CHAPTER 2

RE-ESTIMATION OF THE PRETEST MODE-CHOICE MODEL WITH THE FULL UTDFP SAMPLE

Introduction

In the previous chapter, five issues were identified as requiring further investigation before a usable mode-choice model is obtained. They are: (1) evaluation of the pretest specification on the full sample; (2) choice set and its effects on model coefficients with particular regard to access modes; (3) considerations of choice alternative availability and their effects on model coefficients; (4) examination of the IIA property and its possible violations; and (5) data acquisition for the independent variables in developing travel demand models. Of these the first three are examined here below but the last two are postponed until Part III, Chapters 1 and 5 because the issues involved are complex and require a thorough discussion.

It will become clear from the material in this chapter that we did not proceed to first test specification, then the effects of choice sets, then alternative availability, and so forth. These tasks were undertaken concurrently while following the lead that at the time looked most promising. The form of the work follows the train of thought we pursued when doing our analyses. The analyses are grouped under headings for ease of presentation and understanding but, with some exceptions clearly noted, were done using the models actually estimated and examined. We thought this was a more realistic and appropriate format of reporting than re-doing the work, to straighten out the path we took before we reached our stopping place--the point in time where model estimation must cease and application begin. Clearly, new results in model form and estimation, new data, and new ideas come continuously. It would have been impossible to incorporate even our current understanding of the models in this volume. However, an attempt will be made to show these new or alternative approaches in those contexts where their application might prove productive.

Tests of Model Specification on the Full Sample

A particular model specification was developed using the WTS pretest sample of 161 workers. Ideally, precisely the same specification ought now to be used as the "basic" model specification on the full UTDFP sample. This was not quite possible: four variables not easily available for the full sample were removed. These were an index of distance to parking at home, an index of population density in neighborhood, a dummy variable for the age of respondent, and a dummy variable for the existence of a child in the household. Of these, the variables for "density" and "child in household" has a reasonably high level of significance in the pretest sample. The decision to drop them was based on definitional difficulties. Another change in the specification was to change the boundaries of the income groups to account for the inflation that occurred between the dates of the WTS and UTDFP samples. The modified model appears in Table 7.

The comparison of coefficients of Model 7 with those of Model 2 or 3 reveals that there is a good consistency in model coefficients. Only the walk time coefficient is substantially different from the full sample model. The second components of the headway and the income coefficients is also different, but the effect of the former is probably captured in the transfer-wait time coefficients. The income coefficients have large standard errors in both samples. The reason for the highly different walk time coefficient may be found in the data calculations. Appendix I contains a brief description of the methods used in calculating the travel time data (for a full description see Reid, *et al.*, 1975).

The results in Table 7 give support to the initial model specification. Thus, there is good reason to believe that it forms a solid foundation for further analyses.

Before examining the other issues mentioned in the introduction, it is desirable to conduct some further testing of the model specification. The most obvious and necessary extension of the model is to expand it to include the "share-ride" mode. Table 8 gives the coefficients of such a model for the "basic" model specification. A brief look at the coefficients in Table 8 shows that, excluding the alternative-specific dummy variables and the coefficient for the number of drivers, they have remained practically unchanged.

TABLE 7 Basic Three-Alternative Model on D-Subscripted Version #4 Variables[1]

(Mode 1 - Auto; Mode 2 - Bus, Walk Access; Mode 3 - Bus, Auto Access)

Model: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable	Estimated Coefficient	T-Statistic
Cost divided by post-tax wage, in cents divided by cents per minute	-.0368	5.90
On-vehicle time, in minutes	-.0182	2.18
Walk time, in minutes a/	-.0555	5.09
Transfer-wait time, in minutes a/	-.0509	2.22
Number of transfers a/	.0184	0.148
Headway of first bus, with a ceiling of 8 minutes, in minutes a/	-.118	3.59
Headway exceeding 8 minutes of first bus, in minutes a/	-.00507	0.453
Family income with ceiling of $7500, in $ per year b/	.000129	1.40
Family income minus $7500 with floor of $0 and ceiling of $3000, in $ per year b/	.000113	0.726

[1]For definitions of D-subscripted version #4 variables, see Appendix I.

