TESTING THE SPECIFICATIONS OF RANDOM UTILITY MODELS

by
Joel L. Horowitz
Departments of Geography and Economics
The University of Iowa
Iowa City, IA 52242

ABSTRACT

Probabilistic choice models, such as logit and probit models, are highly sensitive to a variety of specification errors, including the use of incorrect functional forms for the systematic component of the utility function, incorrect specification of the probability distribution of the random component of the utility function, and incorrect specification of the choice set. Specification errors can cause large forecasting errors, so it is of considerable importance to have means of testing models for the presence of these errors. A number of tests based on the likelihood ratio statistic have been developed. These tests and available information on their power are summarized in this paper.

The likelihood ratio test can entail considerable computational difficulty, owing to the need to evaluate the likelihood function for both the null and alternative hypotheses. Substantial gains in computational efficiency can be achieved through the use of a test that requires evaluating the likelihood function only for the null hypothesis. A Lagrangian multiplier test that has this property is described, and numerical examples of its computational properties are given.

An important disadvantage of conventional specification tests is that they do not permit comparisons of models that belong to different parametric families in order to determine which model best explains the available data. Thus, these tests cannot be used to compare models whose utility functions have substantially different functional forms or models that are based on different behavioral paradigms. A practical method dealing with this problem is described.

TESTING THE SPECIFICATIONS OF RANDOM UTILITY MODELS

by

Joel L. Horowitz

1. Introduction

In recent years, considerable progress in the field of travel demand modeling has been made possible through the use of probabilistic models of discrete choice. The multinomial logit model is the simplest and best-known example of a probabilistic choice model (Domencich and McFadden, 1975; McFadden, 1974). Other examples are the multinomial probit model (Daganzo, 1979; Hausman and Wise, 1978; Daganzo, Bouthelier and Sheffi, 1977) and the generalized extreme value (GEV) model (McFadden, 1980). Compared to other formulations of travel demand models, probabilistic choice models have the important advantages of being based on a clearly defined (though not necessarily generally accepted) principle of human behavior, namely that of utility maximization; being able to treat a wider variety of travel and policy options than other operational modeling approaches can treat, and being able to make efficient use of data.

The translation of these qualitative advantages into quantitative models that explain or forecast travel behavior accurately requires, among other things, achieving correct specifications of the functional relations between the aspects of behavior that are being explained and the relevant explanatory variables. The axioms of probabilistic choice theory provide useful guidance for the development of functional specifications, but this guidance is by no means sufficient to prevent the occcurence of potentially serious specification errors. Specification errors in probabilistic choice models, as in other types of econometric models, can cause forecasting errors that are large enough to destroy a model's usefulness (Horowitz, 1981c; Williams and Ortuzar, 1979). Therefore, it is of considerable importance to have means available for testing models for the presence of these errors.

The purpose of this paper is to describe several specification tests for probabilistic choice models, summarize what is known about the power of these tests, and indicate areas where further research is needed. This information is presented here in summarized form. More detailed information is available in references that are cited in the text. Many of the tests that are described here apply only to the multinomial logit model. The logit model is based on more restrictive assumptions and, hence, is subject to a larger number of specification errors than other probabilistic choice models are. In addition, the logit model has been available for routine use for a longer time and has been the object of more research efforts than have other models.

The remainder of the paper is organized as follows. Section 2 summarizes the properties of probabilistic choice models that are relevant to understanding the causes of specification errors in these models and the structures of the specification tests that are discussed in the paper. Section 3 discusses informal specification tests that are based on examination of the signs, ratios and t-statistics of the estimated values of models' parameters. Likelihood-ratio specification tests for the multinomial logit model are discussed in Section 4. Section 5 describes some Lagrangian multiplier tests. The problem of comparing non-nested choice models is discussed in Section 6, and concluding comments are presented in Section 7.

2. Summary of Properties of Probabilistic Choice Models

In probabilistic choice models of travel demand, it is assumed that an individual's preferences among the available travel alternatives can be described by a utility function and that the individual selects the alternative with the greatest utility. The utility of an alternative is represented as the sum of two components: a systematic (or deterministic) component and a random component. The systematic component accounts for the effects of the average tastes of the population and the observable characteristics of the alternative and the individual. The random component accounts for the effects of unobserved characteristics of the individual and the alternative. The probabilistic choice model then forecasts the probability that an individual will choose a particular alternative (i.e., the probability that the utility of the particular alternative is greater than the utilities of all other alternatives) as a function of the observable characteristics of the individual and the available alternatives.

Mathematically, let individual q face a set $\underline{A}(q)$ of J mutually exclusive alternatives. Let U_{qj} be the utility of alternative j to this individual. It is assumed that U_{qj} can be expressed in the form:

$$U_{qj} = \overline{U} (\underline{X}_{qj}, \underline{\Theta}) + \varepsilon_{qj} (\underline{X}_{qj}), \tag{1}$$

where $\overline{U}$ is the systematic component of the utility function, $\underline{X}_{qj}$ is a k-dimensional vector of observed characteristics of individual q and alternative j, $\underline{\Theta}$ is a vector of constant parameters, and ε_{qj} is the random component of the utility function. The probability P_{qj} that individual q chooses alternative j is given by:

$$P_{qj} = \text{Prob}[U_{qj} > U_{qi}, \text{ for all } i \in \underline{A}(q), i \neq j]. \qquad (2)$$

Given equation (1), functional specification of a probabilistic choice model (that is, defining a specific functional form for P_{qj}) involves three basic steps: specifying the probability distribution of ε_{qj}, specifying the functional form of $\bar{U}$ and specifying the set of alternatives $\underline{A}(q)$ among which individual q can choose. After the functional specification of a model has been made, the values of the parameters $\underline{\Theta}$ and of any unknown parameters of the distribution of ε_{qj} are estimated statistically by fitting equation (2) to observations of choice. If an error is made in one or more of the functional specification steps (for example, if the probability distribution of ε_{qj} is specified incorrectly), then the fitted model usually will have an incorrect functional form and, depending on the nature of the specification error, may produce highly erroneous forecasts.

