Testing Disaggregate Travel Demand Models by Comparing Predicted and Observed Market Shares

JOEL L. HOROWITZ

ABSTRACT

An intuitively appealing and popular method for testing a disaggregate choice model of travel demand, such as a logit model, consists of comparing the model's predictions of the market shares of travel alternatives in population groups with observations of these shares. Excessively large differences between the predicted and observed shares indicate that the model being tested is incorrect. In current practice, the decision whether differences between predictions and observations are large is made judgmentally, thereby raising the possibility that a correct or approximately correct model will be rejected because of the effects of random sampling errors. A statistical test is described that enables one to distinguish between the effects of random sampling errors and those of true model errors when predicted and observed market shares are compared. Five easily programmable steps for implementing the test are given, and commercially available software that can help with the computations is identified. A numerical example of the application of the test is presented, and the role of the test in practical model development is discussed.

An intuitively appealing way of testing any model of travel behavior is to compare its predictions with actual observations. In the case of disaggregate choice models, such as logit and probit models, that predict individuals' choices among sets of discrete alternatives, this approach to testing often consists of comparing predictions of market shares of alternatives in population groups with observations of the actual market shares in the same groups. Large differences between predicted and observed shares constitute grounds for rejecting the model. For example, to test a model of mode choice the population of interest might be grouped according to characteristics such as income, automobile ownership, location of residence or work, and so forth. The model's predictions of the proportions of individuals in each group that use each mode would be compared with the observed proportions. The model would be rejected as incorrect if the differences between the predicted and observed proportions were excessively large.

In current practice, the decision whether differences between predictions and observations are excessive is made judgmentally. This is unsatisfactory because differences between predicted and observed market shares are subject to random sampling errors. These errors are not relevant to the question of whether the model under consideration is correct. However, depending on the details of the model, selection of population groups, and size of the data set being used, they can produce differences between predictions and observations that are large by reasonable judgmental standards, even if the model being tested is correct. In other words, when predictions and observations are compared judgmentally, random sampling errors may cause a correct model to be rejected.

To minimize the likelihood of this undesirable outcome, it is necessary to have a method for distinguishing between random sampling errors and true model errors in comparisons of predicted and observed market shares. In the terminology of statistics, it is necessary to have a test of the statistical significance of differences between predictions and observations. The main objectives of this paper are to describe such a test and to present a numerical example illustrating its use. Subsidiary objectives are to discuss briefly two important questions relating to the use of the test. These are as follows:

1. In carrying out the test, should the data used for testing (i.e., for computing predicted and observed market shares) be independent of the data used for model estimation or should all of the available data be used for both estimation and testing?
2. The test based on comparisons of predicted and observed market shares is only one of several procedures that are available for testing disaggregate choice models. How should one choose among these procedures in practical, empirical work?

The remainder of this paper is organized as follows. The test statistic is described in the next section and the first question posed previously is answered. Then the numerical example of the use of the test statistic is presented. In the final section the second question posed previously is answered and some concluding comments are presented.

THE TEST STATISTIC

The most frequently used form of disaggregate choice model is the multinomial logit model with a linear-in-parameters utility function. Therefore, to minimize the complexity of the discussion, it will be assumed here that the model being tested has this form. The test statistic for a general choice model is given by Horowitz (1).

In the multinomial logit model with a linear-in-parameters utility function, the probability $P(i|m,\theta)$ that individual m chooses alternative i from a set of I available alternatives is (2,3)

$$P(i|m,\theta) = (\exp X'_{mi}\,\theta)/\sum_k (\exp X'_{mk}\,\theta) \qquad (1)$$

where X'_{mk} (k = 1, ..., I) is a row vector of explanatory variables evaluated for individual m and alternative k, θ is a column vector of constant parameters, and the sum in the denominator is over all available alternatives. In practice, the values of the parameters θ usually are not known a priori and must be estimated by fitting the model to data. In accordance with usual practice, it is assumed in this paper that θ is estimated by the method of

maximum likelihood using a disaggregate data set consisting of observations of the choices and X-values corresponding to M randomly selected individuals. This data set will be called the estimation data set. Let $\hat{\theta}$ denote the estimated values of the parameters θ. Then the estimated probability that individual m chooses alternative i is $P(i|m,\hat{\theta})$.

