A Disaggregate Modal-Split Model for Work Trips Involving Three Mode Choices

Jeffrey Ganek,* American Telephone and Telegraph Company, New York
Raymond Saulino,* Office of Budget and Management Systems, District of Columbia

This paper describes a disaggregate mode choice model with three travel modes: drive alone, car pool, and transit. A number of alternative model specifications were tested and the results analyzed. In general, the coefficients of in-vehicle time, out-of-vehicle time, and costs agree closely with the results of similar studies. Estimated coefficients of variables not included in previous logit model studies are also presented. Of these, convenience, comfort, and flexibility influenced mode choice but mode unavailability and household income did not. Work location, cars per driver, and sex were the only socioeconomic variables for which statistically significant coefficients were found. Coefficients and models were also estimated for various subpopulations of commuters. The determinants of mode choice for CBD workers were different from those of non-CBD workers. Differences in the cost and time coefficients among travel corridors and income classes were also examined. The estimated models were validated by successfully predicting the mode choices of the commuters for whom the model was estimated, of other commuters, and, finally, of commuters for whom changes in the levels of service were made available.

This paper discusses a study that explored the determinants of commuter mode choice in the six-county region around Pittsburgh. The study was part of an evaluation of the car pool–public transit program administered by the Southwestern Pennsylvania Regional Planning Commission (SPRPC). Models of the multinomial logit form were estimated on disaggregate data to predict the short-run mode choices of commuters in the region, given the current work locations, residential locations, and automobile ownership patterns. The models predicted these choices from a set of alternative modes that include driving alone, car pooling, and riding on public transit as functions of the socioeconomic status (SES) of the commuters and the service attributes of the three modes. Variances in the influences of particular SES characteristics and mode attributes were determined by testing alternative specifications on subpopulations defined by work location and SES characteristics of the commuters. The objectives of this work were to learn why commuters choose the modes they do and to suggest how they could be enticed to choose shared ride modes.

The region around Pittsburgh has several concentrations of commercial and industrial employment. The largest of these is the CBD, which has five major transportation corridors leading to it from suburban areas. The quality of transportation services varies among these corridors. The southwest and east corridors have limited access highways that run to the CBD. The southern corridor is plagued by bottleneck problems as automobiles attempt to pass through Mt. Washington and over the Monongehela River. The northern corridors have a network of well-traveled streets where traffic flow is regulated by stop lights. Bus service to the CBD is available along all corridors, and streetcar lines, some of which have a right-of-way, run from the south. Commuter rail service is limited, and transit service to non-CBD locations is not extensive. The region served by these transportation systems has a population of over 2 million, with an average density of 300 people/km². The SES characteristics of the population are not different from the average characteristics of populations of other large standard metropolitan statistical areas. The region under study, then, has a diverse and well-established transportation system serving a dense population.

METHODOLOGY

The methodology used to estimate mode choice models was the multinomial logit form (1). This methodology was chosen because of its use for policy analysis due to its base in behavioral theory, its capacity to consider choice sets greater than two, and its successful application in other studies (2, 3).

DATA

The data used in this study were obtained from commuter surveys conducted as part of the study and from the Southwestern Pennsylvania Regional Planning Commission (SPRPC) travel time and cost networks. The complete data base included the SES characteristics, commuting preferences, and transportation services available to 740 commuters in the six-county region. This sample

* Both authors were at Carnegie-Mellon University when this research was performed.
population was generally representative of the working population of the region in terms of SES characteristics and available transportation services although there was a slight overrepresentation of commuters who work in the CBD. Additional data were obtained from a survey of commuters matched with potential car poolers by SPPRC.

The data used to estimate the model are discussed below.

In-vehicle time (IVT) is the time in minutes that commuters spend in their primary vehicle on a one-way work trip. These were obtained from the SPPRC network data.

Access time (ATIME) is the time in minutes that commuters spend between leaving home and reaching the primary vehicle for their work trip. Commuter estimates of the ATIME of their chosen modes were obtained from the surveys. Estimates of the ATIMEs of nonchosen modes were also derived from these survey results. ATIMEs for the transit alternative for nontransit riders were estimated to be 6.58 min, the mean ATIME report by transit riders. ATIME for the drive alone alternative was assumed to be 0 for all commuters. All car poolers were assigned a 4-min ATIME for the car-pool alternative, the mean time reported by car poolers. It was assumed that all commuters those with lower car-pool ATIMEs would tend to car pool more often than those with higher; therefore, noncar poolers were assigned 6-min car-pool ATIMEs.

