\theta$ and $x_2\theta$, respectively. The choice probabilities have the form

$$
\begin{aligned}
P_1 &= \exp[(x_1 - x_2)\theta]/\{1 + \exp[(x_1 - x_2)\theta]\} \\
&= \exp(W\theta)/[1 + \exp(W\theta)];\ P_2 = 1 - P_1 \\
&= 1/[1 + \exp(W\theta)]
\end{aligned} \tag{18}
$$

and the optimization objective function (Equation 16) for this model is given by

$$
\begin{aligned}
F(\underline{N}) &= 1/\sum_{k=1}^{K} N_k E^{(k)}[W^2 Q] = 1/\sum_{k=1}^{K} N_k E^{(k)}\{W^2 \exp(W\theta) \\
&\quad \div [1 + \exp(W\theta)]^2\}
\end{aligned} \tag{19}
$$

The minimization of $F(\underline{N})$ in Equation 19 is equivalent to the maximization of the reciprocal of $F(\underline{N})$, i.e.,

$$
\underset{N_k}{\text{Min}}\ F(\underline{N}) = \underset{N_k}{\text{Max}}\ F'(\underline{N}) = \sum_{k=1}^{K} N_k E^{(k)}[W^2 Q] \tag{20}
$$

For a problem with a simple budget constraint (such as Equation 9b), the solution of this SO program is to sample all observations from the group (k) with the largest value of

$$
\alpha^{(k)} = E^{(k)}[W^2 Q]/c_k \tag{21}
$$

The total sample size will, of course, be B/c_ℓ, where ℓ is the group sampled. From Equation 21 it is clear that if the expectations $E^{(k)}[\cdot]$ are similar in all groups, the sample should include observations from the group with the lowest sampling cost. If the expectations differ considerably, however, a group with higher sampling cost may contribute more to the estimation accuracy and should therefore be chosen for sampling.

The accuracy of the initial parameter guess, denoted by θ_0, needed in computing the $\alpha^{(k)}$'s is important only if it can cause the sampling from the wrong group. In other words, as long as the values of $\alpha^{(k)}$ computed by using θ_0 suggest the same choice of group as would happen with the true parameter (θ), the optimal sample composition is not affected by inaccuracies in θ_0.

For example, assume that there are only two groups, and that the sampling costs are the same in both. If the true $E^{(1)}[W^2Q]$ is 10 times larger than the true $E^{(2)}[W^2Q]$, computing $\alpha^{(k)}$ with even a bad guess of θ will still probably suggest sampling from group 1. If, on the other hand, the true $E^{(1)}[W^2Q]$ is only 10 percent larger than the true $E^{(2)}[W^2Q]$, a slight inaccuracy in θ_0 may reverse the choice initiated by Equation 21. In this case, however, the contribution of both groups to the estimation accuracy is similar, and sampling from the wrong group would not introduce a large increase in the estimation error (F).

In summary, Equation 21 indicates that if the group attribute distributions (and hence the group expectations) are considerably dissimilar, sampling from the wrong group may cause a large estimation error, but the correct group for sampling may be relatively easy to determine. In cases when this determination is more difficult (i.e., when the groups are similar), the cost of an error is not high. Thus this analysis leads to the conclusion that SO should be useful in this case, even with questionable prior estimates of θ.

TWO-PARAMETER MODEL

A similar analysis can be applied to a slightly more complicated model, which includes two alternatives and two parameters. In this case the choice probabilities have the form

$$
P_1 = \exp(W_1\theta_1 + W_2\theta_2)/[1 + \exp(W_1\theta_1 + W_2\theta_2)];
$$

$$
P_2 = 1/[1 + \exp(W_1\theta_1 + W_2\theta_2)] \tag{22}
$$

where W_1 and W_2 are the two elements of attribute differences vector $\underline{W} = (W_1, W_2)$. The SO objective function (Equation 16) in this case is given by

$$
F(\underline{N}) = \underline{z}^T \left\{ \sum_{k=1}^{K} N_k E^{(k)} [Q\, W^T W] \right\}^{-1} \underline{z} \tag{23}
$$

where the single element of the matrix Q is

$$
(Q)_{1,1} = \exp(W_1\theta_1 + W_2\theta_2)/[1 + \exp(W_1\theta_1 + W_2\theta_2)]^2 \tag{24}
$$