Probabilistic choice models usually are classified according to the probability distribution of ε_{qj}. In the multinomial logit model, ε_{qj} is assumed to be independently and identically distributed across individuals and alternatives with the following cumulative distribution function:

$$F(\varepsilon) = \exp[-\exp(-\varepsilon)] \qquad (3)$$

In the multinomial probit model, the row vector $\underline{\varepsilon}_q' = (\varepsilon_{q1}, \ldots, \varepsilon_{qJ})$ is assumed to have been drawn from a J-dimensional, multivariate normal distribution whose mean is zero and whose covariance matrix, $\underline{\underline{\Sigma}}(\underline{X}_{q1}, \underline{X}_{q2} \ldots \underline{X}_{qJ})$, may depend on the explanatory variables. The multinomial probit model is considerably more general than the multinomial logit model is. Unlike the logit model, the probit model permits the ε_{qj} to be correlated across alternatives, it permits the variances of the ε_{qj} to depend on j, and it permits random taste variations across individuals.*

In the GEV model, the joint cumulative distribution function of the components of $\underline{\varepsilon}_q'$ is:

$$F(\varepsilon_1, \ldots, \varepsilon_J) = \exp\left\{-G[\exp(-\varepsilon_1), \ldots, \exp(-\varepsilon_J)]\right\} \qquad (4)$$

where G is a non-negative, homogeneous-of-degree-one function that satisfies certain regularity conditions. These conditions have been given by McFadden (1980). Equation (4) implies that in the GEV model, the ε_{qj} may be correlated

*For the purposes of this paper it is assumed that the available data are cross-sectional and contain only one observation of choice per individual. Hence, the possibility of random variation of ε_{qj} in repeated choices by the same individual does not arise. A generalization of the probit model to include such variation has been described by Daganzo and Sheffi (1979). See, also, Heckman (1981a, 1981b) for discussions of the problem of modeling repeated choices by the same individual.

across alternatives. However, the ε_{qj} must have equal variances for all alternatives, and random taste variations are not permitted. Thus, the GEV model is more general than the logit model but less general than the probit model. The nested, or sequential, logit model is a special case of the GEV model (McFadden 1980).

In the multinomial logit model, P_{qj} is given by (McFadden, 1974):

$$P_{qj} = \exp \overline{U}(\underline{X}_{qj}, \underline{\theta}) / \sum_{i \in \underline{A}(q)} \exp \overline{U}(\underline{X}_{qi}, \underline{\theta}). \tag{5}$$

In the multinomial probit model, P_{qj} is given by:

$$P_{qj} = \int_{-\infty}^{\infty} d\xi_j \Big(\prod_{r \neq j} \int_{\infty}^{\overline{U}_{qjr} + \xi_j} d\xi_r \Big) \phi_J(\xi_1, \ldots, \xi_J; \underline{\underline{\Sigma}}) \tag{6}$$

where ϕ_J is the J-dimensional normal density function with mean vector zero and covariance matrix $\underline{\underline{\Sigma}}(\underline{X}_{q1}, \underline{X}_{q2}, \ldots, \underline{X}_{qJ})$, and $\overline{U}_{qjr}$ is defined by:

$$\overline{U}_{qjr} = \overline{U}_{qj} - \overline{U}_{qr} \tag{7}$$

In the GEV model, P_{qj} is given by (McFadden, 1980):

$$P_{qj} = \partial \log G(\overline{U}_{q1}, \ldots, \overline{U}_{qJ}) / \partial \overline{U}_{qj}. \tag{8}$$

Statistical estimation of the parameters of the foregoing models typically is carried out by the method of maximum likelihood. The log-likelihood function, L, is given by:

$$L = \sum_{q} \sum_{j \in \underline{A}(q)} g_{qj} \ln P_{qj} \tag{9}$$

where g_{qj} equals one if individual q chooses alternative j, and g_{qj} is zero otherwise

A probit model with $\underline{\underline{\Sigma}} = (\pi^2/6)\underline{\underline{I}}$ where $\underline{\underline{I}}$ is the identity matrix, will be called an "identity probit model" in this paper. All other probit models will be called "general probit models." An identity probit model is virtually equivalent to a logit model. The logit and identity probit estimates of the parameters $\underline{\Theta}$, the maximum values of the logit and identity probit log-likelihood functions, and the logit and identity probit choice probabilities usually are nearly equal. This near equivalence of identity probit and logit forms the basis of several specification tests for the logit model. These tests are discussed in Sections 4 and 5.

In the logit model, the ratio of the probabilities that two alternatives i and j are chosen by an individual q depends only on the attributes of i, j and q. The ratio is independent of the attributes of all other alternatives and, in particular, does not depend on whether other alternatives are included in or excluded from the choice set $\underline{A}(q)$ (McFadden, 1974). This property of the logit model is called "independence of irrelevant alternatives" (IIA) and is used in some of the logit specification tests that are described later in this paper.

3. Informal Specification Tests

The statistical estimation of the values of the parameters of probabilistic choice models typically includes examination of the signs, ratios and t-statistics of the estimated parameters. If the signs or ratios of these parameters are inconsistent with prior information or if important parameters fail to achieve satisfactory levels of statistical significance, then the utility function specification under consideration usually is deemed inadequate, and another specification is sought. Although it is generally recognized that these informal procedures can provide only rough indicators of the quality of models, they often are the only diagnostic procedures that are carried out during model estimation. Therefore, it is of interest to evaluate the ability of these procedures to identify erroneous models.

Numerical studies by Horowitz (1980 , 1981a) have indicated that the informal procedures do not constitute powerful tests for specification errors. In these studies it was found that logit models can yield asymptotically correct estimates of parameter ratios, signs that are correct both asymptotically and in typical-size samples, and good t-statistics with typical-size samples, even though the models are erroneously specified and give estimates of choice probabilities that are both statistically inconsistent and highly erroneous. The inappropriateness of using coefficient ratios as indicators of the validity of estimated models also has been noted by McFadden (1976).

Another informal test procedure consists of comparing observed aggregate choices by population subgroups (for example, the proportion of individuals within a given income range who choose a particular alternative) with the aggregate choices predicted by the model under consideration. Substantial differences between observed and predicted aggregate choices constitute evidence that the model is incorrectly specified. The statistical properties of this test procedure have not been investigated.

4. Likelihood-Ratio Tests for the Multinomial Logit Model

A large number of likelihood ratio tests for specification errors in logit models can be devised. Three such tests that currently available evidence suggests may be particularly useful are discussed here. Other likelihood ratio specification tests are discussed by McFadden, Tye and Train (1976), McFadden, Train and Tye (1977), and Horowitz (1981a). The tests are described in Parts a through c of this section. Illustrations of the power of the tests are given in Part d.

a. Test based on conditional choice (McFadden, Tye and Train, (1976); McFadden, Train and Tye (1977)):

The IIA property of the logit model implies that consistent estimates of the utility-function parameters can be obtained using estimation data sets in which individuals' choices are restricted to a subset of the full set of available travel alternatives. In the test based on conditional choice, estimates of the utility-function parameters obtained using the full choice set and estimates obtained using a subset of the full choice set (referred to hereafter as the restricted choice set) are tested for equality by means of a likelihood ratio test. Rejection of the hypothesis that the utility function parameters are the same with the two choice sets implies that the logit model under consideration is erroneous.

Specifically, let L_F and L_R, respectively, be the maximum values of the log-likelihood function in independent samples based on the full and restricted choice sets. Let L_{FR} be the maximum value of the log-likelihood function for the combined samples. Then, if the logit model under consideration is the correct specification, the quantity

$$LR = 2(L_F + L_R - L_{FR}) \tag{10a}$$

has asymptotically the chi-square distribution with K degrees of freedom, where K is the number of utility function parameters. It is necessary that the full (F) and restricted (R) choice samples be independent. Otherwise, LR does not have the chi-square distribution.

