CHAPTER 5

TRANSFERABILITY OF MODE CHOICE MODELS OF URBAN TRAVEL

Introduction

Transferability is an important issue in travel demand analysis and forecasting. Without transferability in time, the use of models for forecasting has no foundation; without transferability between regions, models developed in one region cannot be applied elsewhere; and without transferability of models from one part of a region to another part of the same region, taste and value differences then would make the use of models a fortuitous exercise. In sum, if transferability of models of travel behavior is not observed, the content of the underlying theory does not have a firm basis. The task in this chapter is to conduct a series of tests to examine whether transferability of extant mode choice models is a reasonably maintained hypothesis.

In order to lay a solid foundation for the statistical tests of model transferability, several issues have to be resolved. Among these are the acquisition of data and their consistency checks, and development and evaluation of good, alternative model specifications that can be estimated with all the data sets. These tasks are described below followed by the transferability tests themselves.

Data and Comparative Model Specification

Four data sets were assembled for this work. These were from the Washington, DC metropolitan area, the Minneapolis-St. Paul area, and the two Urban Travel Demand Forecasting Project surveys from the San Francisco Bay Area conducted before and after the introduction of BART service.

The Washington, DC data are a standard Home Interview survey conducted by the Washington Council of Governments in 1968. The file contained transportation network level-of-service system attributes, and land use and employment variables at the zonal level. The data were processed to select "pure" home-based round trips from home to work and back. Of the 70,000 round trips in the file, approximately 24,000 were single-stop work-trip tours.

The Twin Cities data, collected in 1970, were in a form similar to the Washington data. However, employment and land use measures were not available in this sample. Of the 16,275 round trip tours in the file, 3,956 were single-stop work-trip tours.

The San Francisco Bay Area survey, conducted before the introduction of BART service, hereafter called pre-BART, was a stratified random sample in which certain zones were sampled more heavily. The file contains 996 work round trips.

The post-BART survey was a re-survey of the pre-BART sample. This survey is only partially the same as the pre-BART survey owing to natural attrition through moving and other causes and to the changes in the employment status of members in the sampled households (Johnson, 1976).[1]

The comparative model specification used as a starting point for data analysis and variable identification was that developed by Atherton and Ben-Akiva (1976) with Washington, DC data. This model, shown in Table 61, is a good starting point because its coefficients have also been transferable, at least in visual similarity, to New Bedford and Los Angeles. These three cities appear superficially quite different, which speaks well for the model specification and estimated parameter values.

[1]So-called APS-sample and the choice-based sample that were added later to augment the original sample are not included, because the former lacks some needed information and because of the choice-based nature of the sample of the latter (Johnson, 1976, Vol. II).

Table 61 Basic Model Specification

Alternatives: 1 - Car, drive-alone
2 - Bus-with-walk-access
3 - Car, share-ride

Mneumonic	Alternatives	Description	Coefficient	(t-value)
OPTC/INC	1,2,3	Out-of-pocket cost divided by household income (cents/$)	-28.8	(2.26)
IVTT	1,2,3	On-vehicle travel time (minutes)	-0.0154	(2.67)
OVTT/D	1,2,3	Out-of-vehicle travel time divided by distance (minutes/miles)	-0.160	(4.08)
$AALD_1$	1	Autos available per licensed driver	3.99	(10.08)
$AALD_3$	3		1.62	(5.31)
CBD_1	1	Central business district (dummy variable)	-0.854	(2.75)
CBD_3	3		-0.404	(1.36)
GW_1	1	Government worker (dummy variable)	0.287	(1.78)
$DINC_{1,3}$	1,3	Disposable income (dollars)	.0000706	(3.46)
$DTECA_3$	3	One way distance (miles) times employees per commercial acre	.000653	(1.34)
$NWORKERS_3$	3	Number of workers in family	.0983	(1.03)
BW_1	1	Breadwinner (dummy variable)	.890	(4.79)
Alt 1	1	Alternative-specific dummies	-3.24	(6.81)
Alt 3	3	Alternative-specific dummies	-2.24	(5.60)

