TRB Paper Number: 910877

FORECASTING INTERCITY RAIL RIDERSHIP USING REVEALED PREFERENCE AND STATED PREFERENCE DATA

by

Takayuki Morikawa[1)]
Moshe Ben-Akiva[2)]
Kikuko Yamada[3)]

1) Department of Civil Engineering, Nagoya University, Chikusa-ku, Nagoya, 464-01, Japan.
(phone) 81-52-781-5111 ext. 4485

2) Department of Civil Engineering, M.I.T., Cambridge, MA 02139.
(phone) 617-253-5324

3) Social System Department, II, Mitsubishi Research Institute, Inc., 2-3-6 Otemachi, Chiyoda-ku, Tokyo, 100, Japan.
(phone) 81-3-3270-9211

April 1991

Submitted to Transportation Research Record

ABSTRACT

The aims of this paper are i) to present a methodology for incorporating revealed preference (RP) and stated preference (SP) data in discrete choice models, ii) to apply the methodology to intercity travel mode choice analysis, and iii) to predict new mode shares for each O-D pair resulting from changes in service levels. The combined estimation technique with RP and SP data is developed to promote advantages of the two complementary data sources. The empirical study of intercity travel demand demonstrates the practicality of the methodology by accurately reproducing observed aggregate data and by applying a flexible operational prediction method.

INTRODUCTION

Travel demand models are usually estimated with observations of actual behavior, or revealed preference (RP) data, using the methods of discrete choice analysis (e.g., Ben-Akiva and Lerman (*1*)). However, in estimating individual choice models RP data may be deficient for the following reasons:

i) it does not provide information on preferences for non-existing services;

ii) the choice set considered by the decision maker may be ambiguous;

iii) some service attributes are measured with error; and

iv) some attributes are highly correlated and/or lack variability.

These drawbacks can be alleviated to a great extent in a survey with hypothetical choice scenarios and fully controlled alternatives. Such experimental data are called stated preference (SP) data and they have been used by a number of travel demand researchers (e.g., Louviere *et al.* (*2*), Bates (*3*), and Hensher *et al.* (*4*)) as well as in marketing research (e.g., Green and Srinivasan (*5*), and Cattin and Wittink (*6*)). However, the applications of SP data in practical transportation studies are still limited due to the uncertain reliability of elicited preferences under hypothetical scenarios. Advantages and disadvantages of RP and SP data and potential biases specific to SP data are discussed in detail by Ben-Akiva *et al.* (*7*).

Since RP and SP data have complementary characteristics, this paper explores the idea of simultaneously using both types of data. The methodology includes explicit consideration of the unknown reliability of SP data and its objective is to yield more reliable travel demand models than those produced by separate or sequential SP and RP analyses. The following contexts exemplifies the main idea of the paper. It is often the case that the trade-offs among certain attributes cannot be estimated accurately from available RP data. For instance, high correlation between travel cost and travel time in RP data may yield insignificant parameter estimates for their coefficients. However, an SP survey with a design based on low or zero correlation between these attributes may provide additional information on their trade-offs. Although the SP responses may not be valid for forecasting actual behavior due to their unknown bias and error properties, they often contain useful information on trade-offs among attributes. Another context where SP data add critically important information on preferences is the introduction of new services such as a new type of high-grade passenger car in rail service. RP data alone cannot provide the information needed to assess the impact of such a new service.

In previous papers, the authors have proposed a methodology for statistically combining RP and SP data in estimating the travel demand models (Ben-Akiva and Morikawa (*8,9*)). The key features of the methodology are:

i) Bias correction: explicit response models for SP data that include both preference and bias parameters;

ii) Efficiency: joint estimation of preference parameters from all the available data; and

iii) Identification: estimation of trade-offs among attributes and the effects of new services that are not identifiable from RP data.

The objective of this paper is to demonstrate the effectiveness of the combined RP/SP estimation method by an application to predict intercity rail ridership in conjunction with changes in service quality. The changes in service considered include the introduction of a high-grade passenger car which could not be evaluated by analyzing RP data only.

Section 2 presents the combined RP/SP model estimation methodology. Estimation results of intercity mode choice models and prediction of mode shares for each O-D pair are presented in Sections 3 and 4, respectively. Section 5 includes the concluding remarks.

