1203

TRANSPORTATION RESEARCH RECORD

Demand Forecasting and Trip Generation–Route Choice Dynamics

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1988

Transportation Research Record 1203
Price: $13.00

modes
1 highway transportation
2 public transit

subject areas
12 planning
13 forecasting

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Demand forecasting and trip generation-route choice dynamics
p. cm.—(Transportation research record ISSN 0361-1981 ; 1203)
Papers from the 67th annual meeting of the Transportation Research Board held in Washington, D.C., 1988.
ISBN 0-309-04769-2
1. Choice of transportation—Forecasting—Congresses. I. National Research Council (U.S.). Transportation Research Board. II. National Research Council (U.S.). Transportation Research Board. Meeting (67th : 1988 : Washington, D.C.) III. Series.
TE7.H5 no. 1203
[HE336.C5]
388 s—dc20 89-13060
[388.4'1314] CIP

Sponsorship of Transportation Research Record 1203

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Chairman: Ronald F. Kirby, Metropolitan Washington Council of Governments

Transportation Data, Economics and Forecasting Section
Chairman: Joseph L. Schofer, Northwestern University

Committee on Passenger Travel Demand Forecasting
Chairman: James M. Ryan, Comsis Corporation
Moshe E. Ben-Akiva, Daniel Brand, Werner Brog, Melvyn Cheslow, Robert M. Donnelly, Christopher R. Fleet, Lawrence V. Hammel, David A. Hensher, Joel L. Horowitz, Peter M. Jones, Ashok Kumar, David L. Kurth, David M. Levinsohn, Hani S. Mahamassani, Fred L. Mannering, James M. McLynn, Eric J. Miller, O. Morellet, Eric Ivan Pas, Joseph N. Prashker, Phillip R. Reeder, Martin G. Richards, Aad Ruhl, Earl R. Ruiter, Gordon W. Schultz, Bruce D. Spear, Peter R. Stopher, Kenneth E. Train, Jeffrey M. Zupan

Committee on Traveler Behavior and Values
Chairman: Mary Lynn Tischer, Federal Highway Administration
James C. Echols, David T. Hartgen, David A. Hensher, Joel L. Horowitz, Ryuichi Kitamura, Lidia P. Kostyniuk, Eric Ivan Pas, Richard H. Pratt, Earl R. Ruiter, Peter R. Stopher, Antti Talvitie, Carina van Knippenberg, Edward Weiner

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1987.

Transportation Research Record 1203

Contents

Predicting Peak-Spreading Under Congested Conditions

WILLIAM R. LOUDON, EARL R. RUITER, AND MARK L. SCHLAPPI

As the resources for expanding street and highway capacity in urban areas have become increasingly scarce, interest has risen in accurately predicting the peak-hour capacity requirements for future years. Much of the travel demand forecasting performed around the country for highway planning has been performed on a twenty-four-hour basis, and peak-hour capacity needs have been estimated by applying a regional factor for the specific facility type or by using the current ratio of peak-hour to twenty-four-hour volume for a specific highway segment under consideration. This method is most often static and does not reflect the reduction in peaking that generally occurs as facilities become congested during the peak hour and trip makers adjust their travel time to avoid the peak. This paper reports the results of research on the peak-spreading phenomenon using traffic data from highway corridors in Arizona, Texas, and California. The data from each corridor covered a period of five to twenty years during which the congestion level in the peak period changed significantly. The research demonstrated that a clear and consistent pattern of peak-spreading emerged for highway facilities as congestion occurred during the three-hour morning and evening peak periods. The relationships derived from the research on peak-spreading have allowed the authors to develop a submodel for the UTPS UROAD assigned package. The new submodel will predict, for each link in the highway network, a peak-hour volume. That peak-hour volume reflects the level of congestion that would result from the predicted three-hour, peak-period volume for the forecast year and the level of capacity planned for each link; and it reflects the effect of peak-spreading that results from the predicted congestion on the facility. More accurate prediction of the peak-hour volumes is also expected to result in better prediction of peak-hour speeds for forecast years. This in turn should result in more accurate forecasting of travel time savings and air quality improvements from highway improvement projects.

In the past ten years, there has been a clear trend away from federal funding of transportation projects. An increasing share of the cost of highways and public transportation must now be borne by state and local governments, for which generating the necessary revenue is far more politically sensitive and, therefore, more difficult. With the shift to state and local finance has come not only an increase in the detail by which capacity needs are evaluated but also an increase in the concern about the benefits that are gained from improvements. Travel time savings, air quality improvements, and improvements in traffic safety are all being examined with greater attention to quantitative estimation. For that reason, it is important that modeling systems accurately predict not only the volume of travel during the periods of peak demand but also the speed at which that travel will occur.

This paper presents the results of research conducted for the Arizona Department of Transportation (ADOT) on the phenomenon of peak-spreading on congested roadways. The research was conducted using a national cross-section of data but with specific application to the Phoenix metropolitan area. The research was designed to result in recommended changes to the UTPS-based forecasting system used by the Maricopa Association of Governments (MAG) Transportation Planning Office that would allow them to reflect peak-spreading phenomena in future-year forecasting.

Considerable research has been conducted on the impacts of time of day on travel volumes, facility speeds, and trip-making behavior. Variations in traffic volumes over the hours of the day have long been observed, and typical patterns for specific types of facilities in a range of urban contexts are provided in the transportation literature (*1*, *2*). Similarly, relationships between facility speeds and volumes have been studied extensively, and a range of mathematical functions has been proposed to represent these relationships (*3*, *4*). More recently, behavioral approaches to travel modeling have focused on how individual travelers make their travel choices, in many cases including considerations of time of day (*5–8*).

There has also been considerable research on the incorporation of travel choice theory into network modeling systems (*9–11*). A major deficiency, however, has been in the area of incorporating peak-spreading as a result of traffic congestion into the large-scale traffic assignment and network equilibrium systems, such as UTPS, required for detailed highway system analysis in major metropolitan areas. Because these modeling systems cannot feasibly be applied at the behavioral or individual traveler's level, the focus in this project was limited to identifying and implementing aggregate representation of peak-spreading phenomena.

The most common practice in modeling of peak-hour travel is to produce a twenty-four-hour assignment and

W. R. Loudon, JHK & Associates, 5801 Christie Avenue, Suite 220, Emeryville, Calif. 94608. E. R. Ruiter, Cambridge Systematics, 222 Third Street, Cambridge, Mass. 02142. M. L. Schlappi, MAG Transportation Planning Office, Transportation Planning Division, Arizona Department of Transportation, 206 S. 17th Avenue, Room 108P, Phoenix, Ariz. 85007.

predict peak-hour trips as a constant percentage of the twenty-four-hour volume (often 10 percent). Some agencies have developed peak-period models by using the percentage of each trip type that occurs during a peak period to create a peak-period trip table (*12*). Even using this approach, however, the percentage of travel occurring in the peak one-hour period is generally a fixed percentage of the peak period, and no effort is made to relate peaking characteristics to the anticipated level of congestion for the assignment. The result is generally an overprediction of the peak-hour volume and often an underprediction of peak-hour speeds.

The need for more accurate modeling of peak-hour volumes was demonstrated by an examination of the variation in peak-hour volume as a percentage of twenty-four-hour volumes for forty-nine freeway and arterial corridors in Arizona, Texas, and California. The data from these forty-nine corridors formed the basis for most of the analysis in the project. The corridors for which historical data were available included:

13 in Arizona
17 in California
19 in Texas

The corridors included thirty-three freeways and sixteen major arterials. The corridors were selected because each had historical hourly count information covering at least a five-year period, and each facility had at least 20 percent growth in traffic during that period. Also available for each facility were number of lanes and capacity.

The range in average a.m. peaking factors (the ratio of maximum one-hour counts to daily counts) across sites was .081 to .126 in the a.m. period and .077 to .123 in the p.m. period. Although the midpoint in the range is almost exactly .100, in both the a.m. and p.m. peak periods (the same as the value most often assumed), the high end of the range of averages in both cases is more than 50 percent larger than the low value in the range. The significant variation strongly suggests the need for more accurate modeling of peak-hour volumes.

APPROACH

The approach in the research for ADOT was to improve the overall modeling of peak-period volumes and speeds by attempting to increase the accuracy of modeling in two areas:

1. Modeling of the peak periods
2. Modeling of peak-spreading within the peak periods as a facility becomes congested

The modeling process that was recommended to implement the research findings of the project is illustrated in Figure 1. The first step is to produce separate trip tables for each of the three time periods: a three-hour a.m. peak, a three-hour p.m. peak, and an off-peak that includes all other times. For MAG, this step was implemented using UMATRIX, but other matrix manipulation programs could be used.

The actual forecasting of peak-spreading in the package implemented for MAG occurs within UROAD. Following development of the peak-spreading models described in this paper, and of peak-hour volume-speed models, an augmented version of the UROAD network equilibrium traffic assignment program was prepared. In this program the new peak-spreading and volume-speed models are applied to each link every time that link speed updating is required. The added steps consist of the following:

- Compute ratio of current assigned (three-hour) volume to three-hour link capacity.
- Apply peak-spreading model to provide peaking factor: the ratio of one-hour volume to three-hour volume.
- Determine peak-hour volume as the product of the peaking factor and the assigned volume.
- Compute ratio of peak-hour volume to hourly link capacity.
- Apply peak-hour speed model to estimate revised link speed.

This link updating process continues throughout the iterative equilibrium procedure.

When the network assignment is complete, link volumes represent peak-period (three-hour) flows, but link speeds correspond to peak-hour conditions. The UROAD modifications provide the option, at this point, either to use the peak-spreading model to determine a final set of peak-hour volumes or to retain the peak-period volumes.

The two key elements of this package of modeling procedures are the peak-period factors that specify the percentage of travel for each purpose that occurs in the peak three-hour period and the peak-spreading model that predicts the percentage of travel in the peak three-hour period occurring in the peak hour. These two elements were the focus of the research, and the results of the research in both areas are reported in the remaining pages of this paper.

MODELING OF PEAK PERIODS

The first step in the production of peak-hour assignments was the division of total daily travel by trip purpose into three periods:

a.m. Peak—6:00 a.m. to 9:00 a.m.
p.m. Peak—3:00 p.m. to 6:00 p.m.
Off-Peak—All other hours

These periods were selected because it was felt that there would be some degree of stability within each period; travelers would not tend to shift out of these peak periods to avoid congestion. As a result, the percentage of travel predicted for each peak period should not vary with the level of congestion. To test this hypothesis, a regression

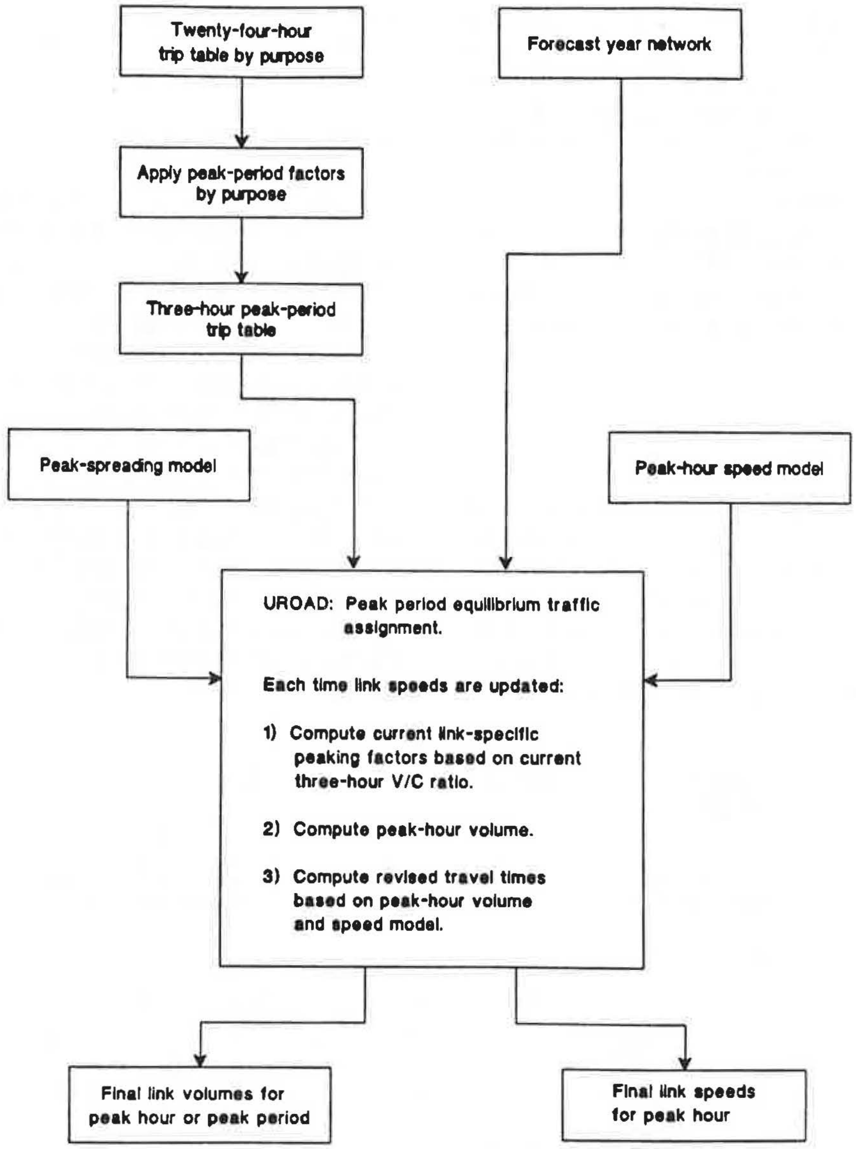


FIGURE 1 Structure of methodology for model enhancement.

analysis was performed for eighteen of the forty-five corridors for which historical data were available. The eighteen corridors were those for which a significant relationship between congestion and peak-spreading within the three-hour peak had been established.

In this analysis, the dependent variable, the ratio of three-hour volume to twenty-four-hour volume, was regressed on the peak three-hour V/C (volume/capacity) ratio as the independent variable. A coefficient for the independent variable that is significantly different from zero would indicate a relationship between the ratio of three-hour volume to twenty-four-hour volume and the three-hour V/C ratio and would lead one to reject the original hypothesis. Only five of the eighteen corridors yielded results that were of the correct sign (negative, indicating a reduction of the fraction of twenty-four-hour volume occurring in the peak three-hour period as a facility becomes congested) and statistically significant at a 95 percent confidence level. Separate regressions were run for the a.m. and p.m. peak periods for each of the eighteen corridors for a total of thirty-six regressions. In twenty-eight regressions the coefficients estimated were of the correct sign, but most could not be considered significantly different from zero at the 95 percent confidence level. This analysis indicates that there is some tendency for peak-spreading to affect hours other than the three hours in the designated peak period, but the tendency does not appear to be a statistically significant one.

The division of trips in Phoenix into the three time periods was based primarily on the reported time of travel by trip purpose by respondents in the Phoenix 1981 Household Survey. In prior work by MAG, each trip record in the household survey was assigned a weight according to the location of the residence of the trip maker and the characteristics of the household from which the trip record was taken. The weights were designed to expand the sample of household trips to represent the full population. The percentage of trips in the a.m. and p.m. peak periods was determined by developing a frequency distribution of trips by trip purpose and by hour of the day, and summing the hourly percentages for each trip purpose in the two peak periods. Table 1 presents the results of those frequency distributions.

For all home-based trips, a distinction was made between trips made from the home (*P* to *A*) and trips made to home (*A* to *P*). The frequency distributions were also calculated using three different measures of travel volume:

- trips
- vehicle miles traveled
- vehicle hours traveled

To obtain the distribution of vehicle miles traveled, each trip was weighted by the length of the trip. Likewise the distribution of vehicle hours of travel was determined by weighting each trip by the time required to make the trip as reflected in a peak-hour skim tree developed from the MAG 1980 base network. Comparison of the results with observed traffic volumes by time of day indicated that the VMT-based distribution provided the best results and was the distribution recommended for use by MAG.

MODELING OF PEAK-SPREADING

The most significant advancement in peak-hour modeling and the primary focus of this paper came in the development of a model to represent the effect of peak-period congestion on the temporal distribution of demand during that period. It has long been recognized that as a facility becomes congested, some trip makers will adjust the time at which they travel to avoid the congestion, and that this leads to some flattening of the peak period. However, this behavior has not been captured in the common UTPS travel forecasting process used by most planning agencies.

In this part of the research the historical data from the forty-nine freeway and arterial facilities in Arizona, California, and Texas were used to estimate a functional relationship between the peak-hour factor (the ratio of the volume of traffic in the single highest hour to the volume during the three highest hours) and the *V*/*C* ratio during the three-hour peak.

TABLE 1 COMPARISON OF ALTERNATIVE PEAK PERIOD TRIP FACTORS

Trip Type	A.M. Trips	A.M. VHT	A.M. VMT	P.M. Trips	P.M. VHT	P.M. VMT
PRODUCTION TO ATTRACTION						
HBW	0.329	0.340	0.344	0.033	0.030	0.029
HBNW	0.067	0.082	0.086	0.086	0.079	0.076
HBShop		0.021	0.022		0.102	0.102
HBO		0.066	0.071		0.072	0.068
HBSchool		0.335	0.329		0.056	0.061
Ext-Int	0.056			0.119		
Airport	0.041			0.113		
ATTRACTION TO PRODUCTION						
HBW	0.011	0.009	0.008	0.285	0.300	0.303
HBNW	0.009	0.008	0.007	0.128	0.085	0.133
HBShop		0.003	0.002		0.172	0.183
HBO		0.010	0.009		0.117	0.119
HBSchool		0.007	0.008		0.107	0.102
Ext-Int	0.077			0.100		
Airport	0.089			0.031		
NHB TOTAL	0.051	0.055	0.057	0.201	0.225	0.231
EXTERNAL-EXTERNAL	0.131			0.205		
TOTAL	0.152	0.179	0.197	0.242	0.243	0.266

Note: The factors for all home-based trips are to be applied to the total number of home-based trips produced; those from the production to attraction and those from attraction to production. The total to which the factor is applied is generally twice the number of trips for the direction indicated.

The functional form chosen for the peak spreading model was:

$$P = \frac{1}{3} + A\ e^{b(V/C)}$$

where

P = the ratio of peak-hour volume to peak-period (three-hour) volume,
V/C = the volume/capacity ratio for the three-hour period, and
a, b = model parameters.

The functional form was chosen because it has the general shape illustrated in Figure 2 and the following desirable characteristics:

- It always has a value of one-third or greater.
- P approaches one-third for large values of V/C.
- Valid values of P are defined for values of V/C greater than one.

The parameters in this equation were estimated using ordinary least squares regression for the transformed equation:

$$\ln(P - \frac{1}{3}) = \ln a + b(V/C)$$

or

$$\ln(P - \frac{1}{3}) = g + b(V/C)$$

where $g = \ln a$.

The regression model was estimated using data from the forty-five corridors for which historical data were available. The parameters g and b in the model were estimated for each corridor for the peak direction using data from both the a.m. and p.m. peak periods. The results are illustrated in Tables 2 and 3. Because the model was to be recalibrated for use in Phoenix by adjusting the value of g, the primary focus in this part of the analysis was on the estimates of b.

The existence of the peak-spreading phenomenon is clearly demonstrated for freeway facilities by the results in Table 2. The peak-spreading would be expected to occur only under congestion conditions, so the regression analysis was performed using only peak direction travel. Of the thirty-two freeway corridors included in the analysis, nineteen had V/C ratios in excess of .75 at some time during the period observed. Of these nineteen corridors, eighteen were of the correct sign (negative), and more than half (eleven) had t-statistics greater than 2 (reflecting statistical significance at roughly a 95 percent confidence level). Three additional corridors had t-statistics between 1.5 and 2.0 and two, between 1.0 and 1.5.

The only corridor for which the estimated coefficient was of the incorrect sign (positive) was the Bay Bridge in the San Francisco Bay Area. This particular corridor, however, reflects the need for close attention to the variation in V/C ratio observed. The range in variation for the Bay Bridge was only .92 to 1.00, indicating severely congested conditions throughout the period of observation. When both peak and nonpeak directions were used for the Bay Bridge, the variation in the V/C ratio was .61 to 1.00 and the estimated coefficient (b) was significantly different than zero at a 90 percent confidence level.

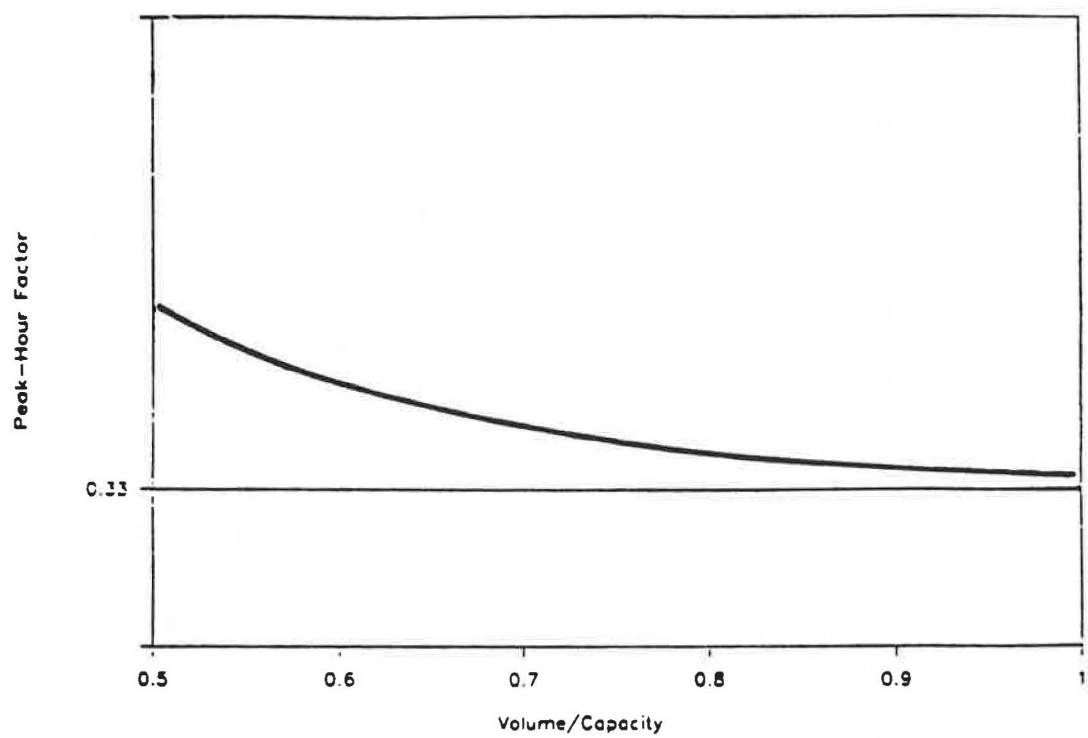


FIGURE 2 Theoretical relationship between peaking factor and volume/capacity ratio.

TABLE 2 FACILITY-SPECIFIC RESULTS OF REGRESSION ANALYSIS OF PEAKING FACTORS—FREEWAYS

CITY	CORRIDOR	b	g	# OBS	R-SQRD	ST ERR b	T-STAT b	V/C Range Low	V/C Range High
ARIZONA									
PHOENIX	I-10 & 49TH ST.	-2.19	-1.14	18	0.33	0.77	-2.83	.27	.41
CALIFORNIA									
BRENER PK	RT. 91 & HOLDER	-2.14	-2.60	8	0.05	3.65	-0.59	.67	.90
CASTAIC	I-5 AT RT. 126 W	-16.43	-1.38	7	0.27	12.16	-1.35	.16	.22
LOS GATOS	HWY 17 AT RT. 9	-8.66	3.26	10	0.65	2.24	-3.87	.73	.85
OAKLAND	HIGHWAY 24	-2.08	-1.08	8	0.18	1.81	-1.15	.82	1.17
PLEASANTON	I-580	-1.15	-2.61	10	0.43	0.47	-2.46	.60	.87
SAN DIEGO	I-5 AT ENCINADOS	-1.11	-2.60	9	0.06	1.71	-0.65	.42	.74
SAN FRAN.	GOLDEN GATE BR	-3.97	-0.46	4	0.09	8.86	-0.45	.57	.60
SAN FRAN.	BAY BRIDGE	0.92	-5.49	10	0.00	5.70	0.16	.912	1.04
SAN LUIS OBISPO	HWY 101 & LOS OSO RD	-5.17	-1.76	10	0.19	3.76	-1.37	.26	.37
SAN RAMON	HWY 580 & HWY 880	-1.85	-2.58	8	0.23	1.37	-1.35	.43	.66
SANTA ROSA	HWY 101 AT RT.12	-3.17	-1.15	10	0.38	1.44	-2.20	.65	.94
WILMINGTON	RT. 110 AT C ST	0.41	-3.22	6	0.01	2.42	0.17	.50	.61
TEXAS									
ARLINGTON	I-30	-2.18	-1.00	8	0.32	1.29	-1.69	.53	.78
AUSTIN	US-183	-1.14	-1.97	22	0.03	1.38	-0.83	.12	.36
AUSTIN	I-35	-2.48	-0.76	8	0.57	0.88	-2.82	.45	.87
AUSTIN	I-35	-1.31	-2.69	22	0.08	1.02	-1.28	.17	.56
DALLAS	I-635	-5.10	0.84	22	0.57	1.00	-5.11	.62	1.05
DALLAS	I-30	-3.22	-0.62	22	0.10	2.13	-1.51	.75	.94
DALLAS	I-35E	-1.58	-1.44	22	0.34	0.50	-3.18	.71	.89
DALLAS	US-75	-0.29	-3.00	22	0.00	1.69	-0.17	.66	.93
DALLAS	I-35E	-2.73	-1.73	22	0.12	1.63	-1.68	.47	.84
HOUSTON	I-45	-0.48	-3.12	22	0.09	0.34	-1.41	.50	1.03
HOUSTON	I-610	-4.78	0.59	20	0.35	1.53	-3.12	.55	.81
HOUSTON	US-59	-0.83	-2.57	22	0.34	0.26	-3.19	.31	.99
HOUSTON	US-59	-3.18	-0.83	18	0.33	1.13	-2.82	.69	.94
HOUSTON	I-10	-2.46	-1.03	22	0.20	1.09	-2.26	.70	.94
HOUSTON	I-610	-2.75	-1.25	22	0.19	1.28	-2.15	.32	.63
SAN ANTON.	I-37	-2.44	-1.34	20	0.26	0.96	-2.53	.30	.66
SAN ANTON.	I-410	-1.61	-1.62	18	0.15	0.94	-1.71	.25	.73
SAN ANTON.	US-281	1.04	-3.17	16	0.06	1.11	0.94	.26	.70
SAN ANTON.	I-410	-2.03	-1.29	20	0.37	0.62	-3.25	.48	1.00

REGRESSION EQUATION: LN(peak vol/3-hr vol - 1/3)= g + b(V/C)

TABLE 3 FACILITY-SPECIFIC RESULTS OF REGRESSION ANALYSIS OF PEAKING FACTORS—ARTERIALS

LOCATION	CORRIDOR	b	g	# OBS	R-SQRD	ST ERR b	T-STAT b	V/C Ratio Low	V/C Ratio High
ARIZONA									
PHOENIX	UNIVERSITY AT STANDAGE	-8.35	-1.80	21	0.31	2.86	-2.92	0.18	0.35
PHOENIX	BROADWAY AT STAPLEY	-6.84	-2.17	18	0.26	2.86	-2.39	0.18	0.32
PHOENIX	SCOTTSDALE AT THOMAS	15.82	1.45	10	0.50	5.60	-2.82	0.27	0.41
PHOENIX	US-60 AT CURRY	-3.41	-1.13	18	0.17	1.89	-1.81	0.30	0.59
TUCSON	WILMOT AT 22ND ST-SB	-1.38	-2.46	5	0.27	1.32	-1.04	0.43	0.74
TUCSON	WILMOT AT BROADWAY-SB	8.22	-10.56	3	0.99	0.81	10.15	0.69	0.82
TUCSON	WILMOT AT BROADWAY-WB	1.45	-3.62	5	0.19	1.70	0.85	0.50	0.68
TUCSON	SPEEDWAY AT CAMBELL-WB	-6.12	-0.05	3	0.71	3.90	-1.57	0.72	0.92
TUCSON	WILMOT AT BROADWAY-NB	-0.64	-4.56	3	0.00	6.66	-0.04	0.43	0.55
TUCSON	WILMOT AT 22ND ST-EB	-0.76	-2.92	3	0.93	0.20	-3.74	0.32	0.69
TUCSON	WILMOT AT SPEEDWAY-WB	-1.08	-2.28	3	0.08	3.59	-0.30	0.35	0.45
TUCSON	WILMOT AT SPEEDWAY-SB	-6.05	-0.51	5	0.43	4.05	-1.49	0.45	0.61
CALIFORNIA									
LOS ANG.	VENTURA AT SEPULVEDA	-2.31	-1.68	6	0.65	0.84	-2.75	0.54	0.85
LOS ANG.	WILSHIRE AT VENTURA	0.72	-4.49	10	0.03	1.53	0.47	0.83	1.06
LOS ANG.	WILSHIRE AT SEPULVEDA	-2.53	-1.65	6	0.16	2.93	-0.86	0.66	0.94
LOS ANG.	WILSHIRE AT WESTWOOD	2.72	-6.41	6	0.06	5.21	0.52	0.57	0.78

REGRESSION EQUATION: ln (peak vol/3-hr vol - 1/3) = g + b(V/C)

Thirteen freeway corridors did not have V/C ratios in excess of .75, and only three of these corridors produced coefficient estimates with t-statistics of 2 or greater. This provided support for the hypothesis that peak-spreading was significant only under congested conditions. For the freeway corridors with values of b significantly different than zero at a 95 percent confidence level, the range of values of b was −0.83 to −8.66 with a mean value of −2.96. Among the arterial corridors the range of values for the corridors with values of b significantly different than zero was −0.76 to −15.82; however, only one corridor had values of V/C that exceeded .75 for the three-hour period, reflecting the presence of congestion. The value of b for that corridor, Ventura Boulevard in Los Angeles, was −2.31.