Table 7, continued

Independent Variable	Estimated Coefficient	T-Statistic
Family income minus $10,500 with floor of $0 and ceiling of $5000, in $ per year b/	-.0000835	1.08
Length of residence in community, in years b/	.128	3.95
Number of persons in household who can drive c/	.871	5.11
Auto alone alternative dummy d/	-4.75	5.99
Bus with auto access dummy e/	-3.29	8.72

Likelihood ratio index: .6206
Log likelihood at zero: -847.0
Log likelihood at convergence: -321.4
Percent correctly predicted: 82.88

Values of time saved as a percent of wage:

On-vehicle time	50
Walk time	151
Transfer-wait time	138

All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 771 .

Table 7, 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 auto-alone and bus 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 carpool alternative and zero otherwise.

TABLE 8 Basic Four-Alternative Model on D-Subscripted Version #4 Variables

(Mode 1 - Auto Alone; Mode 2 - Bus, Walk Access; Mode 3 - Bus, Auto Access; Mode 4 - Carpool)

Model: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable	Estimated Coefficient	T-Statistic
Cost divided by post-tax wage, in cents divided by cents per minute	-.0386	6.90
On-vehicle time, in minutes	-.0180	2.31
Walk time, in minutes a/	-.0578	5.43
Transfer-wait time, in minutes a/	- .0491	2.25
Number of transfers a/	.0596	0.519
Headway of first bus, with a ceiling of 8 minutes, in minutes a/	-.127	4.20
Headway exceeding 8 minutes of first bus, in minutes a/	-.00896	0.843
Family income with ceiling of $7500, in $ per year b/	-.00000667	-.0879
Family income minus $7500 with floor of $0 and ceiling of $3000, in $ per year b/	.000110	.925
Family income minus $10,500 with floor or $0 and ceiling of $5000, in $ per year b/	-.0000714	1.32
Length of residence in community, in years b/	.0784	3.05

Table 8, continued

Independent Variable	Estimated Coefficient	T-Statistic
Number of persons in household who can drive c/	.236	2.28
Auto alone alternative dummy d/	-2.99	4.56
Bus with auto access dummy e/	-2.36	7.44
Carpool alternative dummy f/	-3.11	6.98

Likelihood ratio index: .3447
Log likelihood at zero: -1069.0
Log likelihood at convergence: -700.4
Percent correctly predicted: 60.96

Values of time saved as a percent of wage:

On-vehicle time	47
Walk time	150
Transfer-wait time	127

All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 771 .

Table 8, 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 auto-alone and bus-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 carpool alternative and zero otherwise.

It is reasonable to suspect that the number of drivers is one of the key variables affecting the choice between drive alone and shared-ride alternatives. Another point to note from the comparison of Models 7 and 8[1] is that the percent correctly predicted is higher for Model 7 than that for Model 8. This is as expected because in a three-alternative model (Table 7) chance would predict 33.3 percent correctly, whereas in a four-alternative model (Table 8) chance would be only twenty-five percent correct.

Studies by Ben-Akiva (1975) and others have shown that there are several other variables that affect mode-choice but were not in the specification of the basic model developed in Train and McFadden (1976). These variables are (1) a dummy variable indicating whether or not the respondent is the head of his household, (2) employment density at the work zone, (3) a discrete variable indicating whether the residence is located in, near, or outside the central business district, and (4) the number of autos available per driver. The first three variables are specific to the auto-alone alternative, while the fourth enters the auto-alone and bus-with-auto-access alternatives.

Model 9 is the estimate of a model including these variables. The coefficients of the variables have the expected signs. A household's head has a higher probability of taking auto-alone than a person who is not the head. Employment density at work and the variable indicating a home location in or near the CBD are proxies for unincluded variables, such as difficulty in finding parking places at work and home, which correlates with the unpleasantness of auto travel. As the number of autos per driver increases in a household, the probability that the worker will choose a mode using auto increases (see Train, 1976a). Comparing Model 9 with Model 8 indicates that including these four variables increases the predictive power of the model and somewhat increases the estimated values of time.

