Modeling Mode Choice in New Jersey

KEVIN NEELS AND JOSEPH MATHER

In this paper, a mode choice model developed for NJ Transit and the Port Authority of New York and New Jersey to assist in the evaluation of proposals for increasing the capacity and use of the existing Hudson River crossing connecting Manhattan and northern New Jersey is described. The model focuses on the choices of a.m. peak period eastbound commuters. It allocates demand across seven primary modes, including automobile, bus, two park-and-ride modes (automobile to bus and automobile to PATH), and three rail modes (commuter rail to Penn Station, commuter rail with transfer to PATH, and local access to PATH). The emerging trans-Hudson crisis that provided the impetus for the model development effort, the planning program of which it was a part, the data sources used in the effort, the specification of the model, the procedures used to estimate the model coefficients, the statistical results of the model estimation, and the model's forecasting performance are also discussed.

In this paper, the mode choice model developed by NJ Transit, the Port Authority of New York and New Jersey, and Charles River Associates is described. This model is explicitly designed to be sensitive to the presence and comparative quality of the large number of travel alternatives available in that market. With the large number of modes it handles (seven) and the flexible way in which it captures intermodal competition, it represents one of the most ambitious efforts to date to forecast travel demand in a complex, multimodal environment.

The model was developed to help NJ Transit and the Port Authority to deal with the trans-Hudson crisis. Over the past several years, the growth in service employment in Manhattan has stimulated a rapid increase in journey-to-work travel. Largely a result of its high-quality and comparatively inexpensive housing stock, New Jersey has provided a growing share of the workers filling these new jobs. According to the Bureau of the Census, in 1980, 10 of every 100 Manhattan jobs were held by New Jersey residents. However, recent Port Authority estimates suggest that of the new Manhattan jobs being created in the late 1980s, 34 of every 100 jobs will be held by New Jersey residents. Already in the first half of the decade, trans-Hudson commuters have experienced lengthening backups and delays at the Hudson River crossings and passenger loadings that strain the capacity of trans-Hudson transit links. The trans-Hudson crisis is due to the system's inability to serve current demand and the constraint this places on New Jersey's economic development.

Because of the problems the model was intended to address, the development team had to strike a balance among a number of distinct and sometimes conflicting goals. There were a number of important features that were incorporated into the model, including

K. Neels, Charles River Associates, 200 Clarendon Street, Boston, Mass. 02116. J. Mather, NJ Transit, McCarter Highway & Market Streets, Newark, N.J. 07101.

- Statistical estimation of model parameters,
- Accurate representation of intermodal competition,
- Appropriate responses to policy changes,
- High levels of forecast accuracy, and
- Ease of estimation and use.

The primary requirement was that the model parameters be estimated statistically from locally collected data. This procedure was the only way to achieve the best fit to the data, to ensure that the model parameters fully reflected local patterns of behavior, and to guarantee the objectivity of the model results.

It was also critically important that the model be able to deal with the large number of modal alternatives that are available in the trans-Hudson commuter market and provide an accurate representation of the complex patterns of competition that exist among them. In this complex, multimodal environment characterized already by extremely heavy transit usage, policy makers and planners had to know not just how many commuters might be attracted to a new service, but also from where they would be drawn. To contribute to the solution of the trans-Hudson crisis, a transportation improvement had to draw commuters out of automobiles and other low-occupancy vehicles, and not simply cannibalize existing high-capacity transit ridership.

It was decided early in the development effort to build into the model appropriate responses to key policy variables. The most important goal of the calibration effort was to produce a model that would provide appropriate and accurate predictions of the responses of trans-Hudson commuters to changes in service levels or modal attributes. To achieve this goal, the process had to build into the model appropriate values for the key behavioral parameters. Specifically, the model had to imply reasonable values for self- and cross-elasticities of demand. It also had to be sensitive to the service attributes that were important from a policy point of view.