The average value of b for each facility type was estimated by aggregating the data from all of the individual corridors but using only observations with a V/C ratio of .5 or greater. This screen was based on the evidence that peak spread occurs only at higher levels of V/C. Table 4 presents the results of the aggregate analysis for freeways and for arterials.

A single model was run using all data for freeways, but differences in the value of the aggregate freeway regression for freeways by number of lanes were also explored. Table 4 includes the results of regressions for freeways segmented by size of facility. Each regression produces values of b that are significantly different than zero, and the regression results together reflect a general trend of decreasing b (more negative) with increasing number of lanes. The regression results, by number of lanes, for the freeway data are presented graphically in Figure 3. Grouping the observations with four and five lanes or grouping those with two and three lanes produces the same pattern of decreasing b with increasing number of lanes, but the difference between the two values of b is relatively small. These values for b for the two size classes of freeways were used in the MAG models.

As reflected in Table 4, the results of the aggregate analysis for the arterial corridors was of limited usefulness primarily because of the lack of data from corridors with high V/C ratios. The aggregate regression did not produce values of b that were significantly different than zero at even a 90 percent confidence level. Because of the lack of useful aggregate results, the results from the Ventura Boulevard corridor in Los Angeles were used to represent arterial corridors in the MAG models.

CALIBRATION TO PHOENIX

To reflect current conditions (1985 and 1986) in Phoenix more closely, the model was recalibrated for different facility type and area type combinations in the Phoenix network using observed 1985 data. The calibration is performed using the relationship

$$a_{af} = (P_o - 1/3)/e^{b_f (V/C)_o}$$

where

P_o = the observed value of the peaking factor for the area type–facility type combination;

$(V/C)_o$ = the value of the V/C ratio for the area–facility type combination;

a_{af} = the value of a calibrated to a specific link type; and

b_f = the value of b previously estimated for the specific area type–facility type under consideration.

The values of P_o and $(V/C)_o$ are the average observed 1985 values for the links in each area type–facility type combination.

An analysis of current peaking factors was performed on 1985 data collected in Phoenix. Traffic counts were available for 517 locations in the metropolitan area. Most locations were two-way facilities and produced count data for two one-way links in the network. In all, 988 links had count data.

An analysis of the current peaking factors and V/C ratios for Phoenix supported the theoretical arguments for

TABLE 4 AGGREGATE RESULTS OF REGRESSION ANALYSIS OF PEAKING

Corridors Included	b	g	# Obs.	R-Sqrd.	St.Err. b	T-stat b
FREEWAYS						
All corridors	-2.207	-1.460	388	.215	.214	-10.6
4 or 5 lanes	-2.369	-1.377	196	.238	.304	-7.8
2 or 3 lanes	-2.003	-1.575	192	.189	.301	-7.6
ARTERIALS						
All corridors	-0.977	-3.112	48	0.040	.706	-1.4

Regression equation: ln(peak vol./3-hr. vol. - 1/3) = g + b(V/C)

Only observations for which the V/C ratio exceeded .5 were in the regressions.

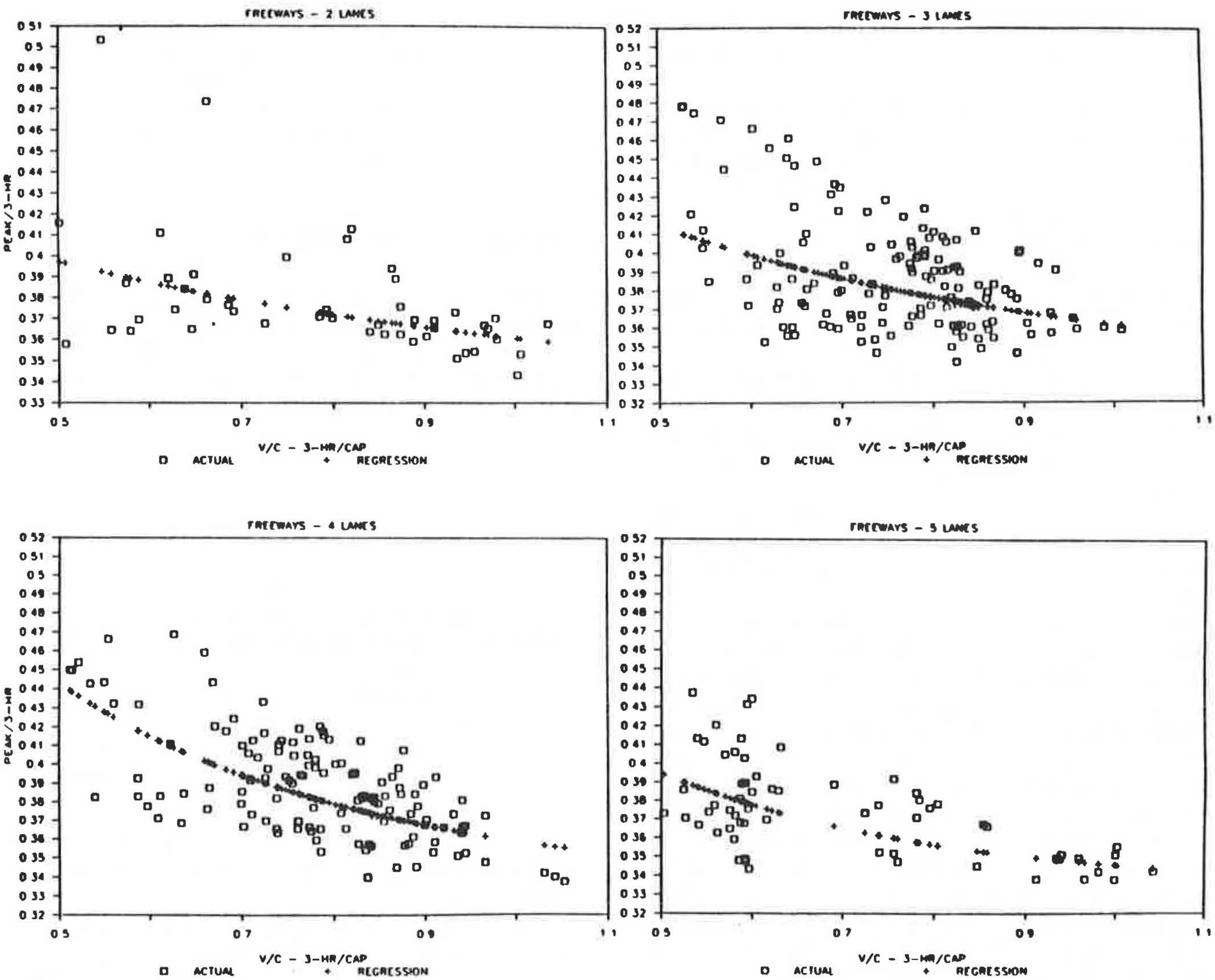


FIGURE 3 Plot of regression analysis of peaking factors for freeways.

the model of peak-spreading—that as facilities become congested, the peaking of traffic is reduced. This was demonstrated by the significantly higher peaking factors in the a.m. period when the average V/C ratio was generally much lower than in the p.m. period. The peaking factors were also generally lower in the denser areas where the V/C ratio was also generally higher.

VALIDATION OF RESULTS

A test of the peak-spreading procedures was performed by preparing assignments for 1985 and comparing the new assignments with observed data on the links where counts and speeds were available. The combination of the procedures described in this paper, the use of a three-hour peak-period trip table and a peak-spreading model, produced a significant increase in the accuracy in the assignment for 1985, when a comparison was made based on the peak-hour prediction of link volumes and speeds. The procedure used in the baseline assignment to which the new forecasts were compared was to produce a twenty-four-hour assignment and to assume that the peak hour was always 10 percent of the twenty-four-hour volume.

Figure 4 illustrates the results of the comparison. One of the principal motivations for the project was to produce better estimates of speeds, because the speeds being predicted by existing procedures were well below those observed on the links. As Figure 4 demonstrates, the desired result was produced with almost a doubling of peak-hour speeds. The resulting improvement in accuracy is illustrated in a reduction in the root mean square error (the square root of the sum of the squared differences between observed and estimated speeds on links) by 35 percent—from 56.0 to 36.6.

The improvement in accuracy in the estimation of link volumes is illustrated by the percent error in the overall estimate of vehicle miles of travel (VMT) on the links for which counts were available. The error was reduced from 16.4 to 2.2 percent.

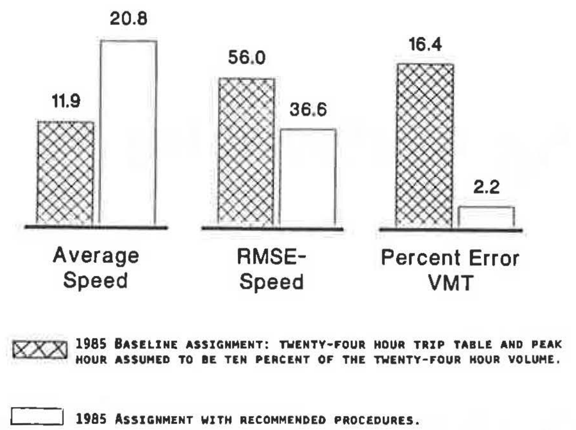


FIGURE 4 Validation results.

CONCLUSION

The research reported in this paper clearly demonstrates the need for accurate modeling of peak-hour travel volumes in planning for new or improved facilities. The standard practice around the country has been to produce twenty-four-hour assignments and to assume the peak hour to be a fixed percentage of the twenty-four-hour volume (usually 10 percent). And yet the variation from thirty-two freeway facilities examined as part of this research showed a range of plus or minus 25 percent of the mean value.

The research has also identified a clear pattern of peak-spreading as facilities become congested. Of nineteen facilities that exceeded a three-hour V/C ratio of .75, all but one produced regression coefficients of the correct sign (indicating peak-spreading), and more than half (eleven) were statistically different from zero at the 95 percent confidence level.

The research findings reported in this paper have been incorporated into the UTPS system of the Arizona Department of Transportation, and peak-hour assignments have been produced. Comparison of the assignment results from the new procedures with the results previously obtained by MAG for the Phoenix area using a twenty-four-hour assignment and constant peak-hour factor shows a significant improvement in accuracy from the new procedures.

There are some limitations to the new procedures. First, there is no guarantee of continuity of flow in the peak-hour prediction. Differences in the three-hour V/C ratio predicted for two adjacent links could result in a different amount of peak-spreading predicted for each. While this could and does occur, its impact is likely to be small because of the calibration of the peaking model on a facility type (rather than link-specific) basis, thereby averaging the effects over a facility.

A second limitation is that the peak-spreading model is applied at the link level, while the peak-spreading on a specific link may occur as a result of a single congestion point on some other link in the network or because of travelers' perception of the average level of congestion in the corridor. To the extent that links in a corridor are fairly homogeneous and the capacity on each link is generally proportional to the peak-period flow on the link, this limitation will not be a serious one.

A final limitation of the recommended procedures for peak-spreading is that they do not reflect spreading of the peak outside of a three-hour period. For the southwestern cities from which data were collected, this does not appear to be a significant limitation, but if the procedure is to be implemented in another part of the country, a broadening of the definition of the peak period may be appropriate.

Despite the procedures limitations, the improvement in modeling of the peak hour that they provide appears to be significant for Phoenix. This improvement has been demonstrated in terms of the prediction of link volumes and speeds. These limitations do, however, represent areas in which procedures might be improved in future research efforts.

REFERENCES

1. *Characteristics of Urban Travel Demand.* Wilbur Smith and Associates; Urban Mass Transportation Administration, U.S. Department of Transportation, Washington, D.C., 1987.
2. *An Analysis of Urban Area Travel by Time of Day.* Peat, Marwick, Mitchell & Co.; Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 1972.
3. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
4. D. Branston. Link Capacity Functions: A Review. *Transportation Research,* Vol. 10, 1976, pp. 223–236.
5. M. Ben-Akiva, A. de Palma, and P. Kanaroglou. Effects of Capacity Constraints on Peak-Period Traffic Congestion. *Transportation Research Record 1085,* TRB, National Research Council, Washington, D.C., 1986, pp. 16–26.
6. S. Cosslett. The Trip Timing Decision for Travel to Work by Automobile. In *Demand Model Estimation and Validation,* The Urban Travel Demand Forecasting Project Phase I Final Report, Vol. V. Institute for Transportation Studies, University of California, Berkeley, Calif., 1977.
7. M. Abkowitz. Understanding the Effect of Transit Service Reliability on Work Travel Behavior. *Transportation Research Record 794,* TRB, National Research Council, Washington, D.C., 1981, pp. 33–41.
8. S. A. Eisheh and F. L. Mannering. A Discrete/Continuous Analysis of Commuters' Route and Departure Time. Presented at the 66th Annual Meeting of the Transportation Research Board, Washington, D.C., 1987.
9. H. Mahmassani and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science,* Vol. 18, 1984, pp. 362–384.
10. M. Ben-Akiva, A. de Palma, and P. Kanaroglou. Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates. *Transportation Science,* Vol. 20, 1986, pp. 164–181.
11. J. L. Horowitz. The Stability of Stochastic Equilibrium in a Two-Link Transportation Network. *Transportation Research B,* Vol. 18B, 1984, pp. 13–28.
12. *1980 Travel Characteristics.* Working Paper No. 8, Metropolitan Transportation Commission, Oakland, Calif., 1983.

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

Traffic Volume Forecasting Methods for Rural State Highways

SUNIL K. SAHA AND JON D. FRICKER

This study builds on previous efforts found in the field of rural traffic forecasting. The study combines careful statistical analysis with subjective judgment to develop models that are statistically reliable and easy to use. This study developed two different kinds of models—aggregate and disaggregate—to forecast traffic volumes at rural locations in Indiana's state highway network. These models are developed using traffic data from continuous count stations in rural locations as well as data for various county, state, and national level demographic and economic predictor variables. Aggregate models are based on the functional classification of a highway, whereas the disaggregate models are location-specific. These models forecast annual average daily traffic (AADT) for future years as a function of present year AADT, modified by the various predictor variables. The use of both aggregate and disaggregate models will provide more reliable traffic forecasts. The number of predictor variables employed in the models was kept to a minimum. The statistical analysis also found that the predictor variables are statistically significant; no other variables will provide significant predictive power to the models. The models developed in this study provide good R^2 values. More refined statistical techniques reinforce the choice of variables used in the models.

The pattern of traffic growth and projected traffic volumes are prime factors in most analyses of highway projects. The traffic growth factor has a significant effect on highway investment decisions, such as whether to increase the capacity of existing highways and the construction of new facilities when funds are limited. Developing future traffic estimates is not an exact science, dependent as it is on so many hard-to-predict variables. Traffic forecasting procedures should be reasonably easy and economical to perform; they should be sensitive to a wide range of policy issues and alternatives and should provide decision makers with useful information in a form that does not require extensive training to understand.

Estimates of future traffic can be obtained by two very different methods: trend projections and forecasts. Analysts can modify extrapolated trends based on their experience and knowledge of the route, state, or region. With trend projections only traffic data are being dealt with; in forecasting techniques, however, a relationship between traffic and explanatory factors must be established. Traffic forecasting techniques therefore are also concerned with predicting the future values of economic and other measures or indicators of person and vehicle travel.

School of Civil Engineering, Purdue University, West Lafayette, Ind. 47907.

Traffic forecasting in urban areas has been extensively explored. Urban forecasting methodologies, based mainly on sophisticated computer modeling programs, are relatively advanced. On the other hand, forecasting traffic for individual rural roads, although widely practiced, is still in its early developmental stages. Standardized methodologies for nationwide use have not been established, and state authorities develop their own procedures to accommodate their needs. Various state departments of highways have developed methods of forecasting rural traffic, but very few of them are well documented.

PREVIOUS STUDIES

Four studies may be considered representative of the approaches used in the field of rural traffic forecasting.

Morf and Houska (*1*), in their 1958 study of the Illinois rural highway network, came to the conclusion that the four factors responsible for traffic growth patterns were (1) geographic location, (2) type and width of pavement, (3) proximity to an urban area, and (4) type of service the roadway provides. Their study also indicated that population is the principal component that affects the trend, followed by persons per vehicle, gasoline use, and vehicle miles per vehicle.

In 1982, Neveu (*2*) developed a set of elasticity-based models to forecast rural traffic. Neveu claimed that the type of service the roadway provides (interurban, interregional, rural to urban, urban to rural) is the only factor that had an appreciable effect on traffic growth rates. Multiple linear regression was used to identify factors that best estimated AADT and their respective elasticities. The factors used were population, number of households, automobile ownership, and employment. The roads were classified according to the type of service they provide: (a) interstates, (b) principal arterials, and (c) minor arterials and major collectors. The R^2 values 0.65, 0.77, and 0.20 for road types (a), (b) and (c), respectively, offer an indication of the explanatory power of the data. Two major problems were associated with the model: (1) the difficulty in obtaining projections of the factors/variables and their questionable accuracy at the level needed and (2) the uncertain degree of applicability of the model in certain areas (i.e., whether a specific area is "rural enough" for the model). Neveu used multiplicative constant elasticity in his model, which implies that the effect of the growth in demand on traffic growth will always be the same. The result is that if the model is estimated during a period of high growth rate, future traffic will be overestimated and vice versa. Thus, such models must be recalibrated as often as practicable. Models with variable elasticities are not very common in traffic forecasting. Such model structures involve more sophisticated and expensive analysis.

The Minnesota Department of Transportation (Mn/DOT) (*3*) computes a route-specific growth factor from a trend analysis of the specific route. After determining base year AADT, ten to twenty years (preceding the base year) of AADT counts are taken from traffic flow maps. By linear regression, a line is fitted to the data and that line is extended to the design year. The overall growth is then the difference between design year AADT and base year AADT. Similar graphical plots of AADT against time for all (or several) major highway segments are done along the proposed project. If the growth rates are uniform, a single rate can be applied to the entire project. If not, forecasters must then use judgment in selecting the appropriate rate for each segment based on their knowledge of the project area.

The New Mexico State Highway Department (*4*) has designed a procedure for forecasting heavy commercial (HC) and average daily traffic (ADT) traffic on the New Mexico Interstate system and then calculating the percentage of HC traffic. This process, called "Trend-line," starts with fourteen distinct geographical sectors on the New Mexico Interstate system. Separate forecasting models were developed for each sector. The disaggregate analysis (a separate analysis for each sector) provides better traffic projections than does aggregate analysis (all sectors together). Eight key demographic and economic indicators are used. In the statistical analysis, linear regressions were conducted using heavy commercial ADT (HCADT) and ADT as dependent variables. Historical data covering a period of six years were used. Regression analyses were conducted to find the best-fit equation, leading to equations that had R^2 values over 80 percent.

COMMENTS ON FORECASTING TECHNIQUES

Armstrong (*5, 6*), in his studies of forecasting, concluded that sophisticated extrapolation techniques have had a negligible payoff in terms of accuracy in forecasting. More sophisticated methods are generally more difficult to understand, and they cost more to develop, maintain, and implement. On the positive side, more sophisticated methods may be expected to produce more accurate forecasts and a better assessment of uncertainty. However, highly complex models may in fact reduce accuracy. Armstrong recommended simple methods and the combination of forecast techniques. He suggested starting with the least expensive method(s) and/or the most understandable method(s), and then investing in successively more expensive methods. He proposed that complexities should be avoided unless they are absolutely necessary.

In this study, considerable effort has gone into the statistical analysis of the relationships between traffic volumes and the predictor variables. Each statistical test, however, is available on most standard statistical packages for computers. More important, the resulting model is intended to be easy to understand and to implement.

PREPARATORY STEPS

Functional Classification of Highways

The standard functional classifications of rural highways (*7, 8*) used in this study are (1) rural interstate, (2) rural principal arterial, (3) rural minor arterial, and (4) rural major collector.

The Data

The traffic volume data comprise all usable observations associated with Automatic Traffic Record (ATR) count stations in rural Indiana between 1970 and 1982. These data were analyzed in two ways—by using disaggregate and aggregate techniques. In disaggregate analysis, each station is analyzed separately. Station- or location-specific models for highways with similar characteristics can be developed. In aggregate analysis, stations within a given category of highway will be analyzed as a group, and a model applicable to any highway classifiable within a certain group will be proposed.

Those stations in a functional category the data for which were clearly well out of the range of values for most of the stations in that category were not used in developing an aggregate model. Instead, these stations were "saved" to test the ability of an aggregate model to "predict" their AADT values. Also, because of an inadequate number of data points, a disaggregate model was not developed for some of the stations.

Resulting cases numbered 26 for rural interstates, 39 for principal arterials, 52 for minor arterials, and 37 for major collectors, respectively. The fact that the cases in a highway category were actually drawn from two to four locations, each with up to thirteen years of AADT counts, made careful statistical analysis of utmost importance. A further complication was that traffic forecasting models tend to involve relationships with predictor variables that are themselves forecasts.

The predictor variables (or background factors) used in this study are listed in Table 1. They were chosen because (1) they reflect conditions at the county, state, and national levels that may affect the amount of travel that takes place and (2) their values are fairly easy to obtain through sources accessible to the state highway department. The variables X_7 to X_{13} were used as candidate background factors only in the case of rural interstates and rural

TABLE 1 VARIABLES FOR TRAFFIC FORECASTS

Symbol	Description of the Variable	Type of Variable
Y	Annual Average Daily Traffic (AADT)	------------
X_1	County Vehicle Registrations	Demographic
X_2	US Gasoline Price in cents per gallon, in 1972 dollars	Economic
X_3	Year	------------
X_4	County Population	Demographic
X_5	County Households	Demographic
X_6	County Employment	Economic
X_7	State Vehicle Registrations	Demographic
X_8	State Population	Demographic
X_9	State Households	Demographic
X_{10}	State Employment	Economic
X_{11}	Consumer Price Index (CPI) -- US	Economic
X_{12}	Gross National Product (GNP), in billions of 1972 dollars	Economic
X_{13}	Per Capita Disposable Personal Income (nationwide), in 1972 dollars	Economic

principal arterials. The variables X_1 to X_6 were candidates in all highway categories.

Model Form

Various forecasting models were examined, and an elasticity-based model (*2*) was adopted to relate future year AADT to present year AADT by means of a number of background factors. The general form of the model is as follows:

$$\mathrm{AADT}_f = \mathrm{AADT}_p\left[1.0 + \sum_{j=1}^{n} \frac{e_j(x_{j,f} - x_{j,p})}{x_{j,p}}\right] \quad (1)$$

or, upon rearrangement,

$$\frac{\mathrm{AADT}_f - \mathrm{AADT}_p}{\mathrm{AADT}_p} = \sum_{j=1}^{n} e_j\left[\frac{x_{j,f} - x_{j,p}}{x_{j,p}}\right] \quad (2)$$

where,

AADT_f = AADT in future year,
AADT_p = AADT in present year,
$x_{j,f}$ = value of variable x_j in the future year,
$x_{j,p}$ = value of variable x_j in the present year,
e_j = elasticity of AADT with respect to x_j,
n = number of associated variables.

The elasticity-based model was selected for several reasons. Most important, it was believed that the range of volumes over which the model would be applied would be much greater than that used in developing the model, making a simple linear regression model relating AADT to the background factors directly inappropriate. Second, the use of present year AADT to estimate future year AADT (as a sort of pivot point) would reduce the problem of nonresident travel. Also, the elasticity portion of the model calculates a growth factor directly. (See the right-hand side of Equation 2.)

The AADT values were obtained from the highway department's continuous count program. Only those stations classified as rural were selected for use in the study. This yielded a total of twenty-three stations throughout the state for the four highway categories.

The elasticities and the appropriate background factors are derived from a linear equation that relates AADT to a variety of the factors in Table 1. It can be shown mathematically that, given an equation of the form:

$$Y_i = a + \sum_{j=1}^{n} a_j X_{ij} \quad (3)$$

where

Y_i = value of dependent variable at ith observation;
$i = 1, \ldots, n_1$,
X_{ij} = value of jth independent variable at ith observation;
$j = 1, \ldots, n$,
a = constant term,
a_j = regression coefficient for jth independent variable,
n_1 = observation number,
n = number of independent variable.

Elasticity measures can be estimated by:

$$e_j = a_j\left[\frac{\bar{X}_{ij}}{\bar{Y}_i}\right] \quad (4)$$

where,

e_j = elasticity of AADT with respect to independent variable x_j,
$\bar{X}_{ij}$ = overall mean of the jth independent variable,
$\bar{Y}_i$ = overall mean value of dependent variable.

Thus, using multiple linear regression, the background factors that best estimate AADT and their respective elasticities can be derived.