[1]Models are identified by their table number, i.e., the model in Table 8 is Model 8.

TABLE 9 Model 8 with Several Variables Added

(Mode 1- Auto Alone; Mode 2 - Bus, Walk Access; Mode 3 - Bus, Auto Access; Mode 4 - Carpool)

MODEL: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable	Estimated Coefficient	T-Statistic
Cost divided by post-tax wage, in cents divided by cents per minute	-.0349	5.64
On-vehicle time, in minutes	-.0202	2.56
Walk time, in minutes a/	-.0566	5.13
Transfer-wait time, in minutes a/	-.0457	2.11
Number of transfers a/	.0715	0.615
Headway of first bus, with a ceiling of 8 minutes, in minutes a/	-.123	3.98
Headway exceeding 8 minutes of first bus, in minutes a/	-.00846	0.805
Family income with ceiling of $7500, in $ per year b/	-.00000500	0.0567
Family income minus $7500 with floor of $0 and ceiling of $3000, in $ per year b/	-.0000632	0.476
Family income minus $10,500 with floor of $0 and ceiling of $3000, in $ per year b/	-.0000826	1.39
Length of residence in community, in years b/	.0574	2.01

Table 9, continued

Independent Variable	Estimated Coefficient	T-Statistic
Number of persons in household who can drive c/	.356	2.90
Auto-alone alternative dummy d/	-4.97	5.94
Bus-with-auto-access dummy e/	-5.17	10.25
Carpool alternative dummy f/	-2.94	6.37
Dummy if person is head of household b/	.503	2.73
Employment density at work location b/	-.00129	1.84
Home location in or near CBD (2=in CBD, l=near CBD, 0 otherwise) b/	-.319	2.73
Autos per driver with a ceiling of one c/	3.18	8.89

Likelihood ratio index	.4031
Log likelihood at zero	-1069.0
Log likelihood at convergence	-638.0
Percent correctly predicted	65.24

Values of time saved as a percent of wage:

On-vehicle time	58
Walk time	162
Transfer-wait time	131

All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 771 .

Table 9, 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 auto-alone and bus 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 carpool alternative and zero otherwise.

Including autos per driver presents statistical problems, because the number of autos owned is probably an endogenous variable, determined simultaneously with the choice of work-trip mode. In econometric models with continuous dependent variables, treating an endogenous variable as exogenous leads to inconsistent estimates. McFadden (1975b) has shown, however, that when two quantal choices are made simultaneously, the inclusion of the outcome of one choice as an explanatory variable in the logit model of the other choice need not produce an inconsistent estimator. This result is dependent upon the assumption that the unobserved components of utility entering the first choice are statistically independent of the unobserved components of utility entering the second choice. If this assumption cannot be made, then an inconsistent estimate results from including the outcome of one choice as an explanatory variable in the model for the other choice. The treatment of some endogenous variables as exogenous variables is discussed more rigorously in McFadden (1975b).

In order to explore to what extent the autos per driver variable may confound the coefficients of the other variables, a model including all the variables of Model 9 except autos per driver was estimated. The coefficients of such a model were practically the same as for Model 9. Thus, the consistency question does not appear to be important in this particular case.

Another specification issue concerns the use of generic cost and on-vehicle time variables, implicitly assuming that auto costs and on-vehicle time are valued the same as bus costs and on-vehicle time, respectively. In Model 10 auto and bus costs and times are allowed to take different values. The estimates in Model 10 indicate that bus cost and auto cost are valued about the same and that, contrary to initial expectations, auto on-vehicle time is considered more onerous than bus on-vehicle time.

This apparent aberration was at first thought attributable to the large amount of auto on-vehicle time spent on congested freeways. To test this, auto time was divided into two components: congestion time and non-congestion time The estimated coefficient of auto congestion time was about thirty percent larger (in magnitude) than that of auto non-congestion time, but auto non-congestion time was still valued more highly than bus on-vehicle time.

