

UC Irvine

Working Paper Series

Title

The Use of Alternative Specific Constants in Choice Modeling

Permalink

<https://escholarship.org/uc/item/5j43q8mj>

Author

Tardiff, Timothy J.

Publication Date

1978-12-01

UCI-ITS-WP-78-11

The Use of Alternative Specific Constants in Choice Modeling

UCI-ITS-WP-78-11

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>

The Use of Alternative Specific Constants in Choice Modeling

by

Timothy J. Tardiff, Assistant Professor

University of California, Davis

ABSTRACT

A specification issue which has been handled differently in various empirical applications is whether or not to include alternative specific constants in models of choice behavior. Some applications have excluded constants, others have included a full set of constants, and a third class of examples uses unique constants for some alternatives, but not all.

In logit models in which each individual has the same set of alternatives, the exclusion of constants in the estimation of models when the correct model actually has alternative specific effects leads to inconsistent estimates of the coefficients of the remaining independent variables. However, the inclusion of constants when no such effects exist does not affect the consistency of the estimates of the coefficients. These results are illustrated by simple hypothetical examples and by empirical examples.

When nonratio scale variables are used in logit models, the coefficients of the independent variables are not invariant under arbitrary scale shifts when alternative specific constants are excluded.

Finally, the use of models to predict the response to new alternatives and the transferability of models which might or might not include alternative specific effects is discussed.

The major conclusion is that the inclusion of a full set of alternative specific constants in logit models estimated with large samples is generally preferred over the exclusion of one or more alternative specific constants.

THE USE OF ALTERNATIVE SPECIFIC CONSTANTS IN CHOICE MODELING

by

Timothy J. Tardiff, Assistant Professor

University of California, Davis

Introduction

The examination of transportation decisions at the individual level and the development of appropriate statistical techniques for these choice problems have been the subject of extensive research effort. The key requirements for this type of modeling are the selection of an appropriate set of independent variables upon which people are assumed to base their choices and an appropriate function relating these independent variables to choice probabilities. The former task is mainly a theoretical and empirical one, while the latter requirement also deals with statistical estimation considerations. In this regard, considerable progress has been made in the development of appropriate techniques for both binary and multinomial choice problems. In recent years binary and multinomial logit models (McFadden, 1973; Domencich and McFadden, 1975) have received the most empirical use in transportation application and therefore, the present discussion will focus on the logit approach.

Although logit analysis can be applied to problems in which there are repeated observations for each individual in the sample, only the case in which there is only one observation per individual will be considered. Here, the task becomes the estimation of the coefficients of the independent variables, usually by maximum likelihood methods, for the following probability function

$$P(i) = \frac{e^{x_i \theta}}{\sum_{j \in C} e^{x_j \theta}}$$

where $P(i)$ is the probability that the i^{th} alternative is selected, x_i is a vector of variables or functions of variables describing the i^{th} alternative and x_j is the corresponding vector for the j^{th} alternative, θ is the vector of coefficients, and C denotes the set of alternatives.

In general, the selection of the appropriate set of independent variables is a theoretical concern. It is obvious that the quality of the resulting model is highly dependent upon this selection process. Among the variables which might or might not be included in the vector of independent variables are alternative specific constants. These are variables which have the value 1 for one alternative and 0 for all others. Mathematically, it is possible to have as many as one fewer alternative specific constants than there are alternatives. Although it might appear that such constants do not have the theoretical or policy significance of variables which in some sense describe the alternatives, the purpose of the paper is to demonstrate that the use or nonuse of constants is very important in the proper estimation of the coefficients of logit models. In an earlier paper, it was shown that the use or nonuse of constants also has implications for the interpretation of some goodness of fit measures (Tardiff, 1976).

The specification issues related to the use of alternative specific constants are special cases of the general specification problem, i.e., the consequences of estimating a model which does not include all of the correct independent variables. The general case is also of interest and is discussed elsewhere (Tardiff, 1978). The special case developed in this paper is of particular interest because, unlike the general case, the researcher always has the option of whether to include constants. Data limitations are not a factor. Consequently, the findings with respect to constants are of immediate practical as well as theoretical usefulness.

An examination of the transportation literature indicates that there have been alternative strategies of using constants. First, it is possible to completely exclude constants. Models using this strategy include the studies of Hauser (1976), Recker and Stevens (1976), Liou and Talvitie (1974), Watson (1974), Recker and Golob (1976), and the shopping trip frequency model of Domencich and McFadden (1975). The usual justification of this strategy is that either the set of independent variables adequately describes each alternative or that the excluded effects are similar for all alternatives.

Another strategy is the inclusion of the full set of alternative specific constants. Examples include the models developed by Koppelman (1976), Train and McFadden (1975), and most binary mode choice models. This strategy is consistent with the assumption that the effects of excluded variables might be different for different alternatives.

Finally, it is possible to have a mixed strategy in which some alternatives share the same constant (Ben-Akiva and Richards, 1976; Recker and Kostyniuk, 1977). Here, it can be assumed that there are groups of alternatives for which the effects of excluded variables are similar within groups, but different across groups.

