PART IV

ISSUES IN DEMAND MODELING AND FORECASTING

CHAPTER 1

THE INDEPENDENCE FROM IRRELEVANT ALTERNATIVES PROPERTY OF THE MULTINOMIAL LOGIT MODEL

Introduction and Background

The multinomial logit model discussed in Parts I and II is today the most widely used mathematical form for choice probabilities in behavioral travel demand analyses. The model is simple, computationally efficient, and permits a simple behavioral interpretation of its parameters. The MNL model also has the property that the ratio of probabilities of choosing any two alternatives (i and h) from the choice set (C) is independent of the attributes or the availability of a third alternative j , or, mathematically:

$$(1) \qquad \frac{\text{Prob}\,(i|C)}{\text{Prob}\,(k|C)} = \frac{e^{v(LOS^{i},SE)}}{e^{v(LOS^{k},SE)}} \quad .$$

This is termed the <u>independence from irrelevant alternatives</u> property. The IIA property has two advantageous implications. The first is that estimation of parameters of $v(LOS^i,SE)$ can be performed using a conditional choice in a small subset of the full choice set, with attendant savings in data collection costs. For example, if a subset D of the choice set C is selected, and a sample is drawn conditioned on choice from D , then the conditional probability of choosing $i \in D$ given a choice in D depends only on the attributes of alternatives in D and the model can be estimated without data on LOS^j for $j \notin D$. The second, converse implication is that a model estimated from choices conditioned on D can be used to forecast choice probabilities for an enlarged choice set C , provided the new alternative attributes LOS^j , $j \notin D$ are specified. We have already used both of these properties in Part II where (1) mode-choice models were estimated with fewer alternatives than actually existed in the choice set and (2) forecasts were made of post-BART mode patronage using a pre-BART model.

The problem with the IIA property is illustrated by a classical example, previously discussed in Part I, Chapter 1. A homogeneous population of consumers have the option of shopping in the central business district (CBD) or the suburban "East Mall;" fifty percent are observed to choose each. Then, a second suburban shopping area, "North Mall," is introduced, with attributes similar to those of East Mall, and such that travel times and costs for each individual are the same to each mall. Consequently, the shoppers previously choosing East Mall now split evenly between the two malls, and the CBD shoppers continue to go to the CBD. The result is fifty percent choice for the CBD, twenty-five percent for each mall. An MNL model fitted to the binary choice between the CBD and East Mall predicts, using the IIA property, that the relative odds of CBD to East Mall choice will remain one-to-one when North Mall is introduced. Because North Mall and East Mall have one-to-one relative odds, the MNL model incorrectly predicts a one-third share for each destination. Whereas North Mall draws solely from East Mall, the MNL model predicts incorrectly that it will draw equally from both the CBD and East Mall. The last effect is a general property of the MNL model: the cross-elasticities of choice probabilities with respect to an alternative attribute are equal. Further discussion of the consequences of this effect is given in Domencich and McFadden (1975) and Charles River Associates (1976).

While the example above is extreme, it does illustrate a tendency of the MNL model to over-predict the choice probabilities for alternatives that are perceived by travelers to be "similar."

In general, imposition of the MNL model upon situations where the IIA property does not hold will yield inconsistent parameter estimates and biased forecasts. The magnitude of these errors will depend on the degree to which the property is violated. In many applications, segmentation of heterogeneous population, more complete specification of the models to reduce the effects of unobserved attributes, and robustness properties of the MNL form may permit reduction of bias due to IIA failure to tolerable limits, thus permitting the MNL model to be used as a practical tool. In this chapter diagnostic statistical tests are developed for investigating the validity of the IIA property. An empirical application of these tests is also conducted.

Before describing the diagnostic tests for the violation of the IIA property, it is instructive to review some basic assumptions on which the behavioral MNL models are based and which cause the IIA property to appear. The examination of these assumptions will, in turn, suggest the diagnostic tests as a natural byproduct.

All individuals are assumed to be utility maximizers with utility functions $U(LOS^i,SE,ULOS^i,USE)$, where i indexes alternatives, and LOS^i and $ULOS^i$ are the observed and unobserved attributes of the alternative i , and SE and USE are the observed and unobserved attributes of the individual and choice environment. Further, for an individual facing a choice set C , define vectors $LOS = \langle LOS^i : i \in C\rangle$ and $ULOS = \langle ULOS^i : i \in C\rangle$. Thus, LOS and $ULOS$ contain all the attributes, observed and unobserved, of all the available alternatives without reference to alternatives they describe.

It may be recalled from Part I that the utility function can be conveniently separated into two parts: mean utility, and a deviation of individual utility from the mean utility; or,

$$(2) \qquad U(LOS^i,SE,ULOS^i,USE) = v(LOS^i,SE) + \varepsilon(LOS^i,SE,ULOS^i,USE) \quad .$$

The choice probabilities could then be expressed in terms of the distribution of the utility deviations ε , or

$$(3) \qquad Prob(i \mid C,LOS,SE) = Prob\ \{\varepsilon \mid v(LOS^i,SE) + \varepsilon^i > v(LOS^j,SE) + \varepsilon^j \text{ for } j \in C\ ,\ j \neq 1\} \quad .$$

Specific mathematical forms for choice probabilities can be derived by assuming distributions for the unobserved variables; in particular, the MNL model can be derived in this way. To lay the background for the diagnostic tests and relate them to the distributional assumptions of the MNL model, let us consider the following simple example for the utility function:

$$(4) \qquad U(LOS^i,ULOS^i,SE,USE) = \alpha \cdot z(los^i,se) + ULOS^i \quad ,$$

where $z(los^i,se)$ is a "variable," say travel cost divided by income, entering the utility function; α is the "importance" weight an individual attaches to the variable $z(los^i,se)$; and $ULOS^i$ summarizes the unobserved attributes of the alternative i . Let β denote the location (mean, or median) of the distribution of z ; then the "mean" utility is defined as

$$(5) \qquad v(LOS^i,SE) = \beta \cdot z(los^i,se) \quad ,$$

and the deviation of individual utility from mean utility is, from equation (2),

$$(6) \qquad \varepsilon^i = (\alpha - \beta) \cdot z(los^i,se) + ULOS^i \quad .$$

If the importance weights α do not vary across individuals for reasons of taste differences, so that $\alpha \equiv \beta$, and unobserved attributes $ULOS^i$ are distributed identically and independently across the alternatives with the extreme value (Weibull) distribution[1], then the choice probabilities have the MNL form:

$$(7) \qquad P(i \mid C,LOS,SE) = e^{\beta' z(LOS^i,SE)} \sum_{j\in C} e^{\beta' z(LOS^j,SE)} \quad .$$

This formula can also be interpreted to hold when β is a <u>vector</u> of "coefficients," and LOS^i and SE are the vectors of observed attributes of alternative i and the characteristics of an individual making the choice.

Thus, the assumptions underlying the MNL model include the absence of unobserved taste variations in the population, the absence of dependence between the unobserved attributes of different alternatives, and the extreme value distribution of the "errors," independent of observed attributes. More concretely, given the utility structure implied by equation (4) and the random effect implied by equation (6), the following conditions must hold to obtain the independently,

---

[1] $\text{Prob}\,[ULOS^i \le \varepsilon] = \exp\,[-e^{-\varepsilon}]$.

identically extreme value ε^i underlying the MNL model. The converse is also true; instances where these conditions do not hold constitute a violation of the IIA property.

1. <u>No variation in the importance weights</u> α. The presence of taste variations will tend to introduce dependencies in ε^i that are correlated with observed attributes of the alternatives, and that depend on the structure of observed attributes. Thus, the presence of taste variations will generally cause IIA to fail.

2. <u>Independence of the unobserved attributes</u> $ULOS^i$ <u>across alternatives</u>. The presence of aspects that are common to different alternatives, such as unmeasured parking convenience common to suburban shopping destinations, will introduce dependence in $ULOS^i$. If the degree of dependence is a function of the mean utilities for the alternatives, IIA may hold even in the presence of dependence. However, it is more likely that the degree of dependence will vary in a pattern unrelated to observed effects, or which is correlated with the structure of observed effects.

3. <u>Observed variables are measured without error, and the structural specification of representative (mean) utility is valid</u>. The presence of measurement errors may lead to a correlation of the factors (LOS^i, SE) and the errors ε^i. This will tend to violate the assumption underlying the MNL model that the distribution of ε^i does not depend on the observed variables. Even in cases where the MNL functional forth is valid, the presence of measurement error will tend to produce inconsistent parameter estimates when usual statistical methods are employed. Structural misspecification will have an effect similar to measurement error, in that the error distribution that results will tend to be correlated with observed variables.

4. <u>Attributes of alternatives are "exogenous," and are not determined as part of the choice process. The choice set is also exogenous to the choice process</u>. If the set of alternatives or their attributes are influenced by the choice process, then there will tend to be a correlation between observed variables and the unobserved variables influencing choice. For example, auto ownership and residential location may be influenced by unobserved tastes that predispose an individual to use transit. Then, transit usage will appear to be strongly responsive to auto ownership and level-of-transit-service, whereas the sensitivity of any individual with given tastes may be low. A similar phenomenon can occur if competition among suppliers of alternative choices influences unobserved

attributes. For example, if a transit agency tends to put the most comfortable buses in corridors where the auto has the greatest relative time advantage, then an unobserved comfort attribute will be correlated with observed relative travel.

5. Errors of aggregation are absent--there are no improperly defined aggregate alternatives, and heterogeneous market segments with differing tastes are distinguished. An aggregate or compound alternative is one representing a class of elemental alternative choices, such as a "transit" alternative when several types of transit choice are available. Consistent aggregation of alternatives in the MNL model requires that the mean utility of the aggregate alternative be related to the mean utilities of the components by the formula

$$e^{v(\text{composite})} = \sum_{\substack{\text{components} \\ i}} e^{v(LOS^{i},SE)} \quad .$$

Any other method of defining composite alternatives introduces some inconsistency. A failure to segment the population properly introduces taste variations that also introduce dependence in the ε^{i} .