METHODOLOGY

Model Specification

Two different model types are considered: RP and SP models. The RP model represents market behavior by some appropriate structure (e.g., random utility model with discrete choices), while SP response is modeled by the SP model. As discussed earlier, although SP data might not be valid for forecasting market behavior due to unknown bias and random error properties, they often contain useful information on trade-offs among attributes and preferences for non-existing services. Thus, the role of SP data is illustrated by the following framework:

<u>RP model:</u>

$$\begin{aligned} u_{in}^{RP} &= \beta' \mathbf{x}_{in}^{RP} + \alpha' \mathbf{w}_{in}^{RP} + \varepsilon_{in}^{RP} \\ &= v_{in}^{RP} + \varepsilon_{in}^{RP},\ i=1,\dots,I_n^{RP},\ n=1,\dots,N^{RP} \end{aligned} \tag{1}$$

$$d_n^{RP}(i) = \begin{cases} 1\text{: if alternative } i \text{ is chosen by individual } n \text{ in the RP data; and} \\ 0\text{: otherwise} \end{cases} \tag{2}$$

<u>SP model:</u>

$$\begin{aligned} u_{in}^{SP} &= \beta' \mathbf{x}_{in}^{SP} + \gamma' \mathbf{z}_{in}^{SP} + \varepsilon_{in}^{SP} \\ &= v_{in}^{SP} + \varepsilon_{in}^{SP},\ i=1,\dots,I_n^{SP},\ n=1,\dots,N^{SP} \end{aligned} \tag{3}$$

$$d_n^{SP}(i) = \begin{cases} 1\text{: if alternative } i \text{ is chosen by individual } n \text{ in the SP data; and} \\ 0\text{: otherwise} \end{cases} \tag{4}$$

where

u_{in}= utility of alternative i to individual n;

v_{in}= systematic component of u_{in};

ε_{in}= random component of u_{in};

$d_n(i)$= choice indicator of alternative i for individual n;

$\mathbf{x}_{in}$, $\mathbf{w}_{in}$, $\mathbf{z}_{in}$= vectors of explanatory variables of alternative i for individual n; and

α, β, γ= vectors of unknown parameters.

The superscript RP or SP indicates the data type.

In the above framework, it is assumed that the SP response is a "choice" or the most preferred alternative presented to the respondent. Even when the SP response is given by other formats such as preference ranking or pairwise comparison with categorical response, the SP model can be based on the same random utility model. A different response format only requires a slightly different estimation method.

The term represented by $\gamma'\mathbf{z}$ is specific to the SP model and may include SP biases and effects of hypothetical new services that are included only in the SP survey. The appearance of β in both models implies that the trade-offs among the attributes in the vector $\mathbf{x}$ are the same in both actual market behavior and the SP tasks.

The level of random noise in the data sources is represented by the variance of the disturbance term ε. If RP and SP data have different noise level, this can be expressed by:

$$\mathrm{Var}\left(\varepsilon_{in}^{\mathrm{RP}}\right) = \mu^2 \mathrm{Var}\left(\varepsilon_{in}^{\mathrm{SP}}\right),\ \forall i,n\ . \tag{5}$$

If SP data contain more random noise than RP data, μ will lie between 0 and 1. μ is also known to represent the "scale" of the model coefficients.

Assuming independently and identically distributed (i.i.d.) Gumbel disturbance terms in the RP model, a logit model is obtained with the choice probability given by:

$$P_n^{\mathrm{RP}}(i) = \frac{\exp\left(v_{in}^{\mathrm{RP}}\right)}{\sum_{j=1}^{J_n^{\mathrm{RP}}} \exp\left(v_{jn}^{\mathrm{RP}}\right)}\ . \tag{6}$$

An i.i.d. Gumbel assumption for the SP utility disturbances leads to the following SP logit model which includes the scale parameter μ:

$$P_n^{\mathrm{SP}}(i) = \frac{\exp\left(\mu \cdot v_{in}^{\mathrm{SP}}\right)}{\sum_{j=1}^{J_n^{\mathrm{SP}}} \exp\left(\mu \cdot v_{jn}^{\mathrm{SP}}\right)}\ . \tag{7}$$