Derivation

Suppose that the model is to be tested by using a test data set consisting of observations of the choices and X-values corresponding to N randomly selected individuals. The test data set is assumed to be either the same as the estimation data set (in which case N = M) or independent of it. Let the individuals in the test data set be organized into J mutually exclusive groups (J > 1) either randomly or according to the values of characteristics such as income, automobile ownership, location of residence or work, and so on and let N_j denote the number of individuals assigned to group j (j = 1, ..., J). Let $z_{ni} = 1$ if individual n in the test data set chose alternative i and let $z_{ni} = 0$ otherwise. Then the observed market share of alternative i in population group j of the test data set ($\hat{Q}_{ij}$) is

$$\hat{Q}_{ij} = \sum_{\substack{n \text{ in} \\ \text{group } j}} (z_{ni}/N_j) \tag{2}$$

The predicted market share of alternative i in population group j according to the estimated model ($\hat{P}_{ij}$) is

$$\hat{P}_{ij} = \sum_{\substack{n \text{ in} \\ \text{group } j}} P(i|n,\hat{\theta})/N_j \tag{3}$$

The difference between the predicted and observed shares ($\hat{D}_{ij}$) is $\hat{Q}_{ij} - \hat{P}_{ij}$ or

$$\hat{D}_{ij} = \sum_{\substack{n \text{ in} \\ \text{group } j}} [z_{ni} - P(i|n,\hat{\theta})]/N_j \tag{4}$$

$\hat{D}_{ij}$ contains two sources of random error that cause it to differ from zero in general, even if the model $P(i|n,\theta)$ with suitably chosen parameter values is correct. First, the individuals in the test data set are sampled randomly, so their choice indicators z_{ni} are random. Second, the estimated parameter values $\hat{\theta}$ are random because they depend on the choices of the randomly selected individuals in the estimation data set. To develop a test statistic that enables one to distinguish between the effects on $\hat{D}_{ij}$ of random sampling errors and those of true errors in the model, it is necessary to know how large the sampling errors might cause $\hat{D}_{ij}$ to be if the model being tested is in fact correct. [Throughout this paper, the term "large" refers to both large positive and large negative values of $\hat{D}_{ij}$.] This in turn requires knowledge of the probability distribution of the random variable $\hat{D}_{ij}$ for the case of testing a correct model.

To obtain this distribution, let θ_0 denote the true (but unknown) values of the parameters. In other words, $P(i|n,\theta_0)$ is a correct model. Then as discussed by Chernoff (4), $P(i|n,\theta)$ is given approximately by

$$P(i|n,\hat{\theta}) = P(i|n,\theta_0) + \sum_k P(i|n,\theta_0)(X'_{ni} - X'_{nk}) \times P(k|n,\theta_0)(\hat{\theta} - \theta_0) \tag{5}$$

where the sum on the right-hand side is over all of the available alternatives. Substituting Equation 5 into 4 yields

$$\hat{D}_{ij} = \sum_{\substack{n \text{ in} \\ \text{group } j}} \{[z_{ni} - P(i|n,\theta_0)]/N_j\} - K'_{ij}(\hat{\theta} - \theta_0) \tag{6}$$

where K'_{ij} is the row vector defined by

$$K'_{ij} = \sum_{\substack{n \text{ in} \\ \text{group } j}} \sum_k P(i|n,\theta_0)(X'_{ni} - X'_{nk})P(k|n,\theta_0)/N_j \tag{7}$$

The first term on the right-hand side of Equation 6 gives the effect on $\hat{D}_{ij}$ of random variations in the z_{ni}'s. This term is normally distributed with a mean of zero by virtue of the central limit theorem and the mean value of z_{ni} of $P(i|n,\theta_0)$. The second term in the right-hand side of Equation 6 gives the effect on $\hat{D}_{ij}$ of random sampling errors in the estimated parameter values. This term also is normally distributed with mean zero because $\hat{\theta}$ is normally distributed with mean θ_0 when $\hat{\theta}$ is obtained by the method of maximum likelihood (5). [The N_j values and the term K'_{ij} that multiplies $\hat{\theta} - \theta_0$ in Equation 6 can be treated as constants in samples of practical size.] Because sums and differences of normally distributed random variables are normally distributed, it follows that for each i and j, $\hat{D}_{ij}$ has the normal distribution with a mean of zero when the model being tested is correct.