PRELIMINARY STATISTICAL ANALYSIS

Although multiple linear regression forms the basis for the elasticity model, several other statistical procedures were applied to develop and verify the intermediate model forms. For the sake of brevity and to illustrate the process used, these techniques are described in the context of primarily one highway category: rural major collectors. The figures and tables are, for the most part, the rural major collector portions of exhibits contained elsewhere (*9*). But the figures and tables for rural principal arterials are also given when doing so helps to clarify the underlying techniques.

Preliminary statistical analyses were conducted to identify any possible relationship between dependent (Y) and independent variables (X's), to check the validity of standard regression assumptions, and to justify the combination of stations under the various highway categories.

Homogeneity of Variance

The homogeneity of variance is a property necessary to permit legitimate linear regression analysis. For the variable AADT, this property was checked using the Cochran and Bartlett-Box tests (*10*) by treating the Y's for each station as a group, since there were an equal number of observations in each station or group. For a highway category with an unequal number of observations at the constituent stations, the Burr-Foster Q-test (*10*) is used to check the homogeneity of variance of Y's.

The test results (Table 2) show that the data for rural principal arterials satisfy both the Cochran and Bartlett-Box tests for the homogeneity of variance condition, having equal observations in each station. The Burr-Foster critical q-value (*10*) shows β-levels for rural major collectors between 0.01 and 0.001. Based on the regression analysis, a linear relationship between Y and the X's is feasible. There is no apparent practical or theoretical reason to transform Y. The distribution of Y's at some stations is sparse. Considering all these factors, homogeneity of variance for rural major collectors was accepted at a β-level of 0.001. In fact, homogeneity of variance was satisfied at reasonable β-levels for all highway categories.

Normality

The Shapiro-Wilk test for normality (*11*) was performed on each station separately and after combining stations in a highway category. The results for rural major collectors are shown in Table 3. In this table, small values of W associated with smaller β-levels are significant; that is, they lead to rejection of the normality assumption. This rejection occurred for the aggregate "all stations" case in every category but rural interstates.

Thus the normality assumption on Y does not support the combination of stations for the other highway categories. As will be seen later, however, other statistical tests justify further examination of each proposed model. When the Y's for individual stations are tested for normality, the assumption is satisfied in every case.

Scattergram

The scatterplots of the dependent variable (AADT) against the independent variables show a general linear trend. Two representative plots are shown in Figures 1 and 2. The clusters found in Figure 1 reflect the fact that the data come from thirteen years at each of three count stations, but linearity is still apparent.

CONCLUSIONS FROM PRELIMINARY ANALYSIS

The homogeneity of variance tests, considering each station as a group, shows equal variances between the stations for each category of highway. The normality hypothesis is accepted for each station separately and for rural interstates as a group. The normality test indicates that analysis for each station separately will yield a better model than that for the combination of stations within a highway category. In the scatterplots for the stations—both sepa-

TABLE 2 RESULTS OF THE TESTS FOR HOMOGENEITY OF VARIANCE

A. Bartlett-Box and Cochran Test (Equal Sample Size)					Homogenity of Variance
Highway Category (No. of station or group)	Cochran c [β-level]	Bartlett-Box f [β-level]	Remarks on β-level: Cochran	Remarks on β-level: Bartlett-Box	
2. Rural Principal Arterial (3)	0.5686 [0.063]	1.949 [0.143]	$\beta > 0.05$	$\beta >> 0.01$	Checked

B. Burr-Foster Q-Test (Unequal sample size)					Homogenity of Variance
Highway Category (No. of station or group)	Calculated q	Critical q: $\beta = 0.01$	Critical q: $\beta = 0.001$	Remarks on β-level	
4. Rural Major Collector (3)	0.4938	0.4827	0.5543	β = .01 - .001	Checked

TABLE 3 RESULTS OF THE TEST FOR NORMALITY

Highway Category	Stations(s)	No. of Cases	Shapiro-Wilk W	B-level	Normality
4. Rural Major Collector	(i) All stations	37	0.8051	<0.01	Unchecked
	(ii) 47A	11	0.9167	0.10 - 0.50	Checked
	(iii) 59A	13	0.9143	0.10 - 0.50	Checked
	(iv) 5420A	13	0.8899	0.10 - 0.50	Checked

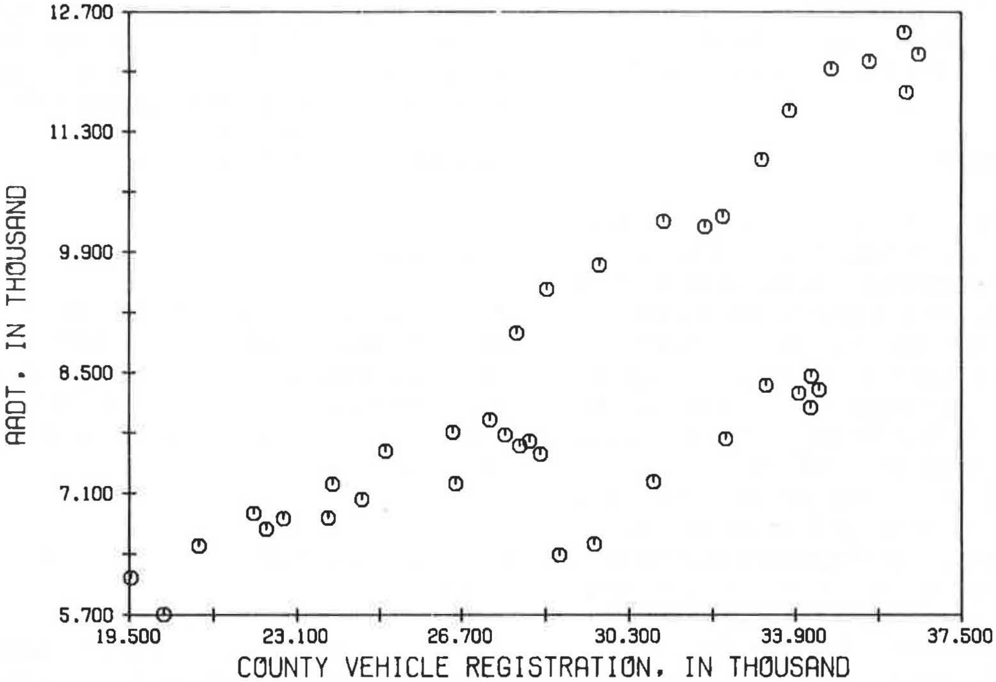


FIGURE 1 AADT versus county vehicle registrations (rural principal arterials).

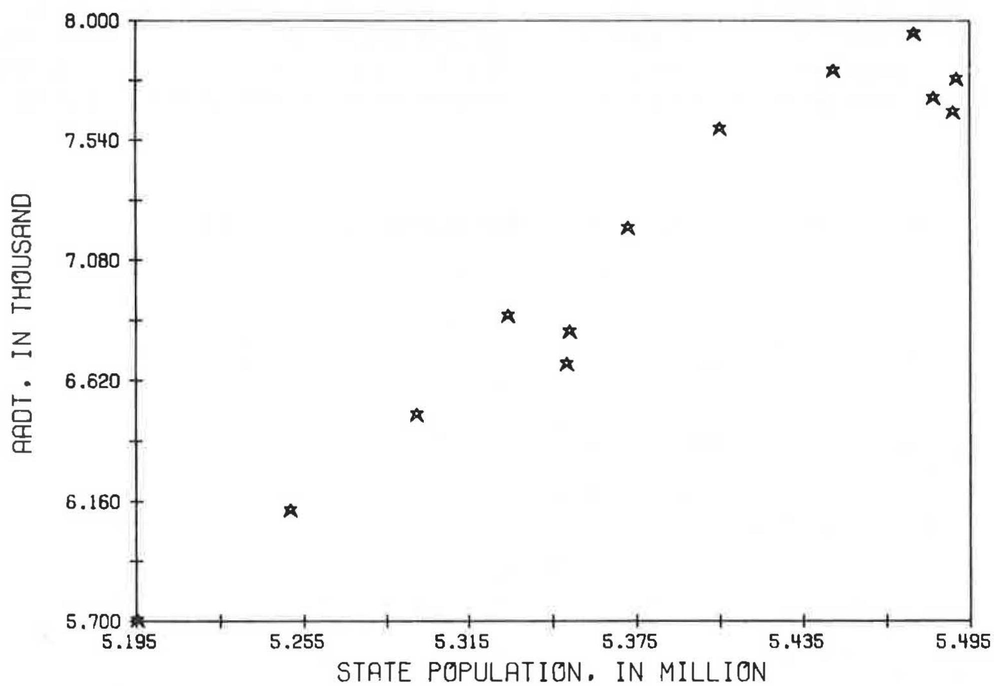


FIGURE 2 AADT versus state population (Station 68A).

rately and together—no recognizable pattern other than linear is noticeable.

Two types of analyses are found appropriate and promising. Aggregate analysis, combining all the stations within a category of highway, is employed to develop an aggregate model for each highway category. Disaggregate analysis of each station separately is also performed, and the resulting models will be location-specific.

AGGREGATE ANALYSIS

In aggregate analysis, a model is sought for each of the four categories of highway, with the count stations pooled by category. The analysis for rural major collectors (and, in some cases, rural principal arterials) is presented here to illustrate the process, but the analysis for other highway categories is similar.

Correlation Matrix

The statistical analysis begins with the study of the correlation matrix for the various factors considered. The correlation matrix for rural principal arterials (Table 4) shows that variables X_4 and X_8 are not highly correlated (r = 0.251). The correlation matrix for rural major collectors (Table 4) shows that X_1, X_4, and X_5 have similar degrees of correlation with Y ($0.731 \leq r \leq 0.915$). But the variables X_1, X_4, and X_5 are highly intercorrelated ($r \geq 0.901$). The existence of this multicollinearity does not invalidate a regression analysis; and neither is the absence of multicollinearity a validation of a particular regression model (*12*). Nevertheless, the use of two or more of these highly correlated variables in the same model should be avoided whenever possible. In case of rural major collectors, county employment (X_6) and AADT (Y) are negatively correlated, which is not the expected relationship; so the selection of variable X_6 will not be considered unless supported by other analyses.

Stepwise Regression

Stepwise regression is the most widely used automatic search procedure. It selects one variable at a time for entry into (or removal from) the model, until a desired subset of variables is selected (Table 5). In designing this regression procedure, four parameters—number of steps, F-value to enter (FIN), tolerance, and F-value to remove

TABLE 4 CORRELATION MATRIX (AGGREGATE ANALYSIS)

(2) Rural Principal Arterial

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	.786												
X2	.275	.519											
X3	.398	.639	.827										
X4	.804	.878	.224	.259									
X5	.881	.938	.364	.429	.974								
X6	.633	.901	.543	.692	.667	.764							
X7	.405	.647	.779	.975	.247	.411	.717						
X8	.402	.642	.797	.974	.251	.415	.721	.993					
X9	.401	.640	.830	.9999	.260	.429	.702	.978	.980				
X10	.354	.547	.626	.763	.191	.318	.697	.857	.847	.781			
X11	.373	.602	.865	.974	.260	.427	.654	.905	.912	.972	.676		
X12	.413	.641	.773	.972	.252	.416	.735	.987	.987	.979	.872	.923	
X13	.412	.643	.754	.974	.251	.414	.727	.993	.990	.979	.857	.915	.994

(4) Rural Major Collector

	Y	X1	X2	X3	X4
X1	.766				
X2	.178	.593			
X3	.164	.687	.801		
X4	.915	.901	.354	.341	
X5	.731	.654	.587	.618	.921
X6	-.453	.212	.452	.568	-.163

XI represents X_I, where i = 1 to 13

TABLE 5 STEPWISE REGRESSION SUMMARY (AGGREGATE ANALYSIS)

Highway Category	Case (*) and Parameter	Step	Variable subscript: Entered	Variable subscript: Removed	F value	Significance level	Last step b-coeff.	R^2	Overall F (**)
RURAL	Case A: Default Parameters	1	4		180.062	0.0	.5988	.837	18.062
		2	6		47.491	0.0	-.0492	.932	233.367
		3	1		2.818	.103	-.0589	.937	164.834
		4	5		1.820	.187	-1.0940	.941	127.151
		5	3		3.125	.087	152.0760	.946	109.102
		6	2		.302	.587	-8.8768	.947	88.921
		Constant term					-303613.9		
MAJOR	Case B: Alpha = .10 FIN = 2.95	1	4		180.062	0.0	.2564	.837	180.062
		2	6		47.491	0.0	-.1979	.932	233.367
	FOUT = 2.80	Constant term					-4908.47		
COLLECTOR	Case C: Alpha = .05 FIN = 4.20 FOUT = 4.10	(SAME AS CASE B)							

(*) A. Default Parameters:
(i) Max. No of Steps = 2 * No. of Independent Variables [For All Cases]
(ii) FIN = .01, FOUT = .005 [For Case A]
(iii) Tolerance level = .001 [For Case A]
B. Tolerance level = 0.1 [For Case B & Case C]
(**) Overall Significance = 0.0

(FOUT)—play important roles in the selection of variables for the models. Case A in Table 5 used the default values provided by the SPSS package (*11*), while cases B and C were defined in terms of the parameter values indicated in the second column of Table 5. For rural major collectors, the case A stepwise regression with default parameters includes all variables, with an R^2 of 0.947. The case B and case C stepwise regressions enter the variables X_4 and X_6 with an R^2 of 0.932. The variable X_4 enters at step 1 with an R^2 of 0.837 in all cases. The inclusion of other variables increased R^2 by only a small amount. It should be mentioned that the order in which a variable enters the regression equation does not indicate its importance. The *b*-coefficients of X_1, X_2, X_5, and X_6 are negative. The negative coefficient of X_2 (gasoline price) is expected. The reason for the negative coefficients of X_1, X_5, and X_6, however, is its high correlation with other variables in the model (for example, $r_{1,4} = 0.901$, $r_{1,5} = 0.954$, $r_{4,5} = 0.921$ in Table 4). Since stepwise regression can be misleading, the results of the stepwise regression were compared with other selection criteria (see Table 8) before establishing the final regression equation.

C_P-Criterion in All Possible Regression

The C_P-criterion considers all possible regression models that can be developed from the pool of potential independent variables and identifies subsets of the independent variables that are "good." The C_P-statistic, R^2, and other values for a reasonable number of subsets of variables were calculated with the help of a program called DRRSQU (*13*). Some of those C_P- and R^2-values for rural major collectors are shown in Table 6. The C_P-criterion is concerned with the total mean squared error (MSE) of the *n* fitted values for each of the various subset regression models. When the C_P-values for all possible combinations of variables in regression models are plotted against *P* (the number of terms in the regression equation), those models with little bias will tend to fall near the line $C_P = P$ (*14*). Models with substantial bias will tend to fall considerably above this line. In using the C_P-criterion, the subsets of *X* variables for which (1) the C_P-value is small and (2) the C_P-value is near *P* are considered for the model. Sets of *X* variables with small C_P-values have a small total mean squared error, and when the C_P value is also near *P*, the bias of the regression model is small. Sometimes the regression model based on the subset of *X* variables with the smallest C_P-value may involve substantial bias. In that case, one may at times prefer a regression model on a somewhat larger subset of *X* variables for which the C_P-value is slightly larger, but that does not involve a substantial bias component. Thus, one should look for a regression model with a low C_P-value about equal to *P*. When the choice is not clear-cut, then it is a matter of personal judgment whether one prefers a biased equation or an equation with more parameters.

Draper and Smith (*14*) recommend the use of the C_P-statistic in conjunction with the stepwise method to choose the best equation. Some statisticians suggest that all possible regression models with a number of *X* variables that is similar to the number in the stepwise regression solution be fitted subsequently to investigate which subset of *X* variables might be best (*15*). The C_P-values for rural major collectors in Table 6 show that the variable set X_3, X_4, and X_5 at $P = 4$ is the best selection, with C_P of 2.40 and R^2 of 0.944. But the selection of X_4 and X_6 at $P = 3$, with $C_P = 7.26$ and $R^2 = 0.932$, is the result of stepwise regression in cases B and C (see Table 5). The variable sets {X_1, X_4, and X_6} and {X_3, X_4, and X_6} at $P = 4$, with C_P of 6.25 and 6.76, respectively, are good for further analysis. Note that X_4 has high correlation with X_5 ($r = 0.921$). Also, X_4 has negative correlation with X_6 ($r = -0.163$), which is not an expected result.

MSE_P or R_a^2 Criterion in All Possible Regression

The adjusted coefficient of multiple determination R_a^2 and mean squared error (MSE_P) are equivalent criteria. The users of the MSE_P criterion seek either the subset of X variables that minimizes MSE_P, or the subset(s) for which MSE_P is so close to the minimum that adding more variables is not worthwhile (*15*). The results of using the MSE_P criterion for rural major collectors are shown in Figure 3. The plot of MSE against P indicates that, according to this criterion, the variable set X_3, X_4, and X_5 at

TABLE 6 SELECTED C_P AND R-SQUARED IN ALL POSSIBLE REGRESSION (AGGREGATE ANALYSIS)

Highway Category	Subscripts of Variables in Equation	C_p Values, in same order	R^2 Values, in same order	P
Rural	4 1 5 6 3	58.7, 199.9, 299.8, 414.9, 515.5	.837, 587, .534, .205, .027	2
Major	4 6, 4 5, 5 6, 1 6, 2 4, 2 5	7.26, 13.46, 18.18, 31.48, 46.87, 178.30	.932, .921, .913, .889, .862, .629	3
Collector	3 4 5, 1 4 6, 3 4 6, 1 4 5, 2 4 6, 4 5 6	2.40, 6.25, 6.76, 8.43, 9.12, 9.19	.944, .937, .937, .934, .932, .932	4
	----------	--------	--------	---
	All Variables	7	.947	7

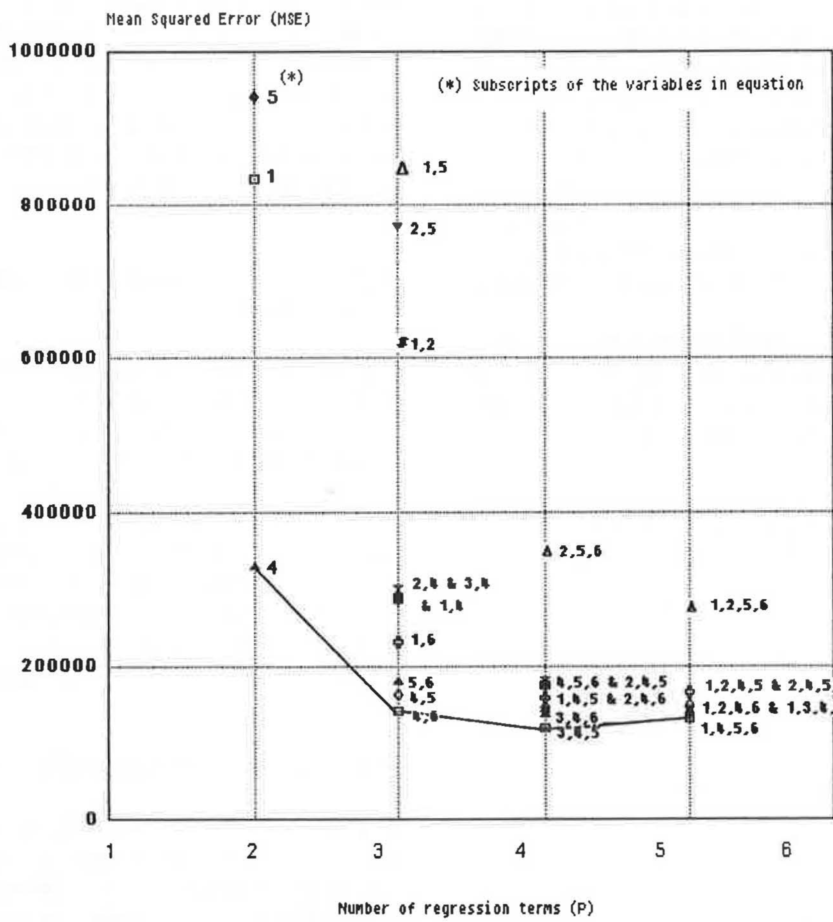


FIGURE 3 *MSE* versus *P* plot for rural major collectors.

$P = 4$ is the best choice. This choice will not be considered further, however, because it involves multicollinearity.

Preliminary Screening of Candidate Variables

The screening of variables to develop the forecasting models was not confined to statistical analysis. The questions listed in Table 7 and the goals in Table 8 provided a basis on which judgment could be applied to the selection process.

Considering all the points discussed above, the following subsets of variables were kept for the final selection process:

1. X_4
2. X_1
3. X_5
4. X_4, X_6
5. X_5, X_6
6. X_4, X_2
7. X_5, X_2

All these choices will provide an R^2 of at least 0.534.

Final Selection of Variables

In the selection of variables for the final regression model, the goals in Table 8 were used as a guide to find the best subset of variable(s) from the preliminary choices for each highway category. Goals 1 to 5 in Table 8 were taken into consideration in the preliminary screening. Final selection of candidate variables from preliminary choices is then made through careful examination of all criteria, subject to subsequent residual analysis and hypothesis testing concerning *b*-coefficients. If the final regression model passes these last two tests, it can be used to build the forecasting model.

In the final selection of variable(s) for a model's equation, the criterion of establishing a high R^2 should not be the only consideration. In addition to high R^2 value, residual plots should also be examined to determine if a "good" fit has truly been obtained. For example, in some situations, data may be quite variable and a large R^2 may not indicate a very good fit. In more controlled situations, however, a relatively small R^2 may indicate a rather good fit (*12*). The value of R^2 will increase if the number of predictor variables increases. Consequently, R^2 is always the maximum for the full set with all predictor variables. So maximizing R^2 should not be the sole selection criterion. However, one can subjectively choose a subset of predictor variables that gives a good value of R^2, so that using any additional predictor variables results in only a marginal improvement in R^2.

TABLE 7 SOME FUNDAMENTAL CRITERIA FOR VARIABLE SELECTION (*12,14*)

1.	Are the proposed variables fundamental to the problem?
2.	Availability of data (variables).
	(a) Are annual data available?
	(b) Are historical data available?
	(c) What is the most recent year of data?
	(d) Will data be available in future?
3.	Cost to obtain the data.
4.	How reliable are the data?

TABLE 8 GOALS OF THE ANALYSIS

1.	The final equations should explain more than 50 percent of the variation ($R^2 > 0.50$).
2.	The C_p value will be lowest and near to P.
3.	Mean squared error (MSE) will be minimum.
4.	The number of predictor variables should be adequate for each model (*).
5.	The selection will respond well to the questions of Table 7.
6.	All estimated coefficients in the final model should be statistically significant at an alpha level of 0.05 or 0.10.
7.	There should be no discernible patterns in the residuals.

(*) As a general rule, there should be about ten complete sets of observations for each potential variable to be included in the model; e.g., if it is believed that the final practical predictive model should have four X-variables plus a constant, then there should be at least forty sets of observations (n = 40) [14].

Regression on Preliminary Choices and Final Selection

Regression on the preliminary choices was done with the help of the SPSS package (*11*), a summary of which is shown in Table 9. This table shows inconsistencies ("-*b*6") in the *b*-coefficient of X_6 for the preliminary choices $\{X_4, X_6\}$ and $\{X_5, X_6\}$. The choice with X_4 has the largest R^2 and lowest C_P among all the choices with one variable. The choices with two variables without inconsistency in regression coefficients do not provide a significant increase in R^2 with respect to the one-variable choices (see Table 9). Considering these aspects, the variable X_4 is the final selection for further analysis for rural major collectors.

Graphic Residual Analysis on Final Selection

Residual plots were generated by the BMDP package (*16*). These plots were done to verify the validity of each regression model. The normal probability plot in Figure 4 falls reasonably close to a straight line, suggesting that the error terms are approximately normally distributed. The plots

TABLE 9 MULTIPLE LINEAR REGRESSION SUMMARY ON PRELIMINARY CHOICES (AGGREGATE ANALYSIS)

Highway Category	Variable Subscripts in Eqn.(*)	b-coefficient in same order	Inconsis-tencies in b's(**)	R^2	Overall F(***)
	4	.2715	---	.837	180.062
	1	.1991	---	.587	49.693
RURAL	5	.7122	---	.534	40.056
MAJOR	4, 6	.2564, -.1979	-b6	.932	233.367
COLLECTOR	5, 6	.8379, -.3988	-b6	.913	177.805
	4, 2	.2892, -34.5641	---	.862	106.030
	5, 2	.9292, -78.3612	---	.629	28.772

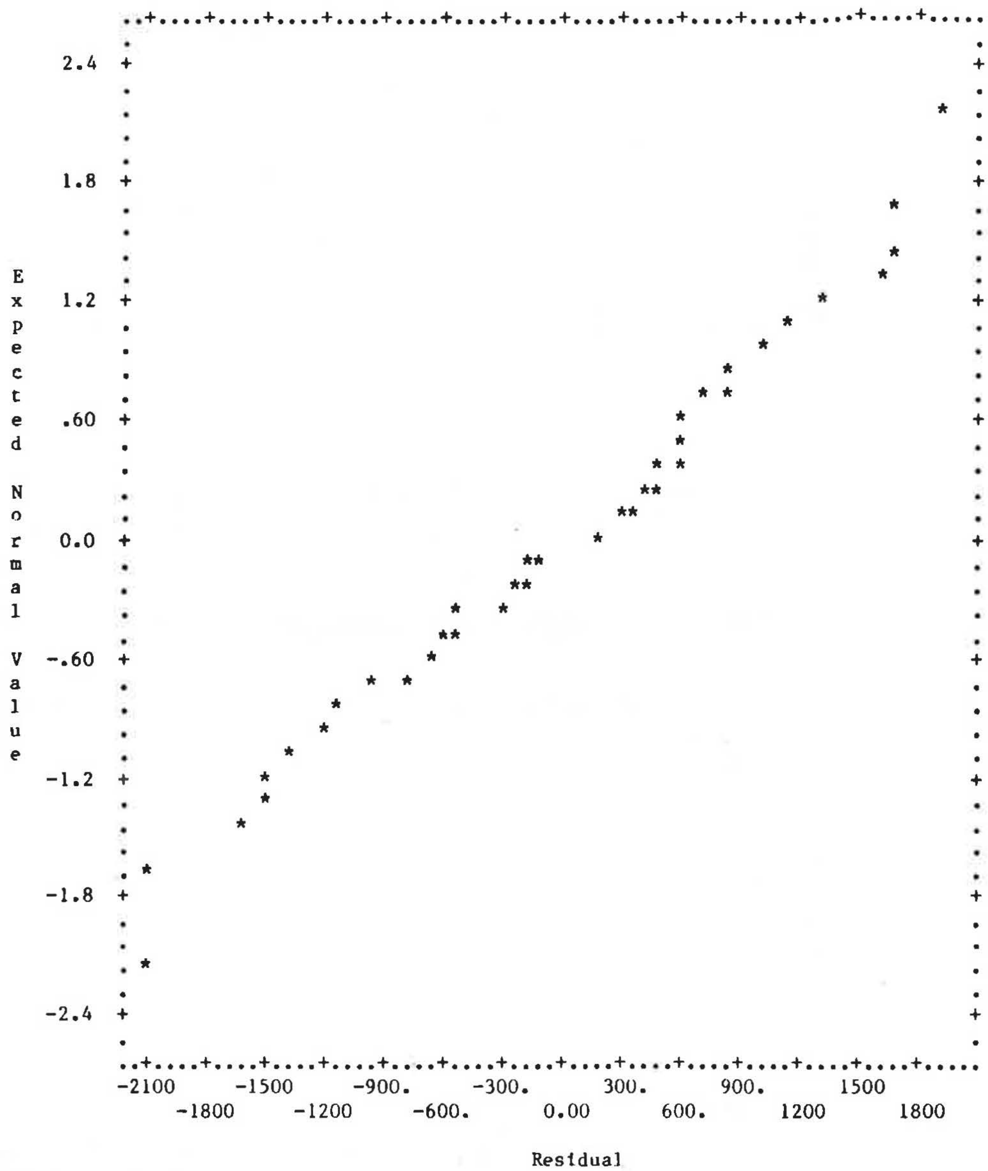


FIGURE 4 Normal probability plot of residuals (rural major collector).

of residuals against the fitted response variable and predictor variable, represented by Figures 5 and 6, indicate no grounds for suspecting the appropriateness of a linear regression function or the constancy of the error variance. The clustering of residuals in some cases (as in Figure 6) is the effect of combining count stations in the analysis. The residual plots against $\hat{Y}$ and the X's do not indicate the presence of any outliers. Residual plots were also generated against variables not included in the model, to check whether some key independent or predictor variables could provide important additional descriptive and predictive power to the model. One such variable is the year (X_3), which has not been included in any model. The plot of residuals against X_3 in Figure 7 does not indicate any correlation between the error terms over time, since the residuals are random around the zero line. Thus, it is confirmed that the most appropriate variable is included in the model and no additional variable will provide significant added explanatory power to the model.