TABLE 10 Model 10: Model 6 with Non-Generic Cost and Time Variables

(Mode 1- Auto Alone; Mode 2 - Bus, Walk Access; Mode 3 - Bus, Auto Access; Mode 4 - Carpool)

Model: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable	Estimated Coefficient	T-Statistic
Auto cost divided by post-tax wage, in cents divided by cents per minute	-.0316	5.42
Bus cost divided by post-tax wage, in cents divided by cents per minute	-.0262	1.84
Auto on-vehicle time, in minutes	-.0443	4.17
Bus on-vehicle time, in minutes	-.0199	2.51
Walk time, in minutes a/	-.0716	6.12
Transfer-wait time, in minutes a/	-.0510	2.40
Number of transfers a/	-.0964	0.798
Headway of first bus, with a ceiling of 8 minutes, in minutes a/	-.0961	3.08
Headway exceeding 8 minutes of first bus, in minutes a/	-.0196	1.72
Family income with ceiling of $7500, in $ per year b/	-.00000103	.0137
Family income minus $7500 with floor of $0 and ceiling of $3000, in $ per year b/	.000117	0.980
Family income minus $10,500 with floor of $0 and ceiling of $5000, in $ per year b/	-.0000601	1.10
Length of residence in community, in years b/	.0769	2.97

Table 10, continued

Independent Variable	Estimated Coefficient	T-Statistic
Number of persons in household who can drive c/	.218	2.118
Auto alternative dummy d/	-2.44	3.61
Bus-with-auto-access dummy e/	-2.67	7.86
Carpool alternative dummy f/	-2.09	4.07

Likelihood ratio index:. .3523
Log likelihood at zero: -1069.0
Log likelihood at convergence: -692.3
Percent correctly predicted: 61.09

All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 771 .

---

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 auto-alone and bus-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 carpool alternative and zero otherwise.

It could be the case that the coefficients of other bus variables, such as bus headway, which has an unexpectedly large coefficient, are picking up the effects of the attributes of bus travel, such as lack of comfort and privacy, which are usually considered to make bus on-vehicle time more onerous than auto time. If these attributes are being accounted for, then perhaps bus on-vehicle time is considered less onerous than auto time because one can read on a bus and need not bother with the driving.

The preceding analyses suggest the following modifications to the initial model specification. First, the models estimated with three and four alternative choice sets indicate that the number of drivers need to be entered on all the alternatives including auto; by indication the same applies to the autos per driver variable. Second, the coefficient for the second component of the headway, headway exceeding eight minutes, has an unstable coefficient. The sample does not appear to be rich enough to permit the estimation of an independent coefficient for that variable. Two courses of action are possible. One, to add the headway exceeding eight minutes together with the transfer-wait variable, or two, to estimate one coefficient for the first headway as a simple variable. This latter course is intuitively more appealing and is followed here. The third modification which is warranted for the initial model specification is the estimation of separate coefficients for auto and transit on-vehicle times. Finally, the variable "length of residence in community" is deleted because of the difficulties involved in its prediction if a model involving such a variable is to be used for prediction, notwithstanding the fact that it appears to have a stable and statistically significant coefficient.[1] The model incorporating these changes was estimated and is shown in Table 11. It is evident from examining the likelihoods at convergence of these models that Model 11 is a statistically superior model to all the others.

This section would be incomplete without a few words being said about so-called "naive" models. Naive models are those that include only a few key "policy" variables, such as cost, on-vehicle time, and "excess" time, (excess time being defined as the sum of walk time, transfer-wait time, and one-half of first headway). These naive models are often employed when time or resources do not permit the acquisition or development of all the socioeconomic variables included in the model specification developed earlier in this chapter.

---

[1]To the extent that the omitted variable is correlated with policy variables, biased coefficients will tend to produce erroneous policy conclusions, and a better method would be to include the variable and utilize crude predictors for its future values. In fact, the correlation of the variable with other variables in the model is low.