These data, together with the comparative model specification and past experience with both pre-BART and post-BART model development, introduced some restrictions upon the choice set and model specification. The "common" choice set can have at most three alternative modes: drive alone, bus-with-walk-access, and shared-ride (two or more occupants in the vehicle). Thus, in the post-BART sample, the respondents choosing BART and the BART alternatives were eliminated. If the logit model is the "true" model, this procedure should cause no bias in the estimated parameters (McFadden, 1976); this argument is supported by the IIA diagnostic tests in Part IV, Chapter 1.

The "common" model specification had to be tailored also. Studies done earlier indicated that statistically significant improvements in a mode choice model were possible with the use of separate walk and wait times rather than out-of-vehicle time, and travel cost divided by personal wage rate rather than household income (Train, 1976a; Talvitie, 1976b). However, these variables were not available in the Washington data; the dummy variable for "government worker" (a proxy for carpooling incentive) that appears in Table 61 had to be excluded altogether because this variable was unavailable in other data sets. The employees per commercial acre was unavailable in the Minneapolis file and the variable employment density times trip distance (DTECA), which entered the shared-ride alternative, could not be included in the Minneapolis models. In principle, if a variable was available in two data sets (thus allowing at least one comparison), it was included.

The bus out-of-vehicle travel time (OVTT) was defined to be the sum of walk time to and from the stop and wait time, defined as one-half the initial headway plus one-half the transfer headways. In the Washington data out-of-vehicle travel time was available only as transit excess time. Lacking more specific information we presumed that the construction of this excess time was done by the above described (UTPS-default) manner. In the drive-alone and shared-ride alternatives OVTT was available as terminal times at the origin and destination zones in all data sets except the Bay Area. For the pre- and post-BART models the terminal times were constructed as a function of employment density.

In-vehicle travel time (IVTT) in the shared-ride alternative was defined to be 12.5 minutes greater than IVTT in the drive-alone alternative.

The breadwinner (BW) was defined as the head of the household if the head worked, or the spouse of the head if the head did not work but the spouse did.

The central business district (CBD) was defined in the Washington and Bay Area samples as zones with more than 100 employees per commercial acre. In the Minneapolis data the CBD's included the downtown Minneapolis and St. Paul zones, whose definition was not included in the description of the received data.

Disposable income was normally defined as household income divided by the number of workers.

The estimation results for the basic model specification with six different data sets were given in Table 62. These six data sets are: two Washington, DC samples, a five percent sample (WASH5) from the entire file and a seven percent sample (WASH7); two Minneapolis samples, a twelve percent sample (MINN12) and a twenty-five percent sample (MINN25) from the entire file, the pre-BART sample, and the post-BART sample.

The alternative availability in all these models was governed by the following considerations. The drive-alone alternative was unavailable if the respondent did not have a driver's license or if his household did not own a car. The bus mode was unavailable when the network pathbuilder was unable to complete a path on bus from the origin to the destination.

Superficial examination of the coefficients of the six models reveals that there are order of magnitude similarities and dissimilarities between the sets of coefficients depending on the observer. Clearly, a more objective and rigorous means of analysis must be employed to do the comparison. The results of such an undertaking are reported in the next three sections.

TABLE 62 Logit Estimation Results (t-values)