Egress time (ETIME) is the time in minutes that commuters spend between leaving home and reaching the primary vehicle and arriving at the workplace. The ETIMEs for car poolers were used in estimating the model. ETIMEs for nonchosen modes were derived from the same data. The ETIMEs were generally homogeneous within commuter subpopulations defined by the workplace location (i.e., CBD or non-CBD). Therefore, the mean ETIME for each mode was calculated for CBD and non-CBD workplaces and the mean time for each mode was calculated for CBD and non-CBD workplaces.

Travel cost data were obtained from SPPRC network data. Transit travel costs were obtained directly from the SPPRC data. These data do not consider the possible purchases of monthly and yearly discount fare passes. Automobile travel costs were estimated from the SPPRC data. They included gasoline, oil, tires, and maintenance costs, which were not included in the SPPRC calculations, were obtained from survey results. The reported parking costs of commuters who drive alone to work were added to the SPPRC cost estimates to obtain total commuting costs. Parking costs for other commuters were assigned according to the traffic zone in which the commuter worked from the average of the daily parking costs of commuters who drive alone to that zone. These costs were added to the SPPRC cost estimates to establish total drive alone costs for commuters who do not drive alone.

The same data were used to estimate car-pool costs. All car pools were assumed to share costs equally among all members. Car-pool costs, then, were obtained by dividing the drive alone costs for the same zone-to-zone trip by the mean size of car pools as reported by car-pool respondents.

Commuter perceptions of the relative levels of comfort and convenience offered by various modes were obtained from their responses to questions that asked them to compare their nonchosen modes with their chosen mode. Their responses were coded on an integer scale of -2 to +2, indicating the service level of the nonchosen mode as compared to that of the chosen mode. A negative number indicated that the nonchosen mode had a lower level of service with respect to either comfort or convenience than the chosen mode with 0 taken as the service level of the chosen mode. Although the data are subject to uncertainties with respect to scaling and definition (4), they are the best available approximation of commuter perceptions, and were used as such.

Flexibility is a measure of the possible irregularity in commuter schedules. It was determined from the survey by asking commuters how often they arrive at or leave early or late. Those who arrive at or leave early or late between 5 and 15 times a month are assumed to have flexible schedule patterns. Others are assumed to have regular schedule patterns. All commuters who commute flexibly were assigned a value of 1 for flexibility; others were assigned a value of 0. This measure of flexibility reflects the preference of a commuter for flexibility, although data from which the measure is constructed may also represent an interaction between the preferences of the commuter and the actual flexibility of the service offered by his chosen mode.

Mode unreliability is a measure of how often each mode is not available for the trip to work. Data describing this service attribute were gathered in the mode choice survey by asking transit riders how many times per month their bus or trolley was more than 0.5 h late, car poolers how many times per month their car pool was unavailable for the trip to work, and those who drive alone how often their car was not available for the trip.

The standard SES attributes (including sex, income, and automobile ownership) of the sample population were obtained from the SPPRC mode choice survey. The mode constants account for the SES characteristics, commuter preferences, and service attributes that influence the choice decision but are not explicitly contained in the model. The mode choice process is highly complex, with many factors influencing the decision. A model was estimated without mode constants, then the factors left unaccounted for could affect the values of the coefficients associated with the variables that are explicitly included in the model.

ESTIMATION RESULTS

A random subsample of 400 observations from the sample population was used to estimate the model parameters for a number of model specifications of which the most interesting are models 1 and 2 in Table 1. Table 1 also includes models 3 and 4, which represent separately estimated specifications for CBD and non-CBD workers respectively.

Some observations about the success of the modeling efforts can be made from Table 1.

1. The signs on all the significant coefficients are as expected.
2. The relative magnitudes of the coefficients are reasonable.
3. In general, the results agree with similar studies made elsewhere.
4. The variances in the magnitudes of the coefficients estimated on different subgroups of commuters were as expected.