To complete the derivation of the probability distribution of $\hat{D}_{ij}$, it is necessary to calculate the variance of $\hat{D}_{ij}$. Before doing this, however, it is important to observe that there are IJ different values of $\hat{D}_{ij}$, one for each alternative-population group pair. To conclude that the model being tested is incorrect, it is necessary for these IJ values to be excessively large collectively rather than individually. As will be seen, carrying out a test of the sizes of the $\hat{D}_{ij}$'s considered collectively requires knowledge of the covariances of D_{ij}'s corresponding to different alternative-population group pairs as well as knowledge of the variances of individual $\hat{D}_{ij}$'s. Thus it is necessary to obtain the covariance terms $\text{cov}(\hat{D}_{ij},\hat{D}_{rs})$ for each combination of pairs ij and rs.

It can be seen from Equation 6 that each $\hat{D}_{ij}$ contains two parts, one associated with random variations in the z_{ni}'s (i.e., the first term on the right-hand side of Equation 6) and one associated with random variations in $\hat{\theta}$ (i.e., the second term on the right-hand side of Equation 6). Let A_{ijrs} denote the covariance of the first term on the right-hand side of Equation 6 and let B_{ijrs} denote the covariance of the second term. Then as shown by Horowitz (1),

$$A_{ijrs} = \sum_{\substack{n \text{ in} \\ \text{group } j}} [P(i|n,\theta_0)\delta_{ir} - P(i|n,\theta_0)P(r|n,\theta_0)]\delta_{js}/N_j^2 \tag{8}$$

where $\delta_{ir} = 1$ if $i = r$, $\delta_{ir} = 0$ if $i = r$, and δ_{js} is defined similarly. B_{ijrs} is given by (1)

$$B_{ijrs} = K'_{ij} V K_{rs} \tag{9}$$

where V is the covariance matrix of the parameter estimates $\hat{\theta}$. Note that A_{ijrs} is associated with random variations in the z_{ni}'s, whereas B_{ijrs} is associated with random variations in $\hat{\theta}$.

The covariance $\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ can be obtained by combining A_{ijrs} and B_{ijrs} in the appropriate way. Consider first the case of independent estimation and test data sets. In this case the two terms on the right-hand side of Equation 6 are independent (because they are computed using different data sets), so $\hat{D}_{ij}$ is the difference of two independent random variables. Therefore, $\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ is simply the sum of the covariances arising from the independent components of $\hat{D}_{ij}$ and $\hat{D}_{rs}$ (5). In other words,

$$\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs}) = A_{ijrs} + B_{ijrs} \tag{10}$$

when the estimation and test data sets are independent.

Derivation of $\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ for the case in which the estimation and test data sets are the same is more complicated because $\hat{\theta}$ depends on the z_{ni}'s in this case, which causes the two terms on the right-hand side of Equation 6 to be correlated. Horowitz has shown (1) that when the estimation and test data sets are the same, $\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ is the difference between A_{ijrs} and B_{ijrs}. Thus,

$$\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs}) = A_{ijrs} - B_{ijrs} \tag{11}$$

when the estimation and test data sets are the same.

To obtain a statistic that tests whether the $\hat{D}_{ij}$'s considered collectively are larger than can be explained by random sampling errors, it is convenient to organize the $\hat{D}_{ij}$'s into a vector and the covariances $\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ into a matrix. The appropriate vector is given in row vector form by

$$\hat{D}' = (\hat{D}_{11}, \hat{D}_{21}, \ldots, \hat{D}_{I1}, \hat{D}_{12}, \hat{D}_{22}, \ldots, \hat{D}_{I2}, \ldots, \hat{D}_{1J}, \hat{D}_{2J}, \ldots, \hat{D}_{IJ}) \tag{12}$$

Let $\hat{D}$ denote the corresponding column vector. The appropriate matrix is most conveniently defined in terms of submatrices. Define the I x I submatrix S_{js} (j,s = 1, ..., J) by

$$S_{js} = \begin{bmatrix} \mathrm{cov}(\hat{D}_{1j}, \hat{D}_{1s}) & \mathrm{cov}(\hat{D}_{1j}, \hat{D}_{2s}) \ldots \mathrm{cov}(\hat{D}_{1j}, \hat{D}_{Is}) \\ \mathrm{cov}(\hat{D}_{2j}, \hat{D}_{1s}) & \mathrm{cov}(\hat{D}_{2j}, \hat{D}_{2s}) \ldots \mathrm{cov}(\hat{D}_{2j}, \hat{D}_{Is}) \\ & \ldots \\ \mathrm{cov}(\hat{D}_{Ij}, \hat{D}_{1s}) & \mathrm{cov}(\hat{D}_{1j}, \hat{D}_{2s}) \ldots \mathrm{cov}(\hat{D}_{Ij}, \hat{D}_{Is}) \end{bmatrix} \tag{13}$$