Model Estimation

The unknown parameter vectors α, β, γ and the scale parameter μ are jointly estimated using both RP and SP data. The log-likelihood functions for the RP and SP data sets are given by:

$$L^{RP}(\alpha,\beta) = \sum_{n=1}^{N^{RP}} \sum_{i=1}^{I_n^{RP}} d_n^{RP}(i) \ln P_n^{RP}(i) \tag{8}$$

$$L^{SP}(\beta,\gamma,\mu) = \sum_{n=1}^{N^{SP}} \sum_{i=1}^{I_n^{SP}} d_n^{SP}(i) \ln P_n^{SP}(i) \tag{9}$$

Separately maximizing (8) and (9) yields maximum likelihood estimators of the RP and SP models, respectively. In that case the scale parameter μ and the coefficients are not separable in the SP model.

By maximizing the sum of (8) and (9) we can force the β coefficients to be the same in the RP and SP models. Thus, the combined RP/SP estimator is obtained by maximizing the joint log-likelihood function:

$$L^{RP+SP}(\alpha,\beta,\gamma,\mu) = L^{RP}(\alpha,\beta) + L^{SP}(\beta,\gamma,\mu). \tag{10}$$

This estimator fully utilizes the information contained in both RP and SP data as discussed above. If the random terms of the RP and SP models for the same individual are assumed to be statistically independent, maximizing (10) will yield the maximum likelihood estimator of all the parameters. If the random terms are not independent, this estimator is consistent but the standard errors of the estimates calculated in the usual way are incorrect (Amemiya (*10*)).

Since the joint log-likelihood function (10) is not linear in parameters due to the introduction of μ, the estimation cannot be carried out using ordinary MNL software packages for logit models. If the response format of the SP data is choice then a program to estimate a Nested Logit model may be employed. Alternatively, the following sequential estimation method by the following method using ordinary software packages may be used to yield consistent but less efficient estimates.

Step 1:

Estimate the SP model (3) by maximizing (9) using the SP data to obtain $\widehat{\mu\beta}$ and $\widehat{\mu\gamma}$. Define $y_{in}^{RP}=\mu\beta'\mathbf{x}_{in}^{RP}$ and calculate the fitted value $\widehat{y_{in}^{RP}}=\widehat{\mu\beta}'\mathbf{x}_{in}^{RP}$ for the RP observations.

Step 2:

Estimate the following RP model with the fitted value $\widehat{y_{in}^{RP}}$ included as a variable to obtain $\hat{\lambda}$ and $\hat{\alpha}$:

$$u_{in}^{RP} = \lambda \widehat{y_{in}^{RP}} + \alpha' \mathbf{w}_{in}^{RP} + \varepsilon_{in}^{RP} . \tag{11}$$

where $\lambda = 1/\mu$.

Calculate $\hat{\mu} = 1/\hat{\lambda}$, $\hat{\beta} = \widehat{\mu\beta}/\hat{\mu}$, and $\hat{\gamma} = \widehat{\mu\gamma}/\hat{\mu}$.

The accuracy of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ can be improved by the following additional step.

Step 3:

Multiply $\mathbf{x}^{SP}$ and $\mathbf{z}^{SP}$ by $\hat{\mu}$ to obtain a modified SP data set. Pool the RP data and the modified SP data and then estimate the two models jointly to obtain $\hat{\hat{\alpha}}$, $\hat{\hat{\beta}}$ and $\hat{\hat{\gamma}}$.

In this paper the joint estimator is employed. It was implemented in a special program written in GAUSS.

Prediction with the RP/SP Models

For prediction only the RP model is used because our concern is actual behavior and not experimental response. Therefore, the systematic utility component used for prediction is given by:

$$\hat{v}_{in} = \hat{\hat{\beta}}' \mathbf{x}_{in} + \hat{\hat{\alpha}}' \mathbf{w}_{in} . \tag{12}$$

Note that $\hat{\beta}$ in the above equation are estimated using both RP and SP data. If some hypothetical services presented in the SP questions are to be included for predicting demand, the corresponding term in the SP model should be added to (12) as follows:

$$\hat{v}_{in} = \hat{\hat{\beta}}' \mathbf{x}_{in} + \hat{\hat{\alpha}}' \mathbf{w}_{in} + \hat{\tilde{\gamma}}' \tilde{\mathbf{z}}_{in} . \tag{13}$$

where

$\tilde{\mathbf{z}}_{in}$ = a subvector of $\mathbf{z}_{in}$, representing hypothetical attributes relevant to the policy changes; and

$\hat{\tilde{\gamma}}$ = estimates of the parameters on $\tilde{\mathbf{z}}_{in}$.

