UCI-ITS-WP-78-12

Specification Issues in Choice Modeling

UCI-ITS-WP-78-12

Timothy J. Tardiff

Department of Civil Engineering
Division of Environmental Studies
University of California, Davis

and

Institute of Transportation Studies
University of California, Irvine

December 1978

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

ABSTRACT

This paper examines problems involved in the specification of the correct set of independent variables in choice models. The analytical approach is similar to Theil's use of auxiliary regressions in the case of standard linear models. The key conclusions are that the inclusion of superfluous independent variables does not affect the consistency of the correct coefficients, but exclusion of independent variables can lead to inconsistent estimates. The sources of bias are the possible correlations between included and excluded independent variables and the change in the structure of the random error terms in the utility functions. Because of the flexibility of its error structure, particular attention is given to the multinomial probit model.

When independent variables are excluded, asymptotic differences among alternative estimators arise because of different implicit error structures. The differences among the alternative estimators and the general effects of under-specification are examined empirically with simulated data.

1. Introduction

The development and application of statistical techniques for analyzing problems involving choice among a finite set of discrete alternatives has occurred in many problem areas (1). In most empirical applications, it is assumed that the model is correctly specified in terms of: a) the definition of the alternatives, b) the selection of the correct independent variables, c) the appropriate assumption about error terms, and d) correct measurement of the independent variables. In addition, it is usually assumed that the sample of observations is random.

Recently, there has been some attention devoted to the effects of violating the usual assumptions. Manski and Lerman (2), and Manski and McFadden (3) relax the assumption of random sampling and observe the consequences of using the usual estimating procedures in these cases. In addition, alternative estimators are developed which yield consistent estimates of the coefficients of the model. McFadden and Reid (4) discuss the effects of measurement error introduced by using variables which are averaged over a geographic zone to approximate variables which should be measured at the individual level. Guttman (5) discusses the consequences of incorrect independent variable selection on the estimates of values of time derived from transportation choice models. Finally, McFadden et al. (6) and Charles River Associates (7) discuss several violations of the assumptions involved in the multinomial logit specification. The violations included categories a) through d). Tests were devised which were designed to reveal violation of multinomial logit specification, and these tests were illustrated by the use of a transportation choice data set and a constructed data set.

It can also be noted that the common empirical practice of testing different sets of independent variables before deciding upon a final specification

is motivated by the recognition that correct specification of the independent variables is very important.

The purpose of this paper is to focus upon the specification issues involved in item b), the selection of the correct independent variables. Guttman dealt with this problem empirically in the special context of value of time calculations and McFadden et al. discussed this issue as a particular violation of the multinomial logit specification. Charles River Associates presented several examples of this type of misspecification and discussed its implications in a qualitative and conceptual manner. However, there has not been a systematic theoretical discussion of the nature of the bias introduced by an incorrect selection of independent variables.

Two major types of specification difficulties can arise: 1) the exclusion of independent variables from the model, and 2) the inclusion of independent variables which are not part of the choice process. Both difficulties are discussed, although the first problem is much more important.

A particular type of independent variable of special interest is alternative specific constants. These are variables which assume a value of one for a particular alternative and zero for all other alternatives. These variables are of particular interest because there have been alternative strategies to the inclusion or exclusion of such variables in particular empirical applications. Specification issues with respect to these variables have been discussed elsewhere (8). However, some of the general findings of this study also have implications for alternative specific constants.

Although there are some parallels between specification issues involving choice models and conventional linear models, there are important differences.

In addition to specification problems arising from correlations between included and excluded variables, changes in the random error structure of choice models also contributes to specification bias.

The general specification problem is developed in the second section. Next, specification analysis for the multinomial probit model is discussed. Like the situation of the use of disaggregate models with aggregate data (9), the multinomial probit specification seems to be quite robust to underspecification. The implications of the preceding analyses on alternative estimation procedures are discussed in the fourth section. Unlike the case of fully specified models, there are asymptotic differences among alternative estimators in the underspecification case. The fifth section presents some empirical results from simulation data and the final section summarizes the paper.

2. The General Case

The standard approach in which the utility of a given alternative is assumed to be a function of its own characteristics and a random error term is used (10). The analysis applies to fixed coefficient models such as the multinomial logit model and Bouthelier and Daganzo's (9) version of the multinomial probit model. The case of random coefficient choice models (11, 12) is not discussed.