6. The errors ε^{i} have the extreme value distribution. Even in the case of independently, identically distributed random variables ε^{i}, their distribution may vary from the extreme value form. Such choice probabilities may satisfy restrictions analogous to IIA[1], but have response surfaces that differ in shape from the MNL form. In practice, this is likely to be unimportant except at extreme probabilities, where the MNL tails may fail to approximate the true response surface accurately.

It is noted again that all the conditions for the IIA to hold are related to the random errors ε^{i} . It is, therefore, natural that the "residuals" from an MNL model provide a ready source for tests of whether the assumptions surrounding the random errors ε^{i} --which the residuals reflect--are upheld.

---

[1]A more general property than IIA that is still implausible in some applications is termed simple scalability or order-independence. A model is defined to be simply scalable if the choice probabilities can be written as generic functions of mean utilities of the alternatives. When ε^{i} have a joint distribution that does not depend on the attributes of the alternatives, then the choice probabilities will be simply scalable. This issue is discussed in further detail in McFadden (1975).

A second source for diagnostic tests is provided by the universal (or "mother") logit model (McFadden, 1975). This model arises as follows. Define $\ln P(i \mid C,LOS,SE) = v^i(C,LOS,SE)$ for any choice model. Then the MNL form is formally satisfied; that is,

$$(8) \qquad P(i \mid C,LOS,SE) = v^{i}(C,LOS,SE) \Big/ \sum_{j \in C} v^{j}(C,LOS,SE) \quad ,$$

with "mean utility" depending on the attributes of <u>all</u> alternatives. McFadden has also shown that $v^i(C,LOS,SE)$ can be approximated to any desired degree of accuracy by a linear function

$$(9) \qquad v^i(C, LOS, SE) = \sum_{\ell=1}^{L} \beta_\ell^i \cdot z_\ell^i(C,LOS,SE) \quad ,$$

where $z_\ell^i(C,LOS,SE)$ is a vector of appropriately chosen continuous known functions. Thus, the only difference between equation (7) and equations (8) and (9) is that in the latter the mean utility of the i^{th} alternative depends also on attributes of alternatives other than i.

This result establishes that any choice model can be formulated formally as a multinomial logit model. Any pattern of non-independence from irrelevant alternatives can be analyzed within the framework of the MNL model of equations (8) and (9). The only generalization of the conventional MNL model, equation (7), necessary to achieve this universality is the introduction of interactions between the attributes of different alternatives. To the extent that these interaction terms provide a significant contribution to "log likelihood" of the model (Part I), they imply a violation of the IIA property.

Yet a third source for the diagnostic tests can be provided by the IIA property itself. It may be recalled that, if the MNL model is the "true" model, then parameter estimates obtained using a subset D of a larger choice set C, $D \subseteq C$, should be approximately equal to the parameter estimates obtained with a larger choice set. Should this not prove to be the case, a violation of the IIA property is indicated.

The diagnostic tests that are about to be described can be modified for application to models constructed by choosing more general distributions for

$ULOS^i$, or for weights α in equation (4), which allow for dependence between alternatives or permit taste variations in the population. The tests of independence then correspond to testing whether the covariance terms of the distributions for ULOS and α are equal to zero. More general choice models have been proposed (Domencich and McFadden, 1975; McFadden, 1977) and are now in the developmental stage. They will be discussed in the next chapter.

Diagnostic Tests for the IIA Property

The diagnostic tests can be divided into two classes: (1) comparisons of the MNL forms with more general specifications[1] or with restrictions on choice sets, and (2) analysis of properties of residuals. Further, within each class tests may be conducted using a single cross-sectional single data set, or "before and after" data. These latter data may be for a situation where "nature" has provided a change in the choice set, much as introduction of a new mode, or may be for comparable populations facing dissimilar choice sets. These types of data are termed comparative data sets. It is noted that single data sets allow the testing of IIA alone. Tests using comparative data sets, while more powerful, imply joint tests of IIA and model transferability (Part IV, Chapter 5).

Diagnostic tests against general specifications and conditional choice sets are described first, followed by the residual tests.

Test against general specifications

This test is based on the universal logit model. Recall that any qualitative choice model with positive choice probabilities can be written in apparent MNL form:

$$(10) \qquad P(i \mid C,LOS,SE) = e^{v^{i}(C,LOS,SE)} \Big/ \sum_{j \in C} e^{v^{j}(C,LOS,SE)} ,$$

which is distinguished from the true MNL form only in that the "mean utility" $v^i(C,LOS,SE)$ now depends on attributes of alternatives other than i, in contrast to the traditional mean utility $v(LOS^i,SE)$ that depends only on the attributes of alternative i. Hence, it is possible to test the true MNL form having mean utilities $v(LOS^i,SE)$ against any apparent MNL model in which the mean utilities contain interactions between alternative attributes.

Specifically, suppose that under the null hypothesis, MNL is the true model; then the mean utilities have the form

[1] This includes the models without IIA to be discussed in the next chapter; these models are not discussed here.

$$(11) \qquad v(LOS^i,SE) = \beta \cdot z\,(LOS^i,SE) \quad .$$

An alternative mean utility $v^i\,(C, LOS, SE)$ can also be approximated in a linear-in-parameters form (McFadden, 1975). One specific example of such a model follows:

$$(12) \qquad v^i(C,LOS,SE) = \beta \cdot (LOS^i,SE) + \sum_{\substack{j \in C \\ j \neq i}} \gamma_{ij} \cdot z(LOS^j,SE) \quad .$$

The model (11) can be tested as a restriction on model (12). With the null hypothesis, $\gamma_{ij} = 0$. This can be tested using the likelihood ratio test (McFadden, 1973). When models (11) and (12) are estimated using the maximum likelihood method, the statistic

$$(13) \qquad X^2 = 2[\text{log likelihood of (12) - log likelihood of (11)}]$$

is asymptotically distributed chi-square with degrees of freedom equal to the number of parameter restrictions imposed by the null hypothesis. Note that the universal logit method is a test of the joint hypothesis of the MNL form and the specification of the mean utility in (11).

Test against the saturated model

Suppose, in the previous test, a sufficient number of interaction effects are introduced so that each choice probability in the likelihood function is effectively a parameter. Then, maximum likelihood estimation of this model will result in estimated probabilities that equal observed sample frequencies. This is called a saturated model. For test purposes, it is unnecessary to specify and estimate the parameters of this model; the only fact needed is that the maximum of log likelihood is given by

$$(14) \qquad L = \sum_{n=1}^{N} \sum_{j \in C} S_{jn} \log\,(S_{jn}/R_n) \quad ,$$

where $n = l,...,N$ indexes the sample; S_{jn} is the number of choices of j at sample point n ; and we take $0 \log 0 = 0$ by convention. When the number of repetitions becomes large, the statistic

(15) $X^2 = 2$[log likelihood, saturated model - log likelihood of (11)]

is asymptotically distributed chi-square with $N(J - 1) - K$ degrees of freedom, where C contains J alternatives and K is the number of parameters in (11).

In the absence of multiple repetitions $(R_n = 1)$, the saturated model has $L = 0$ and $X^2 = -2\cdot$[log likelihood of (11)] . In this case, X^2 does not have an asymptotic chi-square distribution.

Tests based on conditional choice

An implication of the IIA property is that the probability of choice of i from a choice set C , conditioned on a choice in a set D , equals the choice probability of i from choice set D . This implies that the mean utility $v(LOS^i, SE)$, estimated from a sample of choices from a set C , should coincide, asymptotically, with the mean utility estimated from a subsample or independent sample of choices conditioned on choice from a set D . The set D can be defined by systematically eliminating one or more alternatives, or by randomly eliminating alternatives. The latter procedure allows identification of all parameters, but may be less powerful against specific structural failures of IIA.

When the samples of choices from C and D are independent, a Chow-test of the IIA property is possible: let L_C and L_D be the maximum log likelihood levels attained for the samples of choice from C and D , respectively, and let $L_{C\cup D}$ be the maximum log likelihood from the pooled sample. Then, $X^2 = 2(L_C + L_D - L_{C\cup D})$ is asymptotically distributed chi-square with K degrees of freedom, where K is the number of parameters.

In the case that the conditional choice from D is observed for a subsample of the sample of choices from C , an intuitive statistic is $X^2 = 2[L_D(\theta_D) - L_D(\theta_C)]$, where L_D is the log likelihood of the conditional choice subsample, evaluated both at the full sample maximum likelihood estimator θ_C and at the conditional choice subsample estimator θ_D . Alternately, the test-statistic for the independent sample case can be calculated. Because of dependence, neither statistic has a simple asymptotic distribution. For this reason these statistics have to be used with qualifications, despite the fact that they often give illuminating results.

Transportation analysts are occasionally presented with comparative data sets for a population of individuals facing dissimilar choice sets, as, for example, in the case of the addition or deletion of a mode. In such a case, the Chow-test suggested earlier for independent conditional choice samples can be applied to the two data sets to provide a powerful test of independence. If the parameter estimates are obtained "externally," that is, from another area facing a dissimilar choice set, and the null hypothesis is that they are correct, then the statistic

$$X^2 = 2[\text{log likelihood, maximized - log likelihood at external parameters}]$$

is asymptotically chi-square with degrees of freedom equal to the number of parameters. When the sampling distribution of the external parameter estimates is recognized, the asymptotic theory is more complex, and test above is incorrect.

Again, it is to be noted that tests where comparative data sets are involved are joint tests of the MNL form and the transferability of model coefficients from one area or environment to another.

Residuals tests

There are three residuals tests for the validity of the MNL model. Before discussing these tests a brief description of residuals and their properties is given; a more complete account of residuals is given by McFadden (1973).