Then the desired IJ x IJ covariance matrix S is

$$S = \begin{bmatrix} S_{11} & S_{12} & \ldots & S_{1J} \\ S_{21} & S_{22} & \ldots & S_{2J} \\ & \ldots & & \\ S_{J1} & S_{J2} & \ldots & S_{JJ} \end{bmatrix} \tag{14}$$

This matrix contains all the covariances $\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ organized in a way that is conformable with the vector $\hat{D}$. In matrix notation $E(\hat{D}\hat{D}') = S$. If the estimation and test data sets are independent, the elements of S are given by Equation 10. If the estimation and test data sets are the same, the elements of S are given by Equation 11.

The matrix S is singular (i.e., its determinant is zero) owing to the existence of exact linear relations among the $\hat{D}_{ij}$'s. For example, for any population group j

$$\sum_{i=1}^{I} \hat{D}_{ij} = 0 \tag{15}$$

because the sums over all alternatives of the predicted and observed market shares must both equal 1. Thus, the elements of $\hat{D}$ are normally distributed with mean zero and singular covariance matrix S when the model being tested is correct.

Define the random variable C by

$$C = \hat{D}' S^{-} \hat{D} \tag{16}$$

where S^{-} is any matrix satisfying

$$SS^{-}S = S \tag{17}$$

S^{-} is called a generalized inverse of S and, as will be explained shortly, can be computed from S using standard computer software. C is a collective indicator of the sizes of the $\hat{D}_{ij}$'s. Roughly speaking, if the $\hat{D}_{ij}$'s are large, C is large, and if the $\hat{D}_{ij}$'s are small, C is small. Moreover, C has a known probability distribution when the model being tested is correct. Specifically, C has the chi-square distribution with degrees of freedom equal to the rank of S (6). As a result, C can be used to test whether the differences between the predicted and observed market shares are larger than can be explained by random sampling errors. If the D_{ij}'s differ from zero only because of random sampling errors, with probability $1 - \alpha$ the value of C will be less than the $1 - \alpha$ quantile of the chi-square distribution with degrees of freedom equal to the rank of S. Therefore, if the value of C exceeds the $1 - \alpha$ quantile of this distribution, the model being tested is rejected at the α significance level (i.e., either the model being tested is incorrect or an event of probability α has occurred). By choosing a small enough value of α (values of 0.05 and 0.01 are used frequently in practice), one ensures that there is little likelihood of rejecting a correct model because of the effects of random sampling errors.

Two problems remain to be solved before the test based on the C statistic of Equation 16 (hereafter called the C test) can be implemented in practice. First, the elements of the matrix S depend on the unknown true parameter values θ_0 and on the covariance matrix V of the parameter estimates (see Equations 7-9). A means must be found for estimating these quantities. Second, a method is needed for computing the matrix S^{-}. The first problem can be solved by using the parameter estimates $\hat{\theta}$ in place of the unknown true values θ_0 in Equations 7 and 8. The matrix V can be estimated by the inverse of the Fisher information matrix of the estimation data set evaluated at the parameter values $\hat{\theta}$ (2,5). This estimate of V is one of the standard outputs of many logit estimation computer programs.

The matrix S^{-} can be obtained in several different ways. One is by using a commercially available software package for computing generalized inverses of matrices. For example, the operator GINV in the SAS procedure MATRIX computes the generalized inverse of a matrix that is supplied as input to the procedure (7). Another way to compute S^{-} is by using the eigenvalues and eigenvectors of S. Let W be the matrix whose columns are the eigenvectors of S and let λ_k (k = 1, ..., IJ) denote the corresponding eigenvalues of S. The λ's and the ma-

trix W can be obtained using commercially available computer software such as the subroutine EIGRS of the International Mathematical and Statistical Library (IMSL) (8) or the operator EIGEN in the SAS MATRIX procedure. For each k define λ_k^* by $\lambda_k^* = 1/\lambda_k$ if $\lambda_k = 0$ and $\lambda_k^* = 0$ otherwise. Then the matrix whose (p,q) element is

$$S^-_{pq} = \sum_k \lambda_k^* W_{pk} W_{qk} \quad (18)$$

is a generalized inverse of S (5). A singular matrix has infinitely many generalized inverses, and not all procedures for computing generalized inverses yield the same results. However, any generalized inverse can be used in carrying out the C test.