Terms from the RP and SP utility functions can be combined, as shown in (13), since the scale of the utilities is adjusted between the RP and SP models by introducing the scale parameter μ.

CASE STUDY — ESTIMATION OF INTERCITY MODE CHOICE MODELS

Description of the Survey Data

The survey was conducted to assess intercity rail ridership in conjunction with a planned replacement of trains with regular cars by trains with high-grade cars. The alternative travel modes in the study corridor are express bus (or coach) service and private cars. The corridor connects two districts between which it takes two to three hours by rail and four to six hours by bus and car. Currently the corridor is covered by 26 daily trains, of which four trains have high-grade cars. Since there is no difference in rail fare between regular and high-grade trains, these four high-grade trains are always fully booked. The rail operator is considering the upgrading additional trains and would like to know how many new rail passengers will be attracted from the competing modes.

A survey of passengers traveling in the corridor was conducted using pure choice based random sampling for the three competing modes. The questionnaire asked for the socioeconomic characteristics of the traveler, the attributes of the chosen mode, and availability of alternative modes. Level of service attributes such as travel time and cost for the chosen and unchosen modes were calculated using network data for the reported origin and destination of the trip.

Each respondent was also asked for a preference ranking of the three alternative modes under the following hypothetical scenarios:

[for rail passengers]

Scenario 1: status quo;

Scenario 2: better access to the bus terminal;

[for bus and car passengers]

Scenario 1: status quo;

Scenario 2: increase in frequency of high-grade trains (13 services daily);

Scenario 3: reduction in rail line-haul travel time by 10%;

Scenario 4: reduction in rail line-haul travel time and increase in frequency of high-grade trains.

The respondent was asked to rank in order the three travel modes under each scenario.

The numbers of usable responses are 274, 89 and 82 from rail, bus and car passengers, respectively. Those who said that they had no other available modes than the chosen one are assumed to be "captive" to the chosen mode. 133 respondents were found to be captive to rail and 17 and 40 were captive to bus and car, respectively. Captives are excluded from the calibration data set but they are included in the prediction of aggregate ridership.

Estimation Results

Three models were estimated: RP model, SP model, and combined RP/SP model, each of which was estimated by maximizing the log-likelihood functions (8), (9) and (10), respectively. The independent variables include:

i) line-haul travel time: line-haul travel time for rail and bus and total travel time for car (in hours);
ii) terminal travel time: travel time for access and egress trips for rail and bus (in hours);
iii) travel cost: travel cost per person (in 1000 yen); and
iv) business trip dummy: =1 if the trip is associated with a business purpose; =0 otherwise. This variable interacts with travel time and cost.

Since the pure choice based sampling was employed, the estimates of the alternative specific constants should be adjusted by the following correction formula (Manski and Lerman (*11*)).

$$\hat{\beta}_0^i = \hat{\hat{\beta}}_0^i - \log\frac{H_i}{W_i} \tag{14}$$

where,

$\hat{\beta}_0^i$: the adjusted estimate of the constant for alternative i;

$\hat{\hat{\beta}}_0^i$: the estimate of the constant for alternative i through the exogenous sample maximum likelihood;

H_i: share of alternative i in the sample (In case of SP models, sample share must reflect the repetitions of the SP questions for each respondent.); and

W_i: market share of alternative i in the population.

First, the RP model estimated from the RP data is shown in the first column of Table 1. The value of line-haul travel time for a business trip is approximately 1.5 thousand yen per hour, or $10/hour.

Estimation of the SP model used the SP data from the bus and car passengers so as to analyze their intention to switch to rail. A choice data set was created by taking the first ranked alternative as the most preferred one, or chosen one. Since few respondents had the full choice set, i.e., three alternatives, information on the second ranking was not used. A dummy variable that indicates the increase in frequency of high-grade trains was added to the rail utility.