The decision to include less than a full set of constants appears to be based upon either a priori assumptions or on tests of statistical significance which result in the exclusion of insignificant constants. The analysis and empirical examples presented in this paper will indicate that both approaches can lead to inconsistency and bias problems and/or differences in interpretation of the coefficients of the model.

It is interesting to contrast the situation of logit analysis to that of regression analysis. In regression analysis, there is very seldom a question of whether or not a constant is used. In fact, the proper covariance properties of the residual with each independent variable requires the presence of

a constant (Rao and Miller, 1971). Further, many standard computer programs do not allow a decision on whether a constant will be used; it is automatically included. The burden of proof is on the modeler to show why a constant should not be used.

This analogy to regression analysis is appropriate because the exclusion of the alternative specific constants is similar to the effect of excluding the constant in regression analysis. That is both situations involve the estimation of a model which might not have the full set of correct independent variables. The fact that the present results parallel similar well known results regarding the more widely used linear models is consistent with the fact that many features of the two methods are analogous.

There has been some discussion of the desirability of including alternative specific constants in the transportation literature. (Stopher, 1974 and 1976; Ben-Akiva and Richards, 1976). Most of the discussion has been qualitative and has been motivated by considerations of full or perfect specification. While full or perfect specification is certainly a sufficient condition for excluding constants, it is not necessary. The analyses in the next section show that when the effects of excluded variables are the same for all alternatives, constants can be excluded.

Despite the similarities to linear models and the brief discussions of previous authors, it appears that the specification issues involving alternative specific constants are not widely recognized by many transportation researchers. The existence of alternative strategies for including or not including constants seems to confirm this observation. Further, it appears that models without constants are being developed for practical planning applications. For example, Spear (1977) reports that a set of multinomial logit modal choice models developed for the Twin Cities Metropolitan Council explicitly excluded constants. Therefore, there appears to be a definite need

for a thorough discussion of the role of alternative specific constants in logit models.

The remainder of this paper will develop and illustrate some conclusions involving the consequences of use or nonuse of constants on the estimates of the coefficients of the remaining independent variables. First, some general results will be stated on the inconsistency and asymptotic bias of models which a) exclude constants when the effects are, in fact, present, and b) include constants when there are no effects. These conclusions will be illustrated by use of a very simple hypothetical choice situation, and observation of some empirical results. Next, the effects of the use of independent variables which do not have ratio scale properties, e.g., a dichotomous variable describing occupation, on the estimate of the coefficients when constants are not used will be discussed. The possible consequences of the use of constants on the transferability of models and the use of models to predict the response to new alternatives will be mentioned. Finally, the implications of these results on research strategies will be discussed.

Consistency and Asymptotic Bias Results

The key mathematical results assume that each individual faces the same choice set. The results can be considered in two cases. The first case occurs when the correct model contains alternative specific constants which are not included in the hypothetical (or calibrated) model. Using analyses similar to those reported by Manski and Lerman (1976) it is possible to prove that, in general, the coefficients of the remaining independent variables estimated by maximizing the likelihood function are inconsistent. The coefficients will be consistent, however, if the alternative specific effects for the alternative without constants are equal for all alternatives in the

correct model.

The proof of this result is as follows. The logarithm of the likelihood function for the estimated model is

$$(2) \quad L^* = \frac{1}{N} \sum_{n=1}^N \sum_{i \in C} f_{in} \log \frac{e^{x_{in}\theta}}{\sum_{j \in C} e^{x_{jn}\theta}}$$

where $f_{in} = 1$ if alternative i is selected and 0 otherwise. As N approaches infinity, the likelihood function converges to

$$(3) \quad L^* = \int_X \sum_{i \in C} \log \left(\frac{e^{x_i\theta}}{\sum_{j \in C} e^{x_j\theta}} \right) \frac{e^{x_i\theta^* + \alpha_i^*}}{\sum_{j \in C} e^{x_j\theta^* + \alpha_j^*}} p(X) dX$$

Where θ^* and α_i^* are true coefficients and alternative specific constants, respectively. The analysis is essentially based upon the argument that the

sample mean of $\log \left(\frac{e^{x_i\theta}}{\sum_{j \in C} e^{x_j\theta}} \right)$ in Equation (2) converges to the population

mean in Equation (3). The integration is over the space defined by the variables in the vector of independent variables X . $p(X)$ is the density function for the vector X and the logit function evaluated at the true coefficients

is the conditional probability of observing the choice of alternative i given the vector X . Therefore, the product of $p(X)$ and this logit function is the joint density function.