The residuals D_{jn} of an MNL model are defined by

$$D_{jn} = (S_{jn} - R_n P_{jn}) / \sqrt{R_n P_{jn}} \quad , \tag{16}$$

where $n = 1,\ldots,N$ indexes the individuals in the sample; $P_{jn} = P(j \mid C, LOS_n^j, S_n)$ for $j \in C$ is the estimated choice probability; R_n is the number of repetitions (possibly one) of sample point n ; and S_{jn} is the number of choices j .

To avoid statistical dependence in the above residuals, it is sometimes more convenient to work with the transformed residuals

$$(17) \qquad Y_{jn} = D_{jn} - D_{1n}\sqrt{P_{jn}}\,(1 - \sqrt{P_{1n}})\,/\,(1 - P_{1n}) \quad ,$$

where $1 \in C$ is a fixed alternative and $j \neq 1$. Suppose the $R_n \to \infty$. Under the hypothesis that the estimated MNL model is correct, the residuals D_{jn} have, asymptotically, zero mean, unit variance, and covariance $ED_{in}D_{jn} = -\sqrt{P_{in}P_{jn}}$. The residuals Y_{jn} are asymptotically independent, with zero mean and unit variance.

The descriptions of the residuals tests follow.

Means test. The mean of Y_{jn} for each alternative will be zero under the hypothesis that the MNL model is correct. Because the variance of these residuals is asymptotically one, the statistics $\frac{1}{N}\sum_{n=1}^{N} Y_{jn}$ are asymptotically standard normal.

It is also possible to use non-parametric methods to perform these tests; e.g., the Wilcoxon-sign test. This could avoid some questions of the appropriateness of the normality assumption in small samples where skewedness of the residuals is substantial. However, the normal approximation should be good for samples of the size normally encountered in transportation demand analysis.

Variance test. Under the null hypothesis, $X^2 = \sum_{n=1}^{N}\sum_{j \in C} D_{jn}^2$ asymptotically distributed chi-square with $N(J - 1) - K$ degrees of freedom, as the number of repetitions becomes large. However, when the number of repetitions is small, this statistic will in general substantially exceed the chi-square random variable above, and will be extremely sensitive to outlying observations.

Tests of association. For alternative j, rank the estimated probabilities P_{jn}, and associate with each P_{jn} the sign of the associated residual D_{jn}. Classify the ranked $P_{j(n)}$ into M cells, with an average probability $\overline{P}_{jm}$ in cell m. Then, a contingency table can be formed as follows:

	m = 1	m = 2		m = M
Positive Residual	S_1	S_2	...	S_m
Negative Residual	$N_1 - S_1$	$N_2 - S_2$		$N_M - S_M$

hwhere S_m and $N_m - S_m$ are the numbers of positive and negative residuals, respectively, observed in cell m . Under the null hypothesis, $E\left(\frac{S_m}{N_m}\right) = \overline{P}_{jm}$ asymptotically. A variety of tests of association and goodness-of-fit can be applied to this table to test the null hypothesis; for example, the goodness-of-fit test

$$X^2 = \sum_{m=1}^{M} \frac{(S_m - N_m\overline{P}_{jm})^2}{N_m\overline{P}_{jn}} \quad ,$$

has an asymptotic distribution bounded by chi-square distributions with $M - 1$ and $M - K - 1$ degrees of freedom, where K is the number of estimated parameters. These test statistics are not independent across alternatives.

All three residuals tests can be applied to choice models evaluated at "external" parameter values or to comparative data sets. The reduction in degrees of freedom due to estimation of the K parameters is eliminated in these cases.

Discussion of the use of IIA tests

In order to see whether the diagnostic tests can detect deviations from the MNL form, a simulation study was conducted where each of the six conditions, identified on the previous pages as being important for the IIA property to hold, were allowed to fail. The results of this simulation study show that the diagnostic tests are able to detect deviations of the "true" model from the MNL form. The most powerful of these tests appear to be the test against the "universal" logit model and the tests based on conditional choice. The simulation study is reported in an appendix to this chapter.

An Empirical Application of the Diagnostic Tests

Table 36 presents an MNL model of the choice of mode for the work-trip. The model has the form of equation (7) with the independent variables in the table being the elements of (LOS^{i}, SE) and the estimated coefficients the elements of β. Estimation was performed by the maximum likelihood method described in Part I on a sample of 641 workers in the San Francisco-Oakland Bay Area.

Seven alternative modes are considered in the model: auto-alone, shared-ride, bus-with-walk-access, bus-with-car-access, BART-with-walk-access, BART-with-bus-access, and BART-with-auto-access. This model is a close variant of the models discussed in detail in Part II; the reader is referred there to refresh his memory about the reasoning underlying the model.

The model of Table 36 seems particularly well-suited to testing violations of the IIA property. Because some of the alternative modes are similar, unobserved attributes of each mode are perhaps correlated across modes. For instance, the comfort of on-vehicle travel is similar for bus-with-walk-access and bus-with-auto-access, and yet no comfort variable is included in the model. Failure of the IIA could also result from some attributes of the alternatives not being exogenous. If the choice of how many autos to own is related to the work-trip mode choice, then the autos per driver variables are endogenous. Furthermore, if the choice of where to live is related to the work-trip mode choice, then the cost and time variables are endogenous. Whether these problems are severe enough to reject IIA is explored by applying the diagnostic tests developed in the previous section.

The first test employs the universal logit method. A model is specified that includes all the variables in Table 36 plus some variables defined such that the attributes of an alternative are allowed to enter the mean utility of another alternative. The hypothesis that the coefficients of all the extra variables are zero is tested. If the hypothesis of zero coefficients is rejected, then the joint hypothesis of the MNL form and the specification of Table 36 is rejected.

TABLE 36 Work-Trip Mode-Choice Model

Mode 1: Auto-Alone	Mode 6: BART, Auto Access
Mode 2: Bus, Walk Access	Mode 7: Carpool
Mode 3: Bus, Auto Access	Mode 4: BART, Walk Access
Mode 4: BART, Walk Access	Model: Multinomial Logit, Fitted by
Mode 5: BART, Bus Access	By the Maximum Likelihood Method

Explanatory Variables	Estimated Coefficients	t-Statistics
Cost divided by post-tax wage, in cents/(cents per minute)	-.0380	6.83
On-vehicle time, in minutes	-.0162	1.91
Walk time, in minutes a/	-.1006	4.25
Transfer wait time, in minutes a/	-.0122	0.923
Headway of first bus, in minutes a/	-.0341	3.51
Autos per driver with a ceiling of one b/	2.38	6.16
Autos per driver with a ceiling of one c/	1.48	1.92
Dummy if person is head of household b/	.494	2.62
Number of persons in household who can drive b/	.5242	4.18
Number of persons in household who can drive c/	.7567	3.82
Family income with ceiling of $7500, in $ per year b/	-.000308	2.18
Family income minus $7500 with floor of $0 and ceiling of $5000, in $ per year b/	.000139	1.05
Family income minus $10,500 with floor of $0 and ceiling of $5000, in $ per year b/	-.0000966	1.78
Auto alone dummy d/	-1.84	1.74

Table 36, continued

Explanatory Variables	Estimated Coefficients	t-Statistics
Bus-with-auto-access dummy e/	-5.38	5.69
BART-with-walk-access dummy f/	1.94	3.18
BART-with-bus-access dummy g/	-.159	0.285
BART-with auto-access dummy h/	-4.06	4.38
Carpool dummy i/	-2.39	5.28

Likelihood ratio undex:	.4119
Log likelihood at zero:	-982.6
Log likelihood at convergence:	-577.9
Degrees of freedom:	2460
Percent correctly predicted:	64.27

Value of time saved as a pecent of wage:

On-vehicle time:	43
Walk time:	265
Transfer wait time:	32

All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 641 .

Table 36, continued

a/ The variable is zero for the auto-alone and carpool alternatives, and takes the described value for the other alternatives.

b/ The variable takes the described value for the auto-alone alternative, and zero otherwise.

c/ The variable takes the described value for the bus-with-auto-access and BART-with-auto-access alternatives, and zero otherwise.

d/ The variable is one for the auto-alone alternative, and zero otherwise.

e/ The variable is one for the bus-with-auto-access alternative and zero otherwise.

f/ The variable is one for the BART-wtih-walk-access alternative, and zero otherwise.

g/ The variable is one for the BART-with-bus-access alternative, and zero otherwise.

h/ The variable is one for the BART-with-auto-access alternative, and zero otherwise.

i/ The variable is one for the carpool alternative, and zero otherwise.

The more general model includes the variables of Table 36 plus the following variables:

1. Cost of auto-alone divided by post-tax wage, taking the described value in the auto-alone and BART-with-walk-access alternatives and zero otherwise.

2. Cost of bus-with-walk-access divided by post-tax wage, taking the described value in the auto-alone and BART-with-walk-access alternatives and zero otherwise.

3. Cost of BART-with-walk-access divided by post-tax wage, taking the described value in the auto-alone and bus-with-walk-access alternatives and zero otherwise.

4.-6. Variables defined as 1-3, respectively, but with "total weighted time" rather than "cost divided by post-tax wage," where total time is the sum of: on-vehicle time; walk time multiplied by 2.5 ; transfer-wait time multiplied by 1.25 ; and first headway multiplied by 1.25 .

The log likelihood at convergence for this model is -567.6 . The log likelihood at convergence for the model of Table 36 is -577.9 . Therefore, the test-statistic (using formula (16)) is 20.6 . The critical (.05 level) value of chi-squared with six degrees of freedom is 12.6 . The joint hypothesis that the MNL form and the specification of Table 36 are correct is rejected.

The signs of the coefficients of the extra variables are consistent with the hypothesis that the value of auto on-vehicle time is higher than that of transit time. Variables 5 and 6 entered with negative signs (the latter with a t-statistic of 3.0), while the coefficient of variable 4 was estimated to be positive. In Train (1976) the value of auto time was found to be higher than that of bus on-vehicle time and the explanation was given that, while autos are more comfortable than transit, the difficulty of driving an auto during rush hour congestion makes auto time more onerous than transit time. The model of Table 36 constrains auto and bus times to be valued equally; this constraint perhaps contributes to the failure of the model in the test against the more general model.