The general analysis of this case cannot be carried out analytically because of the complexity of Equation 24. The approach followed here was to analyze a specific sample design case with known true parameters. The problem setup included two groups with the following attribute distributions:

$$
W_1^{(1)} = W_1^{(2)} = 1;\ W_2^{(1)} \sim N(0.5, 0.25);\ W_2^{(2)} \sim N(-0.5, 0.25)
$$

The true parameters (see Equation 22) were set to $\theta_1 = \theta_2 = 1.0$. The true group expectations can be calculated by using the simulation method, explained by Equation 17, as follows:

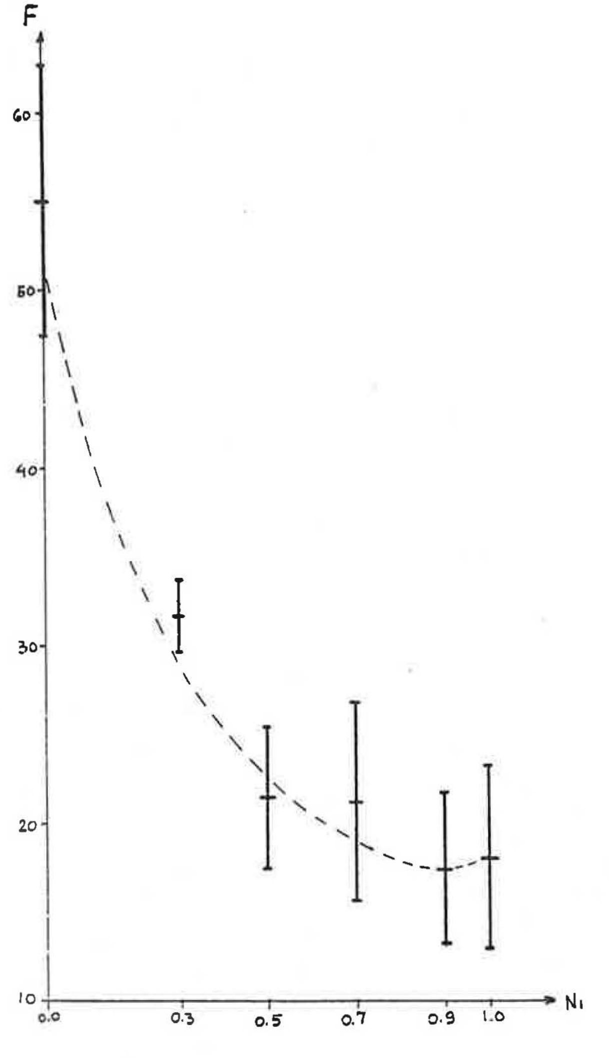
$$
E^{(1)}[QW^TW] = \begin{bmatrix} 0.1496 & 0.0549 \\ 0.0549 & 0.0545 \end{bmatrix};\ E^{(2)}[QW^TW] = \begin{bmatrix} 0.2234 & -0.122 \\ -0.122 & 0.1176 \end{bmatrix}
$$

The budget constraint was set to $N_1 + N_2 \leq 1$, which implies that $c_1 = c_2 = 1$ and that the N_k's can be looked on as sample shares rather than number of observations. Because the budget constraint is always binding in these problems, the sample composition can be represented by the single variable N_1, and N_2 can be replaced by $1 - N_1$.

The dependence of the estimation error on the sample composition was determined by evaluating the objective function (Equation 23) at different values of N_1. The resulting curve is shown as the dotted line on Figure 1. The estimation error has a distinct minimum at $N_1 = 0.908$, which corresponds to the value $F^* = 17.567$. It rises sharply for values of N_1 less than 0.69 (the 10 percent deviation mark).

Each sample composition is associated with a

Figure 1. Intervals $F \pm \sigma_F$ plotted versus the analytical curve.



unique value of the objective function in Figure 1. The sampling process, however, introduces a randomness that may cause the actual estimation error to deviate from the one indicated in Figure 1. This is because once the group size is determined, the actual observations are still randomly sampled within each group. Thus different samples with the same composition may result in different estimation errors. To verify the relationships shown in Figure 1, a simulated data set was generated. Attribute observations were generated from the previously mentioned distribution of the explanatory variables within each group. The chosen alternative was determined by simulating the total utilities of the alternatives to each individual and recording the alternative with the largest utility as the chosen one. This simulation was carried out by generating a Gumbel-distributed random variable (by using the cumulative distribution inversion method) and adding it to the observed utility.

