RailRider—A Comprehensive Commuter Rail Forecasting Model

HOWARD L. SLAVIN, ZVI TAREM, ERIC A. ZIERING, AND ROBERT BRICKMAN

Presented in this paper are the results of research performed to develop a modeling capability for the Long Island Rail Road (LIRR) that would make it possible to forecast the impact of changes in parking supply, price, and other commuter rail service characteristics on LIRR ridership and on the access mode choice, station choice, and parking lot use of riders. In this project, Caliper Corporation and the Long Island Rail Road developed and implemented a unified network modeling approach that incorporated these travel choices and was applied within a modified stochastic user equilibrium framework. This model was implemented in a user-friendly microcomputer software package to provide easy access to this forecasting capability.

Presented in this paper are the results of research performed to develop a modeling capability for the Long Island Rail Road (LIRR) that would make it possible to forecast the impact of changes in parking supply, price, and other commuter rail service characteristics on LIRR ridership and on the access mode choice, station choice, and parking lot use of LIRR riders. Although it is commonly recognized that parking availability can be a key determinant of rail ridership, quantification of the impact of specific changes in parking supply and price has generally been beyond the reach of current forecasting techniques. Nevertheless, the forecasting problem posed is one that is faced by virtually all urban rail passenger transit systems, both existing and proposed.

The challenging nature of the task stemmed from numerous methodological and empirical difficulties. First, the small-scale changes that were of interest to the LIRR were numerous, were not confined to any particular geographic subarea, and could not necessarily be anticipated. Using subarea study approaches, therefore, was not feasible. Second, the forecasting capability had to be able to deal with major changes in system configuration and operation that are being implemented as part of the LIRR’s capital program. This was particularly important with respect to anticipated major changes in the rail network, especially the electrification of the LIRR Main Line through central Long Island, which involved major track reconstruction, the relocation of several commuter stations, and significant improvements in overall level of service. As a result, the model also had to be comprehensive enough to be able to forecast the impact of these changes.

Third, the forecasting model had to account for the multiplicity of travel choices faced by commuters in the Long Island-Manhattan travel market. These choices include whether to ride the LIRR, and, if choosing to do so, station, access mode, and parking lots selected (for LIRR riders who use park access only). Travelers do not consider these decisions separately in a predetermined sequence. The access mode a rail traveler will select depends on the station used; the decision to commute by automobile is often motivated (in the LIRR’s case) by unavailability of parking at a traveler’s preferred station. This interdependence of travel choices is often ignored or overly simplified in demand forecasting, resulting in unrealistic forecasts.

Fourth, the model had to be able to treat explicitly the capacity constraints and congestion effects that influence rail ridership and parking. To the extent that more travelers want to use travel paths than can be accommodated or to the extent that travel costs increase with travel volume, inconsistent forecasts are likely to be produced. Thus, achievement of consistent supply-demand forecasts was considered essential.

TECHNICAL APPROACH

The study approach integrated significant technical efforts in market research, travel demand modeling, network equilibrium analysis, and software development. Existing LIRR data were supplemented by a parking inventory that obtained base data on parking supply and use at the LIRR’s 110 commuter stations. An onboard survey containing a stated preference experiment was conducted so that travelers’ behavioral responses to changes in parking conditions could be quantified with choice models. Extensive analysis and reconciliation of data from multiple sources were required to obtain base-case aggregate data on passenger use of the LIRR by origin zone and on passenger flows on the various rail network links.

A unified travel-demand forecasting model was developed within the framework of stochastic user equilibrium on a network. In this framework, the various choices open to commuters (including choice of mode, access mode, LIRR station, and parking lot) were jointly analyzed in terms of a supernetwork (1) made up of links representing these alternative travel paths. Estimated monetary values of level-of-service attributes from prior- and new-traveler preference models were used as the basis for the perceived generalized costs of rail network links of various types.

A stochastic user equilibrium assignment methodology (2), modified to incorporate capacity constraints, was used to assign travelers to the various modes, access modes, LIRR stations, and parking facilities. The equilibrium formulation ensures that
consistency is achieved between forecast demand and level of service despite the flow dependence of the latter.

In stochastic user equilibrium, travelers vary in their perceptions of link costs and no user believes that he can unilaterally reduce his disutility of travel by selecting an alternative travel path (1). Because perceptions of link costs vary among travelers, stochastic user equilibrium produces much more realistic assignments than does deterministic user equilibrium, in which only the least-cost paths are used. The basic feasibility of the stochastic user equilibrium approach in building an empirical rail demand model had been established in previous internal research and development by Caliper Corporation and in a small-scale application to the problem of forecasting LIRR riders’ destination terminal choices under alternative scenarios (2).