Because the output of the model would be used to evaluate the financial feasibility of alternative capital improvements and to make key engineering decisions regarding capacity and station location, it was essential that the model be able to reproduce and forecast patterns of travel behavior with a high degree of accuracy.

Because this model was to be a working model that would be used on an ongoing basis to analyze and solve practical planning problems, it was important that the model be easy to estimate and easy to use. The goal was to develop an easily applied forecasting tool that could be used by all agencies to analyze trans-Hudson travel. It was also necessary to develop a model that could be updated by NJ Transit or Port Authority staff or reestimated as better or more recent data became available. These goals led to a decision to rely on microcomputers to build and run the model.

The remainder of this paper is divided into four sections. The section that follows presents the form and specification of the model that emerged from this effort. The third section of the paper describes the sources of data that were used in the model estimation. The fourth section describes procedures used to estimate the model coefficients. The final section presents estimation results and summarizes what was learned.

MODEL SPECIFICATION

The model was formulated as a set of logistic regression equations estimated across origin-destination (O-D) pairs *(1)*. The dependent variable in each equation consisted of the log of the ratio of the transit share for the mode in question for that O-D pair, divided by the corresponding automobile share. Six equations were estimated—one for each transit mode. Automobiles were thus used as the reference mode, and the automobile share was computed from the log-odds ratio predictions using the constraint that the estimated shares had to sum to one. The mathematical form of the resulting model is shown in Equation 1.

$$\log(S_i/S_a) = a_0 + a_1X_1 + \ldots + a_nX_n \qquad (1)$$

where

S_i = share for Transit Mode i;
S_a = share for automobile mode,
X_i = explanatory Variable i and
a_i = estimated Coefficient i.

Each demand equation is composed of three sets of independent variables: measures describing the service offered by the subject mode, measures describing the service offered by competing alternatives (which include the automobile reference mode), and measures describing characteristics of the O-D pair itself. The last category includes selected socioeconomic variables, as well as dummy variables specifying whether or not specific modes are available for trips between an origin and destination.

The principal advantage of this formulation is its explicit representation of the attributes of the competing modes. The presence of these variables permits a pair of modes to be either close or distant substitutes. The degree of competition between them can vary continuously between these two extremes, and can be estimated empirically. Thus, both the IIA problem that characterizes the multinomial logit (MNL) model and the sometimes arbitrary groupings that are often found in nested logit models can be avoided. In this respect, the trans-Hudson model represents a considerable advance in the analysis of travel behavior in multimodal environments.

One of the thorniest problems lay in the definition of the modal alternatives. Mode definition was difficult not only because many different transportation technologies were available in the trans-Hudson market, but also because these technologies were used by commuters in such varied and complex ways. A standard technology-based approach to mode definition in this region might have resulted in only three modes: automobile, bus, and rail. However, a close look at the way in which these traditional modes manifest themselves within the region quickly reveals the inadequacy of this simple trichotomy.

Consider, for example, rail. A substantial number of commuters drive long distances to access PATH, the rapid transit system connecting northern New Jersey and Manhattan. The PATH systems serves two other distinct markets as well: local walk-on or bus access riders, and commuter rail riders who transfer to PATH for the final trans-Hudson leg of their journey. The commuter rail riders transferring to PATH, in turn, make up a different market from that of the commuters who travel directly to Penn Station, New York, on NJ Transit or Amtrak trains.

In order to understand patterns of travel demand in this market it was important to account both for the characteristics of the technology and the way in which it was used by commuters. For this reason, the model was based on modal definitions that reflect distinct patterns of travel behavior, rather than distinct vehicle or guideway technologies. The model allocates travel demand across seven distinct travel modes. These include automobile, three combinations of conventional transit (bus, commuter rail with a PATH trans-Hudson link, and commuter rail to Manhattan), two fringe park-and-ride modes (using either bus or PATH for the trans-Hudson segment), and local PATH (which as a mode in itself is defined to be available only within an inner core area along the Hudson River).