An important case of non-independent F and R samples occurs when the R sample consists of those observations in F in which an alternative in the restricted choice set is chosen. In this case, the test based on conditional choice can be carried out using the test statistic

$$S = (\hat{\underline{\theta}}_R - \hat{\underline{\theta}}_F)' (\hat{\underline{\underline{V}}}_R - \hat{\underline{\underline{V}}}_F)^{-1} (\hat{\underline{\theta}}_R - \hat{\underline{\theta}}_F), \tag{10b}$$

where $\hat{\underline{\theta}}_R$ and $\hat{\underline{\theta}}_F$, respectively, are column vectors of maximum-likelihood parameter estimates obtained using the R and F samples, and $\hat{\underline{\underline{V}}}_R$ and $\hat{\underline{\underline{V}}}_F$, respectively, are the estimated covariance matrices of $\hat{\underline{\theta}}_R$ and $\hat{\underline{\theta}}_F$. If the logit model under consideration is the correct specification, then S is asymptotically chi-square distributed with degrees of freedom equal to the rank of $\hat{\underline{\underline{V}}}_R - \hat{\underline{\underline{V}}}_F$ (Hausman and McFadden, 1980).

The test based on conditional choice is computationally straightforward, as it relies on standard logit estimation procedures. However, the test tends to give results that depend on which alternatives ~~are~~ included in the restricted choice set. Thus, it may be desirable to carry out the test several times, using a different restricted choice set each time.

The test based on conditional choice does not yield information on the causes of error in models it identifies as being erroneous. (An illustration of this is given in part d of this section.) Nor does it provide alternatives to models it finds to be erroneous.

b. <u>Test based on the universal logit model (McFadden, Tye and Train, 1976; McFadden, Train and Tye, 1977):</u>

The utility of an alternative normally is considered to depend only on attributes of that alternative and of the traveler. In the test based on the universal logit model, the utility of an alternative is allowed to depend on the attributes of other alternatives, as well. This generalized utility function is constructed so that it includes the utility function of the logit model being tested as a parametric special case. This special case is then tested against the generalized utility function by means of a likelihood ratio test. Rejection of the parametric special case implies that the logit model being tested is erroneous.

Mathematically, let $\bar{U}(\underline{X}_j, \underline{\theta})$ denote the systematic component of the alternative j utility function for the logit model being tested. Let $\bar{U}(\underline{X}_j, \underline{\theta}; \{\underline{X}_k; k \neq j\}, \underline{\Psi}_j)$ denote the generalized systematic component of the utility function for alternative j, where $\{\underline{X}_k; k \neq j\}$ denotes the attributes of all alternatives except j and $\underline{\Psi}_j$ is a vector of parameters associated with $\{\underline{X}_k; k \neq j\}$. Let $\underline{0}$ denote the null vector, and assume that

$$\bar{U}(\underline{X}_j, \underline{\theta}; \{\underline{X}_k; k \neq j\}, \underline{0}) = \bar{U}(\underline{X}_j, \underline{\theta}) \tag{11}$$

for all j. Given an estimation data set, let L_R be the maximum value of the log-likelihood function when $\underline{\Psi}_j = \underline{0}$ for all j, and let L_U be the maximum value of the log-likelihood function when $\underline{\Psi}_j$ is unconstrained. If the logit model being tested is specified correctly, then the quantity

$$LR = 2(L_U - L_R) \tag{12}$$

has asymptotically the chi-square distribution with as many degrees of freedom as there are components in the vectors $\underline{\Psi}_j$

The test based on the universal logit model can be carried out using standard logit estimation methods and does not require separate samples for the restricted and unrestricted estimations. However, in models with large numbers of alternatives or attributes, the vectors $\underline{\Psi}_j$ collectively will tend to have a very large number of components if the generalized utility function of each alternative includes all of the attributes of the other alternatives. This makes estimating the unconstrained model computationally cumbersome, and it may cause the test to have low power, owing to the large number of degrees of freedom that large numbers of $\underline{\Psi}_j$ components produce. These problems can be avoided by allowing only a small subset of the attributes $\{ \underline{X}_k;\ k \neq j \}$ to enter the generalized utility function for alternative j. Systematic procedures for selecting an appropriate subset have not been developed.

The universal logit test does not give information on the sources of error in logit models it identifies as erroneous. This point is discussed further in Part d of this section. The test does yield alternative models to erroneous logit models that it identifies, namely logit models with generalized utility functions. However, these models are not consistent with the standard axioms of utility theory, since the models permit the utility of an alternative to depend on attributes of other alternatives.

The test based on the universal logit model can be generalized easily to apply to non-logit models that are based on the utility-maximization paradigm. The generalization consists of using the desired non-logit choice probability functions for evaluating LR in equation (12). However, if the vectors $\underline{\Psi}_j$ have large numbers of components, then evaluating L_U for a non-logit model may present severe computational difficulties.

c. <u>Test against the probit model (Horowitz, 1981a; Hausman and Wise, 1978):</u>

Because the logit and identity probit models are virtually equivalent, an approximate specification test for the logit model can be obtained in the following way. Let $\bar{U}(\underline{X}_j, \underline{\Theta})$ be the systematic component of the utility function of the logit model that is under consideration. Estimate the parameters $\underline{\Theta}$ and $\underline{\underline{\Sigma}}$ of a general probit model with the same systematic utility function component, and test the hypothesis that $\underline{\underline{\Sigma}} = (\pi^2/6)\,\underline{\underline{I}}$. If the hypothesis is rejected, then this implies that logit is not a correct specification for the particular $\bar{U}$ function and data set under consideration.

Mathematically, let L_P and L_L, respectively be the maximum values of the general probit and logit log-likelihood functions. Then, if the hypothesis $\underline{\underline{\Sigma}} = (\pi^2/6)\, \underline{\underline{I}}$ is true, the quantity

$$LR = 2(L_P - L_L) \tag{13}$$

has approximately the chi-square distribution. The distribution is approximate (even asymptotically) because L_L is not exactly equal to the log-likelihood value that would be obtained from an identity probit model. The number of degrees of freedom in LR is equal to the number of independent components of $\underline{\underline{\Sigma}}$ that must be constrained to achieve identity probit. There are certain complexities involved in counting the independent components of $\underline{\underline{\Sigma}}$. These complexities, which are due to the presence of redundant parameters in $\underline{\underline{\Sigma}}$, are discussed by Albright, Lerman and Manski (1978).