The second column of Table 1 shows the estimates of the SP model. The high-grade train dummy has a significantly positive coefficient. The rail and bus constants are significantly different from those of the RP model, which may be ascribed to the use of only the bus and car passengers' SP data and/or to some SP biases. The value of line-haul travel time for business trips is approximately four hundred yen per hour, or $3/hour.

The third column of Table 1 shows the estimation result of the RP/SP combined model. The parameters are calibrated through the joint estimation method. Alternative specific constants are estimated separately from the RP and SP data because the two models show significant difference in those constants. This implies that alternative specific constant terms belong to $\alpha'\mathbf{w}$ and $\gamma'\mathbf{z}$ in the framework of Section 2.

The high-grade train dummy has a significantly positive coefficient. The value of line-haul travel time for business trips is approximately 5.6 hundred yen per hour, or $4/hour. The scale parameter μ holds 1.33 but it is not significantly different from 1.0, which suggests that the variances of the random terms in the RP and SP models are approximately the same.

PREDICTION FROM THE ESTIMATED MODELS

In this section, two types of aggregation techniques, "sample enumeration" and "representative individual", are applied to the estimated model to predict demand for policy changes.

Sample Enumeration Method

The fitted values of systematic utilities are given by equation (13) and then the fitted choice probabilities are calculated by substituting these values in the MNL form.

Aggregated demand in the population can be obtained by the sample enumeration method as follows: Assumed here is that the ratio of captives for each mode in the population is the same as in the sample. Here $C(i)$ is defined as the number of captives in the population, then the predicted aggregate demand of alternative i is calculated by equation (15).

$$
\begin{aligned}
N(i) &= C(i) + N \cdot S(i) \\
&= C(i) + N \sum_{j=1}^{I} W_j \frac{1}{N_{sj}} \sum_{n=1}^{N_{sj}} \hat{P}_{nj}(i) \\
&= C(i) + N \sum_{j=1}^{I} \frac{N_j}{N} \frac{1}{N_{sj}} \sum_{n=1}^{N_{sj}} \hat{P}_{nj}(i) \\
&= C(i) + \sum_{j=1}^{I} \frac{N_j}{N_{sj}} \sum_{n=1}^{N_{sj}} \hat{P}_{nj}(i) \\
&= C(i) + \sum_{j=1}^{I} E_j \sum_{n=1}^{N_{sj}} \hat{P}_{nj}(i), \quad i = 1, \ldots, I
\end{aligned}
\tag{15}
$$

where

N_{sj} = number of observations choosing alternative j in the estimation sample;

N_j= observed number of individuals choosing alternative j in the population;

N = total number of non-captive individuals in the population;

$S(i)$ = predicted share of alternative i;

W_j = observed share of alternative j in the population; and

$\hat{P}_{nj}(i)$ = predicted choice probability of alternative i for individual n sampled on alternative j.

And the expansion factor E_j is defined by,

$$E_j = \frac{N_j}{N_{sj}} . \tag{17}$$

Table 2 shows predicted aggregate demand by this method under the same four scenarios as used in the SP questions. Observed aggregate numbers are obtained from on/off counts for rail and bus trips and screen-line counts for car trips. It should be noted that the observed and predicted numbers under Scenario 1 (status quo) match perfectly because the full set of alternative specific constants estimated from the *RP data* are used in the predicted utilities. This desirable property of MNL models is obtained by separately estimating alternative specific constants from RP and SP data and using the RP constants for prediction. The table shows that high-grade trains significantly increase rail ridership.

Representative Individual Method

Another aggregation technique employed here is "representative individual" method. This method approximates aggregate shares by the choice probabilities of the "representative" individual. The representative individual can be created by calculating averages of attributes in the sample or assigning appropriate attribute values. This method is very operational when the model is transferred to some other places where disaggregate data are unavailable. It is known, however, that aggregate predictions through this method have an aggregation bias.

The fitted utility functions are also calculated by equation (13) with "representative" attribute values. This case study predicts prefectural level O-D trip tables between the two districts. Each O-D pair is treated as a market segment and average attribute values for each O-D pair in the sample are used for representative individuals.