The explanatory variables used to define the level of service along each trip segment are those traditionally found in mode choice models. These include variables describing ease of access and egress, wait time, transfer time, cost, and line-haul time. In a further effort to take into account the multimodal trans-Hudson environment, separate coefficients for the different types of line-haul time were incorporated into the model to capture the distinctly different characteristics of the different line-haul technologies.

Measures of ease of access were constructed using a parallel impedance formulation. This formulation, which is based on an analogy to electrical circuit theory, was used because of its ability to deal with situations in which multiple-access modes are available. The parallel impedance formula reflects both the number of access options available as well as their quality. It has the property that the addition of a new access mode always improves ease of access, regardless of the quality of the new option. Hence it avoids the feeder bus paradox in which the introduction of a new but inferior access mode increases average access time and decreases the share of the line-haul mode that has been improved. The exact formula used is shown in Equation 2 for a two-access mode example.

$$A = \left(\frac{1}{T_w} + \frac{1}{T_a}\right)^{-1} \qquad (2)$$

where

A = ease of access,
T_w = walk access time, and
T_a = automobile access time.

In exactly the same way, egress parallel impedances were calculated for representation of the egress alternatives in this region.

DATA SOURCES

The data set used for estimation of the model coefficients was constructed from two primary sources. A comprehensive set of travel surveys administered by the Port Authority and NJ Transit provided information on patterns of demand and selected socioeconomic characteristics of commuters. A combination of published schedules, time tables, and field measurements provided travel times, cost, frequency, and other level-of-service measures.

The funneling of the entire target travel market through the Hudson River crossings created an environment in which surveying commuters was relatively easy. Partly for this reason, a large body of recently collected travel survey data was available for the model development effort. Within the 2 years preceding the initiation of the project, on-board surveys were administered to bus riders, commuter rail passengers, and users of the PATH system. In addition, comprehensive surveys of users of automobile facilities were available. Usable responses were obtained from approximately 50 percent of all eastbound peak-period trans-Hudson commuters.

Level-of-service variables were developed from schedules, timetables, and field measurements. Starting with times, costs, and frequencies for individual bus and rail lines, the data were first summarized to the minor civil division level for bus and the station level for rail. Subsequent aggregations summarized the information at an O-D level using zone definitions developed specifically for this project.

An early decision was made to rely on an aggregate approach to model development and forecasting. In contrast to many recent model development efforts, the project was carried out in a data-rich environment. Hence, the economies in estimation that disaggregate modeling can offer were not needed. The use of data based on zonal level averages offered a number of advantages. First, the aggregate data structure made it possible to carry out all calibration and forecasting on a microcomputer and thereby realize significant time and cost savings. Second, the small datasets and microcomputer-based processing permitted by an aggregate approach gave the model the potential for wide distribution and easy use. Third, the use of aggregate data permitted the manual generation of much of the initial input data. This last feature was a great advantage in the early stages of the effort, before much progress towards automating the process of developing model inputs was made.

Because the focus of the modeling effort was entirely on peak-period trans-Hudson travel, the trip table was structured to contain one-way (eastbound) trip flows from origins west of the Hudson to destinations east of the Hudson. Working within a practical limit of approximately 1,000 trip interchanges, the study region was divided into a relatively coarse zone system. This zone system used the region's transportation network as a skeletal framework. The commuting region west of the Hudson River was divided into 23 radial corridors. Each corridor was defined around either a rail line, a bus service corridor, or a concentration of automobile users. Within each corridor, variations in residential density and demographic characteristics were used to define three to four concentric sectors, as appropriate. The final zone system in New Jersey was composed of 68 origin zones, each containing an average of 2,916 peak-period trips in an area of 74 mi^2.

The destination area east of the Hudson River was segmented into 10 Manhattan central business district (CBD) analysis zones, with four additional external destination zones to maintain consistency with overall trip control totals. The 10 destination zones considered in the analysis are all in Manhattan, south of 60th Street. These zones were defined from smaller Port Authority zones primarily on the basis of proximity to Manhattan's various transportation terminals.