The second test is against the saturated model. Because there are no repetitions, the saturated model has log likelihood equal to zero. With the log likelihood of the model of Table 36 equal to 577.9 , the test-statistic is 1155.8 . The critical (.05 level) value of chi-squared with 2460 degrees of freedom is slightly more than 2460 . The joint hypothesis of the MNL form and the specification of Table 36 is not rejected in the test against the saturated model. We note, however, that in the absence of repetitions, the asymptotic distribution of this test, and hence its actual significance level, are unknown.

The next group of tests investigate the implication of the IIA property that the coefficients of mean utility estimated on a subsample of choices conditioned on choice from a subset of alternatives are the same, asymptotically, as the coefficient estimated on the full sample. Estimation is performed on the subsample of individuals who chose an alternative in the subset of alternatives relevant for the test. The coefficients of mean utility are estimated on the subsample and the log likelihood at convergence is calculated; in addition, the log likelihood is calculated on the subsample with the coefficients restricted to the values in Table 36. Using the test statistic 'on page 25, the hypothesis that the coefficients estimated on the subsample are the same as those in Table 36 is tested. The results of the tests for various subsets of alternatives are shown in Table 37. The subsets chosen for testing were those which seemed most likely to result in rejection of the hypothesis of equal coefficients. For example, models similar to that of Table 36 estimated on a sample taken before BART was providing service greatly over-predict the use of BART-with- walk-access; hence, the subset consisting of all alternatives except BART-with-walk-access seemed particularly relevant for testing based on conditional choice.

The hypothesis that the coefficients estimated on the subsample are the same as those of Table 36 is accepted for each subset of alternatives except the subset excluding carpool. The failure for this last subset is at least partly the result of measurement errors in the observed attributes of the carpool alternative. The exact attributes of the carpool mode depend on such factors as the number of people in the carpool, each person's home and work locations, and the allocations of costs among carpool members. Because these variables cannot be determined for persons who do not choose carpool, crude rules-of-thumb were applied in calculating carpool attributes. Also, the specification of the mean utility for carpool mode is incomplete. There exists a lack of knowledge of the factors that encourage carpooling.

TABLE 37 Results of Tests Based on Conditional Choice

Alternatives included in the subset of alternatives	Log likelihood at convergence for sub-sample choosing an alternative within subset of alternatives	Log likelihood with coefficients restricted to values of Table 1	Test-Statistic	Degrees of Freedom	Critical (.05 level) value of chi-squared with appropriate degrees of freedom	Result
All alternatives except the three BART modes	-452.6	-454.6	4.0	16	26.3	Accept
All alternatives except the two bus modes	-400.1	-403.3	6.4	17	27.6	Accept
All alternatives except the bus-with-auto-access and BART-with-auto-access modes	-452.3	-455.0	5.4	17	27.6	Accept
All alternatives except the BART-with-walk-access mode	-557.8	-557.9	0.2	18	28.9	Accept
All alternatives except carpool	-230.7	-247.3	33.2	18	28.9	Reject

Lastly, tests of association are applied to the residuals and estimated probabilities of the model of Table 36. For each alternative a contingency table is constructed in the manner described in the previous section. The estimated probabilities for the alternative are ranked and classified into thirty cells, with each cell containing approximately the same number of cases. The number of positive residuals associated with the probabilities in a cell is counted, as is the number of negative residuals. These counts are recorded in Table 38 along with the average probability for cases in each cell. (The number of positive and negative residuals summed over all cells for a particular alternative is different for different alternatives because the number of people in the sample who have a given alternative available varies across alternatives.)

If the MNL form and the specification of Table 36 are accurate, then the number of positive residuals is expected to be higher for low-numbered cells than for high-numbered cells. This pattern emerges for each alternative. The statistic on page 236 tests association and goodness-of-fit. The value of this statistic for each alternative is recorded in Table 38. Because there are thirty cells and nineteen parameters, the test-statistic has an asymptotic distribution, under the hypothesis that the MNL form and the specification of Table 36 are correct, bounded by chi-square distributions with twenty-nine and ten degrees of freedom. The critical (.05 level) of chi-squared with twenty-nine degrees of freedom is 42.56 ; that with ten degrees of freedom is 18.31 . The values of the test-statistic for all alternatives except the carpool alternative fall below the lower of the two bounding critical values, and therefore the hypothesis is accepted for those alternatives. For the carpool alternative, the test-statistic falls between the two bounding critical values; the test is therefore inconclusive. As in the failure of the test based on conditional choice, measurement and specification errors in the carpool attributes are probably the cause of the inability to pass unambiguously the test of association for the carpool alternative.

As another example of the application and use of the diagnostic tests, consider the BART rider's choice of access mode and station. A brief description of these data, alternatives, and variables is the following.

TABLE 38 Tests of Association

(a) Auto-Alone Alternatives

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	17	21	17	20	16	17	20	19	17	18	14	19	14	18	13
Number of Negative Residuals	5	1	5	2	6	5	2	3	5	4	8	2	7	3	8
Average Probability for Cell	.93	.89	.87	.85	.82	.81	.80	.78	.77	.75	.74	.72	.69	.67	.65
Cell	16	7	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of Positive Residuals	14	12	12	14	11	11	7	10	11	3	5	5	5	4	3
Number of Negative Residuals	7	9	9	7	10	10	14	11	10	18	16	19	19	17	18
Average Probability for Cell	.63	.62	.60	.57	.53	.49	.47	.42	.38	.35	.32	.27	.20	.13	.05

Test-Statistic: 17.51

Table 38, continued (Tests of Association)

(b) Bus-with-Walk Access Alternative

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	14	10	7	7	4	7	1	3	2	2	3	1	3	2	0
Number of Negative Residuals	3	7	10	10	13	10	16	14	15	15	14	16	14	15	17
Average Probability for Cell	.66	.52	.43	.37	.31	.26	.22	.19	.17	.15	.13	.11	.10	.085	.072

Cell	16	17	18	19	20	21	22	23	24	25	6	27	28	29	30
Number of Positive Residuals	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Number of Negative Residuals	15	17	17	16	17	16	17	17	17	17	17	17	16	16	16
Average Probability for Cell	.059	.051	.044	.038	.034	.028	.022	.019	.015	.013	.010	.008	.006	.004	.001

Test-Statistic: 14.14

Table 38, continued (Tests of Association)

(c) Bus-with-Auto-Access Alternatives

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	1	0	2	2	2	0	0	1	0	0	1	0	0	0	0
Number of Negative Residuals	16	17	15	15	15	17	17	16	17	17	16	17	17	17	17
Average Probability for Cell	.090	.053	.045	.038	.034	.030	.028	.025	.022	.020	.018	.016	.014	.013	.012

Cell	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of Positive Residuals	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of Negative Residuals	17	17	17	17	17	17	17	17	17	17	17	17	16	16	16
Average Probability for Cell	.011	.009	.008	.008	.007	.006	.005	.005	.004	.003	.003	.002	.001	.001	.000

Test-Statistic: 15.75

Table 38, continued (Tests of Association)

(d) BART-with-Walk-Access Alternatives

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0
Number of Negative Residuals	12	12	11	12	11	12	12	11	12	12	11	12	12	12	12
Average Probability for Cell	.057	.037	.029	.025	.021	.019	.017	.016	.014	.012	.011	.010	.008	.007	.007

Cell	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of Positive Residuals	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of Negative Residuals	12	12	12	11	11	11	11	11	11	11	11	11	11	11	11
Average Probability for Cell	.006	.005	.005	.005	.004	.004	.003	.002	.002	.002	.002	.001	.001	.001	.000

Test-Statistic: 15.83

Table 38, continued (Tests of Association)

(e) BART-with-Bus-Access Alternatives

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0
Number of Negative Residuals	4	4	4	4	4	5	5	4	4	4	4	4	4	4	4
Average Probability for Cell	.266	.148	.119	.103	.088	.080	.067	.060	.052	.044	.034	.031	.027	.023	.018
Cell	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of Positive Residuals	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of Negative Residuals	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Average Probability for Cell	.016	.014	.014	.012	.010	.008	.007	.006	.005	.005	.004	.003	.002	.001	.000

Test-Statistic: 5.32

Table 38, continued (Tests of Association)

(f) BART-with-Auto-Access Alternatives

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	6	4	2	5	3	1	2	2	1	0	2	1	1	2	1
Number of Negative Residuals	6	8	10	7	9	11	10	10	11	12	10	11	11	10	11
Average Probability for Cell	.390	.289	.222	.195	.177	.160	.138	.124	.113	.103	.093	.086	.077	.069	.065

Cell	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of Positive Residuals	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of Negative Residuals	12	12	12	11	11	11	11	11	11	11	11	11	11	11	11
Average Probability for Cell	.060	.055	.050	.046	.042	.038	.034	.030	.028	.025	.021	.018	.015	.010	.004

Test-Statistic: 13.81

Table 38, continued (Tests of Association)

<u>(g) Carpool Alternative</u>

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of Positive Residuals	6	9	11	8	6	6	4	4	5	6	4	2	5	6	2
Number of Negative Residuals	16	13	11	14	16	16	18	18	17	16	18	19	16	15	19
Average Probability for Cell	.59	.45	.39	.36	.34	.31	.28	.26	.25	.23	.22	.21	.20	.19	.18

Cell	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of Positive Residuals	5	6	7	1	1	4	4	6	3	1	5	4	2	0	4
Number of Negative Residuals	16	15	14	20	20	17	17	15	18	20	16	17	19	21	17
Average Probability for Cell	.17	.164	.158	.154	.147	.141	.135	.129	.122	.116	.110	.104	.085	.080	.054

Test-Statistic: 38.63

The data are a subset of the Urban Travel Demand Forecasting Project's (UTDFP) sample survey of about 800 workers in the San Francisco Bay Area in 1975. This subset consists of 134 BART users, of whom twenty walked, twenty-two used bus, twenty-eight drove and parked, and twenty-four were dropped off at a BART station; of the twenty-four who were dropped off, less than half involved a special trip by someone, normally a family member. These modes are labeled by W, B, DD, DP, respectively. The three stations that were considered to be alternatives for the BART users were based on the "suggestions" obtained from the data analysis. Generally, the alternative stations were adjacent stations even though exceptions occurred. Most prominent of these exceptions was the choice, by a handful of BART riders, to drive to Oakland West station and gain the advantage of six-minute BART headways and escape the need to queue at the Bay Bridge entrance. Taken together, the four alternative modes and three alternative stations make up a choice set of twelve alternatives.