If the maximum likelihood estimators are consistent, then the derivative of Equation 3) with respect to θ evaluated at θ^* will be the zero vector. But

$$\begin{aligned}
 \frac{\partial L^*}{\partial \theta} \Big|_{\theta^*} &= \int_X \sum_{i \in C} \frac{\partial}{\partial \theta} \left(\log \left(\frac{e^{x_i \theta}}{\sum_{j \in C} e^{x_j \theta}} \right) \right) \Big|_{\theta^*} \frac{e^{x_i \theta^* + \alpha_i^*}}{\sum_{j \in C} e^{x_j \theta^* + \alpha_j^*}} p(X) dX \\
 &= \int_X \sum_{i \in C} \frac{\partial}{\partial \theta} \left(\frac{e^{x_i \theta}}{\sum_{j \in C} e^{x_j \theta}} \right) \Big|_{\theta^*} \frac{\sum_{j \in C} e^{x_j \theta^*}}{e^{x_i \theta^*}} \frac{e^{x_i \theta^* + \alpha_i^*}}{\sum_{j \in C} e^{x_j \theta^* + \alpha_j^*}} p(X) dX \\
 &= \int_X \sum_{i \in C} \frac{\partial}{\partial \theta} \left(\frac{e^{x_i \theta}}{\sum_{j \in C} e^{x_j \theta}} \right) \Big|_{\theta^*} \frac{\sum_{j \in C} e^{x_j \theta^* + \alpha_j^*}}{\sum_{j \in C} e^{x_j \theta^* + \alpha_j^*}} p(X) dX \\
 (4) \quad &= \int_X \sum_{i \in C} W_i \frac{\partial}{\partial \theta} \left(\frac{e^{x_i \theta}}{\sum_{j \in C} e^{x_j \theta}} \right) \Big|_{\theta^*} p(X) dX
 \end{aligned}$$

where

$$W_i = \frac{\sum_{j \in C} e^{x_j \theta^* + \alpha_j^*}}{\sum_{j \in C} e^{x_j \theta^* + \alpha_j^*}}$$

Equation (4) is equivalent to the results of demonstrating that choice-based samples yield inconsistent estimates of the coefficients when the standard maximum likelihood procedure is used. The results of Manski and Lerman show that Equation (4) is zero when all W_i are equal to 1. In general, this will happen only when all α_j^* are equal, i.e., when all alternatives have the same alternative specific effects. When the W_i are not all equal to 1, Manski and Lerman show that Equation (4) is not equal to zero for almost every set of weights. This means that the logarithm of the likelihood function coverages to its maximum at a point other than the true values of the coefficients, thus proving inconsistency.

The nature of the inconsistency, i.e., the asymptotic bias, can be examined by observing the logarithm of the likelihood function and its first and second order derivatives (Manski and Lerman, 1976; McFadden and Manski, 1976). Maximizing L^* of Equation (2) with respect to the coefficients, θ , yields estimates for these coefficients, $\hat{\theta}_N$.

The deviation of $\hat{\theta}_N$ from θ^* , the correct coefficients, can be approximated by using a Taylor series expansion of the first derivative of the logarithm of the likelihood function with respect to θ :

$$\left(\frac{\partial L^*}{\partial \theta} \right)_{\hat{\theta}_N} \approx \left(\frac{\partial L^*}{\partial \theta} \right)_{\theta^*} + \left(\frac{\partial^2 L^*}{\partial \theta \partial \theta'} \right)_{\theta^*} (\hat{\theta}_N - \theta^*)$$

Since $\hat{\theta}_N$ maximizes L^* , the term on the left hand side is zero. By rearranging the remaining terms, the approximation of the asymptotic bias becomes

$$\hat{\theta}_N - \theta^* \approx \left(\frac{\partial^2 L^*}{\partial \theta \partial \theta'} \right)_{\theta^*}^{-1} \left(- \frac{\partial L^*}{\partial \theta} \right)_{\theta^*}$$

As N goes to infinity, the first term on the right hand side converges in probability to the inverse of the matrix of the expectations of the second derivatives evaluated at the true values of the coefficients of the independent variables, and the second term is the negative of the gradient vector, again evaluated at the true values. Since $\hat{\theta}_N$ is inconsistent, as N approaches infinity the gradient has a nonzero value at θ^* , leading to the asymptotic bias.

The logit model yields specific formulae for the two matrices on the right hand side (8). If x_{jn} is the vector of independent variables for respondent n on alternative j , P_{jn} the predicted probability from the logit model, without constants, i.e.,

$$P_{jn} = \frac{e^{x_{jn}\theta}}{\sum_{k \in C} e^{x_{kn}\theta}}$$

and f_{jn} equals 1 if alternative j is chosen and 0 otherwise, the matrix of second derivatives evaluated at θ^* is

$$\left(\frac{\partial^2 L^*}{\partial \theta \partial \theta'} \right)_{\theta^*} \equiv A_N = -\frac{1}{N} \sum_{n=1}^N \sum_{j \in C} (x_{jn} - \sum_{i \in C} x_{in} P_{in})' P_{jn} (x_{jn} - \sum_{i \in C} x_{in} P_{in})$$

and the vector of first derivatives

$$\left(\frac{\partial L^*}{\partial \theta} \right)_{\theta^*} \equiv B_N = \frac{1}{N} \sum_{n=1}^N \sum_{j \in C} (f_{jn} - P_{jn}) x_{jn}$$

where P_{jn} is evaluated at θ^* for all j . Therefore we can rewrite the asymptotic bias approximation as

$$(5) \quad \hat{\theta}_N - \theta^* \approx -A_N^{-1} B_N$$

As the sample size, N , goes to infinity, the above expression approximates the asymptotic bias.