Table 3 is the observed aggregate O-D table and the predicted one is shown in Table 4. Those two tables show fairly close agreement, which can be ascribed to good parameter estimates through the proposed method. Although not shown in this paper, predicted O-D tables under different scenarios were calculated.

CONCLUDING REMARKS

This paper presented the method of combined estimation of discrete choice models from RP and SP data. An empirical case study of intercity travel demand analysis demonstrated the practicality of the method. This case study predicted rail ridership under hypothetical scenarios such as introduction of high-grade trains.

When the RP and SP data were used simultaneously to estimate the mode choice model, alternative specific constants were estimated separately from each data set. Using the MNL estimates of the constants from the RP data enables us to reproduce the aggregate shares through the sample enumeration method. Aggregation by the representative individual method also accurately reproduced the observed O-D table. This is an encouraging result for using the combined estimation method and predicting demand under hypothetical scenarios.

The work presented in this paper and two previous studies by the authors (Ben-Akiva and Morikawa (*8,9*)) has shown the effectiveness and practicality of the methodology of combined estimation with RP and SP data. This paper is aimed at providing further evidence. However, more empirical work in different contexts may be needed to justify the methodology conclusively.

In addition, in ongoing work we are developing more efficient estimators that explicitly treat potential correlation between the random utilities of RP and SP models for the same individual.

REFERENCES

1. M. Ben-Akiva and S.R. Lerman. *Discrete Choice Analysis.* MIT Press, Cambridge, Massachusetts, 1985.
2. J.J. Louviere, D.H. Henly, G. Woodworth, R.J. Meyer, I.P. Levin, J.W. Stones, D. Curry and D.A. Anderson. Laboratory-Simulation Versus Revealed-Preference Methods for Estimating Travel Demand Models. In *Transportation Research Record 794*, TRB, National Research Council, Washington, D.C., 1981, pp. 42-51.
3. J. Bates. Stated Preference Technique for the Analysis of Transportation Behavior. *Proc. of the 3rd WCTR*, Hamburg, W. Germany, 1983, pp. 252-65.
4. D.A. Hensher, P.O. Barnard and T.P. Thruong. 1988. The Role of Stated Preference Methods in Studies of Travel Choice. *Journal of Transport Economics and Policy*, Vol. XXII, No. 1, 1988, pp. 45-58.
5. P.E. Green and V. Srinivasan. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, Vol. 5, 1978, pp. 103-23.
6. P. Cattin and D.R. Wittink. Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing*, 46, 1982, pp. 44-53.
7. M. Ben-Akiva, T. Morikawa and F. Shiroishi. Analysis of the Reliability of Stated Preference Data in Estimating Mode Choice Models. *Selected Proc. of the 5th WCTR*, Yokohama, Japan. Vol. IV, 1989, pp. 263-277.
8. M. Ben-Akiva and T. Morikawa. Estimation of Switching Models from Revealed Preferences and Stated Intentions. *Transportation Research A,* Vol. 24A, No.6, 1990, pp. 485-495.
9. M. Ben-Akiva and T. Morikawa. Estimation of Travel Demand Models from Multiple Data Sources. In *Transportation and Traffic Theory (Proc. of the 11th International Symposium on Transportation and Traffic Theory)*, M. Koshi, ed., Elsevier, 1990, pp. 461-476.
10. T. Amemiya. *Advanced Econometrics.* Harvard University Press, Cambridge, Massachusetts, 1985.

11. C. Manski and S. Lerman. The Estimation of Choice Probabilities from Choice-Based Samples. *Econometrica*, 45, 1977, pp. 1977-1988.