The availability of alternatives to users was governed by the following considerations. If an individual had no driver's license or a household owned no cars, the drive-park mode was not available; the bus mode was unavailable if the nearest bus line providing service to any station was farther than the nearest BART station. Walk and drive-drop modes were always considered to be available.

The independent variables that were associated with each user were decided on the basis of "state-of-the-art" mode-choice models that have proven to be successful predictors and, to some extent, transferable in time or space (Atherton and Ben-Akiva in 1976; McFadden, 1974; McFadden and Train, 1976; Train, 1976a,b,c). The variables are: walk time (minutes), on-vehicle time (minutes), headway (minutes), cost/wage (¢/¢/minute), drivers in household, household income ($), cars per driver in household. Walk access was identified by a dummy variable when walk-time-to-station was less than ten minutes. All the system performance variables were derived by locating the individual's residence on the map and routing him to the various stations (Talvitie, 1976).

The base model appears in Table 39, column 1; the models in the other columns will be discussed shortly. The variables were entered into the alternatives indicated in brackets after the variable name; they take the value of zero for the other alternatives.[1]

It was suggested earlier that of all the IIA tests, a test against the universal logit model and the tests based on restricted choice sets are the most powerful for detecting deviations from the MNL form. Many different general models or restricted choice sets can be devised for testing purposes. Some of these tests were conducted with the base model.

The tests against a more general model than the MNL (but less general than the saturated model) were not very successful; of the many models tried only two converged acceptably. The first, in which the bus on-vehicle time was also entered to the drive modes, yielded a χ^2 test statistic of 3.2, which is below the critical value (.95 level) of 3.8 and thus upholds the MNL model. In the other, more general model, the bus on-vehicle time was again entered in the car alternatives and drive time was entered in the bus mode alternatives. The χ^2 statistic is 7.4, which is above the critical value (.95 level) of 5.99; the test thus rejects the joint hypothesis of MNL-form and the presumed model specification.

The test experiments with restricted or conditional choice sets--which means dropping one or more alternatives from the choice set, computing the value of the likelihood function at the coefficients obtained with the full choice set, and comparing that to the value of the likelihood function maximized for the conditional choice set--provided more interesting and telling results. Eight different choice sets were constructed, excluding modes one-at-a-time; excluding station choice altogether; excluding some modes and some stations; and so forth. The test results appear in Table 40.

---

[1] The walk modes are alternatives 1, 5, and 9. The bus modes are alternatives 2, 6, and 10. Drive-park occupies alternatives 3, 7, and 11; and drive-drop occupies the remaining alternatives, 4, 8, and 12.

TABLE 39 Coefficients (t-values) Estimated Using Restricted Choice Sets

Choice set / Variable	Full Choice Set (12 alt) (1)		No Drive-Drop Mode (9 alt) (2)		No Station Choice or Drive-Drop (3 alt) (3)		No Bus Mode (9 alt) (4)		No Station Choice All access modes (4 alt) (5)		Aggregated auto mode with drive park attributes (9 alt) (6)	
Walk Time (all)	-.0651	(3.6)	-.1037	(4.6)	-.0726	(2.7)	-.109	(3.4)	-.491	(2.5)	-.0706	(3.9)
In-vehicle Time (all)	-.0652	(2.3)	-.1157	(2.9)	-.0904	(1.5)	-.0589	(1.5)	-.0464	(1.0)	-.0783	(2.6)
Bus Headway (B)	-.0322	(1.4)	-.0789	(2.9)	-.0479	(1.6)	------		-.0292	(1.1)	-.0529	(1.9)
Cost/Wage (B, DP)	-.120	(4.2)	-.292	(5.7)	-.121	(1.9)	-.133	(3.9)	-.0515	(1.9)	-.155	(5.3)
Drivers (DP, DD)	-680	(2.3)	.672	(2.0)	.505	(1.5)	.898	(1.8)	.671	(2.3)	.491	(1.6)
Income (DP, DD)	.0000594	(1.7)	.0000509	(13)	.0000648	(1.7)	.0000273	(0.5)	.0000643	(1.8)	.0000465	(1.2)
Cars/Drivers (DP)	3.142	(3.2)	2.456	(2.2)	1.943	(1.8)	1.654	(1.3)	2.882	(2.9)	1.232	(1.3)
Cars/Drivers (DD)	-1.014	(1.1)	------		------		-3.174	(2.3)	-.591	(0.6)	------	
Walk Access (W)	4.166	(3.5)	3.779	(3.1)	3.339	(2.6)	2.787	(2.3)	3.710	(3.1)	4.268	(3.5)
Dummy (W)	3.473	(2.9)	3.100	(2.0)	2.676	(1.7)	3.162	(2.2)	2.715	(2.2)	1.389	(1.0)
Dummy (B)	4.478	(4.0)	4.900	(3.1)	3.826	(2.3)	------		3.548	(3.0)	2.768	(2.1)
Dummy (DP)	-.648	(0.8)	------		------		-.937	(1.1)	-.992	(1.2)	------	
L(β*)	-211.7		-113.3		-57.0		-155.6		-111.9		-159.1	
L(O)	-308.7		-213.3		-108.8		-235.2		-175.0		-260.3	
% Right	50.0		62.7		78.9		52.2		66.4		56.7	
ρ	.31		.47		.48		.34		.36		.39	
Degrees of freedom	1857		659		181		796		353			
Sample Size	134		110		109		112		134		134	

TABLE 40 Tests on the IIA-Property Based on Conditional Choice

Alternatives included in the choice subset	Log likelihood for choice subset	Log likelihood with coefficients restricted to values obtained with full choice set	Test-statistic	Degrees of freedom	Critical χ^2 (.95)	Result
All except walk mode	-172.9	-180.0	14.20	10	18.3	Accept
All except bus mode	-155.6	-160.3	9.4	10	18.3	Accept
All except drive-park mode	-78.7	-82.9	8.4	10	18.3	Accept
All except drive-drop mode	-113.3	-125.1	23.6	10	18.3	Reject
No station choice--all access modes	-111.9	-116.9	10.0	12	21.0	Accept
No station choice, no drive-drop	-57.0	-59.8	5.6	11	19.7	Accept
No station choice for drive-drop	-193.4	-186.0	14.8	12	18.3	Accept
No station choice for drive-drop or walk	-182.6	-190.4	15.6	12	18.3	Accept

The results in Table 40 indicate that alternatives connected with the drive-drop mode do not meet the IIA assumption when the choice of access station is included in the choice set. This outcome makes common sense. We may recall that more than one-half of those dropped off at the station entailed no special trip but rather they were--and are--specific "arrangements" between the drivers and those dropped off. On the other hand, the utility function of the drive-drop mode includes no variables that purport to account for circumstances where ride-sharing could occur unless we count "cars per driver" as such a variable.

The result that the shared-ride-access mode does deviate from the MNL for this similar to the result obtained earlier that shared-ride in the traditional "main" mode-choice model also deviated from the MNL form. These results, obtained with independent data sets, indicate that more and better variables need to be used to improve the specification of the alternatives attached to either ride-sharing or carpooling modes.

On the positive side, it may be noted that the MNL model is quite robust with respect to the IIA property and can be used well in practical applications involving many alternatives that superficially would appear to be quite similar, such as alternative stations.

Before concluding this chapter it is of interest to take a look at the useful byproducts of the IIA tests: these are coefficients obtained by varying the choice sets and are given in Table 39, columns 2-5. While it is difficult to make a meaningful summary of these numbers, it is fair to say that the choice set does have a marked effect on the alternative-specific dummy variables. A direct consequence of this is that efforts should be made for improving the specification of the utility functions to lessen the share that the alternative-specific constants account for in the model's "explanatory power."

A system coefficient that seems to be unstable beyond its two standard errors is the coefficient of the "cost/wage" variable. Two remedies for the problem were considered. The first was to see if assigning a round-trip driving cost for the drive-drop mode would help matters. Not unexpectedly, the cost coefficient decreased (in absolute value) somewhat in all the models tried, as compared to the models developed using the zero cost number for the drive-drop mode; and the log likelihood at convergence was virtually unchanged ($\pm .5$).

That the cost coefficient is more unstable than its standard error points again toward misspecification of the model, although parts of that instability are likely attributable to the way travelers assign costs on the auto mode. Given that the cost coefficient plays an important role in the (economic) benefit computations, it is necessary to pin it down more precisely.

As discussed earlier in this chapter, the choice sets can also affect the models via the definition of alternatives. It has been common practice, for example, to aggregate drive-park and drive-drop into one alternative characterized by drive-park trip attributes (Liou and Talvitie, 1974; Train, 1976c). A model embodying these assumptions was estimated with the current data; the coefficients are given in Table 39, column 6. An examination of the coefficients reveals that, compared to the "base model" in column one, the alternative-specific constants or variables multiplied by them (e.g., cars-per-driver) are significantly affected. Keeping in mind that these constants critically affect forecast accuracy, the aggregation of the alternatives must be done very carefully in either developing the model or predicting with it. In particular, it appears best to use the most disaggregate definition of choice sets (alternatives). If some alternative or alternatives then have only few observed choices, they are better left out from the model estimation or aggregated "logitly" (see the following chapter), rather than lumped together with some popular and seemingly similar alternative.