The probit test has the considerable computational disadvantage of requiring estimation of the parameters of a general probit model. This makes the test infeasible for models with large numbers of parameters or alternatives. In addition, the probit test may lose power due to a large number of degrees of freedom if $\underline{\underline{\Sigma}}$ is a large matrix. This problem can be mitigated by constraining some of the components of $\underline{\underline{\Sigma}}$ in the general probit estimation, although systematic procedures for selecting the components that should be constrained have not been developed. The probit test provides information concerning sources of error in logit models it rejects if the assumption that the true model is probit can be accepted. The specification error in the logit model is then caused by erroneously constraining one or more components of $\underline{\underline{\Sigma}}$. However, if the assumption that probit is the true model cannot be accepted, then this interpretation of the results of the probit test is not reliable. The probit test provides alternatives to logit models it rejects. These alternatives, which are probit models, are consistent with utility maximization.

d. The power of the likelihood ratio tests:

Wald(1943) has shown that if the null hypothesis in a likelihood ratio test is false, then the asymptotic distribution of the test statistic is non-central chi-square with degrees of freedom equal to the number of parameters that are constrained by the null hypothesis. If LR_n denotes the value of the likelihood ratio test statistic for a sample of size n, then the non-central parameter Z_N for a sample of size N can be written as:

$$Z_N = N \lim_{n \to \infty} [E(LR_n)/n] \tag{14}$$

If the hypothesis being tested is true, then Z_N is zero, and the likelihood ratio test statistic has the usual central chi-square distribution. Thus, in general the asymptotic cumulative distribution function of LR_N can be written as:

$$\text{Prob}[LR_N \leq X] = \text{Prob}[\ \chi^2\ (Z_N,\ d) \leq X], \tag{15}$$

where $\chi^2\ (Z_N, d)$ denotes a non-central chi-square variate with non-central parameter Z_N and d degrees of freedom, d being the number of parameters that are constrained by the null hypothesis.

If it is assumed that the true choice model is known, then equation (15) can be used to estimate the power of a likelihood-ratio specification test to reject an erroneously specified logit model. Alternatively, the equation can be used to estimate the sample size needed to achieve a given rejection probability at a given significance level. The details of these procedures have been described by Horowitz (1980b, 1981a).

Horowitz (1980, 1981a) has used these procedures to estimate the ability of the previously described logit specification tests to reject erroneous logit models in a number of test cases. In each test case, there were three alternatives in the choice set and two explanatory variables in the true model. The sources of specification error in the test cases were as follows:

1. The true model was a probit model in which the random components of the utility function were correlated across alternatives.
2. The true model was a probit model in which the random components of the utility function had different variances in different alternatives.
3. The true model was a probit model with random taste variations. (Random taste variations cause the random utility components ε to be correlated across alternatives, to have different variances in different alternatives, and to have a covariance matrix that depends on the explanatory variables.)
4. The true model was a logit model, but one of the explanatory variables was omitted from the erroneously specified model.
5. The true model was a logit model, but only group-average values of one of the explanatory variables were available in the estimation data set. This case simulates the estimation of a disaggregate model with zonally averaged data.

The results obtained in examples of each of the test cases are summarized in Tables 1 and 2. (More detailed results are available in Horowitz (1980, 1981a), and the results of additional examples are presented in Horowitz (1980).) Table 1 shows the magnitudes of the errors in choice probabilities that were produced by each of the 5 erroneous logit models. The tabulated errors are the root-mean-square (RMS) absolute and fractional errors for 100 sets of values of the explanatory variables. These errors have been computed as large-sample limits and, therefore, are free of sampling errors. In test cases 3, 4 and 5 particularly, the errors in the choice probabilities are large enough to reduce substantially or destroy the practical value of a model.

Table 2 shows two indicators of the ability of the likelihood ratio logit specification tests to reject each of the erroneous test models. The first indicator is the sample size needed to achieve a probability of 0.95 that the erroneous model would be rejected at the 0.05 significance level. The second indicator is the probability of rejection at the 0.05 significance level if the sample contains 1000 observations. In the test based on conditional choice, the tabulated results are based on the restricted choice sets that yielded the most powerful tests. All of the results are based on the use of equations (14) and (15).

Referring to Table 2, it can be seen that the test based on the universal logit model and the test against a probit model both have high power in the cases that were tested. In contrast, the power of the test based on conditional choice is highly variable. Comparison of the results in Tables 1 and 2 suggests that the power of the various specification tests to reject erroneous logit models is not closely related to the magnitudes of the errors in choice probabilities that would result from accepting an erroneous model.*

In 4 of the 5 test cases, the most powerful test based on conditional choice was obtained by omitting alternative 3 from the restricted choice set. This indicates that the diagnostic information provided by the test based on conditional choice (consisting mainly of the restricted choice set that causes rejection of the model that is being tested) is likely to be the same for a variety of different specification errors. Hence, the results of the conditional choice test cannot be used to infer causes of specification errors.

Although they are not shown in Table 2, the values and statistical significance levels of the parameters $\underline{\Psi}_j$ in the universal logit test were found not to be related to the source of specification error, thereby indicating that this test also does not provide useful information on causes of specification errors (Horowitz, 1981a).

As noted earlier, the probit test gives reliable information on causes of specification error only if the true model is a probit model. Otherwise, the information on causes of specification error that the test provides (this information consists of the estimated values of the components of $\underline{\underline{\Sigma}}$) can be misleading. For example, zonal averaging of explanatory variables can create the appearance of random taste variations in a probit model, even

*However, when the true model is a probit model, the power of the test against a probit model does appear to increase as the sizes of the choice probability errors made by the erroneous logit model increase (Horowitz, 1980b).

if the true model is a fixed-taste logit model (Horowitz, 1981a). In such a case, uncritical acceptance of the diagnostic information obtained in a probit test would lead to the erroneous conclusion that random taste variations are important in explaining the behavior being modeled.

The foregoing results are based on a small number of test cases and, therefore, are exploratory and suggestive, not conclusive. In further research, it would be particularly useful to consider test cases in which there are more explanatory variables or alternatives than there are in the cases described here. This would provide a means of assessing the loss of power that may take place as a result of the increase in degrees of freedom that occurs as the numbers of explanatory variables and alternatives in the tested model increase. It would also be useful to determine whether there are systematic ways to constrain *a priori* the parameters $\underline{\Psi}$ and $\underline{\Sigma}$ in the universal logit and probit tests so as to minimize this loss of power.

5. Lagrangian Multiplier Tests

The likelihood ratio test of a logit model against a probit model suffers from the considerable disadvantage of requiring estimation of the values of the parameters of a general probit model. The difficulty of this estimation arises from the need to integrate the right-hand side of equation (6) numerically. Techniques for approximating the integral have been described by Daganzo (1979), Daganzo, Bouthelier and Sheffi (1977), Alhright, Lerman and Manski (1978) and Hausman and Wise (1978). Computational difficulties also can arise in the universal logit test and its generalization to non-logit models, owing to the potentially large numbers of parameters that must be estimated in maximizing the unconstrained log-likelihood function.