In the discussion of possible sources of specification bias in this and the next sections, it is assumed that the coefficients of various utility functions are estimated with maximum likelihood techniques. Therefore, conclusions on the nature of bias apply in large samples, i.e., asymptotic bias is implied.

Since specification problems with independent variables involve the improper exclusion or inclusion of a set of variables, it is useful to partition

the set of independent variables into two groups. This results in the following expression for the utility of an alternative.

$$U_i = X_i^1\beta^1 + X_i^2\beta^2 + \varepsilon_i \qquad (1)$$

U_i is a scalar utility score for alternative i, X_i^j are $1 \times n_j$ vectors of independent variables, β^j are $n_j \times 1$ vectors of coefficients, and ε_i are the random error terms. The utility maximization assumption and the assumption of a distribution for the ε_i lead to a particular choice model.

The case of the inclusion of superfluous independent variables is straightforward. In Equation 1, if X^1 represents the correct set of variables and X^2 the superfluous variables, then the correct model implies that β^2 is a zero vector. The standard consistency arguments, e.g., Manski and Lerman (2), can be used to show that the maximum likelihood estimators of β^1 converge in probability to the true values and that the maximum likelihood estimators of β^2 converge in probability to zero.

The case of the exclusion of independent variables is analyzed using procedures similar to Theil's (13, 14) auxiliary regression approach. The key assumption is that the excluded variables may be correlated with the included variables. Specifically, define

$$X^* = (X_1^1\ X_2^1 - - - - X_m^1)$$

where m is the number of alternatives. The row vector X^* contains the included independent variables for all alternatives. Then the possible correlations between included and excluded variables can be represented by

$$X_i^2 = X^*\beta_i^3 + \varepsilon_i' \qquad (2)$$

where

$$\beta_i^3 = \begin{matrix} \beta_i^{11} & \beta_i^{12} & - & - & - & - & \beta_i^{1n_2} \\ \beta_i^{21} & \beta_i^{22} & - & - & - & - & \beta_i^{2n_2} \\ - & & & & & & \\ - & & & & & & \\ - & & & & & & \\ \beta_i^{m1} & \beta_i^{m2} & - & - & - & - & \beta_i^{mn_2} \end{matrix} \qquad (3)$$

Each β_i^{jk} is a $n_1 \times 1$ matrix. ε_i' is a random error term.

In words, Equation 2 states that each excluded variable for a given alternative is a linear combination of all the included variables for the alternatives. Examples of this pattern of correlation are given by McFadden, et al. (6) and Charles River Associates (7). If physical exertion is an excluded variable and this variable is correlated with travel time, a correlation between included and excluded variables for the same alternative arises. Similarly, if transit managers place the most comfortable buses on routes where the automobile has the best travel time and if comfort is an excluded variable, an example of correlation between the included variable of one alternative and the excluded variable of another emerges.

If Equation 2 is substituted into Equation 1, the following expression results

$$U_i = X_i^1\beta^1 + X^*\beta_i^3\beta^2 + (\varepsilon_i'\beta^2 + \varepsilon_i) \qquad (4)$$

It should be noted that the linear relationships between included and excluded variables represented in Equation 2 may contain constant terms. In this case, the model in Equation 4 would have alternative specific constant terms, even if the original model in Equation 1 did not. Therefore, in order

to correctly estimate the coefficients of the independent variables in Equation 4, alternative specific constants must be specified.

Equation 4 has two important features. First, it is possible that the utility of one alternative is a function of characteristics of the other alternatives, even though this is not the case for the fully specified model. The universal logit model proposed by McFadden (15) is an example of such a model. This model has been used to test for violation of the independence from irrelevant alternatives assumptions of the logit model (6, 7).

Second, the error structure of Equation 4 differs from that of Equation 1. Depending on the distribution of $\varepsilon_i'\beta^2$, this could result in models of the same general family or completely different families. For example, the common empirical strategy of sequentially including new variables in the same general model is consistent with the assumption that the $\varepsilon_i'\beta^2$ are multivariate normal when the probit model is used and are independently and identically distributed with the following distribution function when the logit model is used:

$$F(x) = \int_0^1 \exp(-(-\log z)^k e^{-kx})dz \qquad (5)$$

where k is the ratio of the standard deviation of ε_i' to the standard deviation of $\varepsilon_i'\beta^2 + \varepsilon_i$. (Equation 5 gives the distribution function for the difference of two extreme value variables with possibly different standard deviations.) An example in which Equations 1 and 4 lead to different families of choice models is when Equation 1 yields a binary logit model and $\varepsilon_i'\beta^2$ is independently and identically normally distributed. Equation 4 then yields selection probabilities from an S_B distribution (16).