TABLE 11 Work-Trip Mode-Choice Model, Estimated Pre-BART

(Mode l--Auto Alone; Mode 2--Bus, Walk Access; Mode 3--Bus, Auto Access; Mode 4--Carpool)

Model: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable (The variable takes the described value in the alternatives listed in parentheses and zero in non-listed alternatives)	Estimated Coefficient	T-Statistic
Cost divided by post-tax wage, in cents divided by cents per minute (1-4)	-.0284	4.31
Auto on-vehicle time, in minutes (1,3,4)	-.0644	5.65
Transit on-vehicle time, in minutes (2,3)	-.0259	2.94
Walk time, in minutes (2,3)	-.0689	5.28
Transfer-wait time, in minutes (2,3)	-.0538	2.30
Number of transfers (2,3)	-.105	0.776
Headway of first bus, in minutes (2,3)	-.0318	3.18
Family income with ceiling of $7,500, in $ per year (1)	.00000454	0.0511
Family income minus $7,500 with floor of $0 and ceiling of $3,000, in $ per year (1)	-.0000572	0.430
Family income minus $10,500 with floor of $0 and ceiling of $5,000, in $ per year (1)	-.0000543	0.907

Table 11, continued

Independent Variable	Estimated Coefficient	T-Statistic
Number of persons in household who can drive (1)	1.02	4.81
Number of persons in household who can drive (3)	.990	3.29
Number of persons in household who can drive (4)	.872	4.25
Dummy if person is head of household (1)	.627	3.37
Employment density at work location (1)	-.00160	2.27
Home location in or near CBD (2=in CBD, l=near CBD, 0 otherwise) (1)	-.502	4.18
Auto per driver with a ceiling of one (1)	5.00	9.65
Autos per driver with a ceiling of one (3)	2.33	2.74
Autos per driver with a ceiling of one (4)	2.38	5.28
Auto-alone alternative dummy (1)	-5.26	5.93
Bus-with-auto-access dummy (3)	-5.49	5.33
Carpool alternative dummy (4)	-3.84	6.36

Likelihood ratio index:	.4426
Log likelihood at zero:	-1069.0
Log likelihood at convergence:	-595.8
Percent correctly predicted:	67.83

Table 11, continued

Values of time saved at a percent of wage (t-statistics in parentheses):

Auto on-vehicle time	227	(3.20)
Transit on-vehicle time	91	(2.43)
Walk time	243	(3.10)
Transfer-wait time	190	(2.01)

Value of initial headways as a percent of wage: 112 (2.49)

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

Number of people in sample who chose

Auto-alone	429
Bus-with-walk-access	134
Bus-with-auto-access	30
Carpool	178
Total sample size	771

A naive model having the specification sketched out above and alternative-specific dummy variables is given in Table 12. It shows the coefficient for excess time (out-of-vehicle time) is comparable to the estimate obtained in the complex model, Model 11, and the coefficient of the on-vehicle time is also comparable to the on-vehicle time coefficients obtained in those complex models having generic specification of the on-vehicle time. However, the coefficient of the "cost/wage" variable has nearly doubled, thus lowering the values of time by nearly fifty percent. Nothing can be said about the magnitudes of the coefficients for the alternative-specific dummy variables because the naive model does not incorporate any socioeconomic variables that, it may be recalled, are always the product of the socioeconomic and alternative-specific dummy variables.

The examination of the summary statistic, the likelihood ratio index, percent right, and especially the log likelihood at convergence, shows that the naive model is a substantially worse predictor than the more complex models, provided that the prediction of the socioeconomic variables can be reliably done for the forecasting date.

The discussion turns next to the examination of the effects of the number of alternatives in the coefficient estimates.

TABLE 12 Naive Four-Alternative Model #3 on D-Subscripted Version #4 Variables: Cost Divided by Wage, On-Vehicle Time, Excess Time and Alternative Specific Dummies

(Mode l--Auto Alone; Mode 2--Bus, Walk Access; Mode 3--Bus, Auto Access; Mode 4--Carpool)

Model: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable	Estimated Coefficient	T-Statistic
Cost divided by post-tax wage, in cents divided by cents per minute	-.0412	7.63
On-vehicle time, in minutes	-.0201	2.78
Excess time, in minutes	-.0531	7.54
Auto alternative dummy d/	-.892	3.38
Bus-with-auto-access dummy e/	-1.78	7.52
Carpool alternative dummy f/	-2.15	8.56

Likelihood ratio index	.3285
Log likelihood at zero	-1069.0
Log likelihood at convergence	-717.7
Percent correctly predicted	58.50

Value of on-vehicle time saved is 49 percent of wage. Value of excess time saved is 129 percent of wage.