In interpreting the expression for the asymptotic bias, it is necessary to realize that it is only an approximation, even when the sample size becomes indefinitely large. This is the case because, unlike the situation in which consistent estimators are obtained, there is no guarantee that the higher order terms in the Taylor series expansion are small compared to the first order term for large samples.

The approximation given in Equation (5) can be used as an indication of the nature of the asymptotic bias. For each coefficient, the bias approximation is a linear combination of the negative of the components of the gradient for all of the coefficients. Each of these gradient components is multiplied by the corresponding partial derivative of the gradient component of the coefficient of interest with respect to each coefficient and added. Consequently, the bias of a given coefficient depends on the magnitude of each component of the gradient and on the rate of change of the given component of the gradient with respect to all coefficients. Therefore, it is impossible to state a priori whether the bias will be in the direction towards or away from zero for any of the coefficients.

It turns out that this inconsistency result also applies to the case in which groups of alternatives are constrained to have the same constant within groups, but perhaps different constants across groups when the correct model has unique constants for alternatives. This occurs because the essential cause of inconsistency in both this and the previous case is the constraining of certain

constants to have values not equal to their true value. The proof of this case involves a slight modification of the proof for the previous case.

The next case to consider is one in which the correct model does not have alternative specific effects (or these effects are identical for each alternative) and the hypothesized model allows such effects. By using a proof very similar to McFadden's proof that only the constants are inconsistent for a multinomial logit model with a full set of alternative specific constants when unweighted maximum likelihood estimates are calibrated with choice-based samples (Manski and Lerman, 1976), it can be shown that the coefficients of the independent variables are consistent. The same result holds when the correct model has sets of identical alternative specific constants.

The proof of this result involves a modification of the previous proof. The expression for the expectation of the likelihood function is

$$(6) \quad L^* = \int_X \sum_{i \in C} \log \left(\frac{e^{x_i \theta + \alpha_i}}{\sum_{j \in C} e^{x_j \theta + \alpha_j}} \right) \frac{e^{x_i \theta^*}}{\sum_{i \in C} e^{x_i \theta^*}} p(X) dX$$

therefore

$$(7) \quad \frac{\partial L^*}{\partial \theta} \Big|_{\theta^*} = \int_X \sum_{i \in C} \frac{\partial}{\partial \theta} \left(\frac{e^{x_i \theta + \alpha_i}}{\sum_{j \in C} e^{x_j \theta + \alpha_j}} \right) \Big|_{\theta^*} \frac{\sum_{j \in C} e^{x_j \theta^* + \alpha_j}}{e^{x_i \theta^* + \alpha_i}} \frac{e^{x_i \theta^*}}{\sum_{j \in C} e^{x_j \theta^*}} p(X) dX$$

The maximum of Equation (6) is obtained when Equation (7) equals zero. When $\alpha_i = 0$ for all i , then

$$(8) \quad 0 = \frac{\partial L^*}{\partial \theta} \Big|_{\theta^*} = \int_X \sum_{i \in C} \frac{\partial}{\partial \theta} \left(\frac{e^{x_i \theta + \alpha_i}}{\sum_{j \in C} e^{x_j \theta + \alpha_j}} \right) \Big|_{\theta^*, \alpha_i = 0 \text{ for all } i} p(X) dX$$

Equation (8) follows from the fact

$$(9) \quad \sum_{i \in C} \frac{e^{x_i \theta + \alpha_i}}{\sum_{j \in C} e^{x_j \theta + \alpha_j}} = 1$$

Therefore,

$$\frac{\partial}{\partial \theta} \sum_{i \in C} \frac{e^{x_i \theta + \alpha_i}}{\sum_{j \in C} e^{x_j \theta + \alpha_j}} = \sum_{i \in C} \frac{\partial}{\partial \theta} \left(\frac{e^{x_i \theta + \alpha_i}}{\sum_{j \in C} e^{x_j \theta + \alpha_j}} \right) = \frac{\partial}{\partial \theta} (1) = 0$$

Equation (8) shows that the maximum likelihood estimators converge to θ^* for the coefficients of the independent variables and to zero for the coefficients of the alternative specific constants. A slight modification of this proof can show that the coefficients of the independent variables are consistent when the correct model has sets of identical alternative specific effects.

A more intuitive way to view this result is to think of the constants of the correct model as not being absent or constrained, but actually present with specific values. In this case, the values just happen to be zero, or perhaps equal for some groups of alternatives. Including the full set of constants in the hypothesized model is then simply an attempt at estimating the true value of the constants.