Silvey (1959) has described a Lagrangian multiplier test that enables hypotheses concerning the values of a model's parameters to be tested without carrying out maximization of the unconstrained log-likelihood function. This test often permits substantial computational savings in comparison to the likelihood ratio test. The general structure of the Lagrangian multiplier test is as follows. Let Y be a random variable with a probability density function $f(y, \underline{\theta})$ that depends on a k-dimensional vector of parameters $\underline{\theta}$ Let $\underline{\theta}_r$ be an r-dimensional vector (r<k) consisting of r components of $\underline{\theta}$. For specificity, assume that the components of $\underline{\theta}$ are ordered so that $\underline{\theta}_r$ consists of the first r components. Let $L(y, \underline{\theta})$ be the log-likelihood function of a sample y on the random variable Y. Suppose the hypothesis is made that $\underline{\theta}_r = \underline{\theta}_r^*$, where $\underline{\theta}_r^*$ is any r-dimensional vector. Denote this hypothesis by H_0. Then the components of $\underline{\theta}$ that are not constrained by H_0 can be estimated by solving the Lagrangian equations:

$$\partial L / \partial \theta_i + \lambda_i = 0; \; i=1,\ldots, r \tag{15a}$$

$$\partial L / \partial \theta_i = 0 \; ; \; i = r+1,\ldots,k \tag{15b}$$

$$\underline{\theta}_r = \underline{\theta}_r^* \tag{15c}$$

where θ_i is the i'th component of $\underline{\theta}$, and $\underline{\lambda}' = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ is a row vector of Lagrangian multipliers. If H_0 is true, then $\underline{\lambda}$ (the column vector corresponding to $\underline{\lambda}'$) will fluctuate around $\underline{0}$, depending on random variations in the sample of Y. Therefore, if $\underline{\lambda}$ is substantially different from $\underline{0}$, this can be taken as evidence that H_0 is false.

More precisely, define the matrix $\underline{\underline{\Omega}}$ by:

$$\Omega_{ij} = \int [\partial \ln f(y, \underline{\Theta})/\partial\theta_i][\partial \ln f(y, \underline{\Theta})/\partial\theta_j] f(y, \underline{\Theta})\, dy \quad (16)$$

where $\underline{\Theta}$ and the derivatives are evaluated at the solution to equations (15). Define the kxr matrix $\underline{\underline{R}}$ by:

$$R_{ij} = \begin{cases} 1, & \text{if } i = j \leq r \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Let N be the sample size. Then, if H_0 is true, the statistic

$$W = (1/N)\, \underline{\lambda}'\, \underline{\underline{R}}'\, \underline{\underline{\Omega}}^{-1}\, \underline{\underline{R}}\, \underline{\lambda} \quad (18)$$

has asymptotically the chi-square distribution with r degrees of freedom (Silvey, 1959). An important property of W is that it depends only on the estimated value of $\underline{\Theta}$ under H_0. Therefore, it is not necessary to estimate the value of $\underline{\Theta}$ without the H_0 constraints or to maximize the unconstrained log-likelihood function in order to evaluate W.

It is a straightforward matter to generalize the statistic W to permit testing of hypotheses of the form $\underline{h}(\underline{\Theta}) = \underline{0}$, where $\underline{h}' = (h_1, \ldots, h_r)$ is any r-dimensional, vector-valued function satisfying certain regularity conditions and r<k. See Silvey (1959) for a discussion of this generalization.

Horowitz (1981b) has derived expressions for $\underline{\lambda}$ and $\underline{\underline{\Omega}}$ for the Lagrangian multiplier test of a multinomial logit model against a multinomial probit model (subject to the approximation that logit and identity probit are equivalent). These expressions are rather lengthy and, in the interest of brevity, will not be presented here. They are available in Horowitz (1981b). Evaluation of the test statistic requires, at most, single-dimensional numerical integration and, in contrast to maximum likelihood estimation, is not iterative. Consequently, the computational effort required to perform a Lagrangian multiplier test of a logit model against a probit model is considerably less than that required to perform the corresponding likelihood ratio test. In numerical examples of the Lagrangian multiplier test Horowitz (1981b) found that W could be evaluated in 3.7 CPU seconds on an IBM 370/168 computer system, whereas evaluation of the corresponding likelihood ratio test statistic would have required over 60 CPU seconds.

The Lagrangian multiplier test procedure does not produce estimates of the values of the parameters of probit alternatives to logit models. Rather, it provides a computationally efficient procedure for testing logit specifications. If, as a result of these tests or other considerations, it is decided that a probit specification would be superior to a logit specification for the behavior being modeled, then it is necessary to undertake maximum likelihood estimation of the probit model's parameters.

Another possible use of the Lagrangian multiplier approach is to test for the presence of nonlinear-in-parameters functional forms in the utility functions of probabilistic choice models. For example, let X_1 and X_2 denote the explanatory variables of a two-variable logit model, and let θ_1, θ_2, α_1 and α_2 denote constant parameters. Suppose that it is desired to test the hypothesis H_0 that the correct specification of the systematic component of the utility function is:

$$\bar{U} = \theta_1 X_1 + \theta_2 X_2 \tag{19}$$

against the alternative H_1 that the correct specification is

$$\overline{U} = \theta_1 X_1^{\alpha_1} + \theta_2 X_2^{\alpha_2} \tag{20}$$

Since H_o is equivalent to the hypothesis $\alpha_1 = \alpha_2 = 1.0$, H_o can be tested with a conventional likelihood ratio test. However, this requires maximizing the logit log-likelihood function with the nonlinear-in-parameters utility function (20). The computations involved in carrying out this maximization can be both cumbersome and time-consuming.

The Lagrangian multiplier test enables H_o to be tested against H_1 without the need for maximizing the logit log-likelihood function with the utility function (20). To develop the Lagrangian multiplier test statistic, define the following notation. Let $\underline{\beta}'$ be the row vector (α_1, α_2, θ_1, θ_2), and let $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively, be the maximum likelihood estimates of the values of θ_1 and θ_2 under H_o. These values can be obtained using standard logit estimation methods. Let $\underline{\beta}_o'$ denote the row vector (1, 1, $\hat{\theta}_1$, $\hat{\theta}_2$). Let P_{qj} be the probability that individual q chooses alternative j (equation (5)) when the utility function is specified as in equation (20). Let $\hat{P}_{qj}$ denote the estimated value of P_{qk} under H_o. Finally, define g_{qj} as in equation (9). Then the matrix $\underline{\underline{R}}$ is:

$$\underline{\underline{R}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{21a}$$

The matrix $\underline{\Omega}$ and the vector of Lagrangian multipliers $\underline{\lambda}$ can be estimated consistently by:

$$\hat{\Omega}_{ij} = (1/N) \sum_q \sum_k (\partial P_{qk}/\partial \beta_i)(\partial P_{qk}/\partial \beta_j)/\hat{P}_{qk};\ i,\ j=1,\ldots,4 \tag{21b}$$

$$\hat{\lambda}_i = (-1/N) \sum_q \sum_k (g_{qk}/\hat{P}_{qk})(\partial P_{qk}/\partial \beta_i);\ i=1,\ 2 \tag{21c}$$

The derivatives are evaluated at $\underline{\beta} = \underline{\beta}_o$. The quantities $\hat{\Omega}_{ij}$ and $\hat{\lambda}_i$ and, hence, the test statistic W can be computed easily once the values of $\hat{\theta}_1$ and $\hat{\theta}_2$ are known.