BACKGROUNf RESEARCH
Survey research was conducted with 900 LIRR passengers to measure their preferences with respect to access mode choice. These data were combined with information collected in previous survey research performed for the LIRR. On-board survey research conducted in 1982 (4) focused on the choice of commutation station by LIRR riders, and measured the relative importance of fares, travel times, parking availability, and access time. Subsequent research by Caliper Corporation developed improved measures of the value of travel time, and also measured the effects of transfers on travel choice (3). On-board survey research conducted in 1985 as part of this study focused on access mode and parking-lot choices.

In these studies, conjoint data collection was used to measure the manner in which current and potential riders trade off selected characteristics of their commutation trip. The data-collection efforts were combined with extensive statistical and econometric analysis using regression, logit, and ordered logit model-estimation techniques. The result of these efforts was an estimate of the dollar values (or marginal rates of substitution) for the various relevant determinants of travel choices. These dollar values were used to determine the appropriate costs on various links of the network, and were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Dollar Value/hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIRR and other mode travel time and wait time at LIRR stations</td>
<td>4.50</td>
</tr>
<tr>
<td>Access time to LIRR stations</td>
<td>5.22</td>
</tr>
<tr>
<td>Walk time from lots to LIRR stations</td>
<td>6.00</td>
</tr>
<tr>
<td>Egress time from LIRR terminals</td>
<td>2.94</td>
</tr>
<tr>
<td>Diesel service penalty</td>
<td>0.48</td>
</tr>
</tbody>
</table>

The diesel service penalty is applied at stations that provide diesel service only, and reflects the inconvenience associated with transferring to an electric train to reach the LIRR’s Pennsylvania Station and Brooklyn terminals.

Significant effort was also invested in developing an accurate estimate of base-case patterns of demand on the LIRR and on competing modes. Total ridership by mode was derived using LIRR ridership data and 1980 U.S. Census Journey-to-Work data for Queens, Nassau, and Suffolk counties. Rail ridership by station was derived by reconciling several different sets of counts and survey data collected by the LIRR.

Base access mode information was compiled through survey tabulations and through a comprehensive parking inventory that obtained parking supply and use statistics at all 110 LIRR commuter stations.

DEVELOPMENT OF FORECASTING MODEL
The forecasting model generates estimates of the travel choices that will be made by commuters in the Long Island–New York City travel market. These forecasts are based on characteristics of the LIRR system and associated parking facilities. The model encompasses all modes of travel, although all non-LIRR travel modes are aggregated within the model. The LIRR system is modeled in great detail, with individual stations and parking facilities included at the most disaggregate level.

The model is network based, with the region represented as a set of nodes and links that connects them. Travel flows can occur between any two nodes along a sequence of links (a path) connecting them. The movement between any pair of nodes along a link may be associated with some impedance or generalized cost. Nodes, on the other hand, do not have any costs associated with them. The base network consisted of 1,930 links and 846 nodes.

The service territory, which covers Queens, Nassau, and Suffolk counties, was divided into 50 zones. Commutation trips originate at these 50 origin zones and terminate at a single destination zone representing New York City. Commuters have hundreds of alternative paths to follow through the network to complete a trip from their origin to their destination, representing various paths through the LIRR system and alternative paths using non-LIRR modes of travel. Certain links represent facilities with a limited capacity (e.g., parking lots), and on these links the capacity limitation is modeled explicitly.

The model evaluates the various paths available to travelers by estimating the total cost of each path. This cost is computed by summing the costs over all the individual links that make up that path. Based on the total cost of the available paths, the model assigns the demand between any two points in the network to appropriate paths.

Because the model is an equilibrium model, the cost of traversing a link may depend on the volume of travelers using it. Therefore, the model reevaluates the costs of alternative paths on an iterative basis, and reassigns some portion of travelers based on the revised estimate of path costs. In addition, because the model is stochastic, the cost associated with traversing a link is represented as a sample from a distribution of costs. At each iteration, a random component is added to the cost of certain types of links to simulate the stochastic effect.

Each link in the model may, therefore, have up to four types of costs associated with it, as follows:

- Out-of-pocket cost (in dollars);
- Time (in minutes);
- Congestion penalties; and
- A random cost element, which simulates the stochastic nature of link costs.

All time costs are converted into dollar equivalents using the marginal rates of substitution presented earlier. Congestion penalties represent the decrease in level of service associated with higher demand levels. Congestion on LIRR service links
reflects lower levels of comfort and decreasing seat availability. Congestion on nonrail links reflects increased highway congestion and discomfort associated with increased subway crowding. The random-cost element represents the variation in link costs and commuter perceptions of link costs described earlier. It is the presence of this random cost element that distinguishes stochastic user equilibrium from deterministic user equilibrium.