The logit estimation routine computes, apart from the parameter estimates, an estimate of the parameter covariance matrix based on the sample. An approximate estimation error may be computed by summing the elements of this matrix (see Equation 6). Five different samples were generated for each selected composition, and the estimation error was computed for each one by using that procedure. An interval of probable values for the estimation error was derived from the mean and standard deviation of the five measurements (i.e., $F = \bar{F} \pm \sigma_F$, where $\bar{F}$ is the average and σ_F is the standard deviation of the five values). These intervals are also plotted in Figure 1. As demonstrated in the figure, the sampling results depict the same relation between the estimation error and the sample composition as shown by the analytical curve.

In the particular example solved here, Figure 1 demonstrates that the SO is worthwhile even when the randomness of the sampling procedure is accounted for. In general, however, this may not be the case if the variance of the attribute distribution is large. Such a case means that the groups are, statistically, quite similar. As in the one-parameter case, this means that SO is not cost effective because the (expected) cost of an error in the groups' composition is not large.

The dependence of the optimal solution on the accuracy of the initial guesses was determined by solving the SO problem by using different values of the initial parameter guesses ($\underline{\theta}_0$) around the true parameters ($\underline{\theta}$). Figure 2 shows contours of equal composition over a range of values of $\underline{\theta}_0$ around the true value of $\underline{\theta}$ = (1.0,1.0). The figure shows that in most of the region, except for the upper right corner, the optimal composition is within 10 percent of the best composition. The best composition is given by N_1 = 0.908, which was computed by using the true parameter values.

Figure 3 demonstrates the same point from a different angle. The relationships between the estimation error and the initial guesses used in the optimization process can be derived by reading, from Figure 1, the values of F that correspond to the sample compositions shown in Figure 2. These values can then be transformed to percentage differences from the minimum error, F^* = 17.567. Figure 3 depicts contours of equal percentage differences over the same range of $\underline{\theta}_0$ used in Figure 2. As shown in Figure 3, most of the region analyzed lies within 10 percent of the minimum error. In summary, it can be concluded that although arbitrary sample compositions may yield large estimation errors (as seen in Figure 1), the use of SO, even with a wide range of possible initial parameter guesses, limits the error to small deviations from the minimum error obtained by using the true parameter values.

Figure 2. Contours of equal N_1 over the range of the initial parameter guesses θ_0.

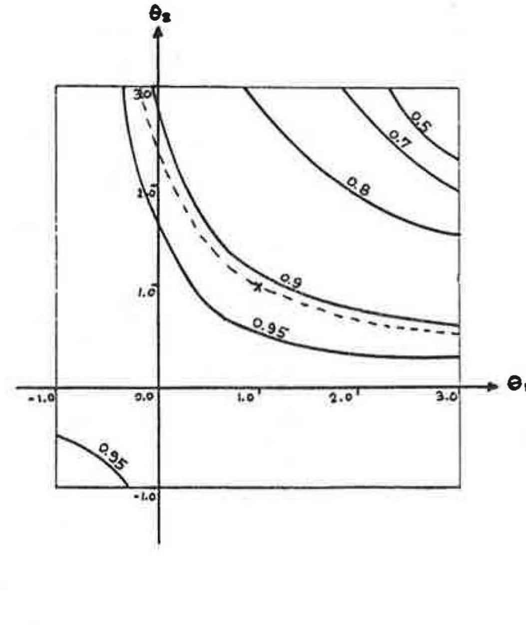
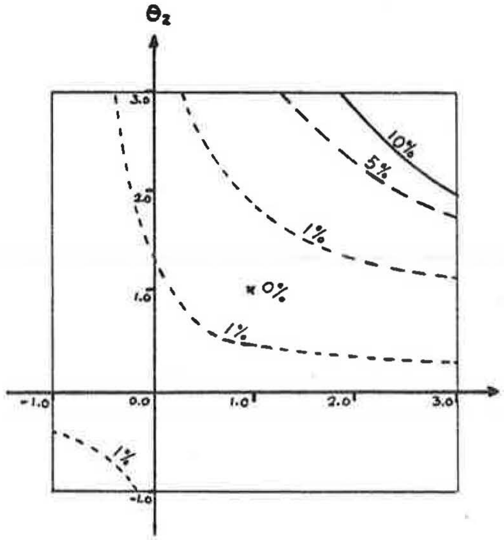


Figure 3. Contours of equal error (F) over the range of initial parameter guesses θ_0.