Testing Hypotheses Concerning Regression Coefficients

The overall F-test for the regression relation explains whether the variables in the model have any statistical

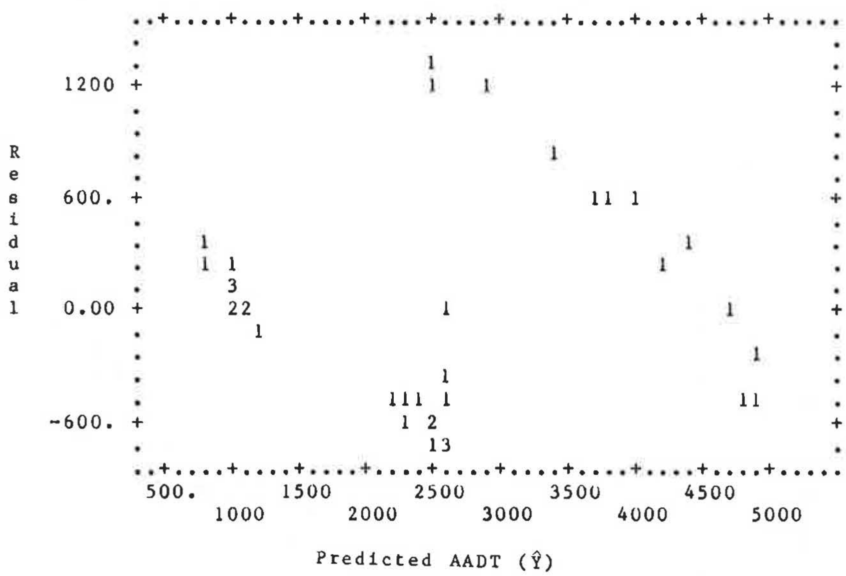


FIGURE 5 Residual plot against $\hat{Y}$ (rural major collector).

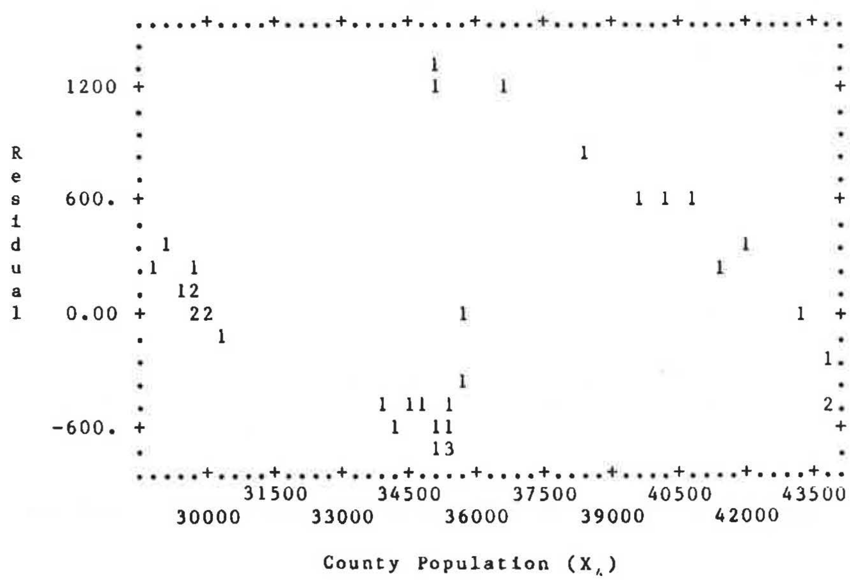


FIGURE 6 Residual plot against X_4 (rural major collector).

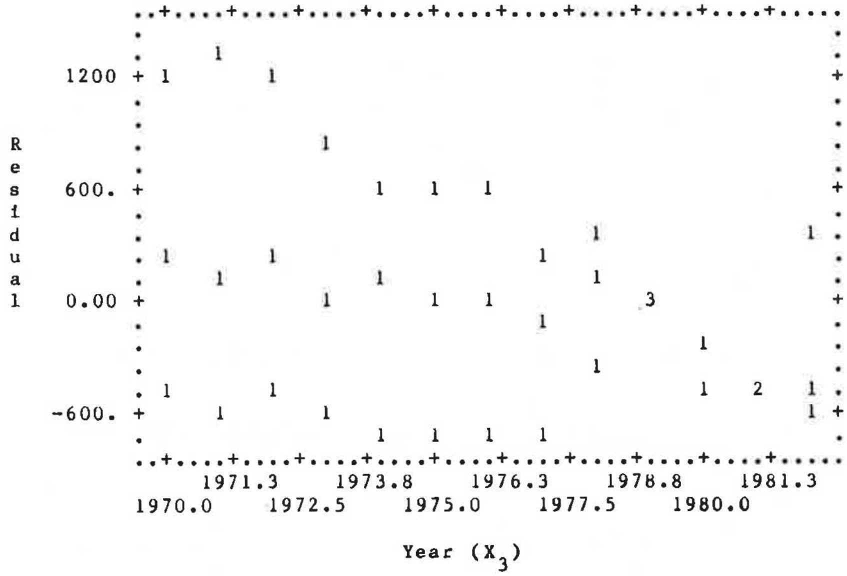


FIGURE 7 Residual plot against X_3 (rural major collector).

relation to the dependent variable. The hypotheses are

H_0: $\beta_1 = \beta_2 = \ldots = \beta_{P-1} = 0$
H_a: all β_k $(k = 1, \ldots, P - 1) \neq 0$.

The test statistic is given by

$$F^* = \frac{MSR}{MSE}$$

If $F^* \leq F(1 - \alpha, P - 1, n - P)$, then H_0 holds, indicating that the variables in the model do not have any statistical relation to the dependent variable. Table 10 shows the result for rural principal arterial and rural major collector at α-levels of 0.05 and 0.10. The test results conclude that hypothesis H_a (i.e., that the relationships between the variables in the models exist) cannot be rejected at an α-level as low as 0.05.

To test the significance of each variable

(H_0: $\beta_k = 0$; H_a: $\beta_k \neq 0$ for $1 \leq k \leq P - 1$)

and each subset with more than one variable

(H_0: $\beta_1 = \ldots = \beta_j = 0$; H_a: all $\beta_j \neq 0$ for $1 < j < P - 1$)

a general linear test (*15*) was employed. The applicable F-statistic is shown in Equation 5.

$$F^* = \frac{[\text{SSE}(R) - \text{SSE}(F)]/(\text{df}_R - \text{df}_F)}{\text{SSE}(F)/\text{df}_F} \qquad (5)$$

where,

F^* = the F statistic,
SSE (R) = error sum of squares for the reduced model,
SSE (F) = error sum of squares for the full model,
df_R = degrees of freedom of the reduced model, and
df_F = degrees of freedom of the full model.

The reduced model was obtained by dropping the element(s) to be tested under H_0 from the full model. Table 11 shows the summary of the partial F-test results for rural principal arterial and rural major collector obtained at α-levels of 0.05 and 0.10. The results indicate that each variable is significant in the model.

Model Development

The final regression equation for rural major collectors is presented in Table 12, along with the R^2 values, overall F values, t-statistics, and elasticities. Using the elasticities obtained from the regression analysis, the forecasting model was developed for each highway category. The models for all four highway categories are presented in Table 13. Each of the models is relatively simple, containing not more than three variables. The use of these models is also straightforward. The input values are the present year AADT and the present and future year value (the year for which the traffic forecast is needed) of the predictor variables. The data needed to predict rural traffic volumes with these models are readily available at the county and state levels. The performance of the models in Table 13 was tested using data for those Automatic Traffic Record (ATR) stations not used in building the models. The results of the trial forecasts for the rural major collector, shown in Table 14, indicate that the models perform satisfactorily. This table shows the present year used in these trial forecasts. Any closest year for which predictor variables are available could be the present year. In that respect, the

TABLE 10 OVERALL *F*-TESTS FOR AGGREGATE ANALYSIS

Highway Category	Variable Subscripts for Full Model	F*	df_R, df_E (*)	α	Is H_a true for	
					α = 0.05?	α = 0.10?
2. Rural Principal Arterial	4, 8	40.017	2, 36	<.001	Yes	Yes
4. Rural Major Collector	4	180.062	1, 35	<.001	Yes	Yes

(*) df_R = degrees of freedom for regression; df_E = degrees of freedom for error

TABLE 11 PARTIAL *F*-TESTS FOR AGGREGATE ANALYSIS

Highway Category	Variable Subscripts for		df_R, df_F (*)	F*	α	Is H_a true for	
	Full Model	Reduced Model				α = .05?	α = .10?
2. Rural Principal Arterial	4, 8	4	37, 36	4.963	.025-.050	Yes	Yes
		8	37, 36	61.223	2.001	Yes	Yes
4. Rural Major Collector	It has only one variable in Full Model.						

(*) df_R = degrees of freedom for SSE or Reduced Model and df_F = degrees of freedom for SSE for Full Model.

TABLE 12 FINAL REGRESSION EQUATION FROM AGGREGATE ANALYSIS

4. Rural Major Collector:	
	AADT = -7048.27 + 0.27151 County Population
R^2 = 0.887	t = 13.41872
F = 180.062	e = 3.77379

closest census year is the suggested present year. The forecasted errors are reasonably small in most of the cases and speak well for reliability of the models. The larger forecast errors in some cases are due to low values of AADT, fewer cases, and large variations in response and predictor variables employed in data tables among the stations and counties.

DISAGGREGATE ANALYSIS

In disaggregate analysis, each station has been analyzed separately and a separate forecasting model has been developed for each. The steps in model development and the criteria for variable selection are the same as those in the aggregate analysis.

Model Development

The final regression equations for rural major collector stations are presented in Table 15, along with R^2 values, overall *F* values, *t*-statistics, and elasticities. In case of station 7047A, the negative sign with county population is not surprising. In fact Rush County is losing population over the years, and the AADT values are also decreasing for the period of analysis. This fact establishes the negative sign with county population for station 7047A. Not all the goals of Table 8 have been set in all of the equations of Table 15. However, these equations are the best possible attempts at meeting the goals of Table 8. The station-specific disaggregate models are presented in Table 16. Each of the models is simple, with not more than two variables in any case. The use of these models is also straightforward. The data needed to predict rural traffic volumes with these models are readily available at the county, state, and national levels. The disaggregate model, however, requires a "similarity test" to determine if a segment whose future AADT is desired has characteristics that permit it to use one of the location-specific models (*9*). As a crude test of the model's performance, forecasts of 1983 and 1984 traffic levels were made for comparison against actual volumes for those years. The disaggregate forecasts (Table 17) were much more competitive than were the aggregate forecasts, but the time frame was too short to be conclusive.

SUMMARY AND CONCLUSIONS

The objective of the study described in this paper was to develop a method of forecasting traffic volumes on rural

TABLE 13 AGGREGATE TRAFFIC FORECASTING MODELS

1. Rural Interstate:

$$AADT_f = AADT_p\,[1 + 4.83314\,(\Delta \text{ State Population})]$$

$R^2 = 0.658$

2. Rural Principal Arterial:

$$AADT_f = AADT_p\,[1 + 1.47809\,(\Delta \text{ County Population} + 2.79623\,(\Delta \text{ State Population})]$$

$R^2 = 0.690$

3. Rural Minor Arterial:

$$AADT_f = AADT_p\,[1 + 0.83377\,(\Delta \text{ County Household})]$$

$R^2 = 0.727$

4. Rural Major Collector:

$$AADT_f = AADT_p\,[1 + 3.77379\,(\Delta \text{ County Population})]$$

$R^2 = 0.837$

(i) R^2 value is for final regression equation.

(ii) Δ represents change in predictor variable with respect to its present value in fraction. For example, $\Delta X = \frac{X_f - X_p}{X_p}$, where X_p and X_f denote present and future values of X.

TABLE 14 PERFORMANCE OF AGGREGATE TRAFFIC FORECASTING MODELS (RURAL MAJOR COLLECTOR)

Traffic Count Station	Present Year	Year	Actual AADT ($AADT_a$)	Forecasted AADT ($AADT_f$)	$AADT_f$ - $AADT_a$	Forecast Error in percent (*)
7047A	1790	1971	257	266	9	3.50
		1972	227	296	69	30.40
		1973	233	292	59	25.32
		1974	226	261	55	24.34
		1975	225	271	46	20.44
		1976	231	271	40	17.32
		1977	204	271	67	32.84
		1978	224	271	47	20.98
		1979	294	241	-53	-18.03
		1980	299	236	-63	-21.07
		1981	288	226	-62	-21.53
		1982	272	205	-67	-24.63
30063A	1980	1979	877	752	-125	-14.25
		1981	824	800	-24	-2.91
		1982	767	793	26	3.38
54382A	1980	1979	1159	1062	-97	-8.37
		1981	973	984	11	1.13
		1982	878	1029	151	17.20
200X	1980	1973	8805	6547	-2258	-25.64
		1974	8834	6823	-2011	-22.76
		1975	9002	7155	-1847	-20.52
		1976	9033	7431	-1602	-17.73
		1977	9079	8038	-1041	-11.47
		1978	9457	8535	-922	-9.75
		1979	9636	8977	-659	-6.84
		1981	9226	9197	-29	-0.31
		1982	9004	9308	304	3.38

(*) "+" sign indicates overprediction and "-" sign indicates underprediction.

$$\text{Forecast Error in percent} = \frac{AADT_f - AAADT_a}{AADT_a} \times 100$$

TABLE 15 FINAL REGRESSION EQUATIONS FROM DISAGGREGATE ANALYSIS

Rural Major Collector			
Station 59A:	AADT = 2772.53 + 0.0481 County Vehicle Registrations		
	R^2 = 0.7222	t = 5.344	
	F = 28.563	e = 0.36063	
Station 200X:	AADT = 6557.79 + 0.0503 County Vehicle Registrations		
	R^2 = 0.635	t = 3.734	
	F = 13.940	e = 0.28407	
Station 5420A:	AADT = 784.34 + 0.1163 County Employment		
	R^2 = 0.681	t = 4.842	
	F = 23.447	e = .59744	
Station 7047A:	AADT = 1122.06 - 0.0433 County Population		
	R^2 = 0.521	t = -3.462	
	F = 11.988	e = -3.48274	

TABLE 16 DISAGGREGATE TRAFFIC FORECASTING MODELS

Rural Interstate {0.706 ≤ R^2 ≤ 0.925}		
Station 172A:	$AADT_f$ =	$AADT_p$ [1 + 5.24231 (Δ State Population)]
Station 3070A:	$AADT_f$ =	$AADT_P$ [1 - 0.44503 (Δ US Gas Price) + 7.74428 (Δ State Pop.)]
Station 5474A:	$AADT_f$ =	$AADT_P$ [1 + 6.18172 (Δ County Population)]
Rural Principal Arterial {0.607 ≤ R^2 ≤ 0.976}		
Station 68A:	$AADT_f$ =	$AADT_P$ [1 + 0.86979 (Δ State Vehicle Registrations)]
Station 134A:	$AADT_f$ =	$AADT_P$ [1 - 0.43949 (Δ US Gas Price) + 0.83878 (Δ State Vehicle Registrations)]
Station 173A:	$AADT_f$ =	$AADT_P$ [1 + 1.47643 (Δ County Vehicle Registrations) - 0.21371 (Δ US Gas Price)]
Station 254B:	$AADT_f$ =	$AADT_P$ [1 + 0.60300 (Δ County Vehicle Registrations)]
Rural Minor Arterial {0.525 ≤ R^2 ≤ 0.867}		
Station 25A:	$AADT_f$ =	$AADT_P$ [1 + 0.90147 (Δ County Vehicle Registrations) - 0.29365 (Δ US Gas Price)]
Station 279A:	$AADT_f$ =	$AADT_P$ [1 - 0.26635 (Δ US Gas Price) + 0.24526 (Δ County Employment)]
Station 301A:	$AADT_f$ =	$AADT_P$ [1 + 0.66731 (Δ County Vehicle Registrations) - 0.26576 (Δ US Gas Price)]
Station 319A:	$AADT_f$ =	$AADT_P$ [1 + 61456 (Δ County Vehicle Registrations)]
Rural Minor Arterial {0.525 ≤ R^2 ≤ 0.867}		
Station 42A:	$AADT_f$ =	$AADT_P$ [1 + 0.49887 (Δ County Vehicle Registrations)]
Station 100X:	$AADT_f$ =	$AADT_P$ [1 + 0.64875 (Δ County Vehicle Registrations)]
Station 256A:	$AADT_f$ =	$AADT_P$ [1 + 0.33059 (Δ County Vehicle Registrations)]
Station 262A:	$AADT_f$ =	$AADT_P$ [1 + 0.28236 (Δ County Vehicle Registrations) - 0.23256 (Δ US Gas Price)]
Rural Major Collector {0.521 ≤ R^2 ≤ 0.722}		
Station 59A:	$AADT_f$ =	$AADT_P$ [1 + 0.36063 (Δ County Vehicle Registrations)]
Station 200X:	$AADT_f$ =	$AADT_P$ [1 + 0.28407 (Δ County Vehicle Registrations)]
Station 5420A:	$AADT_f$ =	$AADT_P$ [1 + 0.59744 (Δ County Employment)]
Station 7047A:	$AADT_f$ =	$AADT_P$ [1 - 3.48274 (Δ County Population)]

R^2 value is for final regression equation; subscripts P and f represent present and future years. Δ represents change (decimal) in predictor variable with respect to its present year value. For example, $\Delta X = \frac{X_f - X_P}{X_P}$, where X_P and X_f denote present and future year values of X.

TABLE 17 PERFORMANCE OF DISAGGREGATE TRAFFIC FORECASTING MODELS (RURAL MAJOR COLLECTOR)

Traffic Count Station	Present Year	Year	Actual AADT ($AADT_a$)	Forecasted AADT ($AADT_f$)	$AADT_f$ - $AADT_a$	Forecast Error in percent (*)
59A	1980	1983	4551	4701	150	3.30
		1984	4769	4718	-51	-1.07
200X	1980	1983	9297	9471	174	1.87
		1984	9950	9568	-382	-3.84
5420A	1980	1983	1979	2117	138	6.97
7047A	1980	1983	281	262	-19	-6.76
		1984	273	262	-11	-4.03

(*) "+" sign indicates overprediction and "-" sign indicates underprediction.

$$\text{Forecast Error in percent} = \frac{AADT_f - AAADT_a}{AADT_a} \times 100.$$

state highways that would be reliable, understandable, easy to use, and well documented. While the reliability of a forecasting model is always difficult to prove in advance, the models in Tables 13 and 16 are each the product of a series of careful statistical tests. The models are understandable: the level and nature of the variables that appear in their final forms have a clear functional relationship. They are not in the model only to provide a better curve fit. They are easy to use: the predictor variables are relatively easy to procure, and the models themselves require only a hand calculator to implement. The two different kinds of models—aggregate and disaggregate—that are offered in the full report (*9*) to forecast traffic volumes at rural locations in Indiana's state highway network can be implemented using any spreadsheet software. And the models are well documented in a report (*9*).

There were (and are) some inevitable problems, however, in developing traffic growth factor values for four functional classes of highways (interstate, principal arterial, minor arterial, and major collector). Although a simple forecasting model was sought, the statistical analyses to develop it were extensive, sometimes complex, and often subject to the analysts' judgment. Most important, there is an even greater need than usual for more data from a larger number of count stations to fill the many gaps in the database used in this study. Fortunately, the Indiana state highway department is installing new, permanent count stations, and a historical record of traffic volumes is now being developed for a wider range of locations.

ACKNOWLEDGMENTS

The work described in this paper was supported by the Federal Highway Administration (FHWA) through the Indiana Department of Highways (IDOH). The authors thank John Nagle of IDOH, the FHWA's Gary DalPorto, and Kumares C. Sinha of Purdue University for their guidance as members of the project's advisory committee. Special thanks are also due Professor Virgil Anderson, who provided advice on the use (and potential abuse) of methods of statistical analysis.

REFERENCES

1. F. T. Morf and V. F. Houska. Traffic Growth Pattern on Rural Areas. *Highway Research Board Bulletin 194*, January 1958, pp. 33–41.
2. A. J. Neveu. Quick Response Procedures to Forecast Rural Traffic. In *Transportation Research Record 944*, TRB, National Research Council, Washington, D.C., June 1982.
3. Minnesota Department of Transportation (MnDOT). *Procedures Manual for Forecasting Traffic on the Rural Trunk Highway System.* Traffic Forecasting Unit, Program Management Division, MnDOT, April 1985.
4. D. Albright. *Abstract on—Trend-Line: Forecasting Heavy Commercial Traffic as a Percent of Average Daily Traffic.* New Mexico State Highway Department, 1985.
5. J. S. Armstrong. The Ombudsman: Research on Forecasting: A Quarter-Century Review (1960-1984). *Interfaces*, Vol. 16, No. 1, January–February, 1986, pp. 89–109.
6. J. S. Armstrong. Forecasting by Extrapolation: Conclusion From 25 Years Research. *Interfaces*, Vol. 14, No. 6, November-December 1984, pp. 52–56.
7. AASHTO. *A Policy on Geometric Design of Highways and Streets*, 1984.
8. M. D. Meyer and E. J. Miller. *Urban Transportation Planning: A Decision-Oriented Approach.* McGraw-Hill Book Co., New York, 1984.
9. J. D. Fricker and S. K. Saha. *Traffic Volume Forecasting Methods for Rural State Highways.* Final Report, Joint Highway Research Project, Purdue University, May 1987.
10. V. L. Anderson and R. A. McLean. *Design of Experiments: A Realistic Approach*; Statistics: Text Books and Monographs, Vol. 5. Marcel Dekker, Inc., New York, 1974.

11. N. H. Nie, C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent. *SPSS—Statistical Package for the Social Sciences*, 2nd ed. McGraw-Hill Book Co., New York.
12. R. J. Freund and P. D. Minton. *Regression Methods—A Tool for Data Analysis.* Marcel Dekker, Inc., New York, 1979.
13. Purdue University Computer Center (PUCC). *DRRSQU (Program for C_p-Statistic—All Possible Regression): Statistical Program.* Statistical Library, PUCC, Purdue University, West Lafayette, Indiana.
14. N. R. Draper and H. Smith. *Applied Regression Analysis.* John Wiley and Sons, Inc., New York, 1981.
15. J. Neter and W. Wasserman. *Applied Linear Statistical Models.* Richard D. Irwin, Inc., Homewood, Ill., 1974.
16. University of California, Los Angeles (UCLA). *Biomedical Computer Programs (BMDP): P-series.* Health Science Computing Facility, UCLA, 1979.

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

Estimation of Trip Matrices: Shortcomings and Possibilities for Improvement

RUDI HAMERSLAG AND BEN H. IMMERS

Since the early 1970s several techniques have been developed for estimating origin-destination (O-D) matrices using all kinds of data. This paper presents a survey of four different O-D matrix estimation techniques. It is shown that some techniques can be applied only in certain restricted conditions and, even worse, may lead to considerable loss of information. To alleviate this problem, it is suggested that elastic, instead of fixed, constraints be used. It is also possible to overcome this problem by using an explicit stochastic estimation technique (binary calibration). Because of the absence of time and space-dependent coefficients, this technique can also be used for making forecasts.

Origin-destination (O-D) matrices contain trips between a number of origins and destinations per time period, means of transport, and the like. O-D matrices—an essential part of the basic information on transport demand—are used for several purposes—for example:

- transport planning and design: the calculation of traffic flows and the prediction of bottlenecks in road networks;
- evaluation of alternatives: sensitivity analyses;
- simulation of traffic flows in networks, design of traffic control devices, and specification of signal settings for controlled intersections.

Depending on the objective of the study, the information stored in the O-D matrix can be specified with respect to:

- size of study area: entries and exits of an intersection or zones in a regional transport study;
- means of transport: per mode or combination of modes;
- time period (time of the day): 24 hours, peak hour, 15-minute intervals, etc.;
- date: past, present, or future situation;
- purpose: home-based work, recreation, shopping, etc.

If the O-D matrix contains information on the present situation, it is called "base year matrix." Such a matrix can be estimated using all kinds of (available) traffic and transport data, such as:

- Complete data: all trips are observed. (Because of organizational and financial constraints, this procedure will never be applied.)
- Incomplete, direct, and indirect data. Incomplete data comes from taking a sample. Direct data results from observing O-D trips. Indirect data is a product of calculating O-D trips using other information sources, such as traffic counts, route assignment, etc. Examples of these kinds of data follow:

  –Household surveys: trips of only those people living in the survey area are observed.
  –Road questionnaires: Trips of only those people passing through are observed. Double counts as well as long-distance trips may bias the estimates.
  –Ticket sales: season tickets and special tickets may lack O-D information.
  –Cordon/screenline/traffic counts: only traffic volumes on a road section are observed.
  –Route choice information.
  –Old O-D matrices.

Problems arise, however, in expanding these data. For example, because all data sources are incomplete, a great number of trips (and O-D pairs) are not observed. Also, the data sources show considerable overlap, which may result in apparently contradictory O-D information (caused by the stochastic properties of the data).

To give a complete and consistent picture of the trip (O-D) pattern, O-D matrix estimation techniques are used. Although the underlying assumptions as well as the applied optimization techniques may differ, the objective of all base-year matrix estimation procedures is to obtain an optimum fit between the estimate and the available set of surveyed data.

Since the early 1970s several techniques have been developed using different approaches with respect to:

- Formulation of the objective function in combination with the use of an underlying distribution model—for example, maximum likelihood (Hamerslag and Huisman [1]), minimum of variance (Smit [2,3]), constrained least

R. Hamerslag, Departments of Civil Engineering and of Mathematics and Information Technology, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands. B. H. Immers, Department of Civil Engineering, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands.

squares (Hendrickson and McNeil [4]), information minimization (Van Zuylen [5], Van Zuylen and Willumsen [6]), entropy maximization (Willumsen [7]), and others (e.g., Nickesen et al. [8]).

• Use of different data sources (direct and indirect data) and a priori information as well as the incorporation of constraints in the calibration process of the model (for example, see Van Zuylen [9,10], Van Zuylen and Branston [11], Willumsen [12], Cremer and Keller [13], Bell [14,15]).