Variable	Alternatives Variables Entered	WASH 5%	WASH 7%	MINN 12%	MINN 25%	Pre-BART	Post-BART
OPTC/INC	1,2,3	-25.5 (-3.14)	-40.8 (5.08)	-16.2 (-1.34)	-10.11 (0.886)	16.75 (-2.47)	-38.2 (02.11)
IVTT	1,2,3	-0.009467 (-2.11)	-0.02598 (-1.33)	-0.0263 (-3.54)	-0.0162 (-2.97)	-0.0388 (-5.34)	-0.0379 (-3.05)
OVTT/D	1,2,3	-0.155 (-3.78)	-0.116 (-3.18)	-0.134 (-4.05)	-0.0758 (-3.64)	-0.168 (-4.26)	-.183 (-2.12)
$AALD_1$	1	2.70 (7.80)	2.50 (7.76)	.0682 (.245)	1.02 (2.91)	1.81 (5.08)	3.43 (4.22)
$AALD_3$	3	.996 (3.58)	1.63 (6.04)	-.853 (-2.64)	-.128 (-.361)	1.46 (4.63)	2.79 (3.62)
BW_1	1	1.59 (7.50)	1.44 (7.84)	1.00 (6.04)	.889 (6.55)	.659 (3.77)	.698 (2.35)
CBD_1	1	-.585 (-1.94)	-.913 (-3.08)	-1.34 (-3.37)	-1.27 (-3.87)	-8.22 (-2.92)	-.953 (-1.75)
CBD_3	3	-.0701 (-.243)	-.596 (-2.09)	-.421 (-1.19)	-.623 (-2.11)	-.0187 (-.0603)	-.186 (-.330)

$DINC_{1,3}$	1,3	.0000437 (1.88)	.0000449 (1.90)	.000126 (3.76)	.000123 (4.51)	.0000465 (2.49)	.0000337 (.704)
$DTECA_1$	3	.0000603 (1.21)	.0000783 (1.45)	--	--	-.0000315 (-.617)	.0000196 (.387)
$NWORKERS_3$	3	-.179 (-1.38)	.162 (1.48)	.150 (.133)	-.0179 (-.192)	.340 (2.66)	.465 (2.40)
Drive Alone Dummy	1	-2.00 (-4.66)	-9.54 (-2.48)	-1.74 (-2.88)	-1.26 (-2.84)	-2.53 (-5.10)	-2.50 (-2.67)
Shared Ride Dummy	3	-.942 (-2.40)	-1.41 (-4.03)	-1.54 (-2.67)	-.873 (-2.03)	-3.27 (-7.15)	-3.27 (-3.46)
Cases		1143	1388	847	1223	858	396
Percent Chosen Alternative	1 2 3	76.2 14.8 17.4	75.7 14.7 17.6	64.7 9.5 29.0	65.9 8.7 28.8	65.2 18.6 23.	67.7 12.9 24.5
Log Likelihood	0 $\hat{\beta}$	-1198. -663.4	-1453. -803.6	-876.2 -578.7	-1274. -858.2	-876.9 -631.8	-392.1 -260.2
ρ		.466	.477	.340	.326	.280	.336
Percent correct		77.6	76.5	68.5	67.3	67.1	70.4

Data Checks for Outliers - Metrical Trimming

Our previous experience with the choice models and the observed dissimilarity in some of the coefficient estimates obtained from the two Washington and two Minneapolis samples led us to suspect that the data "outliers" were also working on the coefficient estimates. A thorough analysis of data for consistency and accuracy would have been a task of enormous magnitude. A short-cut for "throwing out" bad data was employed. One technique for obtaining robust estimators in the presence of "contaminated" data is metrical trimming. A metrically trimmed estimator can be briefly described as follows. Assume that we have an estimate $\hat{\theta}$ of the true parameter θ . A metrically trimmed estimator about $\hat{\theta}$ is one that "throws away" observations that are outside a specified but arbitrary interval around $\hat{\theta}$.

The trimmed logit coefficients were obtained in the following manner. Independent point estimates (from different samples) were used to calculate the choice probabilities for each observation in each sample. Then, if a person's probability of choosing any alternative was less than 0.05 , that person was eliminated from the sample. Specifically, the coefficient estimates from the pre-BART sample were used to trim the Washington and Minneapolis samples, while the coefficient estimates of the Washington seven percent sample were used to trim the Bay Area samples.