Comparison of Equations 1 and 4 illustrates the versatility of the multinomial probit specification when it is assumed that there are important excluded variables. This model allows both differences in variances and correlations among the $\varepsilon_i'\beta^2$ terms. Since a given set of included variables may describe one alternative better than another, the possibility of heterogeneous variances is attractive. Further the normality assumption on the $\varepsilon_i'\beta^2$ is quite standard, whereas the distribution required to preserve the logit model, say, is unconventional.

The multinomial logit model requires that the $\varepsilon_i'\beta^2$ terms have equal variances. When this assumption is relaxed, the standard logit model no longer is appropriate. By generalizing McFadden's (10) derivation of the logit model, it can be shown that the following selection probabilities emerge from the assumption of independently distributed extreme value error terms with possibly unequal variances

$$P_i = \int_0^1 \exp\left(-\sum_{j\neq i} (-\log z)^{k_j} \exp(k_j(V_j - V_i))\right) dz \qquad (6)$$

where $V_j = X_j^1\beta^1 + X^*\beta_j^3\beta^2$ and k_j is the ratio of the standard deviation of $\varepsilon_i'\beta^2 + \varepsilon_i$ to the standard deviation of $\varepsilon_j'\beta^2 + \varepsilon_j$. If Equation 6 is evaluated with numerical integration techniques, a maximum likelihood estimation program for this model is possible. The program would yield estimates of the coefficients of the independent variables plus estimates of the variance terms, k_j.

Equations 1 and 4 compare the utility functions of the fully specified and underspecified models. The selection probabilities of the underspecified model are related to those of the fully specified model in the following way. Let $P_i(X^1, X^2, \beta^1, \beta^2)$ be the selection probability for the i^{th} alternative for the

fully specified model (X^1 and X^2 are matrices of characteristics for all alternatives ; therefore, the subscript has been dropped). Then the selection probability for the underspecified model is

$$\bar{P}_i(X^1, \beta_*^1) = E\left(P_i(X^1, X^2, \beta^1, \beta^2)\right)$$

where the expectation is over X^2.

One of the desirable outputs of a specification analysis is statements on the biases in coefficient estimates. Since the general case involves consideration of correlations of included and excluded variables as well as changes in the error structure of the underlying model, it is very difficult to develop general conclusions. However, the multinomial probit structure offers a sufficiently general class of models from which specific conclusions can be derived.

3. The Multinomial Probit Case

Since most applications of the probit model include only characteristics of the same alternative in the utility functions, Equation 4 will be simplified to reflect this assumption. This is done by setting the $\beta_i^{jk} = 0$ for $j \neq i$ in Equation 3. Alternatively, Equation 2 can be modified to

$$X_i^2 = X_i^1 \beta_i^4 + \varepsilon_i' \tag{7}$$

where β_i^4 is the i^{th} row of β_i^3.

Inserting Equation 7 into Equation 1 yields

$$\begin{aligned} U_i &= X_i^1 \beta^1 + X_i^1 \beta_i^4 \beta^2 + (\varepsilon_i' \beta^2 + \varepsilon_i) \\ &= X_i^1 (\beta^1 + \beta_i^4 \beta^2) + (\varepsilon_i' \beta^2 + \varepsilon_i) \end{aligned} \tag{8}$$

When both ε_i' and ε_i are multivariate normal, the multinomial probit model results from both Equations 1 and 8. However, the variance-covariance matrices of the two models are obviously different. This fact introduces a source of bias which does not occur in standard linear regression models and which does not appear to have been recognized in previous qualitative discussions of specification bias (7).