The coefficients, then, appear to be stable estimates of the true parameters.

The coefficients for the IVT are consistently negative and statistically significant, reflecting the dislike of in-vehicle time. Further analysis of these coefficients suggests that different groups of commuters have different sensitivities to IVT. For example, analysis of a model that included two IVT variables, one for all commuters and one for commuters with annual household incomes
greater than $15,000, showed that all commuters have an IVT coefficient of -0.023 and that wealthier commuters have an additional IVT coefficient of -0.015. While the additional factor of -0.015 is not statistically significant, it does suggest that income level may have some influence on how commuters value IVT. The hypothesis that the value placed on IVT depends on the quality of service available to the commuter was tested by the estimation of five corridor-specific IVT coefficients; these coefficients were not statistically different from one another.

Separate coefficients, which were different from each other, were estimated for ATIME and ETIME. This difference may explain the disparity between these estimates and the coefficients for out-of-vehicle time (OVT) estimated elsewhere, i.e., the higher ATIME here may reflect a division of the independent influences of ATIME and ETIME on mode choice that are aggregated in single OVT coefficients. These variables, however, do not discern whether commuters disvalue ATIME or ETIME differently for different modes but the data were insufficient to estimate mode-specific ATIME coefficients. However, the coefficient here has the appropriate sign and agrees with other results (2, 3, 5) that changes in OVT have a greater effect on mode choice than similar changes in IVT.

The ETIME coefficient for commuters who work in non-CBD locations is close in magnitude to the OVT coefficients reported (2). It is not significant for CBD workers for whom the ETIME differences for different modes were too small to significantly affect mode choice decisions. Table 1 also shows that cost has a negative coefficient and that, although it is relatively small, its difference from zero is statistically significant. The magnitude of the coefficient is not incongruous with other cost coefficients. The cost elasticities derived from the coefficients shown below are similar to those reported elsewhere (5).

<table>
<thead>
<tr>
<th>Percent Change in Ridership</th>
<th>Transit</th>
<th>Car Pool</th>
<th>Drive Alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>From a 1 percent change in</td>
<td>-0.17</td>
<td>+0.09</td>
<td>+0.09</td>
</tr>
<tr>
<td>travel costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From a 1 percent change in</td>
<td>+0.06</td>
<td>-0.14</td>
<td>+0.06</td>
</tr>
<tr>
<td>car-pool costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From a 1 percent change in</td>
<td>+0.18</td>
<td>+0.18</td>
<td>-0.48</td>
</tr>
<tr>
<td>drive alone costs</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Therefore, the relatively small magnitude of the coefficients may be accepted as a reasonable estimate.

Since it was not possible to estimate mode-specific differences in the cost coefficients, a generic cost variable was used in the model. Attempts to identify differences in the sensitivity to cost of commuters in different income classes gave a coefficient that was not statistically significant. This may be due to the small number of low-income commuters in the sample.

The comfort associated with a mode is important to commuters with household incomes >$15,000/year, but is not important to those with smaller incomes. This is indicated by the statistically significant coefficient estimated for comfort ($15,000 only). The coefficient for comfort ($15,000) was not statistically different from zero. Commuters also highly value the mode characteristic identified as convenience. The statistically significant coefficient for convenience is very important in mode choice decisions, but, because of similarities in meaning, comfort and convenience were not included in the same equation. Qualitative factors like comfort and convenience can be included in choice models along with more quantified variables such as time and cost.

The preference for flexibility has been included in the model as a mode-specific variable because car pooling is regarded by the commuters in the sample as very inflexible relative to the other modes (6). Therefore, commuters are assumed to take special consideration of their preference for flexibility only when they assess their car-pool alternative. The statistically significant and negative value of the estimated coefficient confirms this assumption.

Car availability is included in the model twice. The two estimated coefficients represent the relation between the number of cars per driver in the household and the value that a commuter gives to the drive alone and car-pool modes relative to transit. These coefficients show that a strong factor in the mode choice is the number of cars per driver in the household. The more cars per driver, the higher the tendency to commute via automobile. High car availability especially favors the drive alone choice. The influence of car availability on mode choice is a short-run effect. In the long run, car availability becomes a function of mode choice, and, in the long run, the tendencies of commuters to change modes as a result of changes in the levels of service offered may be greater than those estimated here. In the short run, few drivers with high car availability can be enticed to change their mode of commuting.