The metrical trimming caused the point estimates to change by more than two standard errors, sometimes eliminated more than half the sample, and doubled some mode shares. However, this did not lead to any greater apparent consistency of point estimates among the samples.

The appropriate statistical test of the effects of the metrical trimming is a nested (Chow-like) likelihood ratio test of the null hypothesis that the coefficients estimated with the trimmed subsample are the same as those estimated with the "outliers" included. Separate logit estimations were performed on the two subsamples. Let L^* be the log likelihood of the loglt estimation performed on the entire sample, and let L_T and L_0 be the log likelihoods obtained from the trimmed and "outlier" subsamples, respectively. Then the statistic $-2(L^* - L_T - L_0)$ is distributed chi-square with degrees of freedom equal to the number of restrictions (coefficients in this case) in the estimation (McFadden, 1973). The results of these nested likelihood ratio tests and the number of individual cases in each sample are given in Table 63.

TABLE 63 Results of Hypothesis Tests of Metrical Trimming

Sample	Cases		χ^2 Statistic	Degrees of Freedom*	Critical Values χ^2	
	All	Trimmed			α=.05	α=.10
MINN 12%	847	392	18.8	12	21.03	18.55
MINN 25%	1223	564	46.0**	12	21.03	18.55
WASH 5%	1143	816	17.0	13	22.36	19.81
WASH 7%	1388	1015	14.6	13	22.36	19.81
Pre-BART	858	373	21.4	13	22.36	19.81
Post-BART	396	220	17.6	13	22.36	19.81

*Minneapolis samples did not have $DTECA_3$.

**Significant at five percent level.

We see that the null hypothesis (that the trimming had no statistically significant effect) can be rejected at the five percent level only for the Minneapolis twenty-five percent sample, while at the ten percent level the null hypothesis can be rejected also for the Minneapolis twelve percent and pre-BART samples.

In general, the dissimilarity of point estimates from different samples does not appear to be caused by "contaminated" data according to the traditionally allowable confidence limits. On the other hand, if we set $\alpha = .20$, that is, the upper limit on chances of wrongly believing the trimmed and untrimmed estimates to be unequal are two out of ten, then only WASH7 sample coefficients are equal.

Tests on Model Specification

Three alternatives to the comparative model specification were tried. These re-specifications are: (1) use the out-of-vehicle time (OVTT) instead of the out-of-vehicle time divided by distance (OVTTD); (2) separate the out-of-vehicle time into walk and wait time components (not possible with the Washington, DC samples); (3) make the on-vehicle time non-generic, that is, examine whether the coefficients of the on-vehicle time are the same for car and bus alternatives.

In substituting out-of-vehicle travel time for the same variable divided by one-way distance we used the change in log likelihood at convergence as the principle indicator of the efficacy of the change, even though this is not the proper statistical test for it. As a result of this change in one variable the following increases in the log likelihood were observed for the different samples:

Pre-BART	10.3	Post-BART	6.0
WASH5	14.9	WASH7	7.5
MINN12	-3.8	MINN25	2.8

It can be seen that the log likelihood increased in all but the Minneapolis twelve percent sample, although the increase is relatively small in the Minneapolis twenty-five percent sample. On the average the re-specification resulted in an improvement of about one-half a percentage point in the percent mode-choice correctly predicted.

The proper statistical test for examining the effects of OVTT and OVTTD would again be the nested likelihood ratio test. That is, the model ought to be estimated with both OVTT and OVTTD included and then one or the other excluded; then compare the differences in the log likelihoods (multiplied by -2.0) with the critical χ^2 values with one restriction (degree of freedom). This test was performed only for one sample where the relative impact of OVTT and OVTTD would be most clearly seen, this was the WASH5 sample. The inclusion of OVTT and OVTTD together resulted in the log likelihood of -648.5 . This can be compared with the WASH5 model in Table 62. The χ^2 statistic is $-2(-663.4-(-648.5)) = 29.80$ when the χ^2 critical value (at .05 level) is 3.84 . Thus, the null hypothesis that the coefficient of OVTT is equal to zero is soundly rejected. Because the model with OVTT alone resulted in a log likelihood exactly as high as the model with both OVTT and OVTTD (-648.5) , the null hypothesis that the coefficient of OVTTD is equal to zero is accepted.[1]