ESTIMATION

In this section, the procedures followed in calibrating the trans-Hudson mode split model are outlined. Calibration is defined as the full process of bringing up an operational model for practical use. Thus, calibration includes but is not limited to the use of statistical procedures to estimate model coefficients. Much of the hard work involved in achieving the ambitious goals set for this effort actually took place in the calibration process.

Ordinary least squares estimation was used in initial exploratory work. This procedure was consistent with the basic form of the model and with the use of aggregate demand and service data. It generated results quickly and cheaply, and permitted both establishing the basic outlines of the model and refining the data procedures.

In an effort to build the desired policy sensitivity into the model, a number of cross-coefficient constraints were imposed on the various demand equations. These constraints typically set the coefficient for one service attribute to be a multiple of the coefficient for a related service attribute. They were made necessary by the limited amount of variation in these service measures contained in the base data set, and the consequent difficulty of obtaining precise coefficient estimates directly.

Such constraints were relied on heavily in the PATH equation and in the equations for the two park-and-ride modes. For example, in the case of PATH it was important for the sake of completeness and consistency with other modes to consider separately line-haul time, wait time, and transfer time. The PATH system is not extensive, however, and has relatively little variation in service frequency. The only transfer in the entire system is an insignificant across-the-platform transfer at the Journal Square Station. Rather than drop these two variables from the model, relationships found in travel demand literature were used and these two coefficients were set at twice the line-haul time. In this way, the desired policy sensitivity was built into the model.

The use of cross-coefficient constraints solved another potential problem associated with this particular model specification. The incorporation of the competing mode variables presented estimation problems in that there was a large number of such variables. Including all of them could have quickly exhausted the available degrees of freedom. Because all that was sought by including these variables was a general indication of how attractive the alternatives were, the detailed level of service variables for each competing mode were combined into a summary measure of the generalized cost of that mode.

This generalized cost was the sum of the travel cost and the dollar equivalent of a weighted sum of access, waiting, line-haul, transfer, and egress travel times for the competing mode. Access impedance was weighted at three times the value of line-haul time; and waiting, transfer, and egress times were tendency to overestimate shares for these minor modes. To compensate for this problem, a set of threshold limits was estimated that set a lower bound for the estimated share for each mode. These thresholds were set at the values for each mode that best distinguished between zero and nonzero share O-D pairs in the baseline dataset. In applications, mode shares below the threshold limit are set to zero. In effect, the mode split model is applied conditionally, given a prior judgment about which modes will have nonzero shares. That judgment, in turn, is based on the relative attractiveness of the different modal options. This procedure is consistent with the way in which the coefficients of the model were estimated, because in using a logistic regression approach O-D pairs with a zero mode share were eliminated from the estimation dataset.

RESULTS

The following list presents goodness-of-fit and summary statistics for the regression carrying out the simultaneous estimation of the six demand equations.

Statistic	*Value*
R^2	0.6530
Corrected R^2	0.6484
F statistic	142.8
Number of observations	1,999

Despite the large number of primary modes and the complexity and diversity of the region the model describes, the percentage of variation explained by the model is relatively high. The model coefficients are highly significant.

Tables 1–6 present the estimated coefficients of the six individual transit demand equations. As a result of the rich set of data available for model estimation and the use of a priori information in the form of cross-coefficient and cross-equation constraints, the individual coefficient estimates are, as a rule, extremely precise. Standard errors are small.