Work location was a mode-specific variable for both the drive alone and car-pool modes. The resulting negative and highly significant coefficients indicate that for CBD workers the problems of driving and the availability of transit service to the CBD strongly discourage use of the automobile modes and encourage use of transit. The magnitudes of these coefficients suggest that commuting to the CBD is very different from commuting to non-CBD locations. Traffic congestion, the difficulty and cost of parking, higher worker density, and transit availability make the determinants of mode choice in the CBD different from those elsewhere. Therefore, two sets of coefficients, one for CBD workers (model 3) and one for all other commuters (model 4), were estimated. The results show the following influences of work location:

1. CBD workers find IVT more onerous than do non-CBD workers. This may result from the greater traffic congestion in the CBD.
2. Non-CBD workers are much more sensitive to ATIME than are CBD workers, and policies that affect ATIME should have greater impact on non-CBD commuters. The relative magnitudes of these coefficients, however, may be due to the method used to assign ATIMEs. However, non-CBD commuters, who do not have to consider the inconveniences of congestion and parking, undoubtedly discriminate between modes on other dimensions, and it is not surprising that ATIME is more important to them.
3. All commuters are relatively insensitive to costs.
4. CBD workers are very sensitive to the convenience of commuting, but the coefficient of convenience for non-CBD workers is not statistically significant. Thus, where the work trip does not require coping with heavy congestion, parking problems, and the like, commuters are not very concerned with convenience.
5. The preference of a commuter for flexibility detracts more from the value of the car-pool alternative for commuters bound for CBD destinations than for non-CBD workers. The greater and more frequent availability of transit to the CBD may make car pooling, with its rigid schedules, a less attractive alternative there.
6. Car availability (as measured by the number of cars per driver in a household) is more influential in the
mode choices of CBD workers. This may be due to the general lack of transit service to non-CBD locations, which forces non-CBD workers to commute by automobile, regardless of their car availability.

7. A number of the estimated coefficients were not significantly different from zero. These include: relative comfort (= $15,000 only), mode unreliability, sex, and household income. A priori, these variables had been considered important factors in the mode choices of commuters; their actual lack of significance, then, is important.

All of the coefficients in models 3 and 4 have the expected signs and are of reasonable magnitudes. Their significance is that they demonstrate the necessity to consider work location when modeling mode choice.

While the coefficients for sex in models 1 and 2 are highly significant they are not significant in the separate CBD and non-CBD models. This is due to the fact that almost twice as many female commuters work in the CBD than in the non-CBD. When variables are correlated, their estimated coefficients are a clue to the total effect of both variables on mode choice. Since sex became unimportant when the population was partitioned by work location, the total negative effect of sex and CBD must be primarily a CBD effect.

Several variables had insignificant coefficients; i.e., the commuters in the sample were insensitive to variations in these variables when discriminating among the alternative modes. Since these variables were previously considered to be important, their failure to be significant here is worthy of note. The failure of household income to discriminate was particularly troubling. None of the attempts to include household income (e.g., cost per household income and household income adjusted for family size) gave coefficients that were significantly different from zero. Household income may be an important discriminator only in the lower ranges (e.g., under $8000) in which the sample was deficient.

Attempts to estimate a statistically significant coefficient for the SES of the job held by the commuter also failed. Blue collar workers did not have any greater or lesser tendencies to choose any one mode over the others. Nor did the number of workers employed by the employer of a commuter have any influence on mode choice.

MODEL VALIDATION

As a test of the validity of model 1, the probabilities of choosing each of the three modes were calculated for each sample included in the sample of 400 commuters used to estimate the coefficients. In 68 percent of the cases, the mode with the highest probability of choice was the one actually chosen. This percentage is similar to the percentages reported elsewhere (2).