• Incorporation of a stochastic component (as data from observations are stochastic) and assumptions about the distribution pattern of the observed data (for example, see Hamerslag and Huisman [1], Smit [2], Hamerslag et al. [16], Hendrickson & McNeil [4], Willumsen [17]).

• Influence of congestion on route choice (for example, see Nguyen [18], Willumsen [17], Gur [19], and Turnquist and Gur [20]).

All kinds of O-D matrix estimation techniques are being used in the transportation planning process. The objective of this paper is to analyze the qualities and possibilities for applying some of the matrix estimation techniques that are most frequently used.

In this paper four different techniques are discussed:

1. The weighted Poisson estimator in combination with partial matrix estimation techniques,
2. Entropy maximizing and information minimizing techniques,
3. Information minimizing technique with elastic constraints,
4. Explicit stochastic estimation technique (the binary calibration model).

This paper does not deal with techniques that allow for alternative route choice assumptions (e.g., LINKOD; see Gur [19]). Description of the model in terms of all-or-nothing assignment implies that comparisons between observed and calculated link flows should preferably be made on screenline level. Incorporation of alternative route choice assumptions and, consequently, comparison of observed and calculated link flows per link may induce errors introduced in trip generation and the trip distribution model that are being corrected for in the assignment model.

A second reason for exclusion of the LINKOD model is that application of the model is possible only if volume counts on all links are available (which is hard to realize in practical situations). Alleviation of this severe constraint by Gur (19) results in a model based on partial volume counts. However, this model also requires quite some additional information—that is, total vehicle-hours of travel and capacities and free travel times for all links.

The paper is organized as follows. First the weighted Poisson estimator is presented. The weighted Poisson estimator can be used for estimating deterrence functions as well as O-D matrices. The estimator can be used even if only part of all O-D pairs has been observed.

Next both entropy and information optimization techniques are discussed. Based on a practical application, the entropy and information optimizing techniques are shown to generate very poor results and to induce a considerable loss of information. Therefore a new technique is introduced (information optimizing model with elastic constraints) that produces considerably better results. This is the next topic described.

Then a description is presented of an explicit stochastic estimation technique. This technique is based on the likelihood estimation theory. The coefficients of a transportation forecasting model are calibrated in such a way that an optimum fit is obtained to all available data. Simultaneously a base-year matrix and the coefficients of a forecasting model are estimated.

Finally, some major conclusions are drawn about the possible applications of all the techniques discussed. A summary of the particular qualities of the techniques appears in the paper's closing table.

WEIGHTED POISSON ESTIMATOR

The weighted Poisson estimator is based on two assumptions. First, it is assumed that all interzonal volumes are independent and Poisson distributed with some expected value. Second, the expected number of trips is multiplicative—the product of some independent variables. Maximizing the likelihood yields the estimation equations.

The following well-known model (see, e.g., Wilson [21]) is an example of a forecasting model with a multiplicative form:

$$\hat{T}_{ijm} = Ca_iO_ib_jD_jF_m(c_{ijm}) \quad \forall i, j, m \tag{1}$$

with

$$\sum_{jm}\sum \hat{T}_{ijm} = G_i \quad \forall i \tag{2}$$

and

$$\sum_{im}\sum \hat{T}_{ijm} = A_j \quad \forall j \tag{3}$$

where:

T_{ijm} = the estimated number of trips from i to j with mode m
C = constant term
a_i, b_j = balancing factors
O_i, D_j = polarities (generation and attraction power) of zone i resp. j
$F_m(c_{ijm})$ = deterrence function for mode m
c_{ijm} = generalized cost for O-D pair ij and mode m
G_i = the generation of trips in zone i (origins)
A_j = the attraction of trips in zone j (destinations).

Let us assume

$$a_iO_i = o_i \qquad b_jD_j = d_j \tag{4}$$

and

$$F_{mk} \stackrel{\text{def}}{=} F_m(c_{ijm}) \tag{5}$$

where c_{ijm} belongs to generalized cost class k.

Substitution of formulas 4 and 5 in formula 1 yields

$$\hat{T}_{ijm} = Co_id_jF_{mk} \quad \forall i, j, m \tag{6}$$

The probability of observing T_{ijm} trips can be given by the equation

$$\Pr[T_{ijm} \mid \hat{T}_{ijm}] = \exp(-\hat{T}_{ijm})(\hat{T}_{ijm})^{(T_{ijm})}/T_{ijm}! \tag{7}$$

where T_{ijm} = estimate of T_{ijm}.

The number of trips T_{ijm} is assumed to be independent for all combinations of i,j, and m. As a result the value of the log-likelihood function (L^*) becomes

$$L^* = \ln(L) = \sum_{ij}\sum \ln(\Pr[T_{ijm} \mid \hat{T}_{ijm}]) \tag{8}$$

The coefficients in formula 1 should be chosen in such a way that the log-likelihood has a maximum value. Substitution of formulas 1 and 7 in 8 yields the log-likelihood function:

$$L^* = \sum_{ij}\sum \{(-\hat{C}\hat{o}_i\hat{d}_j\hat{F}_{mk}) + T_{ijm}[\ln(\hat{C}) + \ln(\hat{o}_i) + \ln(\hat{d}_j) + \ln(\hat{F}_{mk})] - \ln(T_{ijm}!)\} \tag{9}$$

The maximum value of the log-likelihood is found by setting the first partial derivatives to zero:

$$\frac{\delta L^*}{\delta \hat{o}_i} = 0 \quad \forall i \tag{10}$$

$$\frac{\delta L^*}{\delta \hat{d}_j} = 0 \quad \forall j \tag{11}$$

$$\frac{\delta L^*}{\delta \hat{F}_{mk}} = 0 \quad \forall m, k \tag{12}$$

$$\frac{\delta L^*}{\delta \hat{C}} = 0 \tag{13}$$

The result is a set of nonlinear equations with which the coefficients can be determined using an iterative method, such as the Gauss-Seidel principle.

$$\hat{o}_i = \frac{\sum_{jm}\sum T_{ijm}}{\sum_{jm}\sum \hat{C}\hat{d}_j\hat{F}_{mk}} \quad \forall i \tag{14}$$

$$\hat{d}_j = \frac{\sum_{im}\sum T_{ijm}}{\sum_{im}\sum \hat{C}\hat{o}_i\hat{F}_{mk}} \quad \forall j \tag{15}$$

$$\hat{F}_{mk} = \frac{\sum_{ij,\ c_{ijm}\in k}\sum T_{ijm}}{\sum_{ij}\sum \hat{C}\hat{o}_i\hat{d}_j} \quad \forall m, k \tag{16}$$

and

$$\hat{C} = \frac{\sum_{ijm}\sum\sum T_{ijm}}{\sum_{ijm}\sum\sum \hat{o}_i\hat{d}_j\hat{F}_{mk}} \tag{17}$$

This set of equations can be solved only if there are no inconsistencies in the data. In general this condition will not be fulfilled if data from more than one home interview or from various cordon interviews are used.

The weighted Poisson estimator was applied for the first time in the early 1970s. At first the model was used for estimating deterrence functions (Evans and Kirby [*22*], Hamerslag [*23*]). Since then, the model has also been used for estimating O-D matrices, even if only a part of all O-D pairs has been observed. For that reason, the model is especially suited for making estimates using survey data from cordon interviews. An estimate of unobserved O-D pairs in the matrix can be obtained using partial matrix techniques (Neffendorf and Wootton [*24*]). This application of the model, however, is not always without problems (see, for example, Day and Hawkins [*25*], Kirby and Murchland [*26*]). Information from road traffic counts or public transport ridership cannot be used as it lacks O-D information. The weighted Poisson estimator is also used for the analysis of multidimensional matrices (Hamerslag et al. [*27*]).

ENTROPY MAXIMIZING AND INFORMATION MINIMIZING MODEL

According to information theory, the most likely trip matrix is a matrix that can arise in the greatest number of ways and that satisfies whatever constraints are placed on the system. The entropy maximizing (EM) and information minimizing (IM) models are equivalent; the only difference between them is that the latter makes use of some initial knowledge about the likely trip matrix (an a priori matrix). The IM model is a generalization of the EM model.

If only the total number of trips in the system is known, the entropy maximizing method will distribute these trips evenly over all cells of the matrix. Adding some constraints might result in a more realistic picture. Wilson (*28*), for example, subjected the entropy to the following constraints: the total number of origin and destination trips per zone and a general function of distance (e.g., generalized cost). The result is a conventional, double-constrained gravity model with an exponential deterrence function. It is also possible to incorporate some extra information into the model—for example, the observed link flows (link counts V_a per direction). It should be noted, however, that each link count (V_a) adds an extra constraint to the estimate that results in an extra Lagrange multiplier and, consequently, an extra coefficient in the model. Of course, it is also possible to make an estimate of a trip matrix without using information that is essential in the conventional gravity model (e.g., the average distance traveled).

The EM model for the estimation of trip matrices from traffic counts was introduced by Willumsen (*7*). Van Zuylen (*9*) derived the IM model. An attractive feature of Van Zuylen's model is the possibility of incorporating extra information (an a priori trip matrix) that might result in a more realistic estimate of the actual trip matrix. Later the possibility of incorporating extra a priori information is also introduced in the EM model. The better the a priori matrix, the better the estimate will be. An old O-D matrix might be well suited for use, but it is also possible to use an O-D matrix estimate from the weighted Poisson model (see preceding section) or a Wilson type of model as a first guess.

Formulation of the Information Minimizing Model

According to the information minimizing theory (Van Zuylen [*9*]), the most likely trip matrix T_{ij} satisfies the following equation:

$$L = \min_{T_{ij}} \sum\sum_{ij} \left[\hat{T}_{ij} \ln\left(\frac{\hat{T}_{ij}}{T_{ij}}\right)\right] \tag{18}$$

subject to

$$\sum\sum_{ij} (\hat{T}_{ij} d_{ij}^a) = R_a \quad \forall a \tag{19}$$

where

L = "distance" between matrices

$\hat{T}_{ij}$ = the number of trips from i to j (to be estimated a posteriori)

R_a = constraint a (e.g., traffic count, total number of arrivals per zone, trip length distribution, etc.)

d_{ij}^a = the fraction of trips ij subjected to constraint R_a (e.g., the number of trips ij that make use of link a)

T_{ij} = a priori matrix. This matrix is based on some initial knowledge (old O-D matrix or model estimate).

If T_{ij} = 1, Equation 18 is equivalent to the entropy maximizing formulation.

Minimizing the information of Equation 18 subject to the constraints of 19 gives an estimate for T_{ij}. The solution can be derived by minimizing the Lagrangian (if the Kuhn–Tucker conditions are fulfilled)

$$L = \min_{\lambda_a, \hat{T}_{ij}} \left\{ \sum\sum_{ij} \left[\hat{T}_{ij} \ln\left(\frac{\hat{T}_{ij}}{T_{ij}}\right)\right] - \sum_a \left(\lambda_a \left[\sum\sum_{ij} (\hat{T}_{ij} d_{ij}^a) - R_a \right] \right) \right\} \tag{20}$$

The solution is found by setting the partial derivatives to zero

$$\frac{\delta L}{\delta \hat{T}_{ij}} = 0 \quad \forall ij \tag{21}$$

$$\frac{\delta L}{\delta \lambda_a} = 0 \quad \forall a \tag{22}$$

Equation 21 yields

$$\frac{\delta L}{\delta \hat{T}_{ij}} = 1 + \ln(\hat{T}_{ij}) - \ln(T_{ij}) - \sum_a \left(\lambda_a \sum\sum_{ij} d_{ij}^a \right) = 0 \tag{23}$$

Suppose

$$\ln(X_o) = -1 \qquad \ln(X_a) = \lambda_a \sum\sum_{ij} d_{ij}^a$$

Substitution in Equation 23 yields

$$\hat{T}_{ij} = T_{ij} x_o \prod_a X_a \quad \forall i, j \tag{24}$$

where

$$\sum\sum_{ij} (\hat{T}_{ij} d_{ij}^a) = R_a \quad \forall a \tag{19}$$

Equation 24 shows that the a priori information (O-D matrix) is multiplied by a number of coefficients X_a. Each constraint (R_a) introduces an extra coefficient X_a into the model. An important difference between the IM and EM models is that in case of more observations (N) per O-D pair, there is a positive correlation between the bias of the EM estimate and the number of times the flow is counted; the IM estimate is free of bias (see Maher [*29*]).

The model described earlier has some specific characteristics:

- The amount of traffic (T_{ij}) is estimated using a model with coefficients that are time- and place-dependent. A change in the volume counts will result in a change of the coefficients (Lagrange multipliers) of the model.

• Equation 18 is not defined in case $T_{ij} = 0$. All relations of the trip matrix T_{ij} that are not observed cannot be used in the estimation process. This may become a severe problem if an observed a priori O-D matrix is used that (in most cases) contains, for the greater part, zero values. All O-D pairs that are not observed should be estimated first, for example, by using the weighted Poisson model (see preceding section).

• If the set of constraints in Equation 19 contains inconsistencies, no feasible solution will be found. To be able to solve this problem, it is necessary to remove all inconsistencies from the data and to estimate the trip matrix from the consistent flows (see Van Zuylen [9], Van Zuylen and Willumsen [6]). This problem often occurs, especially if cordon surveys are used in which O-D pairs might be observed several times. Therefore, some techniques have been developed to cope with this problem (see, e.g., Roos [30], Ellis and Van Ammers [31], and Smit and Te Linde [32]).

To improve the estimation technique, the entropy maximizing model has been extended in several ways—for example, the incorporation of a route information component (Van Maarseveen and Ruygrok [33], Van Maarseveen et al. [34], Willumsen [12]), as well as the introduction of time-dependent trip matrices (Willumsen [17]).

Loss of Information

Van Vuren [35] used the IM model to estimate an O-D matrix of bicycle trips in Delft. The a priori information consisted of an O-D matrix of partly observed, partly modeled trips; bicycle counts were added as extra information. As can be seen from Figure 1, the model estimate (a posteriori matrix) differed considerably from the a priori matrix, which indicates that quite some a priori information got lost.

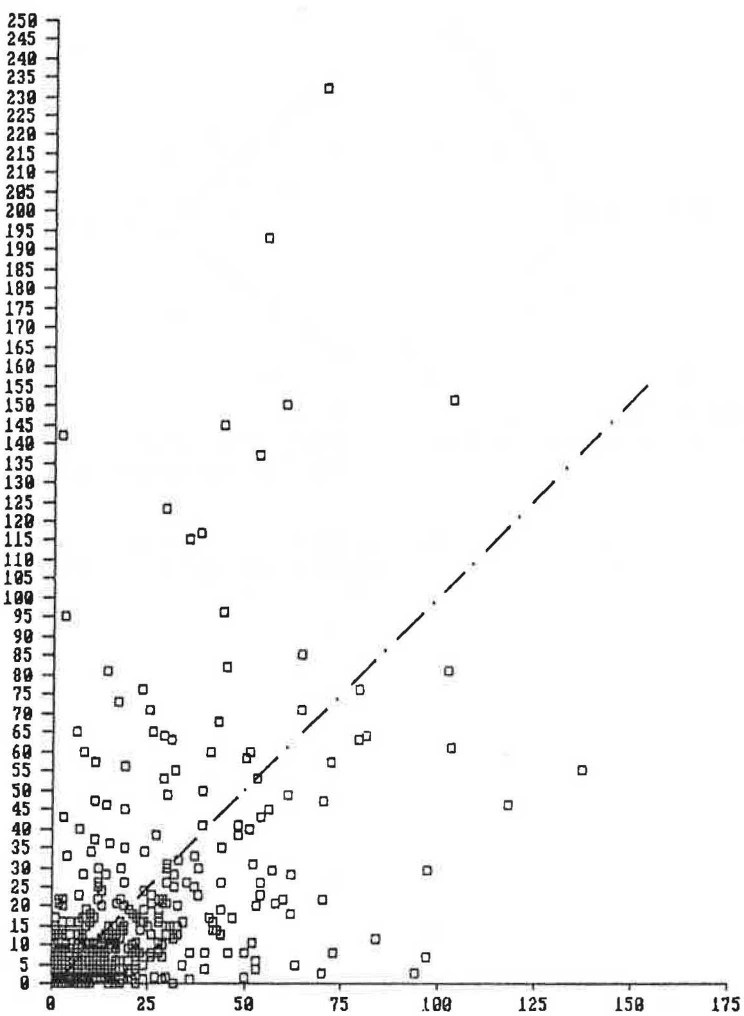


FIGURE 1 Random sample of the a priori and a posteriori values of O-D pairs estimated using the information minimizing model (*35*).

The tremendous change in the trip matrix might have been caused by the following factors:

- incorrect a priori matrix,
- specific properties of the information minimizing model,
- inconsistent data.

Unfortunately, it is not possible to trace the effects of all factors. However, an indication of the effects of the specific properties of the model can be given using the following example in which the model as been applied to a small network. The network is shown in Figure 2. Trips are generated between the zones 1, 2, 3, and 4 only. The observed trip pattern (T_{ij}) is shown in Table 1. This trip matrix will be used as a priori information. The assignment of all trips to the network results in link flows as indicated with *C* (calculated) in Figure 2.

The observed link flows are indicated with *O*. There is some dissimilarity between the calculated and observed flows, but the deviations are kept within bounds. As is usual in practice, only the network flows on a limited number of links are observed.

The results of the estimation of the O-D matrix using the information minimizing model are also shown in Table 1; the calculated flows are shown in Figure 3. As can be seen from Figure 3, the observed and calculated network flows are practically identical, which implies that there are no inconsistencies in the data. Otherwise, no feasible solution would have been obtained. A closer look at Table 1, however, makes it evident that the estimated O-D matrix has changed considerably.

This example clearly illustrates that because of the necessary adjustments that had to be made to meet the link flow constraints, the information stored in the a priori matrix has changed considerably.

In fact, the entropy maximizing model as introduced by Willumsen (*7*) will render almost the same results. The introduction of constraints with regard to the total number of incoming and outgoing flows per zone as well as the total distance traveled will result in the entropy model as formulated by Wilson. This very model, however, can also be considered an a priori matrix. The incorporation of extra constraints will modify the results of the estimation process in the same way as has already been indicated (except for the adaptation to the incoming and outgoing

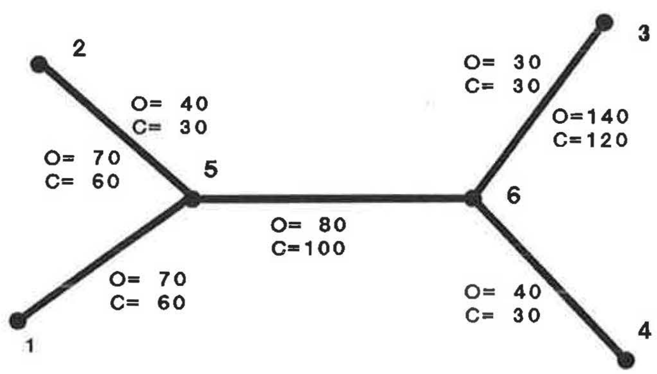


FIGURE 2 Observed (O) and calculated (C) link flows. Calculated link flows are based on assignment of observed O-D matrix (Table 1).

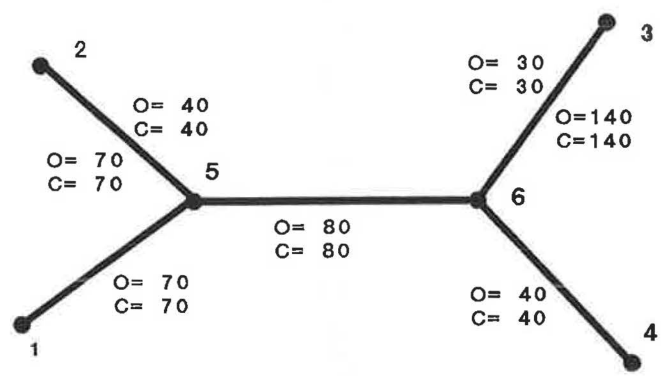


FIGURE 3 Observed and calculated link flows using the information minimizing model.

TABLE 1 OBSERVED (A PRIORI MATRIX T_{ij}) AND ESTIMATED ($\hat{T}_{ij}$) O-D MATRIX USING THE INFORMATION MINIMIZING MODEL

from \ to	1		2		3		4			
	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	$\sum_j T_{ij}$	$\sum_j \hat{T}_{ij}$
1	-	-	10	27	40	30	10	11	60	68
2	10	31	-	-	40	28	10	10	60	69
3	10	6	10	5	-	-	10	19	30	30
4	10	10	10	8	40	81	-	-	60	99
$\sum_i T_{ij}; \sum_i \hat{T}_{ij}$	30	47	30	40	120	139	30	40	210	266

flows per zone). Figure 1, being the result of a practical application of the model, clearly shows which adjustments in the a priori matrix were necessary to cope with all constraints. Actually the model formulation places less importance or less confidence in the a priori information. The concept of giving weights (e.g., in accordance with the reliability of different types of information) might be more appropriate.

INFORMATION MINIMIZING MODEL WITH ELASTIC CONSTRAINTS

The example in the preceding section clearly indicates that application of the information minimizing model may lead to considerable loss of information. The reason for this (unexpected) result can be found in the omission of the stochastic properties of the constraints (traffic counts and the like). All constraints that are incorporated into the model behave deterministically. To solve this problem of "information loss," it is suggested that the model be reformulated using elastic instead of fixed constraints (analogous to the transportation model with elastic constraints; Hamerslag [*23*]).

Model Specification

Application of elastic constraints in the optimization problem yields the following equations:

$$\hat{T}_{ij} = T_{ij} X_o \prod_a X_a \quad \forall i, j \tag{24}$$

$$\sum_{ij}\sum (\hat{T}_{ij} d_{ij}^a) = S_a \quad \forall a \tag{25}$$

$$S_a = X_a^{-g_a} R_a \quad 0 \leq g_a \leq \infty \tag{26}$$

So Equation 19 is replaced by Equations 25 and 26. But in case $g_a = 0$, S_a equals R_a.

In case of a constraint b, substitution of Equations 24 and 25 in 26 yields the following equation:

$$X^{-g_b} R_b = X_b \sum_{ij}\sum \left(T_{ij} X_o \prod_{\substack{a \\ a \neq b}} (X_a) d_{ij}^b \right) \tag{27}$$

which, after some conversion results in:

$$X_b = \left\{ \frac{R_b}{\sum_{ij}\sum \left[T_{ij} X_o \prod_{\substack{a \\ a \neq b}} (X_a) d_{ij}^b \right]} \right\}^{1/(1+g_b)} \tag{28}$$

The exponent $[1/(1 + g_b)]$ in Equation 28 is defined as the elasticity of constraint b. If the elasticity $[1/(1 + g_b)]$ equals 0, then $X_b = 1$, which implies that constraint b does not have any influence on the model estimate. If the elasticity $[1/(1 + g_b)]$ equals 1, however, the model estimate will correspond completely with constraint b. In the latter case, the information minimizing model is being dealt with.

An interesting feature of the model with elastic constraints is that (choosing correct values of the elasticities) the loss of a priori information that will occur has been reduced considerably. This can easily be demonstrated using the example given earlier. The results of an application of the model using an overall value of the elasticities of 0,5 are shown in Figure 4 and Table 2.

As can be seen from Table 2 the estimated O-D matrix is very similar to the a priori matrix; this implies that the information stored in the a priori matrix influences the model estimate (no loss of information). Still some differences between the observed and calculated link flows will remain. This phenomenon is completely in accordance with the nature of observed link flows, however, showing quite a bit of variation over time. Therefore the amount of extra information added to the problem by observed link flows (traffic counts) is limited; consequently, the application of fixed constraints would be erroneous.

The estimate of the O-D matrix strongly depends on the values of the elasticities (as is shown in Table 3), especially for values close to 1. Although it is not yet possible to give a clear indication of which values should be given to the elasticities of the constraints, it is obvious that two factors are predominant:

- the additive amount of O-D information and
- the reliability of the information.

Apart from these factors, it is obvious that the values of the elasticities should be chosen in such a way that the sum (or weighted sum) of chi^2 (for flows and O-D pairs) has a minimum value. (See Figure 5 and Table 4.)

As has been shown, the application of the information-minimizing model with elastic constraints might improve

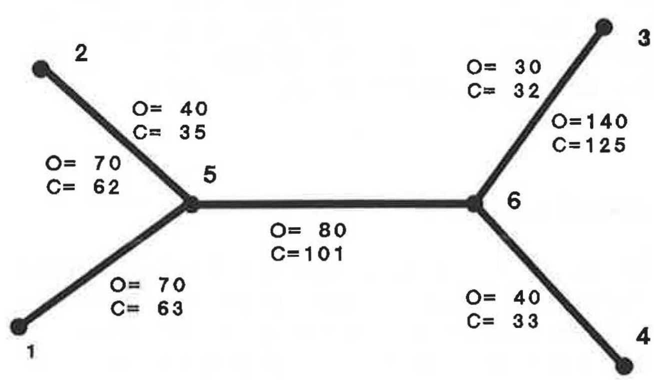


FIGURE 4 Observed and calculated link flows using the information minimizing model with elastic constraints.

TABLE 2 OBSERVED (A PRIORI MATRIX T_{ij}) AND ESTIMATED ($\hat{T}_{ij}$) O-D MATRIX USING THE INFORMATION MINIMIZING MODEL WITH ELASTIC CONSTRAINTS

from \ to	1		2		3		4			
	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	$\sum_j T_{ij}$	$\sum_j \hat{T}_{ij}$
elasticity	-	0,5	-	0,5	-	0,5	-	0,5		
1	-	-	10	13	40	40	10	11	60	64
2	10	11	-	-	40	40	10	11	60	62
3	10	9	10	11	-	-	10	11	30	31
4	10	10	10	11	40	45	-	-	60	66
$\sum_i T_{ij};\sum_i \hat{T}_{ij}$	30	30	30	35	120	125	30	33	210	223

considerably the estimate of an O-D matrix using a priori information and traffic counts. Nevertheless, some shortcomings (some of them indicated in the section on the weighted Poisson estimator) will remain:

- Each observation is formulated as a constraint and consequently results in the incorporation of an extra coefficient in the estimation model. These coefficients are time- and place-dependent, which precludes the possibility of using the model for predicting medium- to long-term changes.
- The estimated O-D matrix depends heavily on the values of the elasticities (see Table 3). If those values equal 0, the a priori matrix will not change. A value of the elasticities equal to 1 will result in the information minimizing model.
- The available information from traffic counts is used to fit only those O-D pairs that pass along the observed links. All other O-D pairs will not change at all. This problem can be overcome by using additional O-D information—for example, trip distance distribution.

BINARY CALIBRATION MODEL

The use of traffic data for estimation (calibration) purposes requires some caution. The continuous variation of traffic flows, the availability of data from various sources (surveys, counts, ticket sales), as well as differences in sample size and observed population groups may result in a considerable amount of stochasticity. Especially if various kinds of survey data are used because of this stochasticity, some inconsistencies in the data may show up (the data seem to be contradictory). In general, use of various data sources does not raise any problems; it is also possible (especially in case of inconsistencies), however, for no feasible solution to be found. The latter is especially true for the entropy type of model. Estimation models do exist, however, that can deal with the stochastic properties of the observations. Like the entropy model, all coefficients are calibrated in such a way that an optimum fit will be obtained between the observed data and the estimated trip matrix. One essential difference, in contrast to the entropy type of model, is that no additional coefficients are introduced to meet all constraints.