OPTIMAL BUDGET ALLOCATION

The initial parameter guesses used in the optimization process may come from two distinct sources. The first one is an external source, such as another study or a set of studies conducted elsewhere or in the past. The second one is an internal source, such as a pilot study conducted on the current population. In this case a small presample may be randomly drawn in order to estimate $\underline{\theta}_0$. The final parameter estimation will be based on a combined sample, including the observations of the presample and the main sample. The relevant question here is what is the appropriate relative investment in the initial sample that will yield the best accuracy of estimation when using the combined sample.

The procedure followed in this research for determining the optimal allocation of the sampling budget was to first allocate some prescribed amount (B_1) to an initial random sample. The parameter estimates based on this sample were used as initial guesses in determining the optimal sampling scheme for the main sample, subject to the remaining budget B_2. The main sample was then drawn and combined with the initial one and used to estimate the model. The estimation error was computed from the estimated parameter covariance matrix of this model. The optimal allocation was determined by parametrically varying the amount spent on the initial sample.

The existence of an optimal allocation stems from the fact that when the budget (B_1) spent on the initial random sample is small, the resulting estimates of the parameters are not accurate. Thus the main sample will not be close to optimality, and the estimation error can be expected to be large. On the other hand, when most of the budget is spent on the initial sample, the resulting initial estimates will be accurate, and the small main sample is close to being optimal. The combined sample, however, will include primarily the random, nonoptimal sample, and the estimation error is again expected to be large. Therefore, there may be some optimal allocation of the budget such that the size of the random sample is sufficient to provide relatively accurate estimates, but the remaining optimized sample is sufficiently large to reduce the error measure.

This procedure was carried out by using a large data set as a population. The data were extracted from the 1977 National Personal Travel Study (NPTS) data base. A simple model of automobile ownership levels was used as an example model in these tests. The model included three alternatives: owning two or more cars, owning one car, and owning no car. The systematic utilities of the alternatives were specified as

$$u_1 = \theta_1 + \theta_3 \cdot \text{INCOME} + \theta_4 \cdot \text{HHSIZE}$$

$$u_2 = \theta_2 + \theta_3 \cdot \text{INCOME}$$

$$u_3 = 0.0$$

where INCOME is measured in \$10,000 units, and HHSIZE is the number of members in the household. The population data set contained 7,393 observations partitioned into three groups along the income dimension, according to the following ranges:

Group	Income Range (\$)	Observations
1	0-7,500	2,565
2	7,500-20,000	3,331
3	>20,000	1,497
Total		7,393

The distributions of the attributes (INCOME and HHSIZE) were estimated from the data.

A budget size of B_1, varying between 40 and 200, was allocated to the initial random sample (assuming a cost of one unit for all observations). The composition of the main sample was determined by solving the optimization problem with the constraint $N_1 + N_2 + N_3 \leq B_2$, where $B_2 = 200 - B_1$. The two samples were then combined to yield a sample of size 200, and the estimation error was computed from the combined sample. This procedure was repeated five times for each value of B_1. The interval $\bar{F} \pm \sigma_F$ of the five measurements is plotted versus B_1 in Figure 4. A shallow minimum can be observed around B_1 = 80, which means that 80 observations should be sampled at random. The results of this estimation should be used to optimize the composition of the remaining 120 observations. The shape of the relationship shown in Figure 4 suggests, however, two hypotheses.

1. The optimal size of the initial sample is fixed, probably because it corresponds to the minimum sample size that yields reasonable initial estimates for the optimization. In this case the optimal initial sample size (B_1) is independent of the total sample size (B).

2. Optimizing a larger sample requires more accurate initial guesses, which implies a larger initial sample. In this case the optimal initial sample size (B_1) is a fixed proportion of the total sample size (B).

To test these hypotheses in the context of the examples analyzed in this section, the test procedure used in this case study was repeated for total sample sizes of B = 400 and 600 observations. The means of the five estimation error measures computed for each selected value of B_1 are plotted in Figure 5. The horizontal axis of the graph is the ratio B_1/B, and the vertical axis represents the estimation error. The measurements obtained from each value of B (i.e., 200, 400, and 600) were normalized for comparison purposes. The figure shows that for all total sample size values, the estimation error does not have a distinct minimum but is flat over the region up to B_1/B = 0.5 and rises thereafter.