In the estimation of any choice model of the type in Equation 1, the magnitude of the coefficients are relative to the variances of the random error terms. Since all utilities can be multiplied by a constant and still convey the same information, it is necessary to arbitrarily fix one of the variances (11, 17), e.g., the variance of the first random error term. When this is done in Equation 1, the variance of the corresponding error term in Equation 8 is larger by a constant, k. Since the interpretation of the coefficients from alternative model specifications requires the same normalization procedure, Equation 8 should be multiplied by 1/k yielding

$$U_i^* = X^1(\beta^1 + \beta_i^4\beta^2)/k + (\varepsilon_i'\beta^2 + \varepsilon_i)/k \qquad (9)$$

where $U_i^* = U_i/k$.

Although the fully specified model assumes that that coefficients of the same variables are equal in different utility functions, i.e., the variables are assumed to be generic, Equation 9 allows coefficients to vary across alternative, i.e., the independent variables may be alternative specific. For example, if comfort is an excluded variable which is correlated with time, but the pattern of correlation differs for auto and transit, alternative specific time variables are appropriate.

Equation 9 shows that there are two sources of bias in the coefficients of the included variables, the $\beta_i^4\beta^2$ term and k. This is different from the linear regression case, where only the first source of bias is present. A simple example illustrates possible directions of bias. If both X_i^1 and X_i^2 are single variables (1 x 1 vectors in Equation 1) and β^1, β^2 are positive, the following are true. First, if X_i^2 is negatively correlated with X_i^1 ($\beta_i^4 < 0$), then the estimated coefficient of X_i^1 is biased downward from β^1. Both sources of bias contribute to the downward bias. If X_i^1 and X_i^2 are uncorrelated ($\beta_i^4 = 0$), the bias is still downward because of the k term. This is definitely different from the linear case where the exclusion of variables which are uncorrelated with the included variables does not bias the coefficients of the included variables. Finally, when the included and excluded variable are positively correlated, the direction of the bias is ambiguous. The $\beta_i^4\beta^2$ component contributes a positive amount and the k term contributes negatively. Again, this is different from the linear case and the qualitative interpretation given for the logit model by Charles River Associates (7), when the bias is positive.

The case of uncorrelated excluded variables is interesting. A consequence is that as model specification is improved by the addition of variables which are uncorrelated with the previously included variables, then coefficients of the original set of variables increase by a constant multiple. However, the ratio of previously included coefficients should not change significantly in large samples, thus resulting in similar estimates for the value of time, or similar measures.

The downward bias in the case of uncorrelated excluded variables is formally equivalent to the downward bias resulting from the estimation of dissaggregate models with aggregate data (4). In this case, the "aggregation

problem" (9, 16) results from aggregating the selection probabilities for people with the same included variables over the excluded variables.

The conclusions in this section depend heavily on the property that the error structure of the multinomial probit model can be parameterized in terms of the variance-covariance matrix of the error terms. Consequently, these results are difficult to generalize. However, it has been noted that the multinomial logit model yields results which are very similar to those of the probit model with independently and identically distributed error terms (11, 17). Since this probit model is a special case of the analyses in this section (the i.i.d. error structures apply to both Equations 1 and 8), it is reasonable to expect that the general findings on the sources and direction of bias also apply to the multinomial logit model. Simulation results to be discussed in a later section tend to confirm this expectation.

4. Implications for Alternative Estimators

The most common technique for estimating choice models has been maximum likelihood. In the binary probit and multinomial logit cases, it is also possible to use least squares to estimate the coefficients of the model when there are repeated observations for each combination of independent variables (4, 10, 18). Two types of repetitions are possible: repeated observations for the same individual and repeated observations over different individuals with the same characteristics. When the choice model is fully specified, i.e., the error component varies randomly over both individuals and for the same individual with repeated observations, maximum likelihood estimators and the two versions of least squares estimators converge in probability to the correct values. However, by comparing the results of the previous section with specification analyses for linear models, it can be shown that the least squares estimators

from observations repeated over the same individual do not converge to the same values as the other two estimators do when there are excluded independent variables.

The least squares procedures for the binary probit and logit models both involve linearization of the appropriate probability functions. The equation for the probit model is

$$\Phi^{-1}(P_{1i}) = (X_{1i} - X_{2i})\beta + \varepsilon_i \qquad (10)$$

where P_{1i} is the probability of selecting alternative 1 for case i, X_{1i} and X_{2i} are vectors of independent variables corresponding to the first and second alternatives for the ith individual, β is a vector of coefficients, ε_i is an error term, and Φ^{-1} is the inverse of the standard normal distribution function.