Hamerslag and Huisman (*1*) formulated the binary calibration model, which is based on the assumption that the observed data are independently Poisson distributed.

Smit (*2*) and Hendrickson and McNeil (*4*) use a normal distribution. Use of both latter models may raise some problems because they allow for negative flows (especially in case of relatively small flows).

Model Formulation

Suppose T_{ij} is the available a priori information (e.g., an old, observed O-D matrix). Instead of using an observed O-D matrix, it is preferable to use estimates of T_{ij} (e.g., an estimate of $\hat{T}_{ij}$ using the weighted Poisson model).

An advantage of the use of estimates is that the O-D matrix T_{ij} will not contain zero values. Analogous to

TABLE 3 ESTIMATE OF O-D MATRIX FOR DIFFERENT VALUES OF THE ELASTICITIES

from	to	1	2	3	4
	elasticities				
1	**0.0**		10	40	10
	0.5		13	40	11
	0.8	-	17	37	11
	0.9		20	34	11
	1.0		27	30	11
2		10		40	10
		11		40	11
		15	-	38	11
		19		35	11
		31		28	10
3		10	10		10
		9	11		11
		8	10	-	13
		8	8		15
		6	5		19
4		10	10	40	
		10	11	45	
		10	12	55	-
		10	11	63	
		10	8	81	

Equation 6 such a matrix can be formulated as follows:

$$\hat{T}_{ij} = Co_i d_j F_k \quad \forall i, j \tag{29}$$

where

- $\hat{T}_{ij}$ = number of trips from i to j (model estimate or a priori matrix)
- C = constant term
- $o_i = a_i O_i$ = product of balancing factor and polarity of zone i
- $d_j = b_j D_j$ = product of balancing factor and polarity of zone j
- F_k = value of deterrence function for a generalized cost class k and $c_{ij} \in k$.

Let us further assume

- Y_a = information from observations (a) (e.g., departures, arrivals per zone, traffic counts, trip distance distribution, etc.)
- d^a_{ij} = a binary variable indicating whether O-D pair ij belongs to observation a ($d^a_{ij} = 1$) or not ($d^a_{ij} = 0$)

and

$$Y_a \overset{\text{def}}{=} \sum\sum_{ij} (\hat{T}_{ij} d^a_{ij}) \tag{30}$$

The relation between the estimate of the base year matrix ($\hat{T}_{ij}$) and the a priori matrix can be represented by

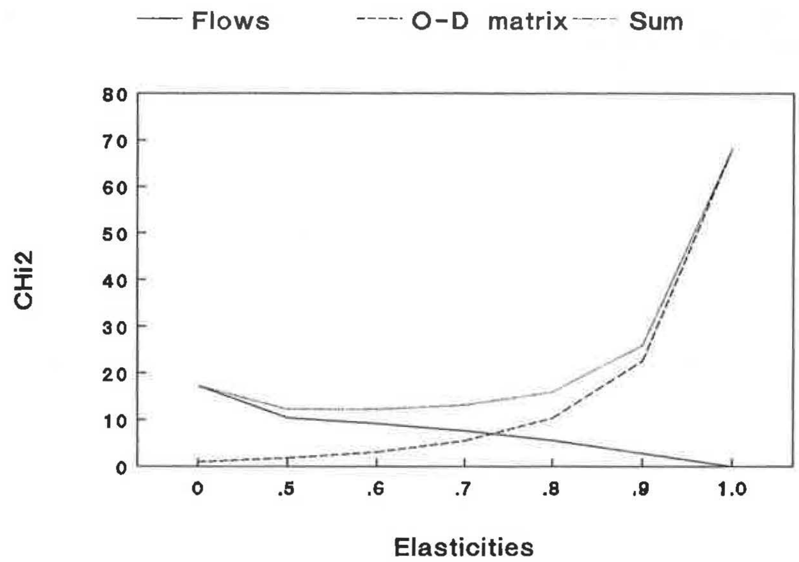


FIGURE 5 Chi-square for different values of the elasticities.

TABLE 4 OBSERVED (A PRIORI MATRIX T_{ij}) AND ESTIMATED ($\hat{T}_{ij}$) O-D MATRIX USING THE BINARY CALIBRATION MODEL

from \ to	1		2		3		4			
	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	T_{ij}	$\hat{T}_{ij}$	$\sum_j T_{ij}$	$\sum_j \hat{T}_{ij}$
1	-	-	10	10	40	40	10	12	60	62
2	10	10	-	-	40	40	10	12	60	62
3	10	10	10	9	-	-	10	11	30	30
4	10	11	10	11	40	40	-	-	60	62
$\sum_i T_{ij};\sum_i \hat{T}_{ij}$	30	31	30	30	120	120	30	35	210	216

the following equation:

$$\hat{T}_{ij} = Co_I d_J F_K T_{ij} \quad (31)$$

where $i \in I$, $j \in J$, and $k \in K$.

The preceding formulation implies that sets of balancing factors as well as discrete values of the deterrence function are modified; however, the general form of the model remains unchanged.

Assuming that the observations (Y_a) are independent and Poisson distributed, the likelihood (L) can be determined as follows:

$$\Pr[Y_a \mid \hat{Y}_a] = \exp(-\hat{Y}_a)(\hat{Y}_a)^{(Y_a)}/Y_a! \quad (32)$$

$$L = \prod_a (\Pr[Y_a \mid \hat{Y}_a]) \quad (33)$$

The logarithm of the likelihood yields

$$\ln(L) = L^* = -\sum_a (\hat{Y}_a) + \sum_a [Y_a \ln(\hat{Y}_a)] - \sum_a [\ln(Y_a!)] \quad (34)$$

An optimum adjustment between $\hat{Y}_a$ and Y_a can be determined by maximizing L^* or as

$$\sum_a [\ln(Y_a!)] = \text{constant}$$

by maximizing

$$L_1^* = -\sum_a (\hat{Y}_a) + \sum_a [Y_a \ln(Y_a)] \quad (35)$$

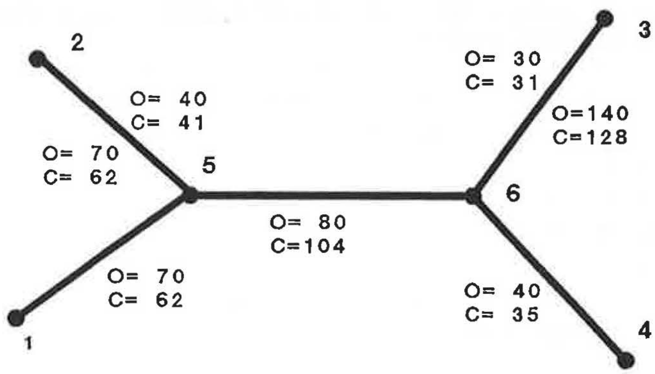


FIGURE 6 Observed and calculated link flows using the binary calibration model.

which yields the following equations:

$$\sum_a \left[\left(\frac{Y_a}{\hat{Y}_a} - 1\right)\frac{\delta Y_a}{\delta o_I}\right] = 0 \quad \forall I \qquad (36)$$

$$\sum_a \left[\left(\frac{Y_a}{\hat{Y}_a} - 1\right)\frac{\delta Y_a}{\delta d_J}\right] = 0 \quad \forall J \qquad (37)$$

$$\sum_a \left[\left(\frac{Y_a}{\hat{Y}_a} - 1\right)\frac{\delta Y_a}{\delta F_K}\right] = 0 \quad \forall K \qquad (38)$$

$$\sum_a \left[\left(\frac{Y_a}{\hat{Y}_a} - 1\right)\frac{\delta Y_a}{\delta C}\right] = 0 \qquad (39)$$

The values of o_I, d_J, F_k, and C will be determined iteratively (analogous to the Gauss-Seidel principle). Inserting those values in Equation 31 yields the estimates for the base year matrix $\hat{T}_{ij}$.

Application of the Model

The results of a practical application of the binary calibration model using the same example as before (network, a priori information, and link traffic counts) are shown in Figure 6 and Table 4.

It can be noted from Table 4 that the trip matrix estimate ($\hat{T}_{ij}$) is practically equivalent to the observed a priori matrix. It can also be noted from Figure 6 that all link traffic flow estimates are not equivalent to the link traffic counts, but there is a great similarity. Both results correspond to the very nature of the data used in the estimation process (see also the preceding section on information optimizing with elastic constraints).

The advantage of using the binary calibration model, especially compared to the information minimizing model with elastic constraints, is that no extra coefficients are incorporated in the model that allows use of the model for making medium- and long-term forecasts.

CONCLUSIONS

This paper deals with four models that can be used for estimating an O-D matrix:

- the weighted Poisson estimator
- the entropy maximizing and information minimizing model
- the information minimizing model with elastic constraints
- the binary calibration model

The weighted Poisson model is especially well suited for estimating deterrence functions and a base year matrix using information from household or cordon surveys (only O-D information). This base year matrix estimate is especially appropriate for use in the information minimizing model with elastic constraints or the binary calibration model (the base year matrix estimate is treated as a priori information).

The entropy maximizing and information minimizing models show some severe shortcomings. First, neither model allows for inconsistent information. Second, all coefficients in the model are time- and place-dependent, which precludes the possibility of using the models for making medium- and long-term forecasts. And, last but not least, use of both models induces loss of information.

The latter problem can be resolved using a model with elastic instead of fixed constraints. As has been shown, the values of the elasticities strongly influence the estimate and, consequently, the loss of information. Besides, the available information from traffic counts is used only to fit those O-D pairs that pass along the observed links. All other O-D pairs will not change at all.

The binary calibration model does not show any of the shortcomings just mentioned. The model, however, has a rather complex structure and has been successfully applied in the Netherlands (see, e.g., Hamerslag et al. [*16*] and Heere and Huisman [*36*]).

Table 5 presents a summary of the main characteristics of the models dealt with in this paper.

TABLE 5 SUMMARY OF MAIN CHARACTERISTICS OF FOUR O-D ESTIMATION TECHNIQUES

model type / characteristic	weighted Poisson	entr.max. inf.min. estimator	inf.min. elastic model	binary calibrat. constraints
estimation unobserved O-D pairs	yes	no	no	yes
apparently contradictory information permitted	no	no	yes	yes
possibility of using traffic counts	no	yes	yes	yes
change of O-D pairs	-	only observed O-D pairs	only observed O-D pairs	all O-D pairs
loss of information	no	yes	dependent on value elasticities	no
time and place dependency of coefficients (if yes, it is not possible to use the model for medium and long term forecasts)	no	yes	yes	no
complex structure of model	no	no	no	yes

ACKNOWLEDGMENTS

This paper is the result of a joint research project of Delft University of Technology and DHV Consultants b.v. Amersfoort, the Netherlands. The authors also would like to thank Piet Bovy (Delft University) for his careful review of the paper.

REFERENCES

1. R. Hamerslag and H. C. Huisman. Binaire Calibratie. De Schatting Van Coefficienten in een Verkeersmodel Met Gebruikmaking van Verschillende Soorten Waarnemingen. *Verkeerskunde*, No. 4, 1978, pp. 166–168.
2. J. G. Smit. Het Schatten van Herkomst- en Bestemmings-Matrices. In *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1980, pp. 647–699.
3. J. G. Smit. Het Schatten van Herkomst-bestemmings Matrices: de Maximum Likelihood Versus de Minimum Variance Methode. In *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1981, pp. 737–741.
4. C. T. Hendrickson and S. McNeil. Matrix Entry Estimation Errors. *Proceedings of 9th International Symposium on Transportation and Traffic Theory*, 1984, pp. 431–430.
5. H. J. Van Zuylen. The Information Minimising Method: Its Validity and Applicability to Transportation Planning. In *New Developments in Modelling Travel Demand and Urban Systems* (G. R. M. Jansen et al., eds.), Saxon House, 1979, pp. 344–371.
6. H. J. Van Zuylen and L. G. Willumsen. The Most Likely Trip Matrix Estimated from Traffic Counts. *Transportation Research*, Vol.14B, No.3, 1980, pp. 281–293.
7. L. G. Willumsen. OD-Matrices from Network Data: A Comparison of Alternative Methods for Their Estimation. *Proceedings of 6th Summer Annual Meeting*, PTRC, Warwick, England, 1978.
8. A. H. Nickesen et al. Ridership Estimation for Short Range Transportation Planning. *Transportation Research*, Vol. 17B, No. 3, 1983, pp. 233–244.
9. H. J. Van Zuylen. A Method to Estimate a Trip Matrix from Traffic Volume Counts. *Proceedings of 6th Summer Annual Meeting*, PTRC, Warwick, England, 1978, pp. 305–311.
10. H. J. Van Zuylen. Some Improvements in the Estimation of an O-D Matrix from Traffic Counts. In *Proceedings of 8th International Symposium on Transportation and Traffic The-*

ory (V. F. Hurdle et al., eds.), Toronto, Ontario, Canada, 1981, pp. 183–190.
11. H. J. Van Zuylen and D.M. Branston. Consistent Flow Estimation from Counts. *Transportation Research*, Vol. 16B, No. 6, 1982, pp. 473–476.
12. L. G. Willumsen. Estimation of Trip Matrices from Volume Counts: Validation of a Model Under Congested Conditions. *Proceedings of 10th Summer Annual Meeting*, PTRC, Seminar Q, Warwick, England, 1982, pp. 311–324.
13. M. Cremer and H. Keller. A Systems Dynamics Approach to the Estimation of Entry and Exit O-D Flows. In *Proceedings of 9th International Symposium on Transportation and Traffic Theory* (J. Volmuller and R. Hamerslag, eds.), Delft, VNU-Science Press, Utrecht, Netherlands, 1984, pp. 431–450.
14. M. G. H. Bell. The Estimation of an Origin–Destination Matrix from Traffic Counts. *Transportation Science*, Vol. 17, No. 2, 1983, pp. 198–217.
15. M. G. H. Bell. The Estimation of Junction Turning Volumes from Traffic Counts: The Role of Prior Information. *Traffic Engineering and Control*, Vol. 25, No. 5, 1984, pp. 279–283.
16. R. Hamerslag et al. Een Toepassing van Binaire Calibratie in een Regionale Studie. *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1981, pp. 713–735.
17. L. G. Willumsen. Estimating Time-dependent Trip Matrices from Traffic Counts. *Proceedings of 9th International Symposium on Transportation and Traffic Theory*, 1984, pp. 377–411.
18. S. Nguyen. *On the Estimation of an O-D Trip Matrix by Equilibrium Methods Using Pseudo-delay Functions*. CRT-publ. 81. Université de Montréal, Centre de Recherche sur les Transports, Montreal, Quebec, Canada, 1978.
19. Y. J. Gur. Trip Table Estimation Based on Partial Volume Counts. Paper Presented at 64th Annual Meeting, TRB, January 1985.
20. M. Turnquist and Y. Gur. Estimation of Trip Tables from Observed Link Volumes. *Transportation Research Record 730*, TRB, National Research Council, Washington, D.C., 1981, pp. 1–6.
21. A. G. Wilson. The Use of Entropy Maximising Models in the Theory of Trip Distribution, Modal Split and Route Split. *Journal of Transport Economics and Policy*, Vol. 2, No. 1, 1969, pp. 108–126.
22. S. F. Evans and H. R. Kirby. A Three Dimensional Furness Procedure for Calibrating Gravity Models. Presented at PTRC Summer Annual Meeting, Warwick, England, 1973.
23. R. Hamerslag. Transportation Model with Elastic Constraints. *Proceedings of 2nd PTRC Summer Annual Meeting*, Warwick, England, 1974.
24. H. Neffendorf and H. J. Wootton. (1974): A Travel Estimation Model Based on Screenline Interviews. *Proceedings of 2nd Summer Annual Meeting*, PTRC, Warwick, England, 1974.
25. M. J. L. Day and A. F. Hawkins. Partial Matrices, Empirical Deterrence Functions and Ill Defined Results. *Traffic Engineering and Control*, 1979, pp. 429–433.
26. H. R. Kirby and J. D. Murchland. Gravity Model Fitting with Both Origin–Destination Data and Modelled Trip-end Estimates. *Proceedings of 8th International Symposium on Transportation and Traffic Theory*, Toronto, Ontario, Canada, 1981.
27. R. Hamerslag. Het Schatten van Herkomst- en Bestemmingsmatrices m.b.v. Schijnbaar Tegenstrijdige Waarnemingen. *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1980, pp. 711–714.
28. A. G. Wilson. A Statistical Theory of Spatial Distribution Models. *Transportation Research*, 1967, pp. 253–269.
29. M. J. Maher. Bias in the Estimation of O-D Flows from Link Counts. Paper presented at UTSG conference, Sheffield, United Kingdom, 1987.
30. J. P. Roos. Het Gebruik Van Verkeerstellingen en Wegenenquetes bij het Opstellen van HB-tabellen. *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1982, pp. 421–431.
31. M. C. Ellis and H. Van Ammers. Forecasting Using Cordon Surveys in the Egyptian National Transportation Study. *Proceedings of 9th Summer Annual Meeting*, PTRC, Warwick, England, 1981.
32. J. G. Smit and J. Te Linde. Dubbelenquetes. *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1982, pp. 375–395.
33. M. F. A. M. Van Maarseveen and C. J. Ruygrok. Het Gebruik van Route-informatie Voor het Opstellen van HB-matrices. *Proceedings Colloquium Vervoersplanologisch Speurwerk*, 1982, pp. 399–420.
34. M. F. A. M. Van Maarseveen et al. Estimating O-D Tables Using Empirical Route Choice Information with Application to Bicycle Traffic. Paper presented at 64th Annual Meeting, TRB, January 1985.
35. T. Van Vuren. Het Schatten van een Herkomst- en Bestemmingstabel: Een Overzicht. Thesis. Delft University of Technology, Delft, Netherlands, 1985.
36. E. Heere and M. C. Huisman. Toepassing van de binaire Calibratiemethode in Zaanstad. *Verkeerskunde*, 1978, pp. 239–242.

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

Data Combination and Updating Methods for Travel Surveys

MOSHE BEN-AKIVA AND TAKAYUKI MORIKAWA

Data from a questionnaire survey that is the principal source of information on individual characteristics are subject to sampling errors and nonsampling biases. In aggregate data, sampling errors are usually small because a large sample is more easily obtainable and the data are also free from nonsampling biases. This paper develops a data combination and updating method that corrects survey data for nonresponse biases and reduces sampling errors by statistically combining survey data with aggregate data. This method is applied to the estimation of an origin–destination (O-D) table stratified by market segments. The O-D matrix entries and parameters of a nonresponse bias model are estimated by the maximum likelihood estimation method. An application for an intercity rail corridor is presented.

Questionnaire surveys are often used in transportation studies as a principal source of data. Most surveys obtain information from an individual in a face-to-face interview, a telephone interview, or a self-administered survey, including mail-out/mail-in, hand-out/hand-in, and hand-out/mail-in surveys. Survey data have the two following major drawbacks:

- High data collection costs and small budgets lead to small sample surveys and, consequently, large sampling errors.
- Nonsampling errors, such as nonresponse bias, result in biased statistics.

These drawbacks are often not present in data obtained by passive collection methods, such as passenger counts or data from utility records. These methods do not require the active participation of individuals and therefore convey information only on groups of individuals. Hence, this kind of data is called "aggregate data." For example, aggregate data in the form of on/off counts give only the numbers of passengers with a common origin or destination. The passengers are "grouped" with respect to origins or destinations, and the on/off counts are the column sums and row sums of the O-D table. Estimation of the cell entries must rely on additional data, such as on-board survey data, that trace each passenger's behavior.

Department of Civil Engineering and Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

A small sample survey can be combined with aggregate data to reduce the sampling errors and correct some of the biases in the survey data. The idea is very similar to the "mixed estimation problem" in econometrics (*1*). Ben-Akiva et al. (*2*) initiated the applications of these ideas to the analysis of survey data. The combined estimation method was employed by Hsu (*3*), who estimated a transit route-level O-D table. The framework and the propositions of these methods are presented in Ben-Akiva (*4*) and are briefly reviewed in the next section.

The purpose of the present research is to develop data combination and updating methods that include the correction of survey errors in estimating O-D tables of intercity rail passengers. The survey error correction method proposed here is to combine statistically the survey data with aggregate data to obtain unbiased and more efficient estimates of unknown population parameters. The O-D tables in this application are stratified by market segments defined by trip characteristics, such as trip purpose. Data sources to be combined are on-board travel surveys and monthly ticket sales. In this application, the principal "parameters" to be estimated are cell entries of O-D matrices by market segment. Response rate parameters specific to both the market segments and the surveys are simultaneously estimated. In addition, weekend factors and monthly seasonal adjustment factors are also estimated.

REVIEW OF DATA COMBINATION AND UPDATING METHODS FOR O-D TABLE ESTIMATION

Estimation of O-D tables is one of the most important elements in transportation planning studies. Common approaches include the direct estimation of the O-D tables by using, for instance, the aggregate gravity model (*5*) or a simulation procedure based on a disaggregate destination choice model (*6*). However, these methods require large-scale surveys and complex model estimation and application procedures.

If an O-D table exists from a previous time or a preliminary estimate is available from a previous study, it can be updated or corrected using less expensive information, such as traffic counts. Here, the concept of the mixed estimation method is used; the unknown parameters to be estimated are the cell entries of the O-D table. Survey data that provide initial estimates of the entries are referred to as "direct measurements" because they give "direct" information on the cell entries. Survey data may be collected by a passenger survey for transit or by a license plate survey for car travel.

Besides the information on every cell entry, aggregate data are also available. For example, data on the column sums and row sums of an O-D matrix are available. On/off counts at bus stops or traffic counts at the cordon line provide such data. This type of information is called "indirect measurement" because it provides "indirect" information on the cell entries.

Given these two types of information, several estimation methods are proposed in the literature. These methods attempt to maximize the "goodness" of fit (or, equivalently, to minimize the "badness" of fit) between the direct and indirect measurements. These methods also include nonstatistical approaches, such as entropy maximization (*7*) or information minimization (*8*).

The methods that have been applied most widely to the estimation of O-D tables are described next.

Iterative Proportional Fitting (IPF)

The Iterative Proportional Fitting (IPF) method is the easiest procedure to apply and has been used in many fields. In the transportation literature it has been referred to as biproportional fitting (*9*), Furness or Fratar procedure (*10*), Kruithof's algorithm (*11*), and Bregman's balancing method (*12*).

The IPF estimators are proportional to the initial matrix entries with a constant of proportionality for every row and every column. These multiplicative factors modify the initial entries to be consistent with the observed row and column sums. This method is computationally advantageous but cannot be extended appropriately to a case with stochastic indirect measurements.

Constrained Generalized Least Squares (CGLS)

The Constrained Generalized Least Squares (CGLS) estimation method is presented in Theil (*13*) and was used to estimate O-D tables by McNeil (*14*) and Hendrickson and McNeil (*15*). This method is based on the assumption that an initial matrix is equal to its corresponding true value plus a random disturbance. The disturbances usually are assumed to be normally distributed with mean zero. In other words, the initial matrix is assumed to be an unbiased estimator of the true matrix. This is a restrictive assumption because the direct measurements may have distinctive bias patterns that vary among O-D pairs.

The true matrix is estimated by the generalized least squares method subject to the equality constraints of the estimated and observed row and column sums (indirect measurements). The CGLS estimator is biased for small cell entries because of the nonnegativity constraints on the cell entries.

Constrained Maximum Likelihood Estimation (CMLE)

Although the Maximum Likelihood Estimation (MLE) method requires specific assumptions on the distribution of the observed values, it is very flexible in the sense that any type of distribution or model specification is feasible. In general, direct measurements are assumed to be independently distributed. In other words, each cell entry observation is collected by an independent sampling process.

Usually, the indirect measurements are assumed to be nonstochastic. In this case, the problem becomes the constrained maximum likelihood estimation (CMLE). Landau, Hauer, and Geva (*16*); Geva, Hauer, and Landau (*17*); and Ben-Akiva, Macke, and Hsu (*18*) applied the CMLE to O-D table estimation.

The GLS and MLE methods can also be applied with stochastic indirect measurements. If statistical independence can be assumed between direct and indirect measurements, the likelihood function to be maximized is simply the product of the likelihood of the direct and indirect measurements. The O-D table estimation method proposed in this paper is in this category, which seems most flexible. Maximum likelihood estimators are consistent and asymptotically efficient (*13*).

Geva, Hauer, and Landau (*17*) also mention that a Bayesian approach could be applied to the O-D matrix estimation if reasonable prior distributions are specified. They suggest estimators that are found by maximizing the likelihood of the posterior distribution.

DEVELOPMENT OF AN O-D TABLE ESTIMATOR WITH SURVEY ERROR CORRECTION

In the previous section the concepts of direct and indirect measurements were introduced. A maximum likelihood estimator is obtained by statistically combining both types of measurements. Direct measurements are often micro-level items and can be obtained only through survey data. Survey data, however, are subject to nonsampling errors (*19,20*). Most types of indirect measurements are obtained from other types of data collection methods and are usually free from nonresponse bias. The survey error correction method proposed here is part of an estimator in which survey data are statistically combined with aggregate data to obtain unbiased and efficient estimates of unknown population parameters.

Although the data combination methods described can be used to correct any kind of survey errors, the development in this paper is focused on the nonresponse bias that dominates all other types of errors in travel surveys. Note that this assumption simplifies the following presentation of the framework but does not imply a loss of generality.

O-D trip tables stratified by market segments are often necessary because different market segments respond differently to changes in levels of transportation services. On-board surveys can provide direct estimates of O-D tables by market segment for a transit service. However, in small samples these direct measurements have large sampling errors and may be subject to large biases due mainly to nonresponse. Ticket sales data, which are usually available for an intercity rail passenger service, provide an estimate of the aggregate O-D table that is free from nonresponse bias. The statistical combination of passenger survey data with ticket sales data can, therefore, be used to yield unbiased and more precise estimates of O-D tables stratified by market segments.

In this application the principal unknown parameters are the number of passengers belonging to a certain market segment and traveling between an O-D pair during a specific time period.

Notation

T_{ijk} = mean value of daily trips from origin i to destination j made by members of market segment k. These are the principal unknown parameters to be estimated.

p_{ijks} = response rate to survey s by individuals in market segment k traveling from i to j. These also are unknown parameters.

t_{ijks} = observed number of trips from i to j by market segment k in survey s. These are the *direct measurements.*

r_{ijm} = number of tickets sold for trips from i to j during month m. These are the *indirect measurements.*

Distributional Assumptions

Individuals in a market segment make trips according to an identical and independent Poisson process with parameter T_{ijk}. Namely, a random variable N_{ijk}, which represents the number of trips by market segment k from i to j during a randomly selected day, is the outcome of a Poisson process with parameter T_{ijk}, as follows:

$$N_{ijk} \sim \text{Poisson}\ (T_{ijk}) \qquad (1)$$

Individual response to an on-board survey is the outcome of a Bernoulli trial. Individuals in a market segment have the same response rate, P_{ijks}, to survey s. Under this assumption, the direct measurement, t_{ijks}, given N_{ijk}, is a binomial random variable with parameters N_{ijk} and p_{ijks}, as follows:

$$t_{ijks} \sim \text{Binomial}\ (N_{ijk}, P_{ijks}) \qquad (2)$$

The compound distribution of t_{ijks} given T_{ijk} is found by deriving the marginal distribution of t_{ijks}. From Equations 1 and 2 it can be shown that t_{ijks} has a Poisson distribution with parameter $p_{ijks}T_{ijk}$, as follows:

$$t_{ijks} \sim \text{Poisson}\ (p_{ijks}T_{ijk}) \qquad (3)$$

The number of trips made in a day is statistically independent of the number made on any other day. Also, the number of trips is statistically independent among market segments.