To illustrate the properties of the Lagrangian multiplier test of H_0 (equation 19) against H_1 (equation 20), a 1000-observation data set was constructed by simulation from a logit model with the utility-function specification (20). The values of the θ and α parameters were θ_1=1.0, θ_2 = 0.2, α_1=0.6 and α_2 = 0.4. The value of the Lagrangian multiplier test statistic was computed using a FORTRAN program that was executed on an IBM 370/168 computer system. Evaluation of the test statistic required 4.4 CPU seconds, exclusive of the time used in logit estimation of the values of the θ parameters. The value of the test statistic for this particular example was W = 58.8. If H_0 is true, then this represents the value of a chi-square variate with 2 degrees of freedom. Hence H_0 is rejected at the 0.001 significance level.

Although the Lagrangian multiplier test of H_0 against H_1, has been illustrated here using a logit model, the test procedure and equations (21) apply to any probabilistic choice model whose choice probabilities can be expressed as functions of $\overline{U}$ in equation (20) and that satisfy certain regularity conditions. Similarly, the Lagrangian multiplier test can be used to test hypotheses concerning the values of the parameters in utility functions with any functional form satisfying the regularity conditions. (See Silvey (1959) for a discussion of regularity conditions.) Thus, for example, a third possible use of the Lagrangian multiplier test in specification testing is for carrying out the test against the universal logit model and the generalization of this test to non-logit models.

The Lagrangian multiplier test and the likelihood ratio test are asymptotically equivalent (Silvey, 1959). However, the two tests can give conflicting results in small samples (Savin, 1976).

6. Comparison of Non-Nested Models

The likelihood ratio and Lagrangian multiplier tests that are described in Sections 4 and 5 all consist of testing a choice model against a parametric generalization of itself. In other words, suppose that the model being tested has explanatory variables $\underline{X}$, parameters $\underline{\Theta}$ and choice probability $P_j(\underline{X}, \underline{\Theta})$ for alternative j. Then it is assumed that the alternative model also has the variables $\underline{X}$ and parameters $\underline{\Theta}$ as well as additional variables $\underline{Y}$ and parameters $\underline{\Psi}$. Let $Q_j(\underline{X}, \underline{Y}, \underline{\Theta}, \underline{\Psi})$ denote the probability that alternative j is chosen in the alternative model. It is assumed in that there exists a value of $\underline{\Psi}$, say $\underline{\Psi}_o$, such that

$$P_j(\underline{X}, \underline{\Theta}) = Q_j(\underline{X}, \underline{Y}, \underline{\Theta}, \underline{\Psi}_o) \tag{22}$$

for all $\underline{X}$, $\underline{\Theta}$, $\underline{Y}$ and j. (In the case of a test of a logit model against a probit model, this equality is approximate.) The hypothesis $\underline{\Psi} = \underline{\Psi}_o$ is then tested against the alternative hypothesis $\underline{\Psi} \neq \underline{\Psi}_o$.

It is easy to conceive of situations in which it would be useful to test a given model against an alternative model that is not a parametric generalization of the given model. For example, suppose that two logit models are under consideration for explaining a particular set of travel data. In one model the systematic component of the utility function is specified as:

$$\overline{U} = \theta_1 X_1 + \theta_2 X_2, \tag{23}$$

where the X's are explanatory variables and the θ's are parameters. In the other model, the systematic component of the utility function is specified as:

$$\overline{V} = \psi Y_1 Y_2, \tag{24}$$

where the Y's are explanatory variables that may be distinct from the X's, and ψ is a parameter. It is desired to test the two models against one another to determine which model best explains the available data. Clearly, there is no ψ value that causes $\tilde{U}$ and $\tilde{V}$ to coincide for all values of the X's, Y's and θ's. Similarly, there are no θ values that cause U and V to coincide for all values of the X's, Y's and ψ. Hence, equation (22) cannot be satisfied for the two models under consideration, and procedures that involve testing models against parametric generalizations of themselves cannot be used (at least directly) to compare the models of equations (23) and (24).

Pairs of models for which equation (22) cannot be satisfied are said to be non-nested. Statistical procedures for comparing non-nested models are discussed in detail in Horowitz (1983a, 1983b). In the remainder of this section the properties of one such procedure that is particularly useful in practice will be discussed. This procedure is the likelihood ratio index.

THE LIKELIHOOD RATIO INDEX

The most commonly used form of the likelihood ratio index is defined as follows. Let L denote the value of a model's log-likelihood function when the values of the model's parameters equal their maximum likelihood estimates. Let L_o denote the value of the log-likelihood function of a model that assigns equal values to the choice probabilities of all alternatives, regardless of the values of the explanatory variables. Then the likelihood ratio index, ρ^2, is defined as

$$\rho^2 = 1 - L/L_o. \qquad (25)$$

If there are N individuals in the estimation sample and each individual chooses among J alternatives, then L_o is given by

$$L_o = -N \log J. \tag{26}$$

(Throughout this discussion it will be assumed that all individuals face the same number of alternatives. Allowing different individuals to face different numbers of alternatives would add complexity to the presentation without changing the results significantly.)

The likelihood ratio index is a goodness-of-fit statistic for probabilistic choice models that is similar in many respects to the coefficient of multiple determination, R^2, in regression models. The larger the value of ρ^2 for a model, the better the model fits the given data. Therefore, two non-nested models P and Q can be compared by comparing the likelihood ratio indices ρ_P^2 and ρ_Q^2 for the two models. If $\rho_P^2 - \rho_Q^2 > 0$, this suggests that model P is superior to model Q, whereas $\rho_P^2 - \rho_Q^2 < 0$ suggests that model Q is superior.

To evaluate the ability of the likelihood ratio index to distinguish between correct and incorrect models, it is necessary to know the probability distribution of $\rho_P^2 - \rho_Q^2$. The following notation will be used in describing this probability distribution. Let P(i,X) denote the true probability that an individual chooses alternative i when the explanatory variables have the value X. (Here, X donotes the entire set of values of all of the explanatory variables of both models. If some elements of X are not variables of one of the models, then the choice probabilities of this model are independent of the values of these elements.) Let Q (i,X) denote the choice probability for alternative i that model Q would yield when the explanatory variables have the value X if there were no random sampling error. P(i,X) and Q(i,X), respectively, are the large-sample limits (i.e., the limits as the sample size approaches infinity) of the maximum likelihood estimates of the choice probabilities of models P and Q. Let $p_x(X)$ denote the proportion of individuals in the population being studied for whom the values of the explanatory variables equal X. Let kp and k_Q, respectively, denote the numbers of estimated parameters in models P and Q. Let N denote the number of individuals in the estimation data set, and let J denote the number of alternatives available to each individual. Finally define Δ^2 by

$$\Delta^2 = \sum_{i,\, X} \{ [P(i, X) - Q(i, X)]/P(i, X) \}^2 P(i, X) p_X(X). \qquad (27)$$

Δ^2 is the weighted mean square fractional error in the choice probabilities that would result from using the incorrect probabilities Q(i,X) in place of the correct probabilities P(i,X). The weight for given i and X values equals the proportion of the population that has explanatory variable values X and selects alternative i. Note that Δ^2 always exceeds zero unless Q(i,X) = P(i,X) for all i and X. In other words, Δ^2 exceeds zero unless model Q is identical to model P and, therefore, is correctly specified.