For the logit models, the situation is represented by

$$\log \frac{P_{ji}}{P_{ki}} = (X_{ji} - X_{ki})\beta + \varepsilon_i \qquad (11)$$

where the symbols are interpreted the same with the exception that the alternatives under consideration are j and k.

When independent variables are excluded, Equation 9 gives the correct expression for the utility function in the probit case. The similarity between multinomial logit and multinomial probit with independently and identically distributed error terms suggests that Equation 9 should also be quite good for the logit model. When repeated observations are taken for different individuals with the same values on the included independent variables, the error structure in Equation 9 is satisfied. Therefore, the usual arguments on the consistency of the least squares estimators apply (4, 10, 18) and these estimates converge in probability to the same values as the maximum likelihood estimators.

When repeated observations for the same individual are made, the error structure in Equation 9 is no longer appropriate (the error term from Equation 1 is the appropriate one). To analyze this case, the specification analysis of Theil (13, 14) is adapted. This approach assumes Equation 10 or 11 represents the correct model and an incorrectly specified model of the following form has been estimated

$$f(P) = X^{o}_{ji} b^{o}_{j} - X^{o}_{ki} b^{o}_{k} \tag{12}$$

where $f(P)$ denotes the estimate of the appropriate dependent variable from Equation 10 or 11, X^{o}_{ji} is the vector of the independent variables used in the estimation and b^{o}_{j} are vectors of estimated coefficients.

Note that Equation 12 allows separate vectors of estimated coefficients, b^{o}_{j}, corresponding to each alternative. This is the case even though the correct model has only a single vector, β. The motivation for this is the same as the motivation for alternative specific variables in the multinomial probit case.

In the case where $b^{o}_{j} = b^{o}_{k} = b^{o}$ an argument similar to Theil's shows that b^{o} converges to a linear function of β. That is

$$\text{plim } b^{o} = R\beta \tag{13}$$

where R is a matrix of the coefficients of the least squares regression of the correct independent variables $(X_{ji} - X_{ki})$ on the variables actually used $(X^{o}_{ji} - X^{o}_{ki})$. These regressions are called auxiliary regressions (14).

A generalization of Theil's procedure involves the use of separate sets of auxiliary regressions corresponding to each alternative, j. That is

$$R_{j} = (X^{o\prime}_{j} X^{o}_{j})^{-1} X^{o\prime}_{j} X_{j} \quad \text{for all } j \tag{14}$$

The X's are now matrices of variables for all individuals rather than vectors corresponding to separate individuals. This yields

$$X_j = X_j^o R_j \quad \text{for all } j \tag{15}$$

This generalization allows for the structures relating observed and correct explanatory variables to vary across alternatives.

Theil's procedure and the generalization of it assume that the independent variables are nonrandom. Therefore, the coefficients of the R matrices are also fixed values. Consequently, the actual values of these coefficients would usually vary across alternatives and for different samples. However, if there is actually a structural relationship between the observed and correct independent variables, the coefficients of the R matrices would be reasonably stable for different samples. Further, if the relationships were similar for the various alternatives the R_i matrices would be similar across alternatives. For large samples, the coefficients in these matrices would converge to the correct structural relationships.

Two special cases of particular interest are the situation in which the observed independent variables are correct, but there are left out variables and the case in which all of the correct variables are included in the estimation, but, in addition, superfluous variables are included.

The consequences of including superfluous independent variables is completely analogous to the usual situation with respect to more standard linear models. The inclusion of such variables in the estimation of the model will not affect the consistency of the coefficients of the correct independent variables. Furthermore, the coefficients of the extra variables will converge to zero for large samples.

Formally, this can be seen by noting the X_j matrix in Equation 14 can be partitioned such that the correct variables are represented by the first k columns and the superfluous variables are in the remaining (m-k) columns. With this partition, the R_j matrix would be

$$R_j = \begin{array}{l} \\ \\ \\ \\ \\ \text{row } k \rightarrow \\ \\ \text{row } m \rightarrow \end{array} \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}$$

A generalization of Equation 13 can be used to show that the b_j^o in Equation 12 satisfy the following

$$\text{plim} \quad b_j^o = R_j^o \beta$$

If b_j^o is partitioned commensurate with R_j, then the result that the first k elements of b_j^o are consistent estimates of β and the last m-k elements converge to zero follows.