These results can be summarized quite simply. For sufficiently large samples (since consistency is a large sample property), excluding or constraining constants incorrectly leads to incorrect results. On the other hand, failure to constrain or exclude constants does not result in inconsistent estimates. In fact, the only effect would be the discovery that some constants were zero or equal for groups of alternatives. Therefore, just like in the regression analysis case, it seems that a full set of constants should be included unless there are strong a priori reasons for excluding them. That is, the burden of proof should be on excluding constants and the usual strategy should be the use of the full set.

Some Examples of the Use and Nonuse of Constants

The conclusions of the previous section can be illustrated by specific examples. First, consider an extremely simple hypothetical model in which there is one binary independent variable with 0, 1 values and a binary dependent variable. The observed data can be completely described by a 2 by 2 table. Let x be the independent variable and y the dependent variable. Then the data yield

		x = 0	x = 1
y = 0		n_{00}	n_{01}
y = 1		n_{10}	n_{11}

Consider two alternative binary logit models

$$P(1) = \frac{e^{a+bx}}{(1+e^{a+bx})} \quad \text{and}$$

$$P(1) = \frac{e^{Bx}}{(1+e^{Bx})}$$

That is, one model uses a constant and the other does not. In this case, the maximum likelihood estimates of the coefficients have a convenient closed form solution. For the first model,

$$a = \log \left(\frac{n_{10}}{n_{00}} \right) \quad \text{and} \quad b = \log \left(\frac{n_{11} \cdot n_{00}}{n_{01} \cdot n_{10}} \right) = \log \left(\frac{n_{11}}{n_{01}} \right) - a$$

The second model yields

$$B = \log \left(\frac{n_{11}}{n_{01}} \right) = b + a$$

For this simple example, the nature of the inconsistency is clear; the coefficients of the independent variable for the two models differ by a , the constant

of the first model. This effect can be further illustrated by supplying actual numbers. The categorical independent variable will be assumed to be sex with males = 0 and females = 1 and the dependent variable will be bus use (0 = no; 1 = yes).

In the first case suppose we have the following

	men	women
no	.45	.45
yes	.05	.05

The cell frequencies are converted to proportions under the assumption that the sample size is large enough so that these proportions reasonably approximate the population proportions.

Inspection of the table reveals that sex has no effect on transit usage, i.e., men and women are users in the same proportions. Calculating the logit model with a constant yields this result. That is,

$$b = 0 \quad \text{and} \quad a = -2.2$$

However, the use of the logit model without the constant would yield

$$B = -2.2$$

indicating an apparent influence of sex on transit usage.

The difficulty can go the other way, also. A second hypothetical test area yielded the following table

	men	women
no	.4	.25
yes	.1	.25

The proper conclusion in this case would be that sex does appear to influence transit usage, since women are more frequent transit users. Again, calibration

of the model with a constant leads to the proper conclusion since

$$b = 1.39 \quad \text{and} \quad a = -1.39$$

However, exclusion of the constant leads to $B = 0$.

It is also informative to look at some actual empirical results. In a previous study (Tardiff, 1975), a modified mode choice model was developed. The data were gathered in the Santa Monica-West Los Angeles area in late 1973. A probability sample of 223 respondents was selected. Because of the heavy automobile orientation of the area, there were not enough bus choosers for the work trip to estimate the usual mode choice model for the work trip. Similar data deficiencies existed with respect to nonwork mode choice. However, there was a variable indicating whether the respondent used the bus in the month preceding the interview. This variable was used as the dependent variable in a binary logit model.

The independent variables included variables describing the trip maker and variables describing the transportation systems available, i.e. bus and car. Since the dependent variable is not trip specific, the latter type of independent variables was also not trip specific. Rather, the variables indicated the general availability of the competing modes. They were distance from the respondent's home to the closest bus line (BDIS), in blocks and the ratio of cars to licensed drivers in the household (C/DL). The former variable is a general description of transit resources and the latter refers to automobile resources. The variables describing the tripmaker include the fairly standard demographic variables of sex (SEX), age (AGE), and socioeconomic status. The first of these is scaled with males as 1 and females as 2 and the second is measured on a seven point age scale (1 = 18-24; 2 = 25-34; 3 = 35-44; 4 = 45-54; 5 = 55-64; 6 = 65-74; 7 = 75 and older). The indicator of socioeconomic status is the

occupation of the household head (OCC) which is measured using Hollingshead's seven point occupation scale. The scale is designed so that as occupational status decreases, the scale value increases. After cases with missing values were eliminated, 217 cases were left for analysis.

Table 1 gives the results of models estimated with and without an alternative specific constant. The linear functions are such that as a variable with a positive coefficient increases, the probability of bus use increases. It appears that the constant in the first model is sufficiently large so that the estimates of the coefficients of the independent variables are affected when the constant is excluded. In the model with the constant both the automobile availability ratio and occupation are statistically significant at least at the .05 level and the sex variable has a positive sign. When the constant is excluded, the auto availability ratio grows larger and even more significant, the occupation variable is reduced enough to lose significance, and the sex variable changes signs, although it is insignificant in both cases. Therefore, it is possible that whether or not the constant is included can affect the interpretation of the magnitude and significance of particular coefficients.