In conclusion, the IIA-tests devised by McFadden, *et al.* (1976) are useful and can be used not only to test the IIA property itself, but also as a check and guide in model specification and further data collection. In retrospect, this is not unsurprising: the IIA property ought to disappear with perfect model specification.

APPENDIX I

A SIMULATION STUDY OF THE INDEPENDENCE FROM IRRELEVANT ALTERNATIVES PROPERTY

Introduction

In this appendix the usefulness of the IIA diagnostic tests is examined. The examination is as follows. Data are generated from various "true" models of choice by simulation. These true models are designed to introduce different types of variations in choice behavior, identified in the main text, each of which will cause the IIA property to fail. MNL models are then "fitted" to these data and diagnostic tests are applied to determine if the tests can detect this deviation of the underlying data from the MNL model.

A travel mode choice among auto, bus, and rail is considered. Level-of-service attributes are created for 100 choice environments. A particular model is designated to be true, and the choice of each of 100 respondents in each of the 100 environments is generated based on this true model. An MNL model is fitted to the data on attributes and choice, and diagnostic tests are applied.

First, an MNL model is fitted to data generated from an MNL model. It is expected, naturally, that the diagnostic tests would uphold the null hypothesis that the IIA assumption is valid. It is assumed that the true utility function has the form:

$$U_i = \beta_1 \cdot COST^i + \beta_2 \cdot TIME^i + \varepsilon_i \quad ,$$

where U_i is utility of mode i ; $COST^i$ is the cost in dollars of mode i ; $TIME^i$ is the time in hours of mode i ; ε_i is standard Weibull; β_1 equals -1 ; and β_2 equals -2.5 . The implied value of time is \$2.50 per hour. The variable CHOICE is obtained by sampling from the MNL model derived from this utility function.

Table 41 presents the estimated parameters of MNL models fitted to these data and the results of diagnostic tests on the models. The estimated parameters of model (1) are close to the true parameters, which is expected given that model (1) has the true specification. Model (2) includes mode-specific dummies that do not enter the true model; their presence, however, improves the estimates somewhat. Models (3) and (4) are estimated on the subsample of respondents who chose either auto or bus. Because the true model is MNL, the estimated parameters based on the conditional choice set are expected to be similar to the true parameters. Both the conditional model without dummies (model (3)) and that with dummies (model (4)) have estimated parameters close to the true ones.

The estimated MNL models pass all the diagnostic tests; that is, the hypothesis that the MNL form and the specification of model (1) or (2) is correct is not rejected in any test. The first tests use the universal logit method. A more general model is specified that includes the variables in the model being tested plus the following six variables:

1. Cost of auto, taking the described value in the bus alternative and zero otherwise.
2. Cost of bus, taking the described value in the rail alternative and zero otherwise.
3. Cost of rail, taking the described value in the auto alternative and zero otherwise.
4. Time of auto, taking the described value in the bus alternative and zero otherwise.
5. Time of bus, taking the described value in the rail alternative and zero otherwise.
6. Time of rail, taking the described value in the auto alternative and zero otherwise.

TABLE 41 Choice Base on MNL Model

Dependent variable is CHOICE

	Parameters (t-statistics in parentheses)			
	(1)	(2)	(3)	(4)
Independent Variables	Choice Set: Auto, Bus, Rail	Choice Set: Auto, Bus, Rail	Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus	Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus
Cost	-.9860 (41.27)	-1.003 (36.0)	-1.031 (34.6)	-1.043 (30.1)
Time	-2.441 (53.5)	-2.452 (48.8)	-2.563 (40.7)	-2.55 (39.4)
Auto Alternative Dummy		-.0268 (.539)		.0415 (.710)
Bus Alternative Dummy		-.0724 (2.16)		
Log likelihood at zero	-10990.	10990.	-5668.	-5668.
Log likelihood at convergence	-6937.	-6935.	-3235.	-3235.
Degrees of freedom	19998	19996	8175	8174

Table 41, continued

Tests: 1A. Test against a more general model:

(a) Log likelihood at convergence of more general model without dummies: -6934.
Test-statistic for model (1): 6.0
Critical chi-squared with 6 degrees of freedom: 12.6 .
Result: PASS

(b) Log likelihood at convergence of more general model with dummies: -6933.
Test-statistic for model (2): 4.0 .
Critical chi-squared with 6 degrees of freedom: 12.6
Result: PASS

1B. Test against the saturated model:

(a) Log likelihood at convergence of saturated model: -22.78
Test-statistic for model (1): 13,828.44
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Log likelihood at convergence of saturated model: -22.78
Test-statistic for model (2): 13,824.44
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

1C. Test based on conditional choice:

(a) Log likelihood at convergence of model (3) with parameters restricted to values of model (1): -.3237.
Test-statistic: 4.0
Critical chi-squared with 2 degrees of freedom: 6.0
Result: PASS

(b) Log likelihood at convergence of model (4) with parameters restricted to values of model (2): -3236.
Test-statistic: 2.0

Table 41, continued

Critical chi-squared with 3 degrees of freedom: 7.8 .
Result: PASS

2A. Means test:

Alternative	Means of Residuals of Model (1)	Result
Bus	-.1662	PASS
Rail	1.737	PASS

2B. Variance test:

(a) Sum of squared residuals for model (1): 12,100.
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Sum of squared residuals for model (2): 12,470.
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

The model estimated with these six variables in addition to the two variables in model (1) attained a log likelihood at convergence of -6934 . Using formula (16), the test-statistic is 6.0 . Under the hypothesis that the coefficients of these six extra variables are all zero, the test-statistic is distributed as chi-square with six degrees of freedom, which has a critical (.05 level) value of 12.6 . Because the critical value exceeds the test-statistic, the hypothesis of zero coefficients is accepted.

A similar test is applied to model (2), with the more general model including the three variables of model (2), plus the six variables listed above. Model (2) passes the test.

The second set of tests are against the saturated model. The log likelihood at convergence of the saturated model, given in equation (14), is -22.78.[1] The test-statistic, using (16), is 13,828.44 . The critical (.05 level) value of chi-squared with 19,998 degrees of freedom is slightly greater than 19,998. Therefore, model (1) passes in the test against the saturated model. Similarly, model (2) passes.

The third set of tests are based on conditional choice. The log likelihood of an MNL model without dummies is calculated on the subsample of respondents who chose either auto or bus with the parameters restricted to the values estimated in model (l); the value of this log likelihood is -3237 . Using the formula of page 25, the test-statistic is 4.0 . The test-statistic is distributed, under the hypothesis that the parameters estimated on the conditional choice subsample are equal to those of model (1), as chi-square with two degrees of freedom, which has a critical (.05 level) value of 6.0 . Because the critical value exceeds the value of the test-statistic, the hypothesis is accepted. A similar test is applied for an MNL model <u>with</u> dummies, and the model with dummies passes the test.

The fourth set of tests are based on the means of the transformed residuals. The means of the transformed residuals of model (1) for the bus and rail alternatives are -.1662 and 1.737 , respectively. Under the hypothesis that the MNL form and specification of model (l) are correct, these means are distributed asymptotically standard normal. The critical (.05 level) values of the standard normal is $\pm$ 1.96 . Therefore, model (1) passes the means test for both alternatives. Means tests are not applied to model (2) because in a model with alternative-specific dummies the means are identically equal to zero.

---

[1] The equation references made in this Appendix refer to equations in Part III, Chapter 1.

Variance tests are the final set of tests. The sum of squared residuals for model (1) is 12,100 . Under the hypothesis that the MNL form and the specification of model (1) are correct, the sum of squared residuals is distributed asymptotically chi-square with 19,998 degrees of freedom. Because the critical (.05 level) value of chi-squared with 19,998 degrees of freedom is slightly more than 19,998, the null hypothesis is accepted. Similarly, model (2) passes the variance test.

The true MNL model thus passed all the diagnostic tests. In the next five models the "forbidden" sources of variation that cause the IIA to fail are introduced one at a time. If the diagnostic tests are able to detect this, then, it is reasoned, they can be used to do the same with data collected by standard data collection methods.

Taste Variation

Assume that the true utility function takes the form

$$U_i = \alpha_1 \cdot \mathrm{COST}^i + \alpha_2 \cdot \mathrm{TIME}^i + \varepsilon_i$$

where ε_i is standard extreme value and α_1 and α_2 vary in the population with (α_1, α_2) equal with probability one-fourth to each of the vectors (.5, 1.25), (.5, 3.75), (1.5, 1.25), and (1.5, 3.75) . The expected value of time is \$3.33 per hour; however, the expected value of each coefficient coincides with the values assumed in the MNL model. The choice variable generated from this model is CHOICEl.

Table 42 presents the estimated parameters of MNL models fitted to these data and the results of diagnostic tests on the models. The estimated coefficient of TIME is lower than the mean of the true coefficients; the estimated value of time is also lower than the mean of the true values of time. The fitted models fail the test against a more general model. As in all the tests in this Appendix (unless otherwise noted), the more general model included the variables in the model being tested plus the six variables listed on page 241. The fitted models pass the test against the saturated model and the test based on conditional choice. The variance test is failed, however, as is the means test for the rail alternative.