Another interesting observation was made when using the out-of-vehicle time instead of out-of-vehicle time divided by distance. In all cases except one, the re-specification caused the coefficient of on-vehicle travel time to become insignificant (.05 level) and even turn positive. The exception is the WASH7 sample where the on-vehicle time coefficient is insignificant (.05 level) in both specifications.

In the second set of tests OVTT was divided into walk and wait time components. The out-of-vehicle travel time versus walk and wait time specification is tested for models with both generic and non-generic on-vehicle times. This set of tests was not possible for the Washington, DC samples because separate values did not exist for walk and wait times.

Again, the test statistic is -2 times the change in log likelihood, which is distributed χ^2 with one restriction. The results of the tests appear in Table 64.

TABLE 64 Chi-Square Statistic: Out-of-Vehicle Time vs. Walk and Wait Times

Sample	Generic Models	Non-Generic Models	Critical χ^2 (1, .05)
Pre-BART	17.2	12.2	3.84
Post-BART	11.8	12.0	3.84
MINN 12%	0.4	0.2	3.84
MINN 25%	0.4	0.2	3.84

[1]Given the data and the results here this is a statistically sound conclusion. Even though the substitution of OVTTD for OVTT yields a "better" (?) coefficient with a plausible sign for an important variable, it would be a contradictory way to use statistical criteria to judge models or model parameters to prefer OVTTD over OVTT. In the final analysis the choice between OVTT and OVTTD, or the rejection of both, must be based on theory arrived at by observation of human behavior. While the data here do not support the postulated behavior embodied in the OVTTD variable, they do not mean it could not exist.

The results in Table 64 tell that the null hypothesis of valuing walk and wait times equally is rejected with a high degree of confidence in the Bay Area samples, while no such conclusions can be drawn from the Minneapolis samples.

The third set of tests was performed comparing generic and non-generic specifications of on-vehicle travel time. The coefficient of IVTT was allowed to vary between alternative two (bus) and alternative one and three (car). The test was performed for models using out-of-vehicle time, and for models using separate walk and wait times. The results are given in Table 65.

TABLE 65 Chi-Square Statistic: Generic On-Vehicle Time vs. Non-Generic On-Vehicle Time

Models with:

Sample	OVTT	Separate Walk and Wait Times
Pre-BART	15.2**	10.2**
Post-BART	0	0.2
WASH 5%	10.2**	--
MINN 25%	3.2*	3.0*
MINN 12%	4.2**	4.0**
	OVTT	
WASH 5%	36.6**	--
WASH 7%	8.4**	--
*Significant at the 10% level. χ^2 - 2.71		
**Significant at the 5% level. χ^2 - 3.84		

The results show that the generic specification can be rejected at the 0.05 level for all data sets except the MINN25 sample, where we can reject at the 0.10 level, and the post-BART sample, where the generic specification appears to be correct.

In summary, some generalizations can be made on the basis of the model specification tests. First, when simple out-of-vehicle time was substituted for out-of-vehicle time divided by distance the log likelihood increased. Even though on-vehicle time becomes insignificant in the re-specification, this re-specified model is accepted as the correct or at least better model, because there does not exist a strong prior behavioral reason to believe that out-of-vehicle time becomes less onerous with increasing distance.

Second, even though walk and wait times appear to be valued similarly in the Minneapolis samples, such is not the case for the Bay Area samples and, in general, the evidence so far is that walk and wait times are valued differently. The reasons for these inconclusive results are unclear. One argument is that the similar valuation of walk and wait times is due to the Twin Cities' climate, which might make waiting as onerous as walking. Another argument was presented in Part IV, Ch. 4 where it was claimed that the walk and wait, and hence also the out-of-vehicle time coefficients are derivatives of the network coding practices and the weights used in building the paths. It is not known who coded the Minneapolis or Washington networks and what rules they obeyed or what weights were used in building the paths.