It should be noted that the constant in the first model is not statistically significant at the .05 level with a two tail test. However, it is significant at the .10 level. Therefore the exclusion of constants which do not meet conventional levels of statistical significance can have important effects on the coefficients of the independent variables. In other words, statistical significance is not necessarily a reliable criterion in deciding whether or not constants should be included.

This situation illustrates the need for clearly distinguishing between statistical significance and model interpretation. Although the latter involves considerations of the former, there is also an element of subjective interpretation.

For example, although the constant is of only marginal significance in the first model and the likelihood ratio test comparing the first and second models is of very similar statistical significance, it is quite likely that most analysts would interpret the two models as being quite different.

It should be further noted that the overall goodness of fit of the two versions does not differ by very much. In effect, the independent variables partially absorb the effect of the constant. The implications of this is that small differences in overall goodness of fit should not be interpreted as indicating the constants have minimal effect. The consistency and asymptotic bias issue is separate from the overall goodness of fit consideration.

It is apparent that the direction of bias is towards zero for some coefficients and away from zero for others. This finding is consistent with the earlier theoretical discussion.

Finally, the results of the previous section show that if a model has a constant with estimated value close to zero, exclusion of this constant has minimal effects on the coefficients of the independent variables. This result has been observed empirically. However, since this result is totally expected and, hence, not very informative, it will not be described further.

The Use of Constants with Nonratio Scale Independent Variables

Variables whose scales can be shifted and still convey the same information pose problems when alternative specific constants are not used. The concern is for shifts in scale after the subtraction implicit in the logit formulation has taken place. That is, the problems arise for alternative specific variables, but not generic variables. The classic example of such variables are categoric independent variables which are entered as alternative specific variables. Examples of such variables which have appeared in transportation studies are sex, income

scale variables, occupation scale variables, and variables indicating trip purpose categories. Some categoric variables are simplifications of variables which could be measured on higher order scales, others are inherently categoric. The essential feature of this type of variable is that scale values can be shifted in certain ways with the information content of the variable remaining the same.

An example of such a categoric variable is the sex variable used in the empirical examples. The variable was coded with males as 1 and females as 2, but coding males as 0 and females as 1 would have conveyed the same information. Therefore, recalibrating the model with the rescaled variable should not effect the coefficients of the independent variables.

The use of alternative specific constants assures invariance of the coefficients of the independent variables under shifts in the variable scales. Only the values of the constants are changed. They, in effect, absorb the shift in scale. The result is not as benign when there is no constant to absorb the shift. An example of the change in coefficients can be obtained by referring to Table 1. Suppose the scale for the sex variable were shifted by the ratio of the constant in Equation (1) to the coefficient of the sex variable. This yields a value of $-.55$ for males and $.45$ for females. Recalibrating the model would result in a value of 0 for the constant and values for the coefficients of the independent variables the same as in Equation (1). Now if the old Equation (1) is recalibrated without the constant, Equation (2) results. However, if the same equation is recalibrated without a constant and with the shifted sex variable, none of the coefficients of the independent variables changes, since the constant is already equal to zero. Therefore, the completely arbitrary shifting of one categoric variable has led to different estimates for the coefficients of the independent variables.

Another example of recent importance is the use of alternative specific attitudinal variables. When a variable such as the satisfaction with bus waiting

time is used, it is almost always measured with some psychological scale such as a seven point scale. Such scales are definitely not ratio scales; they are at best interval scales. An example of the use of such variables in choice models is the study of Recker and Golob (1976). The fact that alternative specific constants were not used in most models reported in the study means that there would probably have been changes in the coefficients of the independent variables if the models were recalibrated with the attitudinal variables shifted.

In the general case, the result is mathematically similar to the inconsistency result reported earlier. This can be seen by viewing the model with the shifted variable(s) as the "correct" model and the nonshifted case as the hypothesized model. Relative to the hypothesized model, the correct model has alternative specific effects induced by the shift. Therefore, the coefficients of the hypothesized model will be "inconsistent" with respect to the shifted model. Similarly, the same treatment of "asymptotic bias" would apply.

Prediction of the Effects of New Alternatives and Transferability

Two ideal properties of the logit model are its potential usefulness in predicting the response to new alternatives and the potential transferability of such models over time and/or geographic region. Both capabilities depend on the correct specification of the model. The former issue has been examined thoroughly by McFadden, et al. (1976) in a general discussion of the independence from irrelevant alternatives property of the logit model.

Since the alternative specific constants capture the mean effects of unobserved variables, it has been argued that models to be used to predict new alternatives and/or transferred over time or place should not have constants (Hausman and Wise, 1976; Hauser, 1976), because of the difficulty of assigning constants to new alternatives or to new areas. In this regard, it is useful to distinguish

between ideal knowledge and current knowledge. In addition, model estimation for purposes of understanding choice behavior can be distinguished from model estimation for purposes of planning analysis.