Thus it can only be concluded that the initial

Figure 4. Estimation error intervals plotted versus the budget spent on the initial sample for total sample size of 200.

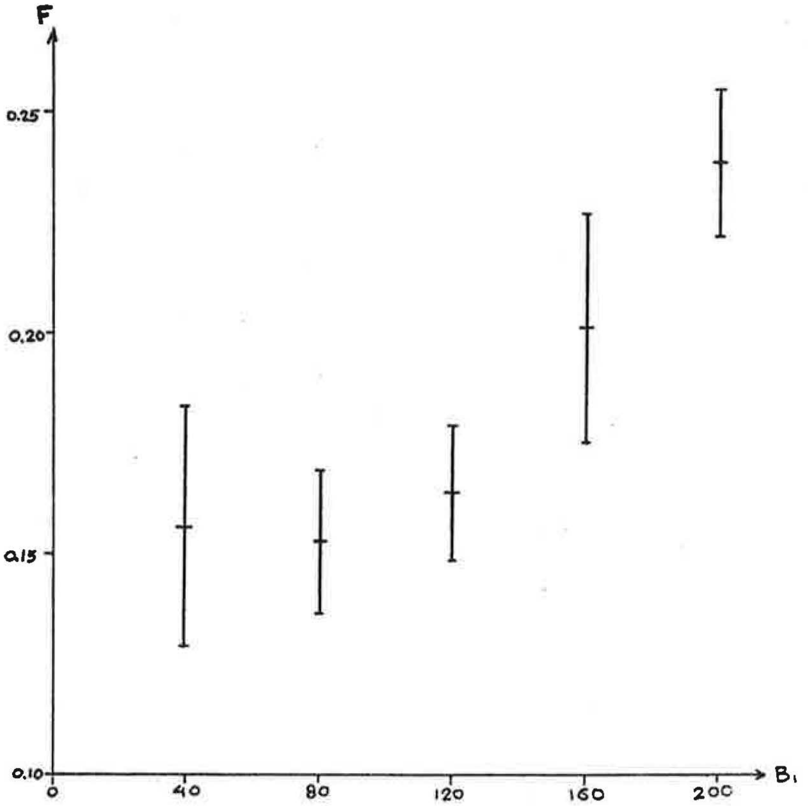
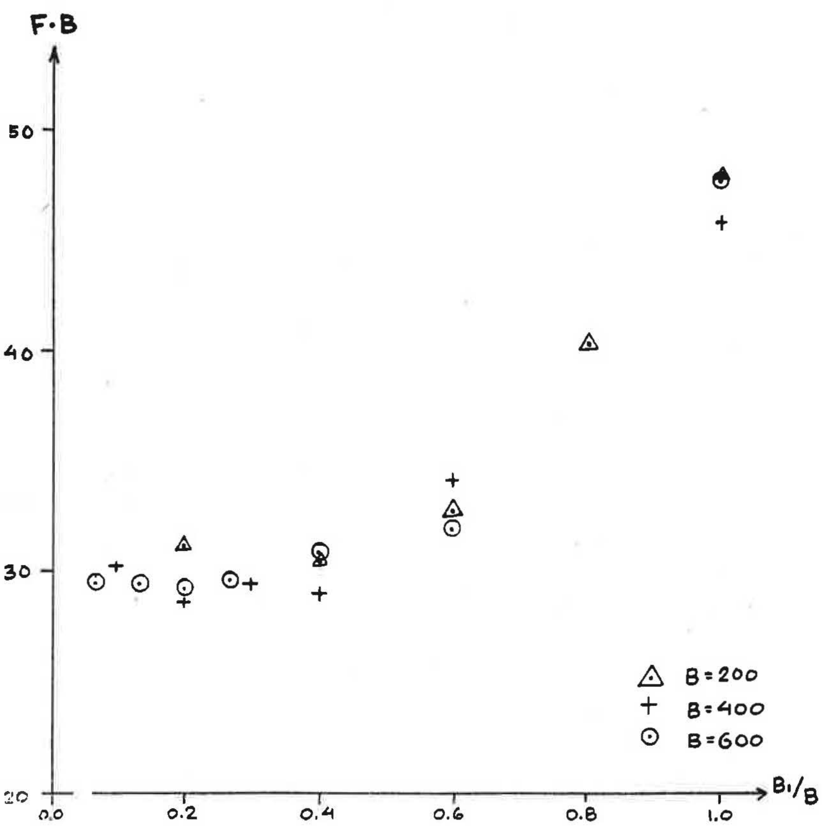


Figure 5. Mean estimation error versus ratio of initial sample size to total sample size.



sample size for this example should be less than one-half of the total sample size. This appears to suggest that, in general, the initial random sample can be small, regardless of the total sample size. The size of this sample may in fact be dictated by the requirements on the estimation of the distribution of the explanatory variables in all the groups. This point was not addressed in this paper, which assumed that this distribution is known.