Table 1 Estimation Results (t-statistics in parentheses)

Variables	RP Model	SP Model	RP/SP Model
Rail constant (RP)	1.66 (5.4)		1.40 (5.1)
Bus constant (RP)	-1.43 (-5.0)		-1.59 (-5.9)
Rail constant (SP)		0.706 (2.4)	0.906 (4.0)
Bus constant (SP)		-3.37 (-1.6)	-3.24 (-1.9)
High-grade train dummy		0.702 (3.1)	0.520 (2.4)
Line-haul travel time × business trip	-0.458 (-1.7)	-0.370 (-0.6)	-0.270 (-1.4)
Terminal travel time × business trip (Rail and Bus)	-0.973 (-1.8)	0.232 (0.3)	-0.143 (-0.5)
Total travel cost	-0.402 (-5.5)	-0.336 (-4.7)	-0.294 (-4.3)
Business trip dummy × total travel cost	0.102 (0.7)	-0.551 (-1.2)	-0.187 (-1.6)
Scale parameter μ			1.33 (3.6)
N	255	434	689
$L(0)$	-191.35	-332.26	-524.61
$L(\hat{\beta})$	-149.25	-271.18	-427.59
ρ^2	0.220	0.184	0.185
$\bar{\rho}^2$	0.189	0.163	0.166

Table 2 Predicted Annual Trips and Modal Shares by Sample Enumeration (difference from the values under Scenario 1 in parentheses)

		Rail		Highway Bus		Car	
Observed	Annual Trips	5,365,865		117,237		1,808,940	
	Modal Share	73.6%		1.6%		24.8%	
Scenario 1		5,357,431		113,046		1,821,565	
(status quo)		73.5%		1.5%		25.0%	
Scenario 2		5,646,818	(+289,387)	80,992	(-32,054)	1,564,232	(-257,333)
(increase in high-grade trains)		77.4%	(+3.9%)	1.1%	(-0.4%)	21.5%	(-3.5%)
Scenario 3		5,369,599	(+12,168)	111,719	(-1,327)	1,810,724	(-10,841)
(reduction of rail time)		73.7%	(+0.2%)	1.5%	(-0.0%)	24.8%	(-0.2%)
Scenario 4		5,656,751	(+299,320)	80,112	(-32,934)	1,555,179	(-266,386)
(Scenarios 2 + 3)		77.6%	(+4.1%)	1.1%	(-0.4%)	21.3%	(-3.7%)

Table 3 Observed O-D Table (Annual Riderships and Shares)

	A1			A2			A3		
	Rail	Bus	Car	Rail	Bus	Car	Rail	Bus	Car
B1	191,768	0	37,230	414,524	2,664	79,570	191,768	1,332	29,565
	(83.7%)	(0.0%)	(16.3%)	(83.4%)	(0.5%)	(16.0%)	(86.1%)	(0.6%)	(13.3%)
B2	1,349,818	0	527,425	1,265,649	27,980	604,805	567,890	13,320	109,135
	(71.9%)	(0.0%)	(28.1%)	(66.7%)	(1.5%)	(31.9%)	(82.3%)	(1.9%)	(15.8%)
B3	475,767	0	146,730	621,082	66,610	218,635	287,600	5,329	55,845
	(76.4%)	(0.0%)	(23.6%)	(68.5%)	(7.3%)	(24.1%)	(82.5%)	(1.5%)	(16.0%)

Table 4 Predicted O-D Table through Representative Individual Method

	A1			A2			A3		
	Rail	Bus	Car	Rail	Bus	Car	Rail	Bus	Car
B1	181,179 (80.0%)	0 (0.0%)	45,819 (20.0%)	408,038 (82.1%)	7,722 (1.5%)	81,498 (16.4%)	179,301 (80.5%)	1,348 (0.6%)	42,016 (18.9%)
B2	1,453,364 (77.7%)	0 (0.0%)	417,541 (22.3%)	1,366,426 (72.0%)	35,672 (1.9%)	496,336 (26.1%)	592,188 (85.8%)	11,131 (1.6%)	87,027 (12.6%)
B3	467,374 (75.1%)	0 (0.0%)	155,122 (24.9%)	659,621 (72.7%)	58,552 (6.5%)	188,154 (20.8%)	300,924 (86.3%)	6,383 (1.8%)	41,466 (11.9%)

TABLE TITLES

Table 1 Estimation Results (t-statistics in parentheses)

Table 2 Predicted Annual Trips and Modal Shares by Sample Enumeration (difference from the values under Scenario 1 in parentheses)

Table 3 Observed O-D Table (Annual Riderships and Shares)

Table 4 Predicted O-D Table through Representative Individual Method