In building the paths for the Bay Area samples the walk and wait times were weighed equally (though there were restrictions in the maximum transfer wait times). However, the headways were later modified to account for the differential headways that occur even within the peak period. The walk times in the pre-BART sample only were also modified after the completion of the path runs. The procedures for making these modifications are too lengthy to be reported here but can be found in Reid, *et al.* (1975). The coding practices observed in coding the Bay Area samples have escaped behavioral analyses. Nevertheless, a general consensus appears to exist that the walk times have been coded optimistically low (Part I, Chapter 3).

If the walk and wait times were given equal weights in buildings the paths in the, Minneapolis samples, the considerations above suggest that this construction could be producing the result. On the other hand were sufficient perturbations to the data in the Bay Area samples to cause the different coefficients, in spite of the equal weights used in path building. Of course, the coding practices could have the most significant effect of all on the model coefficients. However, such effects are too lengthy to be considered here.

Finally, the generic specification of on-vehicle time seems to be uniformly rejected, while the acceptance of the generic hypothesis in the post-BART sample, if not anomalous, gives the reason to discuss the facts which are of interest regarding genericity. It is generally agreed that genericity of the IVTT would be a desirable property of the model, but, it is argued, such a model will come about only after comfort and convenience and other attitudinal variables are accounted for. Thus, the non-generic on-vehicle time coefficients account also for some unobserved attributes that are dissimilar between the modes. And herein lies the problem with non-genericity: it can be a statistical artifact created by the violation of the independence of irrelevant alternatives (IIA) property of the logit model. With better specification, e.g., inclusion of attitudinal variables, the IIA problem would disappear and with it also the non-genericity. Thus, any non-generic specification of the model should be associated with the diagnostic tests for the IIA property and the non-generic specification should be restricted to alternatives which fail to satisfy the IIA property.[1]

[1]Even then there will be problems if the forecasting situation involves new modes.

Transferability of the Models

The empirical examination of model transferability will be done from three vantage points: (1) within regions; (2) in time; and (3) between regions. The results of the statistical tests follow.

Transferability within regions

Transferability of these models within regions is tested in three ways. The first way is to test whether the coefficients estimated from the different samples within the same region are identical. The second way is to investigate whether the coefficients estimated for non-CBD and CBD bound trips within the same region are statistically equivalent. And the third way to test model transferability within a region was to split the sample into two parts and to test whether the coefficients estimated from these two parts are the same. Because the home interview survey results are often in origin zone order this splitting of the samples effectively tests whether the urban travel behavior is similar to sub urban travel behavior. The Washington, DC data were in the origin zone order but the Minneapolis data were not. Thus, for the tests mentioned last, which were conducted on the Washington, DC data, the urban/suburban notion applies; we are not sure whether it applies also to Minneapolis data because there the origin zones appeared to have no particular order.

The WASH5 and WASH7 samples are independent random samples from the same region, as are the MINN12 and MINN25 samples. The nested likelihood ratio test is the proper test of the null hypothesis that the coefficients of each sample are identical.

The Washington models tested here used OVTTD, while the Minneapolis models used the model with separate walk and wait time coefficients. The chi-square statistic with the Washington data obtained a value of 16.0 and with the Minneapolis data a value of 11.0 . The critical chi-square value at .05 level with thirteen degrees of freedom is 22.36 and the null hypotheses that the coefficients from the two samples in each of the cities are the same are accepted.

The next set of tests were designed to examine whether the mode-choice behavior of travelers going to the CBD is different from that of the non-CBD travelers, and whether the urban residents have the same mode-choice coefficients as the suburban residents. Again, the nested likelihood ratio test was performed with the results being shown in Table 66.