CONCLUSIONS

The two major conclusions from the work described here may be stated as follows:

1. The SO procedure can introduce a significant increase in parameter estimation accuracy, and
2. This optimization need not be based on accurate initial parameter guesses; only a small pilot sample is needed to produce sufficiently accurate guesses.

It should be emphasized, however, that these conclusions result from a specific set of tests performed on prespecified models. Even though these models were chosen without any regard to the final results, these results can be generalized only with caution. The results are, however, encouraging in that the SO procedure appears to be worthwhile in cases where it can be applied. It requires nonlinear optimization software, which may not be easily used in many environments.

REFERENCES

1. W.G. Cochran. Sampling Techniques. Wiley, New York, 1963.
2. C.F. Daganzo. Optimal Sampling Strategies for Statistical Models with Discrete Dependent Variables. Transportation Science, Vol. 14, No. 4, Nov. 1980.
3. Y. Sheffi and Z. Tarem. Experiments with Optimal Sampling for Discrete Choice Models. Office of University Research, U.S. Department of Transportation, Aug. 1982.
4. S.R. Lerman and C.F. Manski. Sample Design for Discrete Choice Analysis of Travel Behavior: The State of the Art. Transportation Research, Part A, Vol. 13A, No. 1, Feb. 1979.
5. C.F. Daganzo. On the Uniqueness and Globality of Optimal Data Gathering Strategies. Transportation Science, Vol. 16, No. 2, May 1982.
6. E. Polak. Theory of Optimal Control and Mathematical Programming. McGraw-Hill, New York, 1979.
7. T. Domeneich and D. McFadden. Urban Travel Demand: A Behavioral Approach. North-Holland, Amsterdam, 1975.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Procedure for Predicting Queues and Delays on Expressways in Urban Core Areas

THOMAS E. LISCO

A procedure that predicts morning inbound and evening outbound queuing delays on express highway facilities in downtown areas is discussed. The procedure is based on the relationships among hourly traffic capacities at bottleneck points, daily volumes at those points, and associated queues and delays. The need for such a procedure arose from difficulties in using traffic assignment or other existing analysis techniques to predict queues and delays associated with alternative highway plans. Empirical delay data for developing the procedure came from nearly 600 speed runs conducted on the express highway system in and near downtown Boston. Fourteen queuing and potential queuing situations were analyzed. The relationships derived appear to be generalizable, and the specific results from the Boston area should apply to other urban areas of comparable size.

A procedure that predicts peak-period queuing and delays on express highway facilities in downtown areas is discussed. The procedure is based on the relationships among hourly traffic capacities at bottleneck points, daily volumes at those points, and associated peak-period queues and delays. (In this paper the term daily volume refers to average weekday traffic.) The procedure was developed by comparing observed bottleneck capacities with empirical delay data for traffic upstream of the bottlenecks. Capacities were derived from traffic counts at bottleneck locations. The delay data were from almost 600 speed runs conducted on express highway facilities in and near downtown Boston, mostly during 1978 and 1979. The procedure was developed for use in detailed evaluations of potential traffic impacts and benefits of alternative highway investments in downtown areas.

The need for such a procedure arises initially from difficulties in using the output from traffic assignment models to predict peak-period operating conditions and cost-benefit statistics associated with alternative highway plans. The basic problem is that the regional traffic assignment process derives speeds for individual links separately based on their individual volume/capacity (v/c) ratios and does not consider the queuing effects of bottleneck locations. Thus in typical downtown area queuing situations, where one bottleneck highway segment can create queues stretching into many other segments, traffic assignments cannot indicate the locations and extents of queues or the delays associated with them. Because queuing can be of major importance in peak-period expressway operations in downtown areas, the assignments can be grossly inaccurate in predicting peak-period operating speeds. Similarly, the associated cost-benefit statistics can miss much of the phenomenon they are intended to measure.