This assumption implies that the indirect measurements, r_{ijm}, are also Poisson random variables because the sum of independent Poisson variables is Poisson distributed. Since r_{ijm} is aggregated through market segments and days of the month, it is given by

$$r_{ijm} \sim \text{Poisson}\left(d_m \sum_k T_{ijk}\right) \qquad (4)$$

where d_m denotes the number of days in month m.

The survey data and the monthly ticket sales data are statistically independent. This assumption is valid if the survey data are collected on very few days in the month. The overall likelihood function is then a product of the likelihood of the survey data and that of ticket sales data.

Response Rate Specification

In the preceding equations, the response rates, p_{ijks}, are also unknown parameters. The number of these parameters can be reduced by expressing the response rate as a parametric function of the passenger's socioeconomic characteristics and the survey administration method. In this application, because of the on-board survey administration, the trip length is expected to affect the survey response rate. Assuming that a market segment is homogeneous with respect to the response rate, the following is a reasonable specification:

$$p_{ijks} = \frac{1}{1 + \exp(a_{ks} - b_{ks}d_{ij})} \qquad (5)$$

where

d_{ij} = travel time from i to j; and

a_{ks}, b_{ks} = unknown parameters.

Equation 5 employs a logistic form that bounds the response rate between 0 and 1.

Under the preceding assumption, the likelihood function can now be written as a function of the data and the unknown parameters. The function is composed of two parts: direct measurements and indirect measurements. The estimated values are obtained by applying a numerical maximization algorithm.

CASE STUDY: DATA, MARKET SEGMENTATION, LIKELIHOOD FUNCTION, AND ESTIMATION TECHNIQUE

The O-D table estimation method just described is applied to the estimation of intercity rail passenger O-D tables of the Los Angeles-San Diego (LOSSAN) corridor.

Data

The Orange County Transportation Commission (OCTC) conducted on-board surveys on the following days in July 1984: 10 (Tuesday), 11 (Wednesday), 13 (Friday), and 15 (Sunday). The survey administration method and exploratory data analyses are well documented by OCTC (*21*). Amtrak also carried out an on-board survey in December of the same year. Although the questionnaire used in the survey is available, the administration method and exact date on which it was performed are unknown.

In the OCTC surveys, four out of eight trains were chosen in each direction on each day to conduct the survey. Since the combinations of those four trains differ from day to day and O-D flows dramatically fluctuate according to the combination of trains, survey data on any particular day do not necessarily mirror the daily ridership. Therefore, the OCTC three weekday surveys are combined into a single data set.

Accordingly, the following three surveys are considered:

- survey 1—combined three weekday surveys by OCTC
- survey 2—survey conducted on Sunday by OCTC
- survey 3—Amtrak survey

Besides the on-board survey data, monthly ticket sales data from October 1981 through September 1985 are available. These provide the indirect measurements.

Market Segmentation

The specification of the market segmentation scheme should also depend on statistical considerations. These include a requirement of a minimal number of observations per cell and an a priori expectation with respect to response rate pattern. Namely, all members of a market segment must have approximately the same response rate and mean value of the number of trips by O-D pair. In terms of using the results for impact studies of policy changes, for example, it is desirable for a market segment to have homogeneous elasticities to the services.

The market segmentation scheme employed in this case study relies on trip purpose and the size of the traveling party as follows:

- market segment 1—commuting trips
- market segment 2—other business-related trips
- market segment 3—personal trips, traveling alone
- market segment 4—personal trips with a traveling party numbering two or more

Note: school trips are included as personal trips.

Since the Amtrak survey does not ask for the party size, it provides only an aggregate measurement of market segments 3 and 4. In other words, it provides indirect measurements of market segments 3 and 4.

Likelihood Function

In addition to the principal and the bias parameters, the model includes weekend and seasonality adjustment factors. The weekend factors are the ratio of a weekday to a weekend day value and are specific to market segment. Ridership also drastically fluctuates by season, and its fluctuation pattern depends on the O-D pair. Hence, all the O-D pairs are categorized into the following three O-D groups, and monthly seasonality factors are defined specific to each group:

- O-D group 1—O-D pairs with either the origin or the destination at Anaheim (the location of Disneyland)
- O-D group 2—O-D pairs with either the origin or the destination at San Clemente (a popular summer resort)
- O-D group 3—all the other O-D pairs

Reflecting the specification of the response rate and the weekend and seasonality factors, the model described below is obtained.

$$t_{ijks} \sim \text{Poisson}\left[\frac{1}{1+\exp(a_{ks}-b_{ks}d_{ij})}\,c_{ijm_s}\,w_{ks}T_{ijk}\right] \tag{6}$$

and

$$r_{ijl} \sim \text{Poisson}\left[c_{ijm_l}\left(E_m \sum_k T_{ijk} + F_m \sum_k w_k T_{ijk}\right)\right] \tag{7}$$

where

T_{ijk} = mean value of daily trips from i to j by market segment k (*principal unknown parameter*);
t_{ijks} = number of *respondents* of market segment k traveling from i to j in survey s (*direct measurement*);
r_{ijl} = number of tickets sold for trips from i to j in lth monthly ticket sales data (*indirect measurement*), $l = 1, \ldots, 48$ (i.e., four years);
a_{ks}, b_{ks} = unknown response rate parameters;
w_k = weekend factor of market segment k—that is, weekend/weekday ratio;
w_{ks} = wk: if survey s is conducted on weekend, 1: otherwise
c_{ijm} = seasonality factor of O-D pairs (i, j) in month m;
E_m = the number of weekdays in month m; and
F_m = the number of weekend days in month m.

Note: m_s and m_l denote the month of survey s and the lth ticket sales data, respectively.

The likelihood function to be maximized is given by

$$L = \prod_s \prod_i \prod_j \prod_k \Pr(t_{ijks}) \times \prod_l \prod_i \prod_j \Pr(r_{ijl}) \tag{8}$$

ESTIMATION RESULTS

In discussing the estimation results, it is useful to summarize first the unknown parameters. The 341 parameters are composed of:

288 T_{ijk}'s: since LOSSAN corridor has 9 stations, one O-D table has 72 cells (9 × 8) and there are 4 market segments (4 × 72 = 288);

8 a_{ks}'s: 4 market segments times 2 surveys (OCTC and Amtrak);

8 b_{ks}'s: ditto;

4 w_k's: 4 market segments; and

33 c_{ijm}'s: 3 O-D groups times 11 months because the parameters' values for December are normalized to one.

The estimation results of the principal parameters, T_{ijk}'s, and the other parameters are shown in Tables 1 through 6. Most of the parameters have large t-statistics.

Figures 1 and 2 compare the response rates. They show that all the response rates except for market segment 1 in the OCTC survey have upward slopes. Upward slopes mean that passengers traveling longer respond to the survey more, which, intuitively, is reasonable. The reason that market segment 1 in the OCTC survey shows the downward slope seems as follows. The observed commuting trip distribution in the OCTC surveys is quite different from the distribution in the Amtrak survey. According to the OCTC surveys, the majority of the commuting trips cover a short distance—for instance, between Los Angeles and Fullerton. The Amtrak survey, however, indicates that there are more long-distance commuting trips, such as

TABLE 1 PRINCIPAL PARAMETERS FOR MARKET SEGMENT 1—COMMUTING TRIPS

	LAX	FUL	ANA	SNA	SNC	SNT	OSD	DEL	SAN
LAX	0	20.1	0.7	13.6	16.7	0.0	11.8	15.7	29.4
		(15)	(2.7)	(14)	(13)	(2.6)	(6.9)	(6.7)	(6.7)
FUL	20.9	0	0.0	0.0	3.0	0.1	2.1	4.5	1.3
	(15)		(0.3)	(0.6)	(5.5)	(1.3)	(3.8)	(5.0)	(2.3)
ANA	0.3	0.0	0	0.0	0.6	0.1	0.6	1.9	0.8
	(1.6)	(0.3)		(0.2)	(2.7)	(1.6)	(2.6)	(4.5)	(2.0)
SNA	10.9	0.0	0.0	0	0.5	0.0	1.4	0.8	0.9
	(10)	(1.2)	(0.6)		(2.1)	(0.2)	(7.3)	(2.6)	(2.2)
SNC	20.5	8.5	0.7	2.2	0	0.0	0.6	1.4	0.0
	(13)	(9.3)	(3.1)	(7.5)		(1.4)	(2.7)	(12)	(0.1)
SNT	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0
	(1.6)	(0.7)	(1.3)	(0.6)	(0.4)		(0.6)	(0.2)	(0.5)
OSD	14.1	4.4	2.9	8.0	1.1	0.0	0	0.0	0.0
	(7.6)	(7.4)	(6.0)	(11)	(4.7)	(1.4)		(2.0)	(2.2)
DEL	31.5	5.2	5.2	3.8	2.5	0.0	0.0	0	0.0
	(10)	(5.2)	(7.4)	(7.1)	(5.8)	(1.7)	(0.3)		(2.8)
SAN	70.0	2.9	2.8	1.6	0.4	0.0	0.0	0.7	0
	(11)	(2.2)	(3.7)	(3.1)	(1.6)	(0.3)	(1.9)	(5.3)	

t-statistics in parentheses

TABLE 2 PRINCIPAL PARAMETERS FOR MARKET SEGMENT 2—OTHER BUSINESS-RELATED TRIPS

	LAX	FUL	ANA	SNA	SNC	SNT	OSD	DEL	SAN
LAX	0	21.4	3.8	9.9	7.8	0.1	7.9	13.2	22.2
		(13)	(3.8)	(5.0)	(5.4)	(1.7)	(6.6)	(9.7)	(15)
FUL	26.7	0	0.0	0.0	1.2	0.0	1.2	1.7	3.6
	(11)		(0.2)	(1.3)	(2.1)	(0.6)	(2.9)	(4.2)	(4.7)
ANA	5.5	0.1	0	0.0	0.0	0.0	0.9	1.9	2.1
	(3.7)	(1.8)		(0.8)	(0.2)	(0.3)	(2.9)	(3.2)	(4.0)
SNA	9.6	0.0	0.0	0	2.0	0.0	2.2	1.1	2.7
	(5.0)	(2.2)	(0.6)		(3.4)	(0.5)	(3.9)	(2.5)	(3.2)
SNC	14.3	2.4	6.2	0.0	0	0.0	0.7	0.6	1.9
	(6.9)	(3.3)	(15)	(0.4)		(9.1)	(1.7)	(2.1)	(2.5)
SNT	0.0	0.0	0.1	0.0	0.0	0	0.0	0.0	0.0
	(1.7)	(0.6)	(2.0)	(0.5)	(5.8)		(0.7)	(1.8)	(0.5)
OSD	13.8	3.5	0.2	6.9	0.0	0.0	0	0.0	0.0
	(8.1)	(5.4)	(1.0)	(4.7)	(0.2)	(1.6)		(0.9)	(1.8)
DEL	24.7	4.8	2.6	6.4	1.9	0.2	0.9	0	0.0
	(12)	(5.7)	(5.5)	(5.3)	(2.3)	(2.9)	(3.0)		(1.9)
SAN	28.0	2.9	3.0	6.3	1.8	0.1	0.0	0.0	0
	(13)	(4.7)	(4.7)	(7.1)	(2.8)	(2.1)	(2.0)	(2.2)	

t-statistics in parentheses

TABLE 3 PRINCIPAL PARAMETERS FOR MARKET SEGMENT 3—PERSONAL TRIPS: PARTY SIZE 1

	LAX	FUL	ANA	SNA	SNC	SNT	OSD	DEL	SAN
LAX	0	82.9	14.3	52.3	51.4	2.9	61.7	68.5	131.0
		(42)	(20)	(41)	(34)	(23)	(38)	(37)	(31)
FUL	62.8	0	0.2	0.0	21.2	2.4	20.7	19.7	38.9
	(24)		(16)	(2.6)	(27)	(24)	(36)	(31)	(23)
ANA	10.1	0.2	0	0.2	4.2	0.3	8.6	8.1	14.8
	(11)	(7.0)		(14)	(19)	(5.2)	(27)	(16)	(12)
SNA	50.9	1.8	0.2	0	7.2	2.1	23.5	23.6	48.1
	(34)	(41)	(16)		(13)	(35)	(42)	(40)	(38)
SNC	44.5	20.7	0.0	12.2	0	0.0	5.2	12.5	21.0
	(36)	(24)	(0.8)	(31)		(1.8)	(19)	(42)	(31)
SNT	2.5	2.4	0.2	0.7	0.0	0	0.2	0.3	0.5
	(38)	(26)	(4.9)	(7.5)	(1.4)		(25)	(8.3)	(9.6)
OSD	64.5	20.9	7.9	32.4	7.9	0.4	0	3.6	24.3
	(33)	(34)	(15)	(30)	(24)	(29)		(48)	(57)
DEL	53.5	23.3	4.2	28.0	11.9	0.6	2.4	0	21.9
	(24)	(25)	(8.8)	(34)	(24)	(14)	(10)		(56)
SAN	124.7	59.9	20.7	58.4	19.6	1.2	21.6	21.7	0
	(23)	(33)	(29)	(43)	(33)	(20)	(54)	(55)	

t-statistics in parentheses

TABLE 4 PRINCIPAL PARAMETERS FOR MARKET SEGMENT 4—PERSONAL TRIPS: PARTY SIZE 2+

	LAX	FUL	ANA	SNA	SNC	SNT	OSD	DEL	SAN
LAX	0	4.0	4.3	3.0	4.8	0.5	9.0	13.1	62.9
		(5.3)	(7.6)	(3.9)	(4.7)	(4.7)	(7.1)	(9.5)	(17)
FUL	15.2	0	0.0	1.5	5.0	0.6	4.4	6.3	40.0
	(7.1)		(1.4)	(39)	(6.1)	(7.3)	(6.8)	(6.7)	(20)
ANA	7.6	0.0	0	0.0	0.8	0.1	0.6	0.5	14.7
	(8.5)	(1.3)		(0.4)	(4.4)	(1.3)	(2.1)	(1.8)	(13)
SNA	10.1	0.0	0.0	0	5.0	0.2	0.3	2.4	17.8
	(7.2)	(0.8)	(0.2)		(8.5)	(9.0)	(1.3)	(4.5)	(15)
SNC	5.5	1.2	1.2	0.6	0	0.0	0.8	0.6	5.1
	(5.9)	(3.1)	(4.5)	(2.8)		(0.4)	(2.9)	(3.0)	(8.3)
SNT	0.0	0.1	0.0	0.5	0.0	0	0.0	0.1	0.2
	(1.7)	(2.1)	(0.3)	(5.1)	(0.7)		(0.6)	(2.4)	(4.2)
OSD	11.8	2.0	1.3	2.2	1.8	0.0	0	0.0	0.0
	(8.2)	(3.5)	(3.1)	(3.7)	(5.1)	(2.9)		(1.4)	(2.5)
DEL	19.1	2.2	3.3	0.8	1.7	0.0	0.0	0	0.0
	(11)	(3.2)	(6.9)	(1.9)	(5.5)	(1.8)	(0.6)		(3.8)
SAN	45.4	14.0	7.1	7.9	6.6	0.1	0.3	0.0	0
	(13)	(8.8)	(9.5)	(8.5)	(11)	(2.3)	(3.6)	(2.6)	

t-statistics in parentheses

TABLE 5 RESPONSE RATE PARAMETERS

Response Rate Parameters -- exp(a_{ks})

	OCTC Surveys	Amtrak Survey
m.s. 1	0.00 (1.0)	1.59 (2.8)
m.s. 2	7.71 (5.5)	7.60 (2.8)
m.s. 3	34.6 (11.1)	23.2 (4.9)
m.s. 4	0.94 (3.8)	∞ (0.6)

Response Rate Parameters -- exp(b_{ks})

	OCTC Surveys	Amtrak Survey
m.s. 1	0.22 (6.4)	2.14 (5.3)
m.s. 2	1.86 (13.0)	2.71 (5.9)
m.s. 3	1.39 (55.9)	1.90 (20.3)
m.s. 4	1.09 (20.8)	0.00 (0.6)

t-statistics in parentheses

TABLE 6 WEEKEND AND SEASONAL ADJUSTMENT FACTORS

Weekend Factors -- w_k

m.s. 1	0.10 (9.5)
m.s. 2	0.23 (10.1)
m.s. 3	1.79 (25.5)
m.s. 4	1.76 (24.5)

Monthly Seasonal Adjustment Factors -- c_{ijm}

	O-D Group 1	O-D Group 2	O-D Group 3
Jan	0.91 (78)	0.94 (38)	0.99 (426)
Feb	0.98 (80)	1.13 (40)	0.95 (423)
Mar	1.48 (87)	1.34 (41)	1.15 (443)
Apr	1.55 (88)	1.89 (44)	1.22 (447)
May	1.63 (89)	4.11 (49)	1.33 (456)
Jun	1.99 (93)	7.58 (51)	1.29 (454)
Jul	1.96 (93)	7.04 (51)	1.35 (459)
Aug	2.31 (94)	6.58 (50)	1.53 (468)
Sep	1.51 (87)	2.52 (46)	1.05 (433)
Oct	0.64 (72)	1.52 (43)	0.91 (413)
Nov	0.90 (77)	0.95 (38)	1.03 (432)

t-statistics in parentheses

from Los Angeles to San Diego, than short-distance ones. This discrepancy seems to have resulted from the choice of surveyed trains and passengers' definition of "commuting." Whatever the reason may be, the MLE ascribed this discrepancy to the response rates in order to fit the data.

Figure 2 shows that there were no responses from market segment 4 in the Amtrak survey. This is because the Amtrak survey provides only the aggregate number of passengers with regard to market segments 3 and 4.

The estimates of the weekend factors look reasonable. They show that on weekends the numbers of commuters and business passengers are 10 percent to 23 percent of weekdays, respectively. Also, on a weekend day there are about 1.8 times as many personal trips as on a weekday.

Monthly seasonal factors show considerably different seasonal fluctuation patterns among O-D groups. The factor of O-D group 2 (either origin or destination is San Clemente, a summer resort) takes as much as 7.6 in June, which means that in June 7.6 times as many people as in December travel to or from San Clemente. The users of San Clemente station, however, are only a very small portion of all the patrons (in fact, only one out of eight trains stops at that station daily). O-D group 1 (either origin or destination is Anaheim, the Disneyland station) also shows higher factors in summer than does O-D group 3 (all the other O-D pairs).

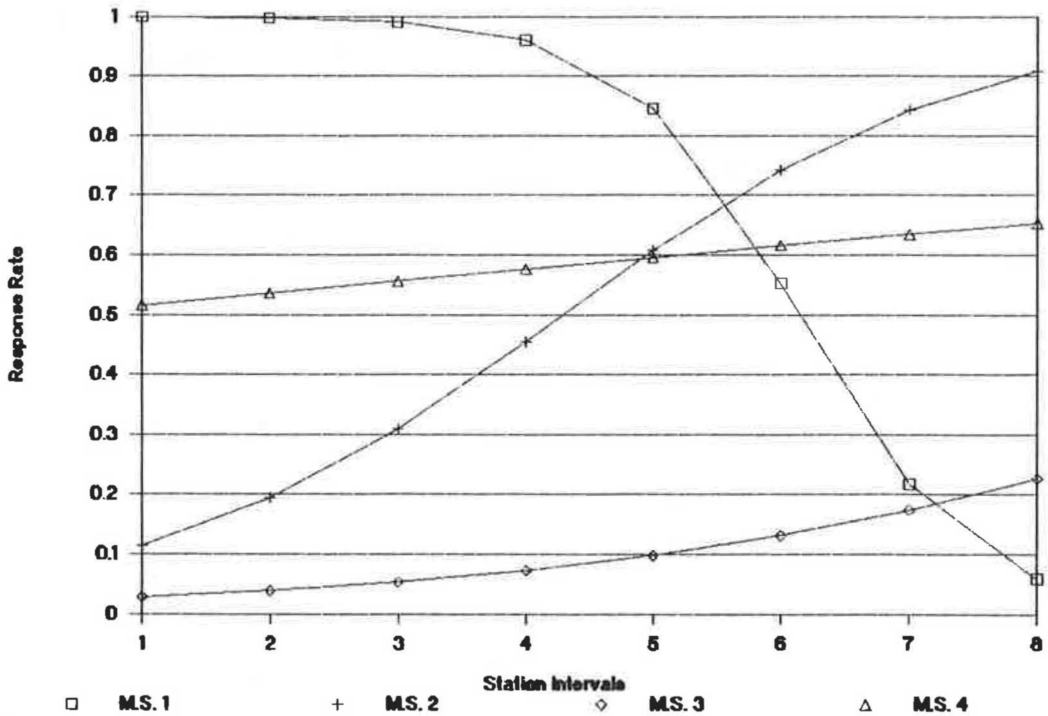


FIGURE 1 Response patterns: unrestricted model—OCTC survey.

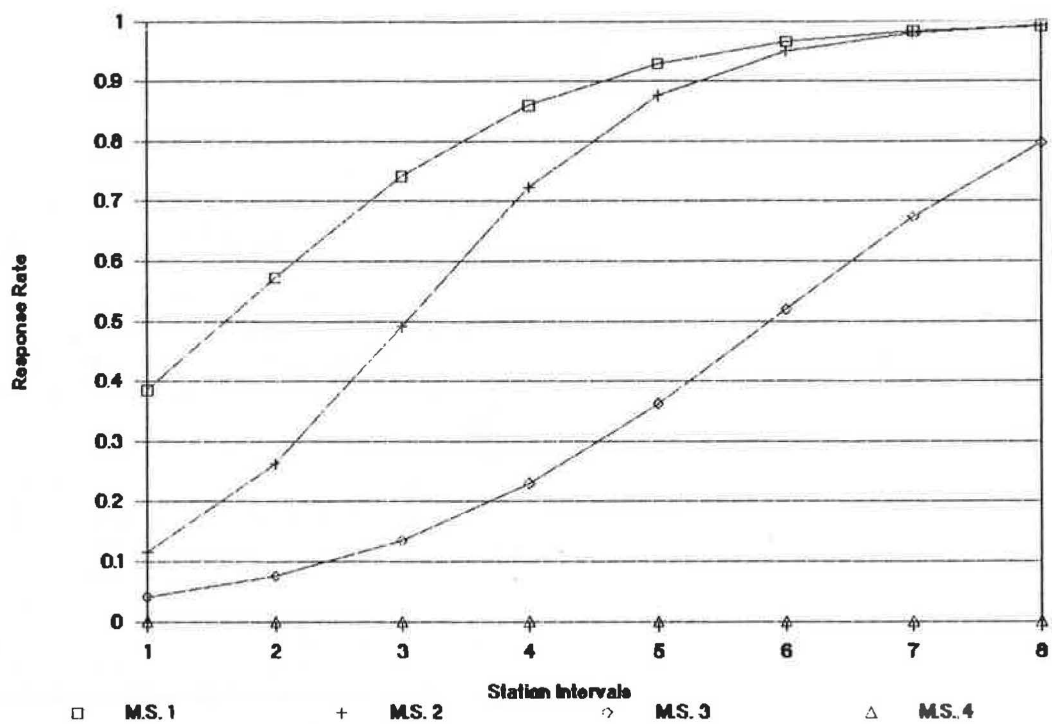


FIGURE 2 Response patterns: unrestricted model—Amtrak survey.

CONCLUSION

This paper proposes a correction method for survey errors by combining different data sources. Although reducing the survey errors by improving the survey administration or by repeating the survey may reduce the nonresponse problem, it is sometimes inefficient in terms of cost and time. Utilization of other available data sources that are inexpensive to collect seems to be an attractive and practical approach. The proposed method statistically combines survey data with aggregate data to obtain unbiased and more efficient estimates.

Although this paper focuses on the nonresponse bias, the proposed method can correct various other types of survey errors. It can also be used to update a database by treating the old database as direct measurements and combining it with up-to-date indirect measurements.

In practice, this methodology enables the utilization of any kind of information, and as much of it as possible. Conversely, it can always be used to improve future data collection methods by designing "optimal" combinations of data collection efforts.

The models described in the last section of this paper are based on some assumptions that should be investigated further. The robustness of the Poisson assumption needs to be tested by employing other distributions, such as the Normal. Secondly, alternative market segmentation schemes should be attempted. Market segmentation that properly reflects the trip generation and response behavior assumptions is required for the model.

The idea of statistically combining data from different sources can be applied in other contexts. For example, parameters of disaggregate and aggregate travel demand models can be estimated by using both survey and external counts or census data. A model transferred from one region to another is often subject to a transfer bias that can be corrected by combining data from both regions (*22*). Another application is the reweighting of survey data using the latest regional demographic data.

ACKNOWLEDGMENTS

This research was sponsored by the Federal Railroad Administration and the Transportation System Center of the United States Department of Transportation. The data for the case study were provided by Amtrak and the Orange County Transportation Commission. The authors are grateful to Stewart Butler and Arrigo Mongini, for their guidance and suggestions, and to John Prokopy, who played a central role in developing the case study. Denis Bolduc assisted with the computational aspect of this research.

REFERENCES

1. G. G. Judge, W. E. Griffiths, R. C. Hill, and T. C. Lee. *The Theory and Practice of Econometrics*. Wiley, New York, 1980.
2. M. Ben-Akiva, H. F. Gunn, and H. Pol. Expansion of Data from Mixed Random and Choice-Based Designs. Prepared for the 2nd International Conference on New Survey Methods in Transportation. Hungerford Hill, Australia, 1983.
3. P. S. Hsu. Estimation of Parameters for Multiple Spatially and Temporally Distributed Populations. Ph.D. dissertation. Massachusetts Institute of Technology, Department of Civil Engineering, Cambridge, 1985.
4. M. Ben-Akiva. Methods to Combine Different Data Sources and Estimate Origin-Destination Matrices. *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 1987.
5. R. E. Alcaly. Aggregation and Gravity Models: Some Empirical Evidence. *Journal of Regional Science*, Summer 1967, pp. 61–73.
6. M. Ben-Akiva, H. F. Gunn, and L. Silman. Disaggregate Trip Distribution Models. *Proceedings of Japanese Society of Civil Engineers*, Vol. 347/IV-1, 1984, pp. 1–17.
7. A. G. Wilson. *Entropy in Urban and Regional Modeling*. Pion, London, 1970.
8. P. Uribe, C. G. de Leeuw, and H. Theil. *The Information Approach to the Prediction of Interregional Trade Flows*. Technical Report 6507. Economic Institute of the Netherlands, School of Econometrics, 1965.
9. M. Bacharach. *Biproportional Matrices and Input-Output Change*. Cambridge University Press, 1970.
10. S. P. Evans and H. R. Kirby. A Three-Dimensional Furness Procedure for Calibrating Gravity Models. *Transportation Research*, Vol. 8, 1974, pp. 105–122.
11. J. Kruithof. Calculation of Telephone Traffic. *De Ingenieur*, Vol. 52, 1937, pp. E15–E25.
12. B. Lamond and N. F. Siewart. Bregman's Balancing Method. *Transportation Research B*, Vol. 15B, No. 4, 1981, pp. 239–248.
13. H. Theil. *Principles of Econometrics*. Wiley, New York, 1971.
14. S. McNeil. 1983. Quadratic Matrix Entry Estimation Methods. Ph.D. dissertation. Carnegie-Mellon University, Department of Civil Engineering, Pittsburgh, 1983.
15. C. Hendrickson and S. McNeil. Estimation of Origin/Destination Matrices with Constrained Regression. *Transportation Research Record 976*, TRB, National Research Council, Washington, D.C., 1984, pp. 25–32.
16. U. Landau, E. Hauer, and I. Geva. Estimation of Cross-Cordon Origin Destination Flows from Cordon Studies. *Transportation Research Record 891*, TRB, National Research Council, Washington, D.C., 1982, pp. 5–10.
17. I. Geva, E. Hauer, and U. Landau. Maximum Likelihood and Bayesian Methods for the Estimation of O-D Flows. *Transportation Research Record 944*, TRB, National Research Council, Washington, D.C., 1983, pp. 101–105.
18. M. Ben-Akiva, P. P. Macke, and P. S. Hsu. Alternative Methods to Estimate Route Level Trip Tables and Expand On-Board Surveys. *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 1–11.
19. W. G. Cochran. *Sampling Techniques*, 3rd ed. Wiley, New York, 1977.
20. E. D. Deming. *Some Theory of Sampling*. Wiley, New York, 1950.
21. Orange County Transportation Commission (OCTC). *An Evaluation of Amtrak's San Diegan and Metroliner Services in the Los Angeles to San Diego Corridor*. Study Report. Orange County, Calif., 1984.
22. M. Ben-Akiva and D. Bolduc. Approaches to Model Transferability: The Combined Transfer Estimation. Presented at 66th Annual Meeting of the Transportation Research Board, Washington, D.C., 1987.