Table 12, continued

All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 771.

---

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 carpool alternative and zero otherwise.

Coefficient Estimates and the Choice Set

The stability of the coefficient estimates with respect to choice set was examined from three different angles: alternative availability to individuals in the sample, aggregation of alternatives in the choice set, and the number of alternatives in the choice set. The tests were conducted using two model specifications: the "basic model specification" (Model 7) and the "final model specification" (Model 11).

Before discussing the results, it is instructive to speculate on what they ought to be. The theoretical underpinnings of the MNL model require that the choice set include only those alternatives that have a positive probability of choice, i.e., at least a small chance of being chosen. The alternative whose availability has concrete restrictions is auto-alone. Unless one owns a car and a license it is impossible to drive to work; for such an individual the probability of choosing auto-alone is zero. Thus, in comparing models with and without restrictions on alternative availability--governed by car (and license) ownership status--it is expected that the coefficients of these governing variables would be affected the most and that their values would be higher in models where no choices were eliminated by these *a priori* considerations. This is because cases where the values of the other attributes would suggest auto use had to be made undesirable by a high coefficient on "cars per driver" or "drivers" variables.

Similar reasoning can be applied when anticipating the effects of aggregation of alternatives on the coefficient estimates. Assume, for instance, that auto-alone and shared-ride alternatives are added together to form one "auto" alternative characterized in terms of auto-alone attributes. For the carpoolers the cost of their trip is now overestimated and the on-vehicle time is underestimated. Because the model is unable to change the values of the explanatory variables it changes the coefficients of the variables instead. In the example of the aggregated auto alternative, the cost coefficient would be decreased and the on-vehicle time coefficient increased.

In the case when some alternative having a non-zero choice probability is dropped from the choice set, no change should take place in the coefficient estimates if the MNL is the true model. This result holds because of the IIA property of MNL model. It is the reverse of the "new mode" forecasting situation: new modes (alternatives) can be added to the choice set without adjusting the coefficients. The reverse is also true, alternatives can be removed from the choice set without any effect on the coefficients if the MNL is the true model.

The results of the tests are shown in Tables 13 and 14. These tables have a "box" for each variable and choice set. In each box there are two numbers. The upper value pertains to the case where all the alternatives are available to every one; the lower value is for the case when auto-alone is not available for those workers who have no cars or licenses in their household.

The numbers in these two tables support the following conclusions: Alternative availability does not appear to have a large effect on the coefficient estimates; the two numbers in each box are reasonably close to each other. Whatever visible differences appear are, as expected, confined to the governing "cars per driver" and "driver" variables. Even these minor differences are compensating by a change in the alternative-specific dummy variable(s). The index of percent correctly predicted is, however, improved by a small amount.

The effect of the number of alternatives still has some minor effects on coefficients. Surprisingly, this holds true whether or not an alternative was simply dropped from the choice set or aggregated with a similar alternative. The only coefficient whose value is changed by aggregation of alternatives is the walk time coefficient for "bus" mode. This mode was characterized in terms of "bus-with-walk-access" attributes, which overstate the walk time for those who have access to bus by car resulting in low walk time coefficient. The other coefficients where instability takes place are the headway and transfer time coefficients. These instabilities are most likely derivatives of specification and data than of the choice set. Note that the instability that occurs in the "drivers" coefficient has its origin either in statistics (large standard error, the low value of the coefficient being compensated by the coefficient of the alternative-specific dummy variables) or in what alternatives the variable took a non-zero value. The experience with the final model specification indicates that the coefficient for "drivers" is a stable one.