In the case of excluded variables, 14 and 15 yield the following

$$X_j^\ell = X_j^{o\ell} \tag{16}$$

if the ℓth vector of X_j is included in X_j^o, i.e., if that variable is included in the estimation. Also

$$X_j^n = \sum_{\ell=1}^{m} a_{\ell j} X_j^{o\ell} \tag{17}$$

That is, each unobserved variable is a linear combination of the observed variables. The linear coefficients, $a_{\ell j}$, are obtained from the appropriate column of R_j.

The information in Equations 16 and 17 can be interpreted as imposing a partition on X_j. That is

$$X_j = \left[X_j^1 \quad X_j^2\right] \quad \text{and}$$

$$X_j^2 = X_j^1 R_j^2 \tag{18}$$

Using these relationships in the correct model $f(P) = (X_j - X_k)\beta + \varepsilon$ (the subscripts denoting individual observations have again been dropped) yields

$$f(P) = (X_j^1 - X_k^1)\beta^1 + (X_j^2 - X_k^2)\beta^2 + \varepsilon \tag{19}$$

where β has been partitioned commensurate with X. Equation 18 implies

$$\begin{aligned} f(P) &= (X_j^1 - X_k^1)\beta^1 + (X_j^1 R_j^2 - X_k^1 R_k^2)\beta^2 + \varepsilon \\ &= X_j^1 (\beta^1 + R_j^2 \beta^2) - X_k^1 (\beta^1 + R_k^2 \beta^2) + \varepsilon \end{aligned} \tag{20}$$

Comparison of Equations 9 and 20, shows that the least squares estimators based upon repeated observations for the same individual differ asymptotically by a factor of k from the estimators obtained from the other two procedures (R_j^2 is equivalent to β_j^4). This difference occurs because of different implicit error structures for the estimators. When repeated observations are taken for the same individual, there is no error variation over excluded variables as in the case of the other two estimators.

When the excluded variables are uncorrelated with the included variables, Equation 20 indicates that the estimators of β^1 converge in probability to their true values for least squares estimation with repeated observations over individuals while the estimators for the other two procedures are biased downward

by a factor of k. However, for purposes of estimating selection probabilities the later probabilities are appropriate, even though the corresponding coefficients are biased. This can be illustrated with the binary probit model. If $V_i = X_i^1\beta^1 + X_i^2\beta^2$, the selection probabilities from the least squares estimation with repetitions over the same individuals satisfy

$$P(1) = \Phi(E(V_1 - V_2))$$

where P(1) is the selection probability for the first alternative, Φ is the standard normal distribution function, and E denotes expectation. For the other two estimators, the selection probabilities satisfy

$$P(1) = E(\Phi(V_1 - V_2))$$

In both cases, the expectation is taken over the excluded variables for cases with identical $X_1^1 - X_2^1$. For the same reason as in the "aggregation problem" (4, 9, 16), the latter selection probability is the appropriate one.

When a data set includes repeated observations over the individuals in the sample, comparison of the alternative estimation procedures can serve as a test for the presence of excluded variables which do not vary over the repetitions. The more important the excluded variables are relative to the specified variables, the larger will be the difference in estimators. It should be noted that when the excluded effects do not vary for the same individual across repetitions, i.e., there is no purely random variation within individuals, the least squares estimation procedure with repetitions over the same individuals breaks down. Mathematically, the inverse functions in Equations 10, 11, 12 cannot be calculated. Conceptually, when there are no purely random effects, repeated observations of the same individual offers no new information.

For maximum likelihood estimators, the difference between repeated observations over the same individuals and repeated observations over different individuals with the same characteristics does not appear to lead to asymptotic differences in estimators. Although elaboration on this point is beyond the scope of the paper, standard proofs of consistency such as those given by Manski and Lerman (2) can be used to yield this result.

5. Simulated Examples

Some of the issues discussed in the previous two sections can be illustrated by a simple example using simulated data. The first independent variable, X, is discretely distributed over the interval (-2, 2) so that there are 101 unique values of X. For each value of X, 10 values of the second variable, Y, are randomly distributed with a standard normal distribution. Therefore, X and Y are essentially uncorrelated. For each combination of values on the independent variables, ten binary responses are simulated. The choice rule used in simulating the data is a binary probit model with linear function

$$.5X + Y$$

Since the binary probit choice rule is virtually indistinguishable from the binary logit rule with appropriate rescaling of the coefficients, estimation of binary logit models from the data is performed. Since the exact mathematical results of the previous sections assumed the probit specification, empirical tests with the logit model are useful in assessing the extent to which those results apply to the logit model.