Implementation

The results of the preceding discussion can be organized into five steps for comparing predicted and observed market shares by means of a C test. These steps, which can easily be programmed for implementation by computer, are as follows:

1. Use the method of maximum likelihood to estimate the values of the parameters $\hat{\theta}$ of the model to be tested and the covariance matrix V of the parameter estimates.
2. Organize the test data set into the desired population groups. For each alternative i and group j, compute the observed market share $\hat{Q}_{ij}$ (Equation 2), the predicted market share $\hat{P}_{ij}$ (Equation 3), and the difference $\hat{D}_{ij}$ between the observed and predicted shares. Organize the elements $\hat{D}_{ij}$ into the vector D (Equation 12).
3. Using the estimate of V and the estimated parameter values $\hat{\theta}$ in place of θ_0, compute the quantities A_{ijrs}, B_{ijrs}, and $\text{cov}(\hat{D}_{ij},\hat{D}_{rs})$ (Equations 7-9 and either 10 or 11, depending on whether the estimation and test data sets are independent or the same). Organize the quantities $\text{cov}(\hat{D}_{ij},\hat{D}_{rs})$ into the matrix S (Equations 13 and 14). Determine the rank of S. An easy way to do this is by computing the eigenvalues of S using one of the methods described earlier. The rank of S equals the number of nonzero eigenvalues.
4. Compute S^- using one of the methods described earlier.
5. Compute the test statistic C (Equation 16). Reject the model under consideration at the α significance level if the value of C exceeds the $1 - \alpha$ quantile of the chi-square distribution with degrees of freedom equal to the rank of S.

A numerical example illustrating the application of these steps is given later is this paper.

Division of Data

A statistical test can identify an incorrect model with high probability only if the effects of random sampling errors on the test statistic are less than those of errors in the model. In other words, to maximize a test's power (or ability to identify erroneous models) it is necessary to minimize the effects of random sampling errors.

In the case of the C test, Equations 10 and 11 indicate that the effects of random sampling errors are likely to be smaller when the estimation and test data sets are the same than when they are independent. There are two reasons for this. First, recall that the covariance components A_{ijrs} and B_{ijrs} respectively represent the effects on the $\hat{D}_{ij}$'s of random sampling errors in the z_{ni}'s and the parameter estimates $\hat{\theta}$. Equations 10 and 11 show that the latter errors are added to the former when the estimation and test data sets are independent, whereas the latter are subtracted from the former when the estimation and test data sets are the same. This causes the sampling errors in the $\hat{D}_{ij}$'s to be smaller when the estimation and test data sets are the same than when they are independent. Second, use of independent estimation and test data sets means that only part of the available data is used in each of the two stages, estimation and testing, whereas all of the data is used in both stages when the estimation and test data sets are the same. This also makes the random sampling errors in the $\hat{D}_{ij}$'s smaller when the estimation and test data sets are the same than when they are independent.

Because the random sampling errors in the $\hat{D}_{ij}$'s are smaller when the estimation and test data sets are the same than when they are independent, the power of the C test usually is larger when the same data are used for estimation and testing than when independent data sets are used. [A more formal presentation of this argument is given elsewhere (1).] In summary, when a model is tested by comparing predicted and observed market shares in population groups, all the available data should be used for both estimation and testing whenever this is feasible (i.e., whenever the full data set is not too large for use in model estimation). The data should not be divided into separate estimation and test data sets.

A NUMERICAL EXAMPLE

The example consists of testing the logit model of mode choice. The modes are automobile and transit. The data used in the example consist of 500 observations of choices and the relevant explanatory variables. The observations were generated by simulation from a logit mode-choice model specified as in Equation 1 with

$$X'_{mi}\,\theta_0 = -0.0865T_{mi} - (0.24881C_{mi}/Y_m) + 3A_mR_i - 4.5R_i \quad (19)$$

where

- i = automobile or transit,
- T_{mi} = individual m's travel time (min) by mode i,
- C_{mi} = individual m's travel cost (cents) by mode i,
- Y_m = individual m's annual income ($10,000s),
- A_m = number of automobiles owned by individual m's household, and
- R_i = 1 if mode i is automobile and 0 if mode i is transit.