The network assignment procedure that was used as the basis for the forecasting model is the stochastic user equilibrium (SUE) assignment procedure described by Sheffi, which uses the method of successive averages (1). The mathematical approach presented in that paper is proved to converge at a true SUE solution. The variant used here assumes that the path flows are logit distributed based on the differences in path costs. This formulation leads to a Gumbel distribution of the stochastic link cost component.

Initial implementation of the SUE algorithm identified several practical complications. First and most significant, the original SUE algorithm did not lend itself well to networks in which a large number of links have fixed capacities. This was a severe problem in the LIRR network, because more than 400 links represented parking facilities with fixed capacities. Simulation of fixed link capacities through exponentially increasing congestion penalty functions yielded an unstable model that would require countless iterations before reaching equilibrium.

As an alternative, an attempt was made to use a deterministic user equilibrium approach with fixed capacity constraints. This approach yielded a model that produced unrealistic solutions, in which many of the possible paths through the network (which were known to be used by travelers on the real LIRR network) were not used. This undesirable characteristic of deterministic equilibrium solutions is well known and was, in fact, a principal motivation for the development of the stochastic user equilibrium approach. Other ad hoc procedures for multipath assignment were also considered but rejected based on their unsound theoretical and empirical properties.

As a result, the SUE algorithm was modified to accommodate links with absolute capacity constraints. The modified algorithm performed well, but resulted in an increased computational burden in producing a network assignment. The slower speed of the forecasting model had a significant impact on the level of effort required to calibrate the model.

Calibration was the final step in the development of the forecasting model. In this step, model parameters are adjusted so that the forecasting model accurately reproduces base case conditions. Conceptually, calibration may be thought of as tuning the model to reflect the contribution of variables that influence travel behavior but are not accounted for explicitly. Calibration of the model was greatly complicated by the multiplicity of travel choices and the high degree of interdependence among flows on alternative paths. The mathematical problem of directly solving for the constants is computationally intractable, and an enormous computational burden was involved in informed trial-and-error tests. Ultimately, a semiautomated approach was developed that was extremely effective, although time consuming. Calibration was ultimately achieved to within 1 percent of the base case flows on most links. For links with very small flows, errors of less than one or two riders were achieved. This calibration was dramatically better than that typically achieved in large-scale urban transportation planning studies, in which link volumes may be off by more than 100 percent.

**RAILRIDER FORECASTING SOFTWARE**

A key element of this study was the incorporation of the forecasting model into a microcomputer software package that made the sophisticated modeling capabilities directly and easily accessible to the LIRR. Many network equilibrium models must be applied by a specially trained analyst who manually develops and codes the transportation network. Changes in fares or in service must be implemented through the time-consuming and laborious process of manually recoding the network representation of the system. Many of these systems produce results in a format that is not immediately accessible and requires considerable post-processing before it can be comprehended. In developing the software, the goal was to eliminate these problems, thereby reducing the likelihood of errors and facilitating use of the system by individuals without specialized training.

The product of the development effort was RailRider, a proprietary microcomputer software product that implements Caliper Corporation's network-based demand forecasting methodology. The RailRider software is a user-friendly, menu-driven microcomputer package that was developed to the standards set by commercial microcomputer software. An on-line context-sensitive help facility makes operating instructions and technical advice on producing forecasts immediately accessible to the user. No specialized technical knowledge of microcomputers or of the specific equilibrium algorithms used by the model is required to make use of the system. The software runs under MS-DOS on 80286- or 80386-based microcomputers.

The RailRider forecasting system has three major components: a file editor, a forecasting module, and a report generator. Each of these is described as follows. A schematic of the RailRider model appears in Figure 1.

**File Editor**

The file editor allows the user to edit seven different types of files that contain information on LIRR service, fares, and parking facilities and on overall commutation demand in the Long Island–New York City travel market. The file editor is designed to manage large numbers of input files effectively, and provides the user with the ability to give each file an alphanumeric label to assist in tracking scenario development and forecast generation. The editor functions in a fashion similar to a spreadsheet program, with special functions for adding and deleting records and for producing formatted printed copies of each file. Editing of alphanumeric fields is simplified by the incorporation of pop-up menus, and range checks are automatically performed on all numeric fields.