The probability distribution of $\rho_P^2 - \rho_Q^2$ is derived in Horowitz (1983a). It is shown there that $\rho_P^2 - \rho_Q^2$ has asymptotically the normal distribution with the following mean (μ) and variance (σ^2):

$$\mu = \Delta^2/2 \log J + (k_P - k_Q)/2N \log J \qquad (28)$$

$$\sigma^2 = \Delta^2/N(\log J)^2. \qquad (29)$$

It follows from equations (28) and (29) that $\rho_P^2 - \rho_Q^2$ exceeds zero, thereby indicating that the correct model is superior to the incorrect one, with the following probability:

$$\Pr(\rho_P^2 - \rho_Q^2 > 0) = \Phi\,[N^{1/2}\Delta/2 + (k_P - k_Q)/2N^{1/2}\Delta], \quad (30)$$

where Φ is the cumulative standard normal distribution function.

It is easy to see from equation (30) that the likelihood ratio index is consistent. In other words, as N approaches infinity, $\Pr(\rho_P^2 - \rho_Q^2 > 0)$ approaches 1. It also can be seen from equations (28) - (30) that with finite samples, adding parameters to an incorrect model (i.e., increasing k_Q) tends to decrease the value of $\rho_P^2 - \rho_Q^2$ and of $\Pr(\rho_P^2 - \rho_Q^2 > 0)$, even if the variables associated with the added parameters are incorrectly specified or irrelevant to the choices being studied. This clearly is an undesirable characteristic of the likelihood ratio index, since it means that in finite samples the index tends to favor models with large numbers of parameters, regardless of whether these models are correct. However, this characteristic can be removed by making a simple modification in the definition of the likelihood ratio index. Define $\bar{\rho}^2$, the modified likelihood ratio index, for a model with k estimated parameters by

$$\bar{\rho}^2 = \rho^2 - k/2N \log J \quad (31)$$

or, equivalently,

$$\bar{\rho}^2 = 1 - (L - k/2)/N \log J. \quad (32)$$

The modified statistic $\bar{\rho}^2$ is used in the same way as ρ^2 for comparing two models. Thus, $\bar{\rho}_P^2 - \bar{\rho}_Q^2 > 0$ indicates that model P is superior to model Q, and $\bar{\rho}_P^2 - \bar{\rho}_Q^2 < 0$ indicates that model Q is superior.

Equations (30) and (31) imply that the probability that $\bar{\rho}_P^2 - \bar{\rho}_Q^2$ exceeds zero is given by

$$\Pr(\bar{\rho}_P^2 - \bar{\rho}_Q^2 > 0) = \phi(N^{1/2}\Delta/2). \qquad (33)$$

It can be seen from equation (33) that $\bar{\rho}^2$ is consistent and, in contrast to the unmodified likelihood ratio index, is not biased in favor of models with large numbers of parameters. This makes $\bar{\rho}^2$ more useful than ρ^2 for comparing models. Accordingly, only the modified index $\bar{\rho}^2$ will be used in the remainder of this discussion.

The ability of $\bar{\rho}^2$ to distinguish between correct and incorrect models can be assessed by computing $\Pr(\bar{\rho}_P^2 - \bar{\rho}_Q^2 > 0)$ for various values of N and Δ. Table 3 shows the results of such a computation. It can be seen that if the sample size exceeds roughly 250, a comparison of two models using $\bar{\rho}^2$ has a probability of at least 0.80 of selecting the correct model when the RMS percentage difference between the two models' choice probabilities (i.e., $100\,\Delta$) exceeds 10 to 15 percent.

The probability distribution of $\bar{\rho}_P^2 - \bar{\rho}_Q^2$ can be used to derive a simple upper bound on the probability that $\bar{\rho}^2$ for an incorrect model (Q) exceeds $\bar{\rho}^2$ for a correct model (P) by an arbitrary amount z. The bound is (Horowitz, 1983a)

$$\Pr(\bar{\rho}_Q^2 - \bar{\rho}_P^2 > z) \leq \phi[-(2Nz \log J)^{1/2}]. \qquad (34)$$

This inequality implies that in moderate size samples, very small differences between the $\bar{\rho}^2$ values of two models indicate with high probability that the model with the lower $\bar{\rho}^2$ value is incorrect. For example, if $N \geq 250$, $z \geq 0.01$, and $J \geq 2$, inequality (15) yields

$$\Pr(\bar{\rho}_Q^2 - \bar{\rho}_P^2 > z) \leq 0.03. \qquad (35)$$

In other words, if $N \geq 250$ and the $\bar{\rho}^2$ values of two models differ by 0.01 or more, the model with the lower $\bar{\rho}^2$ value almost certainly is incorrect.

7. Conclusions

One important conclusion that is suggested by the foregoing discussion is that specification testing should become a standard part of the development of multinomial logit models of choice behavior. These models are subject to a large number of specification errors owing to the restrictive stochastic assumptions of logit, and the specification errors can cause large errors in forecasts of behavior. A relatively large group of specification tests for logit models is available. Although information on the power of these tests is extremely limited, it seems reasonable to speculate that many of the tests will be able to identify erroneous logit models often enough to be useful. In addition, the accumulation of practical, empirical experience with these tests is essential to the development of a better understanding of their value.

Another important conclusion is that there is a need for considerably more research in the area of specification testing of probabilistic choice models. Some research topics that may be especially useful include:

1. Estimating the power of the logit tests when there are more than two variables and three alternatives in the tested model. This should include attempting to develop ways of preventing loss of power due to degrees of freedom as the number of variables and alternatives in the tested model increase.

2. Developing methods for testing non-logit choice models. Here the main difficulty appears to be identifying suitable alternatives against which these models can be tested. For example, with the possible exception of the probit generalization of the universal logit model, it is not yet known

what classes of alternative models might provide useful (and tractable) specification tests for general probit models.

3. Developing methods for identifying specific causes of specification errors. The tests discussed here either give no information on specific causes or give information that is reliable only if certain *a priori* alternative hypotheses concerning the correct model are true. For example, the test of a logit model against a probit model gives reliable information on causes of logit specification error only on the hypothesis that the correct model is probit. It would be useful to have procedures for identifying causes of specification error that are less dependent on such hypotheses.