Five sets of coefficients are estimated: 1) maximum likelihood estimators of the full model, 2) maximum likelihood estimators of the model with X and a constant, 3) least squares estimators of the full model, 4) least squares

estimators of the underspecified model in which repetitions are over individuals, and 5) least squares estimators of the underspecified model in which repetitions are over the identical values of the included independent variable. The first two estimators are based upon 10100 cases. For the next two least squares estimators, ten repetitions are taken over each of the 1010 combinations of values on the independent variables, resulting in 1010 cases. For the last set of estimators, 100 repetitions are taken over the 101 unique values of X, resulting in 101 cases. For the third and fourth estimators, when none or all of the choices are of a single alternative for a given data point, the value of .5 or 9.5, respectively, is assigned for the number of choices. This procedure was suggested by Berkson (19).

Table 1 lists the models resulting from the alternative estimation procedures. Qualitatively, the results are as expected. For the maximum likelihood estimators and the least squares estimators based upon repetitions over the same value of X, the magnitude of the coefficient of X is biased downward as expected. Further, the two alternative estimators yield very similar results. The least squares estimation based upon repeated observations over the same individuals yields a coefficient of X very similar to that of the full least squares model, as expected. It should be noted that the alternative estimators for the full model are somewhat different. This is probably the result of the assignment of values for the cases in which all repetitions resulted in the selection of the same alternative.

Since the maximum likelihood estimators do not require arbitrary assignment of values for cases in which the same alternative was always selected, it is informative to compare the two maximum likelihood models. The ratio of

the X coefficient of the fully specified model to the underspecified model is 1.47. For the binary probit model from which the data were simulated, it can be shown that the k term of Equation 9 equals $\sqrt{2} \simeq 1.41$. (If the variance of ε_i is one half in Equation 1 and $X_i^2\beta$ are independent and identically distributed normal variables with variance equal σ^2, then k in Equation 9 is $\sqrt{1+2\sigma^2}$. In the simulated data, Y is actually the difference in the values for the two alternatives; hence, the variance of Y corresponds to $2\sigma^2$. Therefore $k = \sqrt{1+1} = \sqrt{2}$.) Thus, the data suggest that the mathematical results for the probit model may be quite applicable to the logit model.

6. Summary and Conclusions

The major conclusion is that the exclusion of independent variables leads to two possible sources of specification bias in the coefficients of the independent variables: 1) the effects of correlations between included and excluded variables and 2) the effects of changes in the structure of the random error components in the utility functions. Because of its robustness in the specification of error terms, the multinomial probit model is especially versatile in handling underspecified models.

A related finding is that in the case of excluded independent variables, least squares estimators with repeated observations over the same individual do not converge to the same value as maximum likelihood estimators or least squares estimators with repeated observations over different individuals with the same observed characteristics. Even though the former type of estimators are consistent when the excluded variables are uncorrelated with the included variables, the other two types of estimators are the appropriate ones for estimating selection probabilities because they adjust for the "aggregation problem" introduced by excluded variables.

The discussion of the nature and sources of bias in coefficient estimators should be useful for understanding the specification issues involved in excluding appropriate independent variables. For practical reasons or because of a lack of theoretical development, underspecified models are estimated and used. The usefulness of such models depends heavily on the stability of the relationships between included and excluded variables. At one extreme, these relationships could be merely a statistical artifact of the given data set. In this case, no stability over different samples or over time can be expected. Consequently, such models would be of limited usefulness both in terms of explaining the relative importance of the included variables and predicting future situations.

On the other hand, the relationships between included and excluded variables may be real and stable over different samples and/or time. In this case, although the pure effects of the included variables are not revealed, the model would be useful for purposes of prediction. That is, the effects of excluded variables would be captured by the influence of these variables on the coefficients of the included variables. Further, the downward bias from the change in the error structure of the utility functions adjusts for the aggregation of selection probabilities over the excluded variables.

This paper has been primarily concerned with the effects of specification problems on the coefficients of choice models. Future research on the effects of specification problems on derived measures such as elasticities would be useful. It would be especially interesting to observe whether the downward bias of coefficients in the case of uncorrelated excluded variables is reflected in elasticity measures.