The predicted distribution over the modes (i.e., the mode split) was calculated in two different ways. First, the mode choice of each individual was predicted by the highest probability method and the percentage of individuals predicted to choose each mode then calculated. Second, the individual probabilities of each mode were summed and averaged to give the expected value estimate of the modal-split distribution. These two distri-
sults are compared with the actual distribution for the estimation sample in Table 2. The proportion of commuters predicted to drive alone is almost equal to the actual proportion in the sample. At the worst, transit ridership is overpredicted by 5.3 percent and car poolers are underpredicted by 5 percent. The model, then, does very well at aggregate prediction of the distribution over the modes and reasonably well at prediction of the individual mode choices for the estimating sample.

To further test the predictive ability of the model, the three mode choice probabilities were calculated for 300 commuters who were not used in the estimation of the model. The model predicted their individual mode choices correctly in 59.9 percent of the cases.

CONCLUSION

This modeling effort has identified numerous characteristics of car pools and potential car poolers that should be considered in managing programs to encourage the use of car pooling. It identifies subpopulations of commuters who have high car-pool potential and service attributes of car pooling that would entice more commuters into car pools if they were improved.

The subgroups of commuters at whom car pool encouragement efforts should be directed are

1. Commuters who have regular (as opposed to flexible) commuting schedules,
2. Commuters who have moderate or low ratios of cars per driver in their households,
3. Commuters who have annual incomes of less than $15,000 (suggested by the coefficient of comfort for commuters having incomes greater than $15,000 and by the fact that commuters who drive alone generally view car pooling as a considerably less comfortable mode), and
4. Commuters who drive alone to the CBD who will not easily be lured into car pools (because CBD commuters are very sensitive to convenience and those who drive alone generally consider car pooling to be a less convenient mode).

The modeling effort also helped to identify improvements in the service attributes of car pools that would have a significant impact on mode choice. These include:

1. Provision of a back-up service for car poolers—This would lessen the influence of its inflexibility on mode choices. Taxi and emergency car-pool matching operations are possible back-up services.
2. Policies that minimize the OVT of car pooling—Here again, preferential parking spots for car poolers would be effective. To minimize the OVT of car pooling, commuter computer matching operations should strive to match only commuters with very similar origins, destinations, and schedules. Also, since the 4-min average ATIME of car poolers in the sample is significantly lower than previous estimates, advertising its real magnitude could alter the perceptions of car pooling held by commuters. The magnitude of the OVT coefficient suggests that such a change could have a sizeable impact on mode choices.
3. Policies, such as exclusive car-pool lanes, that influence IVTs.
4. Policies that improve the convenience of car pooling—Such policies include preferential parking or flexible work hours for car poolers. Another effective policy would be an advertising campaign designed to dispel subjective notions about the inconvenience of car pooling.

In addition to these conclusions, the model casts doubts on some efforts that previously have been effective in drawing commuters to car pools. These include:

1. The economy of car pooling does not substantially encourage many commuters to car pool. Therefore, advertising campaigns that emphasize cost will not be as effective as the themes suggested above.
2. The estimated cost coefficients show that in the short run, non-transportation incentives (e.g., taxes on driving alone) should not be expected to yield substantial changes in mode choices.

Finally, the modeling effort has defined some areas needing further research. Further work on the long- and short-run influences of costs, the significance of OVT, and the importance of comfort and convenience will be useful.

ACKNOWLEDGMENT

This paper was part of the systems synthesis project course taken by MS students in the School of Urban and Public Affairs at Carnegie-Mellon University. Other students participating in the course were Pearl Davis, Alan Stadning, Ahmed Thomas, Douglas Gunwaldsen, Douglas Doxey, and Richard Kuzminak. Alfred Blumenthal was the faculty advisor supervising the project. Significant assistance was provided by Charles Manski and Martin Wohl of Carnegie-Mellon University and Moshe Ben-Akiva of Massachusetts Institute of Technology. Additional assistance was provided by Daniel Nagin and Jacqueline Cohen of the Urban Systems Institute of Carnegie-Mellon University. The effort was supported by the Southwest Pennsylvania Regional Planning Commission. Wade Fox of the Southwest Pennsylvania Regional Planning Commission was extremely helpful in providing guidance and suggestions. The support of all these individuals is very much appreciated.

REFERENCES