TABLE 42 Choice Based on Model with Taste Variations
Dependent variable is CHOICE1

Independent Variables	(5) Choice Set: Auto, Bus, Rail	(6) Choice Set: Auto, Bus, Rail	(7) Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus	(8) Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus
Cost	-.9767	-1.059	-.9784	-1.082
	(41.9)	(38.2)	(35.4)	(32.3)
Time	-2.268	-2.081	-2.259	-2.194
	(53.7)	(46.87)	(41.57)	(39.8)
Auto Alternative Dummy		.5398		.3198
		(11.1)		(5.81)
Bus Alternative Dummy		.1776		
		(5.16)		
Log likelihood at zero	-10990.	-10990.	-5868.	-5868.
Log likelihood at convergence	-7167.	-7103.	-3647.	-3630.
Degrees of freedom	19998	19996	8464	8463

Table 42, continued

Tests:

1A: Test against a more general model:

(a) Log likelihood at convergence of more general model without dummies: -7127.
Test-statistic for model (5): 80.0 .
Critical chi-squared with 6 degrees of freedom: 12.6 .
Result: FAIL

(b) Log likelihood at convergence of more general model with dummies: -7076.
Test-statistic for model (6): 54.0 .
Critical chi-squared with 6 degrees of freedom: 12.6 .
Result: FAIL

1B. Test against the saturated model:

(a) Log likelihood of saturated model: -23.11
Test-statistic for model (5): 14,305.78
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Log likelihood of saturated model: -23.11
Test-statistic for model (6): 14,159.78
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

1C. Test based on conditional choice:

(a) Log likelihood at convergence of model (7) with parameters restricted to values of model (5): -3647.
Test-statistics: 0.
Critical chi-squared with 2 degrees of freedom: 6.0
Result: PASS

(b) Log likelihood at convergence of model (8) with parameters restricted to values of model (6): -3632.
Test-statistic: 4.0 .

Table 42, continued

Critical chi-squared with 3 degrees of freedom: 7.6 .
Result: PASS

2A. Means test:

Alternative	Means of Residuals of Model (5)	Result
Bus	.05338	PASS
Rail	11.94	FAIL

2B. Variance test:

(a) Sum of squared residuals for model (5): 1,266,000 .
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: FAIL

(b) Sum of squared residuals for model (6): 582,500 .
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996 .
Result: FAIL

Dependence of Unobserved Attributes with Observed Attributes
or Across the Alternatives

Assume the true utility function takes the form

$$U_i = \beta_1 \cdot COST^i + \beta_2 \cdot TIME^i + \beta_3 \cdot NOISE + \varepsilon_i \quad ,$$

where ε_i is extreme value; β_1 equals -1; β_2 equals -2.5; β_3 equals -0.2; and $NOISE^i$ is an unobserved attribute of mode i that is correlated with $TIME^i$ but is uncorrelated across modes once the common effect of TIME is removed. The choice variable generated from this model is CHOICE2 .

Table 43 presents the estimated parameters of MNL models fitted to these data and the results of diagnostic tests on the models. Models (9) and (10) are fitted with NOISE included as an observed variable. The estimated parameters are fairly close to the true parameters, though all of the parameters are underestimated. Models (11) and (12) are fitted without NOISE being included as an observed variable. The estimated coefficient of time is closer. The models fitted with NOISE unobserved fail the test against a more general model and the tests based on conditional choice, but pass the test against the saturated model and the means and variance tests.

A similar model in which an unobserved variable causes violation of IIA is constructed by assuming that the true utility function is

$$U_i = \beta_1 \cdot COST^i + \beta_2 \cdot TIME^i + \beta_3 \cdot EXERTION^i + \varepsilon_i \quad ,$$

where all terms are the same as before but where $EXERTION^i$ is an unobserved attribute of mode i that is correlated with $TIME^i$ and is closely correlated between bus and rail even after the common TIME effect is removed. CHOICE3 is the choice variable generated from this model.

Table 45[1] presents the MNL estimates and test results for this model. The estimates are closer to the true values in models (15) and (16) that include EXERTION , than in models (17) and (18), which do not include EXERTION . The models fitted without EXERTION fail all the tests except the test against the saturated model and the means test for the bus alternative.

[1] There is no Table 44.

TABLE 43 Choice Based on Model with Unobserved Attributes: NOISE

Dependent variable is CHOICE2

Parameters (t-statistics in parentheses)

Independent Variables	(9) Choice Set: Auto, Bus, Rail	(10) Choice Set: Auto, Bus, Rail	(11) Choice Set: Auto, Bus, Rail	(12) Choice Set: Auto, Bus, Rail	(13) Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus	(14) Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus
Cost	-.8913 (36.5)	-.8901 (30.5)	-.7570 (34.0)	-.7221 (27.56)	-.8544 (29.95)	-.8132 (24.5)
Time	-2.30 (48.0)	-2.31 (44.1)	-2.390 (52.2)	-2.435 (47.6)	-2.645 (38.96)	-2.688 (38.1)
Noise	-.1909 (26.4)	-.1909 (26.3)				
Auto Alternative Dummy		-.0194 (.379)		-.0962 (1.92)		-.1402 (2.41)
Bus Alternative Dummy		-.0114 (.319)		.0372 (1.08)		
Log likelihood at zero	-10990	-10990	-10990	-10990	-5779	-5779
Log likelihood at convergence	-6481	-6481	-6872	-6868	-3230	-3227
Degrees of freedom	19997	19995	19998	19996	8336	8335

Table 43, continued

Test:

1A: Test against a more general model:

(a) Log likelihood at convergence for more general model without dummies: -6817.
Test-statistic for model (11): 55.0 .
Critical chi-squared with 6 degrees of freedom: 12.6 .
Result: FAIL

(b) Log likelihood at convergence of more general model with dummies: -6800.
Test-statistic for model (12): 136 .
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

1B: Test against the saturated model:

(a) Log likelihood at convergence of saturated model: -21.24.
Test-statistic for model (11): 13,701.52 .
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Log likelihood at convergence of saturated model: -21.24
Test-statistic for model (12): 13,693.52 .
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

1C. Test based on conditional choice:

(a) Log likelihood at convergence of model (13) with parameters restricted to values of model (11): -3238
Test-statistic: 16
Critical chi-squared with 2 degrees of freedom: 6.0
Result: FAIL

(b) Log likelihood at convergence of model (14) with parameters restricted to values of model (12): -3234
Test-statistic: 14.0

Table 43, continued

Critical chi-squared with 3 degrees of freedom: 7.6 .
Result: FAIL

2A. Means test:

Alternative	Means of Residuals of Model (11)	Result
Bus	.1291	PASS
Rail	1.357	PASS

2B. Variance test:

(a) Sum of squared residuals for model (11): 9,602 .
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Sum of squared residuals for model (12): 11,240.
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

TABLE 45 Choice Based on Model with Unobserved Attributes: EXERTION

Dependent variable is CHOICE3

Independent Variables	(15) Choice Set: Auto, Bus, Rail	(16) Choice Set: Auto, Bus, Rail	(17) Choice Set: Auto, Bus, Rail	(18) Choice Set: Auto, Bus, Rail	(19) Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus	(20) Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus
Cost	-.9124 (33.4)	-.9178 (29.4)	-.7786 (30.7)	-.8804 (29.6)	-.8329 (26.9)	-.9702 (26.4)
Time	-2.17 (30.54)	-2.168 (30.1)	-3.116 (50.03)	-2.928 (44.6)	-3.178 (37.6)	-3.05 (35.6)
Exertion	-.1902 (22.6)	-.1896 (20.9)				
Auto Alternative Dummy		.00731 (.128)		.4119 (7.84)		.4304 (7.08)
Bus Alternative Dummy		-.0123 (.320)		.0436 (1.14)		
Log likelihood at zero	-10990	-10990	-10990	-10990	-6024	-6024
Log likelihood at convergence	-5648	-5648	-5936	-5904	-2974	-2949
Degrees of freedom	19997	19995	19998	19996	8689	8688

Table 45, continued

Tests:

1A. Test against a more general model:
- (a) Log likelihood at convergence of more general model without dummies: -5888
  Test-statistic for model (17): 96.0
  Critical chi-squared with 6 degrees of freedom: 12.6
  Result: FAIL

- (b) Log likelihood at convergence of more general model with dummies: -5869
  Test-statistic for model (18): 70.0
  Critical chi-squared with 6 degrees of freedom: 12.6
  Result: FAIL

1B. Test against the saturated model:
- (a) Log likelihood at convergence of saturated model: -18.42
  Test-statistic for model (17): 11,835.16
  Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
  Result: PASS

- (b) Log likelihood at convergence of saturated model: -18.42
  Test-statistic for model (18): 11,771.16
  Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
  Result: PASS

1C. Test based on conditional choice:
- (a) Log likelihood at convergence of model (19) with parameters restricted to values of model (17): -2977
  Test-statistic: 6.0
  Critical chi-squared with 2 degrees of freedom: 6.0
  Result: NOT CLEAR

- (b) Log likelihood at convergence of model (20) with parameters restricted to values of model (18): -2952
  Test-statistic: 6.0

Table 45, continued

Critical chi-squared with 3 degrees of freedom: 7.6
Result: FAIL

2A. Means test:

Alternative	Mean Residuals of Model (17)	Result
Bus	.0573	PASS
Rail	251.1	FAIL

2B. Variance test:

(a) Sum of squared residuals of model (17): 630,000,000
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998
Result: FAIL

(b) Sum of squared residuals of model (18): 258,700,000
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996
Result: FAIL

Measurement Error in Observed Attributes

The true utility function is assumed to have the form of that specified for the MNL model for the first section of this appendix. The variables COST and TIME are measured with error, however, as NCOST and NTIME , respectively. The variables with errors might be interpreted as network approximations to the true cost and time variables.

Table 46 presents the fitted MNL models and diagnostic tests using CHOICE as the dependent variable and NCOST and NTIME as the independent variables. The parameters are underestimated, as is expected for an errors-in-variables model. The models fitted to the data with errors fail the test against a more general model and the tests based on conditional choice, but pass the test against the saturated model, the means test, and the variance test.