TABLE 66 Chi-Square Statistics: CBD-bound Travelers vs. Non-CBD-bound Travelers (1) and for Urban vs. Suburban Residence of Travelers (2)

Sample	χ^2 - Stat (1)	Critical χ^2 .05 (df)	χ^2 - Stat (2)	Critical χ^2 .05 (df)
WASH5	21.4*	16.92 (9)	48.6*	22.36 (13)
MINN25	5.6	16.92 (9)	23.8*	22.36 (13)
Pre-BART	22.02*	18.31 (10)	not done	
Post-BART	7.0	18.31 (10)	not done	
*Null hypothesis of equal coefficients rejected at .05 level.				

The best specification of the system variables was used with each data set (i.e., OVTT with Washington data and walk and wait times with the other data sets, and the other variables as before).

The results in Table 66 show that for the Washington sample both tests, equality of coefficients for the CBD and non-CBD travelers and equality of coefficients for urban and suburban dwellers, are soundly rejected. For the Minneapolis sample the equality of coefficients for CBD and non-CBD bound travelers is accepted while the test for the urban vs. suburban dwellers is rejected. The results are mixed for the two Bay Area samples.

Because most of the above tests reject the null hypothesis of equal coefficients for different market segments it is of interest to learn if the rejection is due to differences in tastes with regard to the system attributes. This test was performed with post-BART data (all modes were included in the choice set) in Part I, Chapter 4, with the result that the urban and suburban dwellers indeed value the system attributes differeretly, they have different values of time. Similar tests here for both urban and suburban dwellers and CBD and non-CBD workers using Washington data accepted the null hypothesis of no taste deviations between these groups. The chi-square statistic for the former test was 3.4 and for the latter 2.0 with the critical value (.05 level, three degrees of freedom: OVTT, IVTT, OPTC/INC) of 7.82 .

In summary, the tests of transferability of mode choice models within the same region give inconclusive results; even though there is more evidence to support non-transferability than vice versa. It appears clear that any research that attempts to reach a sound conclusion in the matter of model transferability within regions or taste variations between different market segments must start with meticulously collected and collated data. The result above, which found taste variations in the Bay Area but not in the Washington area, is symptomatic. This may be a real effect; however, it is more likely the result of the methods used in collecting and calculating the values of the trip attributes: time, cost, and so forth. The discussion at the end of the previous section is also relevant here.

Transferability in time

The pre- and post-BART samples, although from the same region, are from different surveys at different times. The likelihood ratio test between the pre- and post-BART samples used models with OVTT/D and had thirteen degrees of freedom. The resulting χ^2 value was 24.4 . Because the critical chi-square values with thirteen degrees of freedom at the five percent and 2.5 percent levels are 22.4 and 24.7 respectively, we can reject the null hypothesis at the five percent level, but not at the 2.5 percent level. A similar test with pre- and post-BART data using an improved specification of the model and complete choice sets was performed in Part I, Chapter 3, with the same results.

Transferability between cities

In the likelihood ratio tests for the transferability of mode choice models between cities only the smaller sample was used from those cities where two samples were drawn. Tests that did not involve the Washington sample used OVTT. Further, the restrained logit estimations (combined sample estimations) did not use $DTECA_3$ while the unrestrained (separate) logit estimations, with the exception of the Minneapolis sample, did. Because the coefficient of $DTECA_3$ was never significant (at the .05 level), this exclusion makes only an insignificant change in log likelihood and, therefore, makes no difference in the results of these tests.

Results of the tests between cities are shown in Table 67.

TABLE 67 Chi-Square Statistics for Between Cities Comparisons

Sample	Chi-Square Statistic	Critical χ^2 .05 (df)
Pre-BART vs. MINN12	60.8	22.36 (13)
Pre-BART vs. WASH5	106.2	22.36 (13)
Post-BART vs. MINN12	46.6	23.68 (14)
Post-BART vs. WASH5	38.6	23.68 (14)
WASH5 vs. MINN12	78.4	22.36 (13)

We can see that the null hypothesis of similar coefficients is soundly rejected in all inter-city comparisons.