4. Investigation of the small sample properties of specification tests. All of the tests described here are based on the large-sample properties of the test statistics. Is is largely unknown whether serious errors are likely to be made by assuming that the large-sample properties apply to the sample sizes normally used in empirical choice modeling. (See McFadden (1974) for some limited results concerning the small sample properties of parameter estimates).

TABLE 1 - MAGNITUDES OF ERRORS IN LOGIT CHOICE PROBABILITIES DUE TO VARIOUS CAUSES

Source of Error	RMS Absolute Error	RMS Fractional Error When the Choice Probabilities are Less than 0.05	RMS Fractional Error When the Choice Probabilities Exceed 0.05
Correlated Random Utility Components	0.044	6,000,000	0.350
Random Utility Components with Unequal Variances	0.049	3,000,000	0.363
Random Taste Variations	0.135	5,000	0.882
Omitted Relevant Explanatory Variable	0.264	60	1.52
Zonally Averaged Explanatory Variable	0.262	64.9	1.52

TABLE 2 - ILLUSTRATIONS OF THE POWER OF LOGIT SPECIFICATION TESTS[a]

Source of Error	Conditional Choice Test		Universal Logit Test		Probit Test	
	Sample Size	Rejection Probability with Sample Size of 1000	Sample Size	Rejection Probability with Sample Size of 1000	Sample Size	Rejection Probability with Sample Size of 1000
Correlated Random Utility Components	1,100	0.92	1,300	0.86	400	1.0
Random Utility Components with Unequal Variances	1,200	0.90	920	0.97	380	1.0
Random Taste Variations	260	1.0	140	1.0	80	1.0
Omitted Relevant Explanatory Variable	6,500	0.25	990	0.95	970	0.95
Zonally Averaged Explanatory Variable	2,800	0.52	580	1.0	840	0.98

[a] "Sample size" is the size of data set needed to achieve a probability of 0.95 that the erroneous model would be rejected at the 0.05 significance level. The rejection probability with a sample size of 1000 is computed for the 0.05 significance level. In the test based on conditional choice, it is assumed that half of the data are used to estimate the model with the full choice set and half are used to estimate the restricted-choice-set model. The test is based on the likelihood-ratio statistic of equation (10a).

TABLE 3 - PROBABILITIES THAT THE MODIFIED LIKELIHOOD RATIO INDEX SELECTS THE CORRECT MODEL (P) IN A COMPARISON WITH AN INCORRECT MODEL (Q)[a]

N	Δ	$\Pr(\bar{\rho}_P^2 - \bar{\rho}_Q^2 > 0)$
100	0.05	0.60
	0.10	0.69
	0.15	0.77
	0.20	0.84
250	0.05	0.66
	0.10	0.79
	0.15	0.88
	0.20	0.94
500	0.05	0.71
	0.10	0.87
	0.15	0.95
	0.20	0.99

[a]N is the size of the estimation data set, Δ is the RMS difference between the large sample limiting values of the choice probabilities of models P and Q, and $\Pr(\bar{\rho}_P^2 - \bar{\rho}_Q^2 > 0)$ is the probability that the correct model is selected.

REFERENCES

Albright, R. L., Lerman, S. R., and Manski, C. F. (1978). Report on the development of an estimation program for the multinomial probit model, paper presented at the 57th annual meeting of the Transportation Research Board, Washington, D. C.

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

Daganzo, C. F., Bouthelier, F. and Sheffi, Y. (1977). Multinomial probit and qualitative choice: a computationally efficient algorithm, Transportation Science, Vol. 11, pp. 338-358.

Daganzo, C. F. and Sheffi, Y. (1980). Estimation of discrete choice models from panel data, paper presented at the 59th annual meeting of the Transportation Research Board, Washington, D. C.

Domencich, T. A. and McFadden, D. (1975). Urban Travel Demand, North Holland/ American Elsevier Publishing Company, New York.

Hausman, J. and McFadden, D. (1980). A specification test for the multinomial logit model, unpublished working paper, Department of Economics Massachusetts Institute of Technology, Cambridge, Massachusetts.

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

Heckman, J. (1981a). Statistical models for discrete panel data, forthcoming in Structural Analysis of Discrete Data (edited by C. Manski and D. McFadden), The MIT Press, Cambridge, Massachusetts.

Heckman, J. (1981b). The incidental parameters problem and the problem of initial conditions in estimating a discrete stochastic process and some Monte Carlo evidence on their practical importance, forthcoming in Structural Analysis of Discrete Data (edited by C. Manski and D. McFadden), the MIT Press, Cambridge Massachusetts.

Horowitz, J. L. (1981c). Sources of error and uncertainty in behavioral travel demand models, in New Horizons in Behavioral Travel Research, (edited by W. Brog, A. Meyburg, and P. Stopher), D. C. Heath and Company, Lexington, Ma.

Horowitz, J. (1980). The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand, Transportation Research, Vol. 14B, pp. 331-341.

Horowitz, J. (1981a). Identification and diagnosis of specification errors in the multinomial logit model, Transportation Research, 15B, 345-360.

Horowitz, J. (1981b). Testing the multinomial logit model against the multinomial probit model without estimating the probit parameters, Transportation Science, 15, 153-163.

Horowitz, J. L. (1983a). Statistical comparison of non-nested probabilistic discrete choice models, Transportation Science, forthcoming.

Horowitz, J. L. (1983b). An evaluation of the usefulness of two standard goodness-of-fit indicators for comparing non-nested random utility models, Transportation Research Record, forthcoming.

McFadden, D. (1981). Econometric models of probabilistic choice, in Structural Analysis of Discrete Data with Econometric Applications, (edited by C. Manski and D. McFadden), the MIT Press, Cambridge, Massachusetts, forthcoming.

McFadden, D. (1976). The theory and practice of disaggregate demand forecasting for various models of urban transportation, paper presented at the Seminar on Emerging Transportation Planning Methods, Daytona Beach, Florida.

McFadden, D. (1974). Analysis of qualitative choice behavior, in P. Zarembka, ed., Frontiers in Econometrics, Academic Press, New York, pp. 105-142.

McFadden, D., Train, K. and Tye, W. B. (1977). An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model, Transportation Research Record, No. 637, pp. 39-46.

McFadden, D., Tye, W. and Train, K. (1976). Diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model, Working Paper No. 7616, Institute of Transportation Studies, University of California, Berkeley, California.

Savin, N. E. (1976). Conflict among testing procedures in a linear regression model with autoregressive disturbances, Econometrica, Vol. 44, pp. 1303-1315.

Silvey, S. D. (1959). The lagrangian multiplier test, Annals of Mathematical Statistics, Vol. 30, pp. 389-407.

Wald, A. (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society, Vol. 54, pp. 426-487.

Williams, H. C. W. L. and Ortuzar, J. D. (1979), Behavioral travel theories, model specification and the response error problem, Working Paper No. 116, Institute for Transport Studies, University of Leeds, Leeds, England.