The ability of the model to replicate the baseline demand data varies somewhat by mode. Automobile and PATH modes are forecast with the highest accuracy. Prediction errors for these modes are about one-third the average number of trips per interchange. The model deals well with these two modes

TABLE 4 REGRESSION RESULTS FOR RAIL-TO-PATH EQUATION IN MODAL-SPLIT MODEL

Variable	Coefficient	Standard Error	T-Statistic
Rail-to-PATH Service Variables:			
Rail-to-PATH Cost	-.2162	.0054	-39.828
Rail-to-PATH Rail Time	-.0301	.0008	-39.828
Rail-to-PATH PATH Time	-.0301	.0008	-39.828
Rail-to-PATH Wait Time	-.0602	.0015	-39.828
Rail-to-PATH Transfer Time	-.0903	.0023	-39.828
Rail-to-PATH Access Impedance	-.1230	.0096	-12.825
Rail-to-PATH Egress Impedance	-.1748	.0217	-8.037
Competing Mode Variables:			
Auto Time	.0364	.0018	20.715
Auto Cost	.2871	.0137	20.927
Bus Generalized Cost	.1840	.0167	11.041
Direct Rail Generalized Cost	.0026	.0004	6.906
Modal Availability Flags:			
Local PATH Market Area Flag	-.1353	.3000	-0.451
Direct Rail Market Area Flag	-.0874	.0127	-6.906
Other Terms:			
Intercept	-2.9703	.5066	-5.863

NOTE: The standard error shown for the intercept term is based upon an approximate calculation that ignores the covariance between the intercept term for the pooled regression and the intercept shift for this equation

SOURCE: Regression Analysis of Travel Demand and Level of Service Data

TABLE 5 REGRESSION RESULTS FOR NONLOCAL PATH EQUATION IN MODAL-SPLIT MODEL

Variable	Coefficient	Standard Error	T-Statistic
Nonlocal PATH Service Variables:			
Nonlocal PATH Cost	-.4835	.0322	15.012
Nonlocal PATH Line Haul Time	-.0623	.0041	-15.012
Nonlocal PATH Wait Time	-.1246	.0083	-15.012
Nonlocal PATH Transfer Time	-.1246	.0083	-15.012
Nonlocal PATH Access Impedance	-.1869	.0124	-15.012
Nonlocal PATH Egress Impedance	-.1246	.0083	-15.012
Competing Mode Variables:			
Auto Time	.0535	.0030	17.639
Auto Cost	.4156	.0236	17.639
Direct Rail Generalized Cost	.0042	.0005	7.970
Modal Availability Flags:			
Direct Rail Market Area Flag	-.1392	.0175	-7.970
Other Terms:			
Intercept	-2.7571	.5361	-5.143

NOTE: The standard error shown for the intercept term is based upon an approximate calculation that ignores the covariance between the intercept term for the pooled regression and the intercept shift for this equation.

SOURCE: Regression Analysis of Travel Demand and Level of Service Data

TABLE 6 REGRESSION RESULTS FOR DIRECT RAIL EQUATION IN MODAL-SPLIT MODEL

Variable	Coefficient	Standard Error	T-Statistic
Direct Rail Service Variables:			
Direct Rail Cost	-.0779	.0254	-3.068
Direct Rail Line Haul Time	-.0104	.0034	-3.068
Direct Rail Wait Time	-.0208	.0068	-3.068
Direct Rail Transfer Time	-.0312	.0102	-3.068
Direct Rail Access Impedance	-.1593	.0141	-11.324
Direct Rail Egress Impedance	-.2367	.0290	-8.176
Competing Mode Variables:			
Auto Time	.0363	.0018	20.663
Auto Cost	.2865	.0137	20.884
Rail-to-PATH Generalized Cost	.0098	.0079	1.248
Modal Availability Flags:			
PATH Market Area Flag	-.1301	.4525	-0.287
Other Terms:			
Northeast Corridor Flag	1.4130	.2180	6.483
Intercept	-.9501	.6480	-1.466

NOTE: The standard error shown for the intercept term is based upon an approximate calculation that ignores the covariance between the intercept term for the pooled regression and the intercept shift for this equation.

SOURCE: Regression Analysis of Travel Demand and Level of Service Data

weighted at twice the value of line-haul time. Time was converted to dollars by using one-half the average hourly wage rate of the users of the competing mode as reported in on-board surveys.