These same tests were performed on the "trimmed" subsamples, with the same results.

Conclusions

The single and most important conclusion regarding transferability of mode choice models can be simply stated: if the transferability of models is judged using statistical criteria, there is little ground to claim that the extant work-trip mode-choice models are transferable.

This disheartening conclusion does not, however, invalidate the approach to modeling behavior adopted in this volume. Rather, the results obtained ought to make us work harder to find better concepts and explanations of mode-choice and trip-making behavior. Two major areas of work can be readily identified. First, work should proceed on replacing the strict assumption of independent and identically distributed unobserved attributes of the multinomial logit model. The elimination of this assumption will increase explanatory power, given the current model specification. However, this increase is likely to be minor compared to the benefits it may give to the area now in need of greatest attention: better model specifications.

Improved model specification means not only an expanded list of variables and choice sets but a better understanding of how travel needs are intertwined with the everyday patterns of the life of households. Apparently, we do not now have sufficient understanding of this phenomenon.

It has been a strong tradition in econometrics to first specify a model structure, embodying a theory that one believes to explain behavior, formulated by observation and thought. Only then should one estimate statistically the parameters of this structure. With the easy availability of computers and "canned" programs, and with usually great emphasis in "quantitative methods" this good and solid tradition is almost forgotten.

In order to develop models of travel behavior there is a need to return to direct observation of the many circumstances which surround individuals' decisions. One good way of doing this is that modelers themselves collect the data for the next few years to gain a better understanding of human behavior; once such understanding is arrived at they can confidently move on to specifying models of travel and estimating them statistically.[1]

What should the planner or policymaker do then? Are the current models good for making a travel forecast? The answer is a qualified yes. The models can be used for making "ballpark" travel prognoses in traditional planning situations involving minor changes in the policy variables included in the model. This answer is based on the belief that, as the models indicate, the traditional policy variables of travel times and costs affect only marginal changes in travel behavior, and that the variables effecting substantial changes in behavior move slowly and are incorporated in the alternative specific dummy variables that account for the largest part of model explanatory power. More precisely, tests of transferability of only those coefficients that change with transportation policy--particularly travel times and costs--are more likely to pass than are tests for all coefficients including socioeconomic and demographic effects. Hence, the impacts of incremental, short-run policy changes are more likely to be correctly calculable from transfered models than are absolute travel patterns. Finally, the generally observed phenomenon that the likelihood function is relatively flat at its maximum in some directions for these transportation data sets--reflecting multicollinearity in the data and causing problems for precise estimation of coefficients--means that evidently disparate coefficients may yield similar predicted probabilities. Thus, apparently different models may yield similar forecasts. Therefore, tests of transferability in terms of forecast probabilities are more likely to pass than tests in terms of coefficients, and such probability tests are the more relevant for many forecasting purposes.

---

[1]Several important issues are raised for discussion here. One is the ability of statistical hypothesis testing to disprove, or fail to do so, the correctness of some postulation of behavioral theory on the basis of data. Two questions are central. First, often variables are included in the model whose t-values (vs. $\beta = 0$) are 1.7 - 2.0 implying α-level of about .05, and forgetting that on the other side of the coin there is about a fifty-fifty chance of accepting a faulty claim. Second, even if a model, whose structure is based on observation, has statistically highly significant coefficients, it does not guarantee its "correctness." This is because "policies" (in a broad sense) may have been in effect that have rewarded certain kinds of behavior and penalized other types of behavior. This is a particularly severe problem in nearly all econometric models of behavior; that is, the right hand side variables are not "irreducibles," but just the left hand side variable in disguise. Thus, means other than statistics and predictions are needed to support a theory (Talvitie, 1976). Therefore, models should be used for the purpose of policy with sensitivity for their shortcomings.