[Those who wish to work through this example in detail may obtain a copy of the data set and the FORTRAN program that generated it from the author on request.]

The logit model that is tested in the example has the specification

$$X'_{mi}\,\theta = \theta_1 T_{mi} + (\theta_2 C_{mi}/Y_m) + \theta_3 R_i \quad (20)$$

This model is incorrect because it does not include the variable A_mR_i. To maximize the probability that the C test will identify the model of Equation 20 as incorrect, the same data set will be used for both estimation and testing.

The C test will now be implemented by carrying out the five steps given in the previous section.

Step 1

The maximum-likelihood estimates of the parameters of the model of Equation 20 and the estimate of the covariance matrix V are given in Table 1. Note that the signs of the time and cost coefficients are consistent with expectation and that the t-statistics of these coefficients are satisfactory. Thus, neither the signs nor the t-statistics suggest that the model is incorrect. The nonsignificance of the estimated coefficient of R_i does not indicate that the model is incorrect because there are no strong a priori expectations concerning the importance of this variable.

Step 2

Two population groups are used in this example: Group 1 consists of individuals whose households own one car, and group 2 consists of individuals whose households own two cars. All individuals in the data set used in the example are from one-car or two-car households. The values of the $\hat{Q}_{ij}$'s, $\hat{P}_{ij}$'s, and $\hat{D}_{ij}$'s are shown in Table 2. Organizing the D_{ij} values into the vector $\hat{D}$ yields

$$\hat{D} = \begin{bmatrix} -0.1124 \\ 0.1124 \\ 0.0942 \\ -0.0942 \end{bmatrix} \tag{21}$$

TABLE 1 Estimation Results for the Logit Model of Equation 20

Parameter	Estimated Value	t-Statistic
θ_1	-0.05056	-2.494
θ_2	-0.2148	-9.382
θ_3	0.3194	1.158

$$V = \begin{bmatrix} 0.4110 & -0.3542 & 4.318 \\ -0.3542 & 0.5242 & -2.388 \\ 4.318 & -2.388 & 76.08 \end{bmatrix} \times 10^{-3}$$

TABLE 2 Values of $\hat{Q}_{ij}$, $\hat{P}_{ij}$, and $\hat{D}_{ij}$ for the Numerical Example

i (Mode)[a]	j (Group)	$\hat{Q}_{ij}$	$\hat{P}_{ij}$	$\hat{D}_{ij}$
1	1	0.3465	0.4588	-0.1123
2	1	0.6535	0.5412	0.1123
1	2	0.6618	0.5676	0.0942
2	2	0.3382	0.4324	-0.0942

[a] Mode 1 is automobile, and mode 2 is transit.

Step 3

The row vectors K_{ij}' defined in Equation 7 are

$$K_{11}' = (-0.9799, 0.4670, 0.09388) \tag{22a}$$

$$K_{21}' = (0.9799, -0.4670, -0.09388) \tag{22b}$$

$$K_{12}' = (-0.8070, 0.2279, 0.08145) \tag{22c}$$

$$K_{22}' = (0.8070, -0.2279, -0.08145) \tag{22d}$$

The covariance terms A_{ijrs}, B_{ijrs}, and $\text{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ are shown in Table 3. Organizing the quantities $\text{cov}(\hat{D}_{ij}, \hat{D}_{rs})$ into the matrix S yields

$$S = \begin{bmatrix} 0.2033 & -0.2033 & -0.1704 & 0.1704 \\ -0.2033 & 0.2033 & 0.1704 & -0.1704 \\ -0.1704 & 0.1704 & 0.1429 & -0.1429 \\ 0.1704 & -0.1704 & -0.1429 & 0.1429 \end{bmatrix} \times 10^{-3} \tag{23}$$

The eigenvalues of S were computed using the subroutine EIGRS of the IMSL and are (0, 0, 0, 6.923×10^{-3}). Because there is only one nonzero eigenvalue, the rank of S is 1.