There is no question that the ideal model both for theoretical and practical purposes would not contain constants. That is, if the choice theory with respect to the types of alternatives in question were sufficiently developed, there might not be any unobserved effects, therefore no constants. The reasons why a model without constants is also a practical ideal have already been mentioned.

It is useful to separate these ideal situations from the current situation in which knowledge is limited, i.e., separate the theoretical and empirical issues (Domencich and McFadden, 1975, p. 118). The results of this paper indicate that the exclusion of alternative specific constants can lead to inconsistent estimates of the coefficients. If this problem is sufficiently severe, then one might question the value of using such a model in new situations.

In developing better understanding of choice behavior, the alternative specific constants might be viewed as a diagnostic tool which indicates whether key effects have not been identified. That is, the existence of strong alternative specific effects is an indication that the hypothesized model describes response to some alternatives differently than others. Since, in theory, response should not depend on arbitrary labels for alternatives, the existence of such difference is an indication of the need for improved theory. That is, rather than assuming that the ideal situation of no alternative specific constants exists, researchers could use constants as a tool in developing models which actually approach the ideal.

The situation with respect to model application is somewhat different. Here the purpose of the application is relevant. That is, a model without constants may give sufficiently accurate predictions of the overall responses to alterna-

tives in either the new alternative or transfer case. Of course, such a model should be examined for its predictive abilities (Koppelman, 1976).

On the other hand, if the application deals with the effects of policies affecting specific variables, e.g., transportation policies which change the cost of modes, then the inconsistency problems might be important. For example, inconsistent estimates of particular coefficients can affect the elasticity estimates.

It should be noted that there have been applications of choice models in new situations where alternative specific constants have been used. For example, McFadden (1976) reports on the use of a model developed before BART was inaugurated to predict choice in the post-BART situation. It was assumed that the constants for the new rail modes were similar to those for the existing public transit modes. Therefore, the existence of constants may not present as large a difficulty as once believed in the application of choice models to new situations. If this is the case, then the case for the general use of constants is strengthened even further.

Summary and Conclusion

The effects of alternative specific constants on model estimation have been examined. In interpreting the results, three qualifying features should be kept in mind. First, all of the consistency and asymptotic bias results apply only to the logit model. Similar results have not been derived for other probability functions at this time. Therefore, conclusions should not be generalized beyond the logit case although these models have been used extensively in transportation applications.

Second, the results are large sample properties. As in other works referring to maximum likelihood models, nothing has been demonstrated with respect to

small samples. As a result, although the inclusion of constants when no such effect exists may be innocuous in large samples, such a strategy may affect the efficiency and bias of estimates in smaller samples. Therefore, when sample size is limited, judicious exclusion of certain constants may still be a preferred strategy.

Third, the mathematical proofs assumed that each individual faces the same choice set. Therefore, the case where individuals have different choice sets has not been covered rigorously. However, the intuitive arguments and examples seem to indicate that similar inconsistency and asymptotic bias results would hold for this case.

Most empirical destination choice models are examples of cases in which individuals are not assumed to face the same choice set. Further, the number of destinations in an urban area is usually too large to allow the use of unique constants for every destination (Ben-Akiva, 1974b). The usual practice has been to completely exclude constants (Domencich and McFadden, 1975) or only to use constants for special destinations such as the CBD (Ben-Akiva, 1974a). The fact that such models have generally yielded poorer statistical fits than mode choice models suggests that the lack of constants may cause serious bias problems. A possible strategy for including constants would be to classify destinations into a fairly small number of groups, based upon assumed similarities in the average values of unobserved attributes. Then each destination in a particular group would be assigned the same constant. Future research would probably indicate that a finer classification than CBD and non CBD destinations is more appropriate for most destination choice problems. Such a classification strategy actually parallels the natural classification implicitly present in mode choice modelling. For example, many different makes and

models of cars are all classified under the automobile mode. An approach similar to this was used in the grocery shopping destination choice model of Recker and Kostyniuk (1977), in which grocery stores were classified into five groups.

Even ignoring some of the issues discussed in this paper, the exclusion of constants might be viewed as a less than optimal situation. Exclusion of constants is consistent with the hypothesis that effects not captured by the excluded independent variables have either zero mean value or constant mean value for all alternatives. A more cautious research approach would not impose this assumption but rather test it through the use of constants. Therefore, the dictates of a more cautious research approach as well as the conclusions of this paper seem to indicate that there generally should be a definite preference for including a full set of alternative specific constants in choice models.

Acknowledgements

The empirical examples in this paper are based on data collected in a project supported by a National Science Foundation Grant for Improving Doctoral Dissertation Research in the Social Sciences. The author has benefited from discussions with Professor William Moss of the University of California, Davis.