Shown in Figure 2 is the RailRider software as it offers the user a choice of files to edit. Note that each file is accompanied by its descriptive label. The RailRider display is shown in Figure 3 in the process of editing the 1986 Base Case Fare Zone file.
Forecasting Module

The forecasting module has two major components: a network builder, and the assignment routine. The network builder is a powerful program module that translates 19 types of input files into an internal representation of an appropriate network. The network builder by itself is an innovation in the application of network forecasting methods to transportation problems because it permits the user to specify the structure and parameters of a network by editing easily understandable files. The network builder also saves time and reduces errors in developing analysis scenarios.

Networks that are used in transportation analysis typically differ from the physical networks that they represent. Invariably, they include additional nodes and links that represent internal connectivity among system components but have no physical parallel. Individual LIRR stations, for example, must be represented in the internal network as two distinct nodes with a link between them. To model accurately the parking restrictions that apply to lots at LIRR stations, RailRider must construct an internal subnetwork of dummy nodes and links that connect origin zones with individual parking facilities.

In traditional transportation network forecasting applications, users are required to manually construct the internal network. RailRider’s network builder makes it possible for the user to edit a set of simple input files, each organized in a manner familiar to the user, and to have these files automatically converted into the complex internal representation required to produce an accurate forecast.

The other major component of the forecasting module is the assignment routine itself. As described earlier, the assignment routine is a stochastic user equilibrium assignment algorithm that has been modified to accommodate fixed-capacity constraints.

Report Generator

The report generator takes the results of an assignment and produces a variety of on-screen and printed reports. The information that can be abstracted from a RailRider network...
forecast is extensive, and the report generator is designed to allow the user to compile desired results efficiently and quickly, and with appropriate documentation.

When the forecasting module completes an assignment, it produces an output file that contains all the information required to produce and fully document sets of on-screen and printed reports.

The output file contains the names, dates, creation times, and descriptions of all the input files that were used to generate that particular forecast, the date and time that the output file was created, and a user-defined label that describes the run. The output file also contains a reproduction of the internal network representation, all of the text labels required for producing output reports, and the resulting network flows. As a result, reports can be produced from the output file without requiring that the original input files also be available. This alleviates a host of file-management problems.

The following reports are produced by the report generator:

- Summary reports
- System summary report
- Origin zone summary report
- County and town summary report
- Branch summary report
- Branch detail reports
- Branch passenger loading reports
- Branch parking summary reports
- Station detail reports
- Station parking detail reports

All reports come in both on-screen and printed formats, with the printed versions supplying slightly more detail than their on-screen counterparts. For printed reports, the user can select any combination through on-screen menus, specifying the branches and stations for which detailed reports are required. All printed reports are automatically accompanied by a header page that describes the run and the input files that were used in preparing the forecast. Several of the on-screen reports produced by RailRider are shown in Figures 4, 5, and 6.

The practical benefit of the report generator lies in the fact that a user can produce a forecast and examine some of the "top-line" results on the screen, and then return to that forecast several days (or weeks or months) later and generate more detailed reports. Even if the input files used to generate the forecast have been edited or deleted, the complete forecast results are still accessible.

RailRider also generates graphic output of forecast results, showing bar charts of demand-by-access mode at each station along each branch. A reproduction of the RailRider graphic display appears in Figure 7.

The RailRider software is designed with a sophisticated user interface that simplifies the use of the various program modules. The entire system is menu driven, with on-screen menus and prompts to help the user select an appropriate course of action. All user keystrokes are automatically screened to prevent illegal entries, and all numeric inputs are checked to ensure that they are within a valid numeric range.

RailRider has an integrated, context-sensitive Help facility, similar to that provided in many microcomputer software products. At any point in the program, pressing the Help key
accesses the on-line Help Manual, which contains over 50 screens of information on how to use the RailRider system. The on-line Help facility can be used, for example, while editing a file to clarify the definition of a particular data file, or while in the report generator to remind the user of the contents of a particular type of report. Shown in Figures 8 and 9 are various screens from the RailRider help system.

Because of this advanced interface, model users do not need to have extensive microcomputer experience to use the RailRider forecasting models, although rudimentary familiarity with MS-DOS and the microcomputer keyboard can be helpful.

Using RailRider

The RailRider forecasting software produces demand forecasts based on a particular scenario, which consists of a designated set of seven input files. These files are as follows:

- Fare zone files, which contain monthly ticket prices in each fare zone;
- Branch files, which provide information about service on LIRR branches;
- Station files, which provide information about the characteristics of individual stations;
- Parking lot files, which contain data on parking capacities, restrictions, and prices;
- Parking lot files, which contain data on parking capacities, restrictions, and prices;
- Local service files, which indicate local service connections on each branch;
- Express service files, which indicate skip-stop, express, or flyer connections between nonadjacent stations; and
- Demand files, which contain information on the total size of the Long Island to New York travel market.