Step 4

The generalized inverse of S as computed from Equation 18 is

$$S^{-} = \begin{bmatrix} 424.2 & -424.2 & -355.6 & 355.6 \\ -424.2 & 424.2 & 355.6 & -355.6 \\ -355.6 & 355.6 & 298.0 & -298.0 \\ 355.6 & -355.6 & -298.0 & 298.0 \end{bmatrix} \tag{24}$$

Step 5

The test statistic C is computed by substituting Equations 21 and 24 into Equation 16. The result is C = 62.15. The 0.95 quantile of the chi-square distribution with 1 degree of freedom is 3.841. Because the computed value of C exceeds this, the model of Equation 20 is rejected as incorrect at the 0.05 significance level.

CONCLUSIONS

Testing a disaggregate choice model by comparing predicted and observed market shares has much intuitive appeal, but it is not the only way in which these models can be tested. A large number of test procedures based on likelihood ratio tests, Lagrangian multiplier tests, and the likelihood ratio index goodness-of-fit statistic are also available (9,10). It is worthwhile to consider how one might choose among these tests in practical model development.

An important difference between the C test discussed in this paper and the other tests is that the other tests require the analyst to specify an alternative model (e.g., a logit model with a different specification of the utility function $X_{mi}'\theta$) against which the model under consideration is to be tested. In effect, these tests attempt to determine whether the alternative model fits the available data better than the model under consideration does, in which case the model under consideration is rejected as being incorrect. In contrast, the C test does not require specification of an alternative

TABLE 3 Covariance Terms for the Example

A_{ijrs}

		s = 1		s = 2	
		r=1	r=2	r=1	r=2
j = 1	i = 1	4.117×10^{-4}	-4.117×10^{-4}	0	0
	i = 2	-4.117×10^{-4}	4.117×10^{-4}	0	0
j = 2	i = 1	0	0	2.995×10^{-4}	-2.995×10^{-4}
	i = 2	0	0	-2.995×10^{-4}	2.995×10^{-4}

B_{ijrs}

		s = 1		s = 2	
		r=1	r=2	r=1	r=2
j = 1	i = 1	2.084×10^{-4}	-2.084×10^{-4}	1.704×10^{-4}	-1.704×10^{-4}
	i = 2	-2.084×10^{-4}	2.084×10^{-4}	-1.704×10^{-4}	1.704×10^{-4}
j = 2	i = 1	1.704×10^{-4}	-1.704×10^{-4}	1.566×10^{-4}	-1.566×10^{-4}
	i = 2	-1.704×10^{-4}	1.704×10^{-4}	-1.566×10^{-4}	1.566×10^{-4}

$\mathrm{cov}(\hat{D}_{ij}, \hat{D}_{rs})$

		s = 1		s = 2	
		r=1	r=2	r=1	r=2
j = 1	i = 1	2.033×10^{-4}	-2.033×10^{-4}	-1.704×10^{-4}	1.704×10^{-4}
	i = 2	-2.033×10^{-4}	2.033×10^{-4}	1.704×10^{-4}	-1.704×10^{-4}
j = 2	i = 1	-1.704×10^{-4}	1.704×10^{-4}	1.429×10^{-4}	-1.429×10^{-4}
	i = 2	1.704×10^{-4}	-1.704×10^{-4}	-1.429×10^{-4}	1.429×10^{-4}

model. In effect, it tests the model under consideration against all alternatives simultaneously.

The ability of a test against a specific alternative to identify an erroneous model and the relative power of the C test and a test against a specific alternative depend on the choice of the alternative. As is discussed elsewhere (1,11), a test of an erroneous model against an alternative that is a good approximation to the correct model is likely to be much more powerful than a C test. However, a test against an alternative that is a poor approximation to the correct model can be less powerful than a C test.

These considerations suggest that the C test and tests against specific alternatives are complements, rather than substitutes, and that both types of tests should be carried out during the process of developing empirical models. A practical approach to this begins by formulating several alternatives to the model currently under consideration. The developer of an empirical model can virtually always do this. Although there can be no assurance that any of these alternatives is correct or approximately correct (if there could be such assurance, the tests being discussed here would be unnecessary), one or more of the alternatives may nonetheless provide a powerful test of the current model if the model is seriously erroneous (9). Accordingly, the current model should be tested against the alternatives using likelihood ratio tests or other appropriate procedures (9,10). However, it is not possible to test a model against all reasonable alternatives, and failure to reject the current model in tests against specific alternatives may be the result of a poor choice of alternatives rather than an indication that the current model is correct. Therefore, if the current model is not rejected in tests against specific alternatives, a C test should be carried out. The C test amounts to a test of the current model against all remaining alternatives and may have a higher probability of detecting errors in the current model than do the tests against specific alternatives if the specific alternatives are themselves highly erroneous.