In conclusion, because of the source of bias from changing the underlying error structure of the utility function, specification analysis for choice models is not completely analogous to that for linear models. Further, there is an important conceptual similarity between the aggregation problem and specification problems which is inherent in the nonlinearity of most existing choice models.

REFERENCES

1. McFadden, D., "Quantal Choice Analysis: A Survey," Annals of Economic and Social Measurement, Vol. 5, No. 4, pp. 363-390, 1976.

2. Manski, C.F. and S.R. Lerman, "The Estimation of Choice Probabilities from Choice Based Samples," Econometrica, Vol. 45, No. 8, pp. 1977-1988, 1977.

3. Manski, C.F. and D. McFadden, "Alternative Estimators and Sample Designs for Discrete Choice Analysis," 1977 (unpublished).

4. McFadden, D. and F. Reid, "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models," Transportation Research Record, No. 534, pp. 24-37, 1975.

5. Guttman, J., "Avoiding Specification Errors in Estimating the Value of Time," Transportation, Vol. 4, No. 1, pp. 19-42, 1975.

6. McFadden, D., W. Tye, and K. Train, "Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," Urban Travel Demand Forecasting Project Working Paper No. 7616, University of California, Berkeley, 1976.

7. Charles River Associates, Disaggregate Travel Demand Models Project 8-13: Phase I Report, Vol. II, prepared for the National Cooperative Highway Research Program, 1976.

8. Tardiff, T.J., "The Use of Alternative Specific Constants in Choice Modeling," 1978 (unpublished).

9. Bouthelier, F. and C.F. Daganzo, "Aggregation with Multinomial Probit and Estimation of Disaggregate Models with Aggregate Data: A New Methodological Approach" Paper presented at the meeting of the Transportation Research Board, Washington, D.C., January, 1978.

10. McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka, P., ed., Frontiers in Econometrics, New York: Academic Press, 1974.

11. Hausman, J.A. and D.A. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences," Econometrica, Vol. 46, No. 2, pp. 403-426, 1978.

12. Cardell, N.S., R. Dobson, and F.C. Dunbar, "Consumer Research Implications of Random Coefficients Models," Charles River Associates, Cambridge, Massachusetts, 1977.

13. Theil, H., "Specification Errors and the Estimation of Economic Relationships," Review of the International Statistical Institute, Vol. 25, No. 1, pp. 41-51, 1957.

14. Theil, H., Principles of Econometrics, New York: Wiley, 1971.

15. McFadden D., "On Independence, Structure, and Simultaneity in Transportation Demand Analysis," Urban Travel Demand Forecasting Project Working Paper No. 7511, University of California, Berkeley, 1975.

16. Westin, R.B., "Predictions from Binary Choice Models," Journal of Econometrics, Vol. 2, No. 1, pp. 1-16, 1974.

17. Albright, R.L., S.R. Lerman, and C.F. Manski, Report on the Development of an Estimation Program for the Multinomial Probit Model prepared for the Federal Highway Administration, 1977.

18. Domencich, T.A. and D. McFadden, Urban Travel Demand: A Behavioral Analysis, Amsterdam: North Holland Publishing Company, 1975.

19. Berkson, J., "Maximum Likelihood and Minimum Chi-Square Estimates of the Logistic Function," Journal of the American Statistical Association, Vol. 50, No. 269, pp. 130-162, 1955.

TABLE 1: ALTERNATIVE MODELS FROM SIMULATED DATA

	Maximum Likelihood		Least Squares		
	Full Model	Underspecified Model	Full Model	Underspecified Model A	Underspecified Model B
X	.899	.611	.772	.783	.643
	(.0250)	(.0190)	(.0214)	(.0436)	(.0358)
Y	1.708		1.436		
	(.0378)		(.0255)		
Constant	-.0111	-.0205	.00822	-.0262	-.0246
	(.0255)	(.0211)			
R^2			.82	.24	.77
ρ^2	.33	.08			
N	10100.00	10100.00	1010.00	1010.00	101.00

ρ^2 is the likelihood ratio index defined by McFadden (10). Underspecified Model A is the least squares model with repetitions over the same individuals and underspecified Model B is the least squares model with repetitions over observations with the same value on X. Standard errors are in parentheses below the coefficients.