Taken together the empirical and theoretical evidence indicates the MNL model is robust against the inclusion of alternatives in the choice set that have zero or a very small chance of being chosen. The MNL model is also robust against possible violations of the IIA property; alternatives could be dropped from the choice set without too much damage to the coefficient estimates. On the other hand, the insensitivity of the model for the aggregation over alternatives may be an indication of some "trouble." This "trouble" is likely to be related to the "shared-ride" mode because its aggregation with "auto-alone" was reflected only in the alternative-specific dummy variable’s coefficient. We will return to these considerations in Part III, Chapters 1 and 2.

TABLE 13 Selected Coefficient Estimates for Different Choice Sets

"Basic Model Specification" (Table 7)

Choice set / Variable	Four alternatives: Auto Alone (1), Shared Ride, Bus with Walk, Bus with Auto	Three alternatives: Auto Alone (1), Shared Ride, Bus with Walk	Three alternatives: Auto Alone & Shared (1) Ride, Bus with Walk, Bus with Auto	Two alternatives: Auto Alone & Shared (1) Ride, Bus	Two alternatives: Auto Alone (1), Bus with Walk	Comments
Cost/wage	-.0386[a] -.0408[b]	-.0329 -.0344	-.0368 N.A.	-.0343 N.A.	N.A. .0499	
On-Vehicle Time	-.0180 -.0180	-.0201 -.0210	-.0182 N.A.	-.0233 N.A.	N.A. .0199*	
Walk Time	-.0578 -.0579	-.0430 -.0422	-.0555 N.A.	-.0240 N.A.	N.A. .0380	
Headway with 8 min max	-.127 -.128	-.0805 -.0808	-.118 N.A.	-.109 N.A.	N.A. -.0965	
Headway Exceeding 8 min	-.00896* -.00734*	-.0273 -.0260	-.00507* N.A.	-.0113* N.A.	N.A. -.0177*	
Transfer Time	-.0491 -.0455	-.0780 -.0723	-.0508 N.A.	-.0536 N.A.	N.A. -.0809	
Drivers in the Household (1)	.236 .0298*	.167 .156*	.871 N.A.	.550 N.A.	.299* N.A.	Entered alt. (1) (Auto Alone) only
Auto Alone Dummy	-2.99 -1.98	-2.21 -1.01*	-4.75 N.A.	-3.80 N.A.	N.A. -3.10	
Percent Right	61.0 64.3	63.4 66.3	82.3 N.A.	83.5 N.A.	N.A. 86.9	

*t - value less than 1.5

a/ upper number: all alternatives available;

b/ lower number: auto alone not available if no cars or drivers in household

TABLE 14 Selected Coefficient Estimates for Different Choice Sets

"Final Model Specification" (Table 11)

	Four alternatives: Auto Alone (1), Shared Ride, Bus with Walk, Bus with Auto	Three Alternatives: Auto Alone (1), Shared Ride, Bus with Walk	Two Alternatives: Auto Alone (1), Bus with Walk	Comments
Cost/wage	-.0284 -.0283	-.0240 -.0239	N.A. -.0311	
Auto On-Vehicle Time	-.0644 a/ -.0650 b/	-.0599 -.0605	N.A. -.0754	
Bus On-Vehicle Time	-.0259 -.0260	-.0280 -.0282	N.A. -.0242	
Walk Time	-.0689 -.0697	-.0533 -.0538	N.A. -.0567	
First Headway	-.0318 -.0318	-.0385 -.0387	N.A. -.0310	
Transfer Time	-.0538 -.0538	-.0895 -.0890	N.A. -.0974	
Drivers in the Household (1)	1.02 .905	.979 .803	N.A. 1.136	Entered alternative 1 (Auto alone) only
Cars per Driver (1)	5.00 4.42	4.91 4.30	N.A. 5.51	Entered alternative 1 (Auto alone) only
Auto alone Dummy	-5.26 -4.55	-5.08 -4.34	N.A. -6.15	
Percent Right	67.8 68.1	69.0 69.8	N.A. 89.8	

a/ b/ See footnotes in the preceding table.