ACKNOWLEDGMENT

The work reported in this paper was supported in part by a cooperative agreement between UMTA and the University of Iowa.

REFERENCES

1. J.L. Horowitz. Testing Probabilistic Discrete Choice Models of Travel Demand by Comparing Predicted and Observed Aggregate Choice Shares. Transportation Research, in press (1985).
2. T.A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. American Elsevier, New York, 1975.
3. D.A. Hensher and L.W. Johnson. Applied Discrete-Choice Modelling. Croom Helm, London, 1981.
4. H. Chernoff. Large-Sample Theory: Parametric Case. Annals of Mathematical Statistics, Vol. 27, 1956, pp. 1-22.

5. C.R. Rao. Linear Statistical Inference and Its Applications, 2nd ed. Wiley, New York, 1973.
6. C.R. Rao and S.K. Mitra. Generalized Inverse of Matrices and Its Applications. Wiley, New York, 1971.
7. SAS User's Guide: Statistics. SAS Institute, Inc., Cary, N.C., 1982.
8. The IMSL Library. IMSL, Inc., Houston, Texas, 1982.
9. J. Horowitz. Specification Tests for Probabilistic Choice Models. Transportation Research, Vol. 16A, 1982, pp. 383-394.
10. J.L. Horowitz. Evaluation of Usefulness of Two Standard Goodness-of-Fit Indicators for Comparing Non-Nested Random Utility Models. In Transportation Research Record 874, TRB, National Research Council, Washington, D.C., 1982, pp. 19-25.
11. A. Wald. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large. Transactions of the American Mathematical Society, Vol. 54, 1943, pp. 426-487.

Evaluation of Heuristic Transit Network Optimization Algorithms

KAY W. AXHAUSEN and ROBERT L. SMITH, JR.

ABSTRACT

Changes in urban land use and travel demand have created the need to restructure many existing mass transit networks. Heuristic network optimization as one of the available methodologies to improve transit networks is described. The characteristics and results of the algorithms developed in Europe are summarized and a short description of the American algorithms is given. The potential for applying network optimization methodologies in the context of small to medium-sized American cities is evaluated. The review and evaluation of 13 heuristic methodologies revealed a wide range of approaches that are generally theoretically sound, have reasonable potential for generating improved networks, and are computationally and otherwise feasible. Application of an unproven new algorithm by Mandl to the bus network for Madison, Wisconsin, and the light rail network for Duesseldorf, West Germany, showed that a fairly complex heuristic algorithm can be implemented quickly and easily. Mandl's algorithm, however, did not generate an improved network, primarily because the initial computer-generated network does not follow demand. Better results were obtained with two other heuristic methodologies that have been applied to the Duesseldorf network. The Madison and Duesseldorf applications form the basis for recommendations for further improvement of heuristic methodologies.

The bus transit networks that are the predominant form of public transit in American cities have changed only slowly since the elimination of the streetcar in the 1930s and 1940s. Often the major bus lines still run on the same streets that the streetcars used. Because of the major shifts in population and employment that have occurred in recent years, the bus networks in many cities could probably be restructured to serve the existing demand better and reduce operating costs at the same time. Transit managers are often reluctant to make major changes in routes because of the almost certain political opposition by those who think they will receive poorer service. Also, transit managers generally do not have analytical tools readily available to aid them in generating and evaluating alternative networks. As the result of the current fiscal crisis in transit, transit managers should be more interested in methodologies for restructuring their bus networks.

Chua and Silcock (1) identify six methodologies for transit network restructuring and optimization: manual approach using service standards and guidelines, systems analysis using standard travel demand and trip assignment models, market analysis using manual trip assignment for corridors or small service areas, systems analysis with interactive graphics, heuristic procedures, and mathematical optimization. The first three methodologies are limited by the number of alternative networks that can be evaluated in a reasonable amount of time. By adding interactive graphics to systems analysis, network development and evaluation are greatly enhanced. Many more networks can be tested in much less time. The methodology, however, tends to be biased toward the existing network, so unconventional solutions may not be examined. Furthermore, there is no guarantee that solutions near the optimum will be found.

In contrast, mathematical optimization using linear programming or general integer programming will produce an optimal network within the specified