TABLE 46 Choice Based on MNL Model: Attributes Measured with Error

Dependent variable is CHOICE

	Parameters (t-statistics in parentheses)			
	(21)	(22)	(23)	(24)
Independent Variables	Choice Set: Auto, Bus, Rail	Choice Set: Auto, Bus, Rail	Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus	Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus
NCOST	-.8107 (38.7)	-.8418 (33.7)	-.8753 (32.8)	-.9137 (28.1)
NTIME	-2.158 (53.5)	-2.123 (48.1)	-2.27 (41.1)	-2.24 (39.7)
Auto Alternative Dummy		.868 (1.83)		.118 (2.12)
Bus Alternative Dummy		-.0306 (.919)		
Log likelihood at zero	-10990	-10990	-5668	-5668
Log likelihood at convergence	-7179	-7177	-3388	-3386
Degrees of freedom	19998	19996	8175	8174

Table 46, continued

Tests:

1A. Test against a more general model:

(a) Log likelihood at convergence of more general model without dummies: -7126
Test-statistic for model (21): 106
Critical chi-squared with 6 degrees of freedom: 12.6
Result: FAIL

(b) Log likelihood at convergence of more general model with dummies: -7126
Test-statistic for model (22): 102
Critical chi-squared with 6 degrees of freedom: 12.6
Result: FAIL

1B. Test against the saturated model:

(a) Log likelihood at convergence of saturated model: -22.78
Test-statistic for model (21): 14,312.44
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Log likelihood at convergence of saturated model: -22.78
Test-statistic for model (22): 14,308.44
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

1C. Test based on conditional choice:

(a) Log likelihood at convergence of model (23) with parameters restricted to values of model (21): -3391
Test-statistic: 6.0
Critical chi-squared with 2 degrees of freedom: 6.0
Result: NOT CLEAR

(b) Log likelihood at convergence of model (24) with parameters restricted to values of model (22): -3389
Test-statistic: 6.0

Table 46, continued

Critical chi-squared with 3 degrees of freedom: 7.6
Result: FAIL

2A. Means test:

Alternative	Means of Residuals of Model (21)	Result
Bus	-.02373	PASS
Rail	.7053	PASS

2B. Variance test:

(a) Sum of squared residuals for model (21): 2292
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998
Result: PASS

(b) Sum of squared residuals for model (22): 2265
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996
Result: PASS

Attributes of Alternatives Not Exogenous

Assume that in the true mode, the value of time varies and people with higher than average values of time tend to locate such that the ratio of auto time to auto cost is lower than average. That is, the attributes of modes are endogenous in that they are chosen by the decision maker in conjunction with his locational choice. CHOICE5 is the choice variable generated from a model in which utility has the form

$$U_i = \beta_1 \cdot COST^i + \beta_2 \cdot TIME^i + \varepsilon_i \quad ,$$

where ε_i is extreme value; β_1 equals -1; and β_2 varies in the population such that persons with low ratios of auto time to auto cost have low values of β_2 (and, hence, high values of time) and persons with high ratios have high values of β_2. The mean value of β_2 is -2.5 .

Table 47 presents the parameter estimates and test results for MNL models fitted to the data generated by this model. The parameters are underestimated in all the models. The models fail the test against a more general model, the variance test, and the means test for the rail alternative; the other tests, however, pass.

Table 47 Choice Based on Model in Which Value of Time Varies and Attributes Depend on the Value of Time

Dependent variable is CHOICE5

	Parameters (t-statistics in parentheses)			
	(25)	(26)	(27)	(28)
Independent Variables	Choice Set: Auto, Bus, Rail	Choice Set: Auto, Bus, Rail	Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus	Conditional Choice Set: Auto, Bus, Rail, Given Auto, Bus
Cost	-.9160 (41.0)	-.9450 (35.8)	-.9024 (33.4)	-.9193 (28.8)
Time	-2.12 (53.6)	-2.065 (47.8)	-2.084 (40.6)	-2.07 (39.3)
Auto Alternative Dummy		.1443 (3.01)		.0546 (.999)
Bus Alternative Dummy		.0197 (.597)		
Log likelihood at zero	-10990	-10990	-5683	-5683
Log likelihood at convergence	-7363	-7359	-3635	-3634
Degrees of freedom	19998	19996	8197	8196

Table 47, continued

Tests:

1A. Test against a more general model:

(a) Log likelihood at convergence of more general model without dummies: -7296
Test-statistic for model (25): 134
Critical chi-squared with 6 degrees of freedom: 12.6
Result: FAIL

(b) Log likelihood at convergence of more general model with dummies: -7278
Test-statistic for model (26): 162
Critical chi-squared with 6 degrees of freedom: 12.6
Result: FAIL

1B. Test against the saturated model:

(a) Log likelihood at convergence of saturated model: -23.71
Test-statistic for model (25): 14,678.58
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: PASS

(b) Log likelihood at convergence of saturated model: -23.71
Test-statistic for model (26): 14,670.58
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: PASS

1C. Test based on conditional choice:

(a) Log likelihood at convergence of model (27) with parameters restricted to values of model (25): -3635
Test-statistic: 0
Critical chi-squared with 2 degrees of freedom: 6.0
Result: PASS

(b) Log likelihood at convergence of model (28) with parameters restricted to values of model (26): -3635
Test-statistic: 2.0

Table 47, continued

Critical chi-squared with 3 degrees of freedom: 7.6
Result: PASS

2A. Means test:

Alternative	Means of Residuals of Model (25)	Result
Bus	.1248	PASS
Rail	6.481	FAIL

2B. Variance test:

(a) Sum of squared residuals for model (25): 377,700
Critical chi-squared with 19,998 degrees of freedom is approximately 19,998.
Result: FAIL

(b) Sum of squared residuals for model (26): 299,700
Critical chi-squared with 19,996 degrees of freedom is approximately 19,996.
Result: FAIL

Aggregation over Alternatives

Assume that the true model has the MNL form of the first section of this Appendix but that the researcher aggregates the bus and rail alternatives into one alternative, called transit. Table 48 presents parameter estimates and diagnostic test results for MNL models fitted to the data aggregated over alternatives in each of three ways. The three aggregation methods are defined by the following formulae for ECOST , MCOST , and ACOST :

$$ECOST^{t} = \log\left(e^{COST^{b}} + e^{COST^{r}}\right) \quad ;$$

$$MCOST^{t} = MIN\left(COST^{b}, COST^{r}\right) \quad ;$$

$$ACOST^{t} = \frac{1}{2}\left(COST^{b} + COST^{r}\right) \quad ;$$

where b and r denote bus and rail, respectively, and t denotes the aggregate transit alternative. For the auto alternative, $ECOST^{a} = MCOST^{a} = ACOST^{a} = COST^{a}$. The three aggregated time variables are created analogously: ETIME , MTIME , and ATIME . ACHOICE is the aggregated choice variable and is equal to one in the transit alternative if the respondent chose either bus or rail and is equal to one in the auto alternative if the respondent chose auto.

The model that gives the best parameter estimates is (30), which includes ECOST , ETIME , and an alternative-specific variable. Model (30) passes all the diagnostic tests, while the other models pass all the tests except the test against a more general model. In this test, the more general model includes the variables of the model being tested plus the following two variables:

1. Cost of auto, taking the described value in the transit alternative and zero otherwise.
2. Time of auto, taking the described value in the transit alternative and zero otherwise.

The conditional choice tests are not applied to these models since the full choice set consists of only two alternatives.

TABLE 48 Aggregation over Alternatives

Dependent variable is ACHOICE

	Parameters (t-statistics in parentheses)					
	(29)	(30)	(31)	(32)	(33)	(34)
Independent Variables	Choice Set: Auto, Transit	Choice Set: Auto, Transit	Choice Set: Auto, Transit	Choice Set: Auto, Transit	Choice Set: Auto, Transit	Choice Set: Auto, Transit
ECOST	-.3521 (28.7)	-1.019 (33.2)				
ETIME	-2.009 (48.4)	-2.345 (47.2)				
MCOST			-1.112 (42.6)	-.9542 (32.2)		
MTIME			-2.697 (45.8)	-2.863 (46.3)		
ACOST					-1.135 (42.8)	-.9542 (32.2)
ATIME					-1.781 (46.3)	-1.981 (46.3)
Auto Alternative Dummy		1.650 (24.8)		-.5663 (11.5)		-.6134 (12.6)
Log likelihood at zero	-6931	-6931	-6931	-6931	-6931	-6931
Log likelihood at convergence	-4744	-4395	-4491	-4425	-4588	-4507
Degrees of freedom	9990	9997	9998	9997	9998	9997

Table 48, continued

Tests:

1A. Test against a more general model

	E-Prefix Variables	M-Prefix Variables	A-Prefix Variables
(a) Log likelihood at convergence of more general model without dummies	-4667	-4466	-4539
Test-statistic	134	50	98
Critical chi-squared with 2 degrees of freedom	6.0	6.0	6.0
Result	FAIL	FAIL	FAIL
(b) Log likelihood at convergence of more general model with dummies	-4935	-4415	-4501
Test-statistic	0	20.0	12.0
Critical chi-squared with 2 degrees of freedom	6.0	6.0	6.0

Table 48, continued

1B. Test against the saturated model:

Log likelihood at convergence of saturated model: -21.60

Model	(29)	(30)	(31)	(32)	(33)	(34)
Test-statistic	9444.8	8746.8	8938.8	8806.8	9132.8	8970.8
Critical chi-squared	9998	9997	9998	9997	9998	9997
Result	PASS	PASS	PASS	PASS	PASS	PASS

2A. Means test:

Model	(29)	(31)	(33)
Mean of residuals of transit alternative	-.8408	.9541	.9807
Result	PASS	PASS	PASS

2B. Variance test:

Model	(29)	(30)	(31)	(32)	(33)	(34)
Sum of squared residuals	1128	986.3	1296	2807	1152	2173
Critical chi-squared	9998	9997	9998	9997	9998	9997
Result	PASS	PASS	PASS	PASS	PASS	PASS

General Conclusions

The diagnostic tests seem to be able to detect deviations of the true model from the MNL form. The most powerful tests seem to be the test against a more general model and the test based on conditional choice. The least powerful test seems to be the test against the saturated model, which every model in this Appendix passed.