There may be several files of each of these types available to the user. For example, there may be a base fare file and two

**FIGURE 7** Sample RailRider graphic display.

Slavin et al.
other fare files that represent fare increases of 3.5 and 7.0 percent. Additional input files of each type can be created by the user using the RailRider file editor module.

To create a scenario, the user must specify one file of each type to be used as input to the model. Because the input files can be used in a variety of combinations, a large number of scenarios can be generated from a limited number of input file types. With two files of each type, for example, 93 different scenarios can be created by mixing and matching the files in all their possible combinations.

After the user specifies a single file of each type, RailRider processes the input files, generates the appropriate internal representation of the network, produces the demand forecast, and creates the output file. All of these actions are performed automatically, without requiring user intervention.

After a forecast is complete, the user has complete freedom to access the results using only the RailRider report generator module.

TESTING AND APPLICATION OF RAILRIDER

Test runs of the RailRider model in its final calibrated form provide substantial evidence that the model performs well. It yields results that are consistent with expectations and external measurements. The fare elasticity exhibited by the model is between -0.13 and -0.18, depending on the level of fare change. These values are consistent with previous measurements. The model exhibits a travel time elasticity of -0.20 for changes of 10 percent or less. This is lower than normal for the transit industry but reasonable based on the LIRR's large mode share and the relatively long travel times experienced by Long Island commuters.

To test the capabilities of the RailRider model, forecasts were developed for 15 different scenarios specified by the LIRR. Most of these were hypothetical and did not represent actual service changes currently being considered by the LIRR. One of these scenarios, however, was a simulation of the electrification of the LIRR's Main Line, an ongoing capital improvement project scheduled for completion in the fall of 1987. Application of the RailRider model to this major system change was considered a critical test of the model's ability to produce reasonable, consistent, and internally valid forecasts of demand. Under this and all other tests, the RailRider model produced forecasts that were consistent with prior expectations and rational in light of the LIRR's understanding of the composition and behavioral characteristics of its ridership.

External validation, however, is a much more appropriate test for a forecasting model. The ongoing capital program projects and related changes to the LIRR system will in the future provide ample opportunity to validate the RailRider model forecasts against real-world behavior.

CONCLUSIONS

The RailRider forecasting model breaks ground in the implementation of a new generation of travel-demand forecasting procedures based on theoretical advances of the last decade. From a methodological perspective, the project demonstrates the feasibility of applying demand models within a stochastic user equilibrium framework to a problem of sufficient scale and complexity to warrant consideration of this approach for solving significant transportation planning problems.

Based on experience in this project, very different and inherently more plausible forecasts come from a modeling framework that deals explicitly with capacity constraints, flow-dependent costs, and supply-demand equilibration than from traditional four-step planning models.

The computer implementation of the model differs from other urban transportation planning models in that whereas the model itself is more complicated, the forecasting system is easier to use, largely because the software was customized for the LIRR system. Although the conceptual approach is transferable and could be applied to other urban transportation properties, the network structure and the behavioral models that are embedded in the model are specific to the LIRR. As a result, the model does not require a large amount of set-up work for a given application, and there is no need for specially trained analysts who are familiar with the arcane set-up procedures of the Urban Transportation Planning System (UTPS) and similar packages.

Significant automation and simplification of the application process was an essential design goal of the project that is believed have been successfully achieved. It is hoped that, through the on-line Help system, the automatic error checking, and the network builder, RailRider will have made progress in reducing or eliminating user errors in the forecasting process.

A concern at the outset of the project was the size of the network equilibrium problem that could be solved on a microcomputer within reasonable running time. Although RailRider strains the limits of current microcomputer performance, taking an hour to run on a fast microcomputer, the performance is still more than competitive with mainframe models, and the microcomputer implementation permits the use of many sophisticated software features that would not otherwise be possible. Also, the continuing speed increases available in low-cost hardware suggest that current limitations on problem size and performance will be short-lived.

Further development activities could increase still more the accuracy and usefulness of RailRider. Better base ridership data could support improved model calibration. Integration of the model with ticket type of choice models would yield a more
flexible forecasting capability. Finally, a more explicit representation of alternative commutation modes and incorporation of a more sophisticated mode choice model would improve the quality of the forecasts and expand the capabilities of the model to determine the impact on the LIRR of changes in service or capacity on competing modes.

REFERENCES


Publication of this paper sponsored by Committee on Rail Transit Systems.