In subsequent refinements of the model a procedure was adopted that permitted estimating the coefficients of all six equations simultaneously. To do this, the estimation datasets for the six demand equations were concatenated and slope shift variables were introduced to allow each equation to take a different set of coefficients. Use of this procedure allowed the imposition of cross-equation constraints on coefficients and use of generalized least squares to correct for cross-equation correlation of error terms in subsequent reestimations.

The ability to impose cross-equation constraints on coefficients permitted more efficient estimation of model coefficients and ultimately improved the policy sensitivity of the model. These constraints were used in two ways: to incorporate prior information about relationships between modes, and to place bounds on the cross-elasticities of demand between modes.

An example of the first use occurred with the two commuter rail modes, where there was ample reason to believe that an extra minute of commuter rail time was viewed in the same way by users of either mode. The coefficients on rail time in the two equations were constrained to be the same, thereby improving the precision of the overall model estimate.

We also used cross-equation constraints to correct a number of instances in which the estimated cross-elasticities of demand between modes were slightly negative. This typically occurred in cases where the modes in question were not close substitutes and where the estimated cross-elasticity was not significantly different from zero. With such constraints, these elasticities could be constrained to remain strictly, though only slightly, positive.

As part of the calibration process, two adjustments to the raw regression results were carried out to improve the model's accuracy in practical applictions.

The first such adjustment corrected for functional form bias. Because the ordinary least squares method was used in connection with a log-odds transformation of the underlying dependent variable, the means of the model's predicted shares did not necessarily equal the means of the raw data. This potential bias was corrected by adjusting the constant terms. A set of mode-specific factors was estimated to adjust the total predicted demand for each mode to the total actual demand found in the base trip table. This procedure resulted in a distribution of over and under predictions at the zone level that summed to zero by mode and were, therefore, unbiased. An iterative process estimated the values of these mode-specific adjustment factors.

The second adjustment improved forecasts of minor share modes. Minor share modes (those attracting less than 2 percent of the trips within an interchange) result from the method used to define the analysis zones. Because these zones are defined around major transportation facilities, they tend to be dominated by a single mode. Hence, at least one of the competing modes typically assumes a small share. For the two park-and-ride modes, which had small shares regionwide, shares at an O-D level would often be zero. Because the log form of the model prevents estimates of a zero share, the model had a

TABLE 1 REGRESSION RESULTS FOR BUS EQUATION IN MODAL-SPLIT MODEL

Variable	Coefficient	Standard Error	T-Statistic
Bus Service Attributes:			
Bus Cost	-.1946	.0164	-11.882
Bus Line Haul Time	-.0182	.0015	-11.882
Bus Wait Time	-.0364	.0031	-11.882
Bus Access Impedance	-.2141	.0303	-7.060
Bus Egress Impedance	-.0364	.0031	-11.882
Competing Mode Variables:			
Auto Time	.0363	.0018	20.629
Auto Cost	.2865	.0137	20.884
Rail/PATH Generalized Cost	.0098	.0079	1.248
Modal Availability Flags:			
Local PATH Market Area Flag	.1921	.1710	1.123
Rail/PATH Market Area Flag	-.3928	.3147	-1.248
Other Terms:			
Percent of All HH's In High Income Category	-.0492	.0052	-9.519
Intercept	-.0233	.3240	-0.072

SOURCE: Regression Analysis of Travel Demand and Level of Service Data

TABLE 2 REGRESSION RESULTS FOR LONG-HAUL AUTOMOBILE-TO-BUS EQUATION IN MODAL-SPLIT MODEL

Variable	Coefficient	Standard Error	T-Statistic
Auto-to-Bus Service Variables:			
Auto-to-Bus Cost	-.5089	.0904	-5.631
Auto-to-Bus Line Haul Time	-.0567	.0101	-5.631
Auto-to-Bus Wait Time	-.1135	.0202	-5.631
Auto-to-Bus Access Impedance	-.1702	.0302	-5.631
Auto-to-Bus Egress Impedance	-.1135	.0202	-5.631
Competing Mode Variables:			
Auto Time	.0545	.0099	5.509
Auto Cost	.4890	.0888	5.509
Other Terms:			
Intercept	-3.8394	.6760	-5.679

NOTE: The standard error shown for the intercept term is based upon an approximate calculation that ignores the covariance between the intercept term for the pooled regression and the intercept shift for this equation.