References

1. Ben-Akiva, Moshe and Richards, Martin G. "A Disaggregate Multi-Modal Model for Work Trips in the Netherlands," Transportation Research Record, No. 569, 1976, pp. 107-123.
2. Ben-Akiva, M. "Structure of Passenger Travel Demand Models," Transportation Research Record, No. 526, 1974a, pp. 26-42.
3. Ben-Akiva, M. Author's Closure to Discussion of "Alternative Travel Behavior Structures," Transportation Research Record, No. 526, 1974b, pp. 59-64.
4. Domencich, Thomas A. and McFadden, Daniel. Urban Travel Demand: A Behavioral Analysis, Amsterdam, North Holland Publishing Company, 1975.
5. Hauser, John. R. "The Information Measure Provides a Mathematically Rigorous and Intuitively Appealing Predictive Test for Probabilistic Demand Models," Transportation Center Working Paper, Northwestern University, 1976.
6. Hausman, Jerry A. and Wise, David A. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependent and Heterogenous Preferences," Department of Economics Working Paper No. 1973, Massachusetts Institute of Technology, 1976.
7. Koppelman, Frank S. "Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models," Transportation Research Record, No. 592, 1976, pp. 17-23.
8. Lerman, Steven R. and Manski, Charles F. "Alternative Sampling Procedures for Calibrating Disaggregate Choice Models," Transportation Research Record, No. 592, 1976, pp. 24-28.

9. Liou, Peter S. and Talvitie, Antti P., "Disaggregate Access Mode and Station Choice Models for Rail Trips," Transportation Research Record, No. 526, 1974, pp. 42-52.
10. Manski, Charles F. and Lerman, Steven R. "The Estimation of Choice Probabilities from Choice Based Samples," to be published in Econometrica, 1976.
11. McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka, P., ed., Frontiers in Econometrics, New York, Academic Press, 1973.
12. McFadden, Daniel. "The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation," Urban Travel Demand Forecasting Project Working Paper No. 7623, University of California, Berkeley, June, 1976.
13. McFadden, Daniel and Manski, Charles. "Alternative Estimates and Sample Design for Discrete Choice Analysis," University of California, Berkeley, June, 1976.
14. Recker, Wilfred W. and Golob, Thomas F. "An Attitudinal Mode Choice Model," Transportation Research, Vol. 10, 1976, pp. 299-310.
15. Recker, W.W. and Kostyniuk, L.P. "Factors Influencing Destination Choice for the Urban Grocery Shopping Trip," to be published in Transportation, 1977.
16. Recker, Wilfred W. and Stevens, Richard F. "Attitudinal Models of Modal Choice: The Multinomial Case for Selected Non-Work Trips," Transportation, Vol. 5, 1976, pp. 355-375.
17. Rao, Potluri and Miller, Roger L. Applied Econometrics, Belmont: Wadsworth, 1971.

18. Spear, B.D. Applications of New Travel Demand Forecasting Techniques to Transportation Planning, U.S. Department of Transportation Federal Highway Administration, 1977.
19. Stopher, P.R. Discussion of paper, "Alternative Travel Behavior Structures," Transportation Record, No. 526, 1974, pp. 56-57.
20. Stopher, P.R. "Derivation of Values of Time from Travel Demand Models," Transportation Research Record, No. 587, 1976, pp. 12-18.
21. Tardiff, Timothy J. "Comparison of Effectiveness of Various Measures of Socioeconomic Status in Models of Transportation Behavior," Transportation Research Record, No. 534, 1975, pp. 1-9.
22. Tardiff, Timothy J. "A Note on Goodness of Fit Statistics for Probit and Logit Models," Transportation, Vol. 5, 1976, pp. 377-388.
23. Tardiff, Timothy J. Specification Issues in Choice Modeling," Department of Civil Engineering and Division of Environmental Studies, University of California, Davis, 1978.
24. Train, Kenneth and McFadden, Daniel. "The Measurement of Urban Travel Demand II," Urban Travel Demand Forecasting Project Working Paper No. 7517, University of California, Berkeley, 1975.
25. Watson, Peter L. "Comparison of the Model Structure and Predictive Power of Aggregate and Disaggregate Models of Intercity Mode Choice," Transportation Research Record, No. 527, 1974, pp. 59-65.

Table 1. Modified Mode Choice Model

(the standard errors for the logit coefficients are in parentheses)

	Equation 1	Equation 2
CONSTANT	-1.59 (.90)	
AGE	-.084 (.11)	-.15 (.10)
SEX	.35 (.38)	-.071 (.29)
OCC1	.24 ^a (.10)	.18 (.094)
C/DL	-1.43 ^b (.46)	-1.89 ^b (.38)
BDIS	.13 (.14)	.11 (.14)
N	217	217
L*(θ)	-99.41	-101.04
L*(0)	-150.41	-150.41
L*(constant)	-110.77	-110.77

^aLogit coefficient significant at $p < .05$ (two tail test)^bLogit coefficient significant at $p < .01$ (two tail test)

L* denotes the logarithm of the likelihood function evaluated at various points.