SOURCE: Regression Analysis of Travel Demand and Level of Service Data

TABLE 3 REGRESSION RESULTS FOR LOCAL PATH EQUATION IN MODAL-SPLIT MODEL

Variable	Coefficient	Standard Error	T-Statistic
Local PATH Service Variables:			
Local PATH Cost	-1.0370	.0917	-11.314
Local PATH Line Haul Time	-.0773	.0068	-11.314
Local PATH Wait Time	-.1545	.0137	-11.314
Local PATH Transfer Time	-.1545	.0137	-11.314
Local PATH Access Impedance	-.2318	.0205	-11.314
Local PATH Egress Impedance	-.1545	.0137	-11.314
Competing Mode Variables:			
Auto Time	.0342	.0015	22.849
Auto Cost	.2831	.0117	24.235
Bus Generalized Cost	.0447	.0046	9.614
Other Terms:			
Intercept	2.6844	.5717	4.695

NOTE: The standard error shown for the intercept term is based upon an approximate calculation that ignores the covariance between the intercept term for the pooled regression and the intercept shift for this equation

SOURCE: Regression Analysis of Travel Demand and Level of Service Data

because each captures a significant number and consistent share of the trips within its market area.

Bus park and ride and PATH park and ride are handled least well. The primary motivation for defining these travel paths as modes was to remove the influence of fringe park-and-ride users from bus and local PATH coefficient estimates. In doing so, two small share modes were created, neither of which had a strong facility orientation. They drew a small market share from a wide region, and were difficult to predict. However, the accuracy of Auto-Bus and Auto-PATH forecasts was judged adequate given the small number of trips these modes attract both across the region and within each interchange. It should also be noted that other modeling efforts are under way at NJ Transit to deal more specifically with the park-and-ride modes and to supplement the more aggregate forecasts of this model.

The uniform forecast accuracy among the conventional transit modes is a positive characteristic of the model. Bus, direct rail, and rail with transfer to PATH are all replicated well by the model. Predictions for these modes are only marginally less accurate than the automobile forecasts. The model is not biased toward any of the conventional transit modes. The model also does not exhibit any strong geographic bias in predictive accuracy. Root mean square errors by mode within the northeast test area (Hoboken Division) are consistent with those throughout the region. Predictive accuracy within the southwest (Newark Division) is generally consistent with the region, though the treatment of automobiles there is somewhat less accurate.

The model has been applied extensively by NJ Transit and the Port Authority of New York and New Jersey, as well as by a variety of consultants to analyze options for improving access, travel times, and capacity in the trans-Hudson corridor. It has proven itself to be a sensitive and flexible tool that has made an important contribution towards resolution of the trans-Hudson crisis.

ACKNOWLEDGMENTS

This mode choice model is the result of a collaborative effort among NJ Transit, the Port Authority of New York/New Jersey, and Charles River Associates, Inc. The success of the effort is due to the participation of Jeffrey Zupan and James Redeker of NJ Transit, Eugene Lessieu and Gerald Pfeiffer of the Port Authority, and Daniel Brand and Kevin Neels of Charles River Associates, Inc. The study was financed in part with a grant from UMTA, U.S. Department of Transportation.

REFERENCE

1. H. Theil. On the Estimation of Relationships Involving Qualitative Variables. *American Journal of Sociology*, Vol. 76, 1970, pp. 103–154.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.