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

Empirical Analysis of Trip Chaining Behavior

KAZUO NISHII, KATSUNAO KONDO, AND RYUICHI KITAMURA

This study is concerned with commuters' trip chaining behavior in which a stop for nonwork activities is introduced to the basic home-work-home travel pattern. An important question is whether a commuter will make the nonwork stop during the commuting trip or, alternatively, will pursue the nonwork activity by making a separate trip chain from home. A theoretical model is formulated to address this question. The analysis indicates that the likelihood of pursuing the nonwork activity in a separate, home-based trip chain will increase with the speed of travel and will decrease as commuting distance, travel cost, or the density of opportunities increases. The empirical analysis, using data sets from two areas and two points in time, generally supports these theoretical relationships. It also shows, however, that trip chaining behavior does not remain stable over time despite the fact that trip rates are very stable.

Extensive knowledge of trip chaining behavior is now available through the accumulation of studies in various disciplines. Several analytical frameworks have been proposed, and common behavioral characteristics have been identified across data sets and over time. Recent developments include attempts to introduce the concept of trip chaining into demand forecasting procedures (see the review in the following section). Not unexpectedly, however, previous analyses and models are subject to limitations. Most notably, the well-recognized spatial and temporal constraints imposed on trip chaining behavior (*1*, *2*) have rarely been incorporated into quantitative analyses. In addition, the conditions in the travel environment that induce multistop chains (i.e., a home-to-home tour that contains more than one stop) have not been determined. These limitations have motivated this study.

The objective of this study is first to determine the conditions in the travel environment that encourage or discourage linking of trips into a multistop chain. Second, the study attempts to determine whether these conditional relationships prevail over time and across metropolitan areas. For this purpose, a model of trip chaining behavior is developed while explicitly incorporating time-space constraints, and using an abstract linear-city setting. The spatial and temporal stability of the behavioral properties derived from this model is examined empirically using travel data from two metropolitan areas at two points in time.

The focus of this study is on linking work trips and other trips in a constrained environment. The probability of combining work and other activities into one home-based trip chain is related to parameters characterizing the travel environment, such as the travel speed and commuting distance. The model is an extension of the one in Kondo and Kitamura (*3*). In the present model, in-home and out-of-home activity durations are endogenously determined assuming a specific functional form to represent the utility of activity engagement. Relationships derived from the model are examined for their empirical validity in order to identify the conditions that induce linked trips. Also evaluated through this process is the effectiveness of the model framework that draws on the notion of time-space constraints. The data sets from two areas and two points in time allow a robust assessment of the model's validity and usefulness.

Although the model enables the derivation of behavioral relationships, its development is based on certain idealizations and simplifications that may limit the generality of the results. Perhaps most critical is the assumption that opportunities are uniformly distributed in the urban area. Because of this assumption, behavioral properties derived from the model may not necessarily be compatible with observed behavior, which is obviously influenced by variations in the type and intensity of land use within an urban area. Another idealization is that the urban area can be represented as a one-dimensional space. This simplification is not restrictive, however, because the study examines trip chaining under the assumption that destination locations are predetermined (a two-dimensional extension appears in Goulias and Kitamura [*4*] in which location choice in trip chains is the subject of analysis). It is also assumed that trip makers are homogeneous and that the properties identified through the model apply uniformly to all individuals. The applicability of the study results to population segments needs to be examined in a future effort. Despite these limitations the results presented in this paper offer behavioral insight and provide a basis for future model building efforts.

The paper is organized as follows. The next section offers a brief review of past studies of trip chaining behavior, with an emphasis on several especially pertinent ef-

K. Nishii, Department of Civil Engineering, University of California, Davis, Calif. 95616. Current affiliation: Department of Civil Engineering, Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400, Japan. K. Kondo, Department of Civil Engineering, Fukuyama University, Fukuyama, Hiroshima 729-02, Japan. R. Kitamura, Department of Civil Engineering, University of California, Davis, Calif. 95616.

forts. The following section presents the framework of the linear-city model and derives relationships between trip chaining and parameters characterizing the urban area. The results of empirical analysis appear next. Temporal and spatial stability in the distribution of path types and time use patterns is discussed. The properties obtained from the model are then empirically examined. The final section is a brief summary.

BACKGROUND

Existing models of trip chaining behavior can be classified into four broad groups: Markovian, activity-based, disaggregate utility-maximization models, and simulation models. Markovian models, which have been applied in trip chaining research since the concept of linked trips was introduced, are frequently used to represent the linkage between trip purposes or between the facilities at trip-ends. An application of Markov chain models in the trip distribution phase of the travel demand forecasting process can be found in Sasaki (*5*). An extension to modal split has also been investigated by Kondo (*6*) and Nishii and Sasaki (*7*). Application of Markov renewal process models has recently been proposed for a more elaborate representation of trip chaining behavior (*8*). This is generalized to a time-dependent, probabilistic process, assuming that the decision underlying trip chaining is influenced by factors related to the time of day (*9*). Extensions of discrete Markovian models can also be found in O'Kelly (*10*) and Borgers and Timmermans (*11*), who propose nonstationary Markov chain models. Many limitations of Markovian models noted in the past (*12*) are overcome in these recent extensions.

The activity-based approach to travel demand analysis has its roots in time geography (*1*) and human activity analysis (*13*). The main subjects for investigation include the interdependencies among household members in daily activities, the effect of life-cycle stages on activity patterns, and characteristics of time-space paths (i.e., the trajectory formed by an individual in the time-space dimension) and constraints imposed on them (*2,14*). Recently, efforts have been made to apply this approach to the examination of alternative hypotheses and development of theoretical frameworks for trip chaining behavior. In particular, Recker et al. (*15*) developed a model system that explicitly incorporates the time-space constraints. This model system promises to be a useful tool for a wide range of transportation planning options. The activity-based approach has contributed importantly to the analysis of trip chaining behavior.

There have been several trip chaining models that take on the utility-maximization approach (*8,16*). Adler and Ben-Akiva (*17*) developed a utility-maximizing model while viewing the decision for nonwork travel behavior as a choice from among a set of feasible daily paths. An effort to develop a stochastic process model using discrete choice models can be found in van der Hoorn (*18,19*) and in Kitamura and Kermanshah (*20,21*). The effect of time-space constraints on trip chaining is evaluated in a simplified context (*3*). In many of these applications, the decision underlying daily travel behavior is decomposed, and discrete choice models are applied to the resulting decision components. Whether these components can be integrated to reconstruct a daily travel pattern properly and meaningfully remains as a future research subject.

Another class of studies consists of simulation analyses of trip chaining. For example, Southworth applies heuristic rules to generate a set of destination locations visited in a trip chain (*22*). Markovian assumptions are applied in other simulation studies (*23,24*).

Of particular importance to this research is a set of studies that share the common interest: How nonwork activities and trips are linked to work trips. This group of work includes Oster (*25*), Hanson (*26*), Damm (*14*), Adiv (*27*), and Kondo and Kitamura (*3*). Damm (*14*) focuses on the interdependence in activity engagements across periods of the day and attempts to develop a quantitative model describing how nonwork activities are linked or not linked to work trips. Damm's tabulation using observations from Minneapolis and St. Paul (Minnesota) indicates that 45 percent of workers who engaged in nonwork activities did so in connection with work, while 48 percent pursued nonwork activities in separate, home-based chains made after work.

The linking of nonwork activities to work activities is further examined by Kondo and Kitamura (*3*) with the use of constrained choice models to quantify the effect of time constraints. Their analysis indicates associations between trip chaining and the duration of nonwork activity, residential location, and work trip mode. The important association between travel mode and trip chaining is also shown by Nishii and Sasaki (*7*) using Japanese data.

The "second role" of the work trip, that is, providing the opportunity to link nonwork trips, is emphasized in Oster (*25*) and Hanson (*26*). In general, the empirical results indicate the importance of this second role. Hanson (*26*) notes that stops at supermarkets, kiosks, and other persons' homes are most frequently made while commuting and points out that stops made during a work journey tend to be unplanned.

The analysis by Adiv (*27*) is motivated by the hypothesis that commuters are reluctant to use public transit because of the convenience driving offers for engaging in extra activities during the work trip. Adiv offers interesting statistics of trips and activities pursued on the way to work, during work, and on the way home from work; he reports that no particular activity type tends to be linked to work more frequently than statistically expected. Adiv's overall observation is that "in spite of the empirical results which showed the existence of extra activities during the work-trip, one has to conclude that most daily activities of working people, as represented by this sample, are still conducted independently of the journey to work" (*27*,p.135). Golob's tabulation of a Dutch data set indicates that approximately 80 percent of home-based trip chains are simple home-work-home chains (*28*), leading to the same question about the significance of trip chaining involving work in the Netherlands.

Adiv's observation may be a reflection of land use development particular to the San Francisco Bay Area, where his observation was taken. It is conceivable that the apparent contradiction between the results of Adiv and those of Hanson (using data from Uppsala, Sweden) may be due to land use differences. Nonetheless, Adiv's observation raises the important question of whether accounting for trip chaining behavior using more complex methods and models is justifiable from a practical viewpoint. Further effort obviously needs to be devoted to trip chaining involving work trips.

Extensive effort has been made in the analysis of trip chaining behavior. Common characteristics of trip chaining behavior have been identified as summarized in recent reviews (*29,30*). The field appears to be entering the state where practical and application-oriented questions need to be addressed. Surprisingly, however, relatively little is known about conditions in the travel environment that induce trip chaining. Consequently, little information is available to help determine whether trip chaining will be more prevalent in the future. This paper's aim is to determine on what conditions workers tend to link trips to form trip chains involving work and other activities.

MODEL OF TRIP CHAINING UNDER TIME-SPACE CONSTRAINTS

In this section a simple model of trip chaining behavior under time-space constraints is formulated, and some properties of the model are derived. The section's intent is to lay an analytical foundation for the empirical analysis of trip chaining behavior presented in the following section. The model is based on the one proposed by Kondo and Kitamura (*3*) and depicts the formation of trip chains in a linear city under constraints represented by a time-space prism (*1*). The study extends the model by Kondo and Kitamura by postulating a specific functional form for the utility of out-of-home and in-home activities and by endogenously determining the amounts of time allocated to out-of-home and in-home activities.

Prism in the Time-Space Coordinates

Consider a linear city in which opportunities for out-of-home activities are distributed with constant density $\Gamma(x) = \Gamma$, $-\infty < x < +\infty$. Suppose a worker in this linear city resides at $x = L$ and commutes to his stationary work location at $x = 0$. The simplest travel pattern the worker can follow is a two-leg, one-stop, home-work-home trip chain. If the worker is subject to a fixed work schedule, the timing of an additional activity is constrained in the time-space coordinates because of the work schedule and also because of the limited speed of the travel mode used. The region in the time-space coordinates that the worker can occupy is often called a "prism." A typical worker's travel behavior is confined by three prisms: one before the scheduled work, one during the lunch break, and one after work (*14,31*).

Figure 1 shows an example of the prism for a one-dimensional urban space. The parameters that define the prism are:

L = distance between work location and home,
t_0 = earliest possible time that the worker can leave a base (home or office),
t_1 = time by which he must arrive at a base (office or home);
v = speed of the travel mode the worker uses,
$T = t_1 - t_0$, total available time, or height of the prism.

The maximum amount of time available for activities, h, is obtained as

$$h = T - L/v.$$

These parameters are shown with the prism in Figure 1.

The analysis of this study is concerned with the case in which the worker pursues a discretionary out-of-home activity in addition to work. If the location and time of the additional activity are not fixed, the activity can be pursued in any one of the three prisms. If the activity is pursued in the first or the third prism, the worker has the options of (1) combining this additional activity with the commuting trip and (2) pursuing the activity in a separate, home-based, one-stop trip chain. A sequence of trips, or a path, shall be called a "multichain" path if the additional activity is pursued in a separate, home-based trip chain and a "single-chain" path if the activity is linked to a commuting trip. Then the worker has five types of paths (*14*): a multi- or single-chain path in which the activity is pursued before work, a path with the activity pursued during work, and a multi- or single-chain path with the activity pursued after work.

Utility Components of Time-Space Paths in the Prism

Let $A(x)$ be the attractiveness of the set of opportunities in a unit-distance interval at x. Suppose that time can be

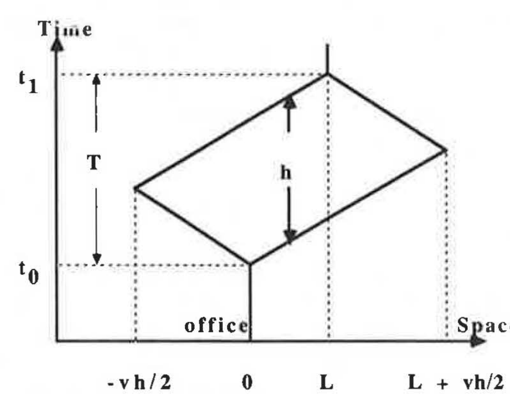


FIGURE 1 A prism in the time-space coordinates.

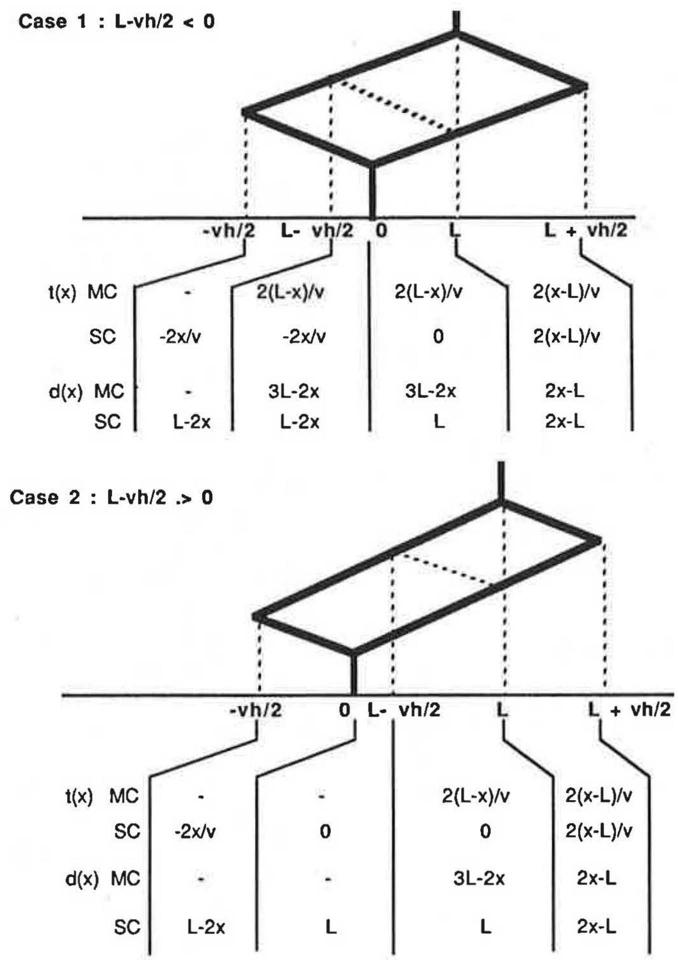


FIGURE 2 ***t*(*x*) and *d*(*x*) by activity location, *x* for *MC* path and *SC* path.**

TABLE 1 RELATIONSHIP BETWEEN *DU* AND PARAMETERS CHARACTERIZING THE LINEAR CITY

	$L-hv/2 \le x < 0$	$0 < x \le L$	$L < x \le L+hv/2$
$\partial DU / \partial L$	< 0	< 0	0
$\partial DU / \partial v$	> 0	> 0	0
$\partial DU / \partial \theta$	< 0	< 0	0
$\partial DU / \partial A$	< 0	< 0	0
$\partial DU / \partial H_0$	> 0	> 0	> 0

freely allocated between in-home and out-of-home activities. In this case, it can be assumed that the duration of the activity at x will be determined such that the total utility of in-home and out-of-home activities will be maximized. Assume as well that the utility of the activity at x increases proportionally with the attractiveness measure $A(x)$ and, further, that utilities of in-home and out-of-home activity durations can be combined into a Cobb-Douglas function. The total utility of engaging in an activity at x and allocating s amount of time can then be expressed as

$$U(x, s) = A(x)s^{\beta}\{h - t(x) - s\}^{1-\beta} \qquad 0 < \beta < 1$$

where

s = the duration of out-of-home activity at x,
$t(x)$ = the reduction in available activity time due to travel, and
h = the maximum activity time available, as defined earlier.

Setting the first derivative with respect to s equal to 0, the optimal activity duration at x, s^*, is obtained as

$$s^* = \beta\{h - t(x)\}$$

The total utility of the activity time in the prism with s^*, $U^*(x)$, can be expressed as

$$U^*(x) = \beta^{\beta}(1 - \beta)^{1-\beta}A(x)\{h - t(x)\}$$

The net utility, then, can be formulated as follows:

$U^*_{SC}(x) - \theta d_{SC}(x)$ if a single-chain path
$U^*_{MC}(x) + H_0 - \theta d_{MC}(x)$ if a multichain path

where

$d(x)$ = total travel distance [$d_{SC}(x)$ for single-chain path, $d_{MC}(x)$ for multichain path],
θ = travel cost per unit distance, and
H_0 = the added utility that arises when the discretionary activity is pursued in a separate, home-based trip chain.

If returning home before engaging in the discretionary activity offers an opportunity to freshen up, rest, or to attend to household chores, the separate home-based trip chain may yield additional utility. This added utility, H_0, is associated with the schedule convenience.

TABLE 2 INDICATORS OF TRAVEL PATTERNS BY YEAR

Indicators		1970	1980	change(%) 1980/1970
Trip rates (trips/tripmaker)	Osaka	2.72	2.92	+7.4%
	Kyoto	2.77	2.92	+5.4%
Trip rates (trips/worker)	Osaka	2.84	3.07	+8.1%
	Kyoto	2.87	3.03	+5.6%
The average number of office-based chains	Osaka	1.21	1.20	-0.8%
	Kyoto	1.20	1.20	0.0%
The average number of office-based sojourns	Osaka	1.69	1.62	-4.1%
	Kyoto	1.78	1.67	-6.2%
Tour length*(km) : C.B.D. tours	Osaka	8.74	8.43	-3.5%
	Kyoto	7.55	7.44	-1.5%
Tour length * (km): all tours	Osaka	14.01	14.45	+3.1%
	Kyoto	12.87	12.33	-4.2%

*) Tour length is defined as the average distance between tripmaker's home and the farthest sojourn location. The figures are for workers who made at least one trip by car. A C.B.D. tour is one which contains at least one stop in the central area.

Formation of Trip Chains

The relation between the likelihood of trip chaining and the model parameters is now derived assuming that the worker chooses between single-chain (SC) and multichain (MC) paths given the location of the additional activity, x.

If both types of paths are feasible (i.e., $L - hv/2 \leq x \leq L + hv/2$), it may be assumed that the choice between MC and SC paths depends on the utility difference, DU,

$$DU = U^*_{MC}(x) - U^*_{SC}(x)$$
$$= \beta^{\beta}(1-\beta)^{(1-\beta)}A(x)\{t_{SC}(x) - t_{MC}(x)\} + H_0 + \theta\{d_{SC}(x) + d_{MC}(x)\}$$

where $A(x)$ is the attraction measure as before, and $U^*_{MC}(x)$ and $U^*_{SC}(x)$ are the net utilities of a sojourn of duration s^* at an opportunity at x in an MC path and an SC path, respectively. The trip distance in the linear city, $d(x)$, and time available for the activity, $t(x)$, can be determined automatically for each path type given the activity location, x. The results are summarized in Figure 2, which is prepared for the third prism. Using this result, the utility difference can be expressed as

$$DU = \begin{cases} H_0 - \beta^{\beta}(1-\beta)^{1-\beta}A(x)(2L/v) - 2\theta L & \text{if } L - hv/2 \leq x < 0 \\ H_0 - \beta^{\beta}(1-\beta)^{1-\beta}A(x)(L-x)/v - 2\theta(L-x) & \text{if } 0 \leq x \leq L \\ H_0 & \text{if } L < x \leq L + hv/2 \end{cases}$$

It can be assumed that the worker is more likely to take on an MC path as the utility difference increases.

Using this relationship, the likelihood of MC and SC paths can be related to the model parameters that characterize the linear city. The results are summarized for L, v, θ, $A(x)$, and H in Table 1. Note that the relations between DU and the model parameters vary depending on the activity location. In particular, most of the model parameters examined here are unrelated to DU if $L < x \leq L + vh/2$ (see Figure 2). In the other regions of the linear city, the likelihood of an SC path increases with the

TABLE 3 DISTRIBUTION OF PATH TYPES INVOLVING TWO STOPS

city	year	Before work MC	Before work SC	During work	After work SC	After work MC	Total
Osaka	1970	193	337	2773	3280	1196	7779
		2.5%	4.3%	35.6%	42.2%	15.4%	100%
	1980	119	305	4072	2450	665	7611
		1.6%	4.0%	53.5%	32.2%	8.7%	100%
Kyoto	1970	77	95	643	875	445	2135
		3.6%	4.5%	30.1%	41.0%	20.8%	100%
	1980	91	162	1282	1005	475	3015
		3.0%	5.4%	42.5%	33.3%	15.8%	100%

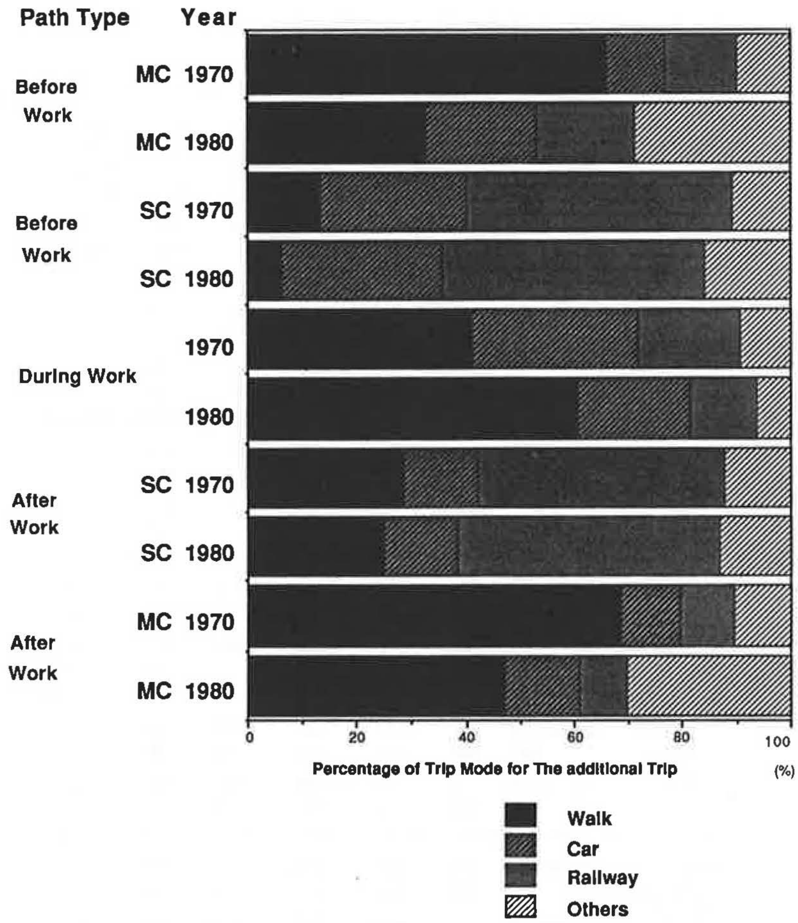


FIGURE 3 Distribution of modes used for additional activity.

commuting distance (L), travel cost (θ), and the mean attractiveness of the opportunities ($A(x)$); that of an *MC* path increases with the speed of travel (v) and the added utility (H_0).

It is important to note that the relationships described are derived using the assumption of a uniform linear city. This assumption is obviously restrictive, and observed relationships may divert from these theoretical relationships because of variations in the distribution of opportunities and other irregularities. Nonetheless, the theoretical results have offered guidance to the empirical analysis, the results of which are presented in the next section.

EMPIRICAL RESULTS: TWO-STOP PATHS

The temporal and spatial characteristics of trip chaining behavior are studied in this section using observations from two time points. The validity of the behavioral properties derived from the theoretical linear city model is then examined. The latter focus reflects the desire to determine the usefulness of the concept of time-space constraints in analyzing travel behavior and also to determine the extent to which these constraints are binding. The analysis is limited to those paths that involve a sojourn (or two sojourns) for work at a fixed location, and a nonwork sojourn.

The data sets used are from the Kyoto-Osaka-Kobe metropolitan area in Japan. Comparable data are available from 1970 and 1980 and are suitable for a comparative analysis between the two time points ten years apart. Both data sets contain records of the entire set of trips made by each member (≥5 years old) of sample households on the survey day. Sample households of the two surveys were selected randomly based on residential location with sampling rates of approximately 3 percent.

Mobility characteristics in these data sets are summarized in Table 2. Trip rates are very similar (within 2 percent differences) between Osaka and Kyoto subareas in both 1970 and 1980. Spatial stability of trip generation is evident from this tabulation. The measures of tour length (i.e., the distance between the home base and the farthest sojourn location), on the other hand, show that a tour tends to be longer in the Osaka area, which has a much larger urban area. Table 2 also indicates that trip rates in both areas increased in 1980, with the Osaka area showing slightly higher rates of increase. Unlike trip rates, tour length does not offer a clear tendency between the two time points.

The analysis presented in the rest of this paper is based on a subsample consisting of those who (1) were employed at the time of survey, (2) made a work trip on the survey day, (3) pursued exactly one out-of-home activity in addition to work, and (4) had a closed home-based path.

TABLE 4 INDICATORS OF TIME USE PATTERNS BY PATH TYPE

Indicators	City Year	Before work MC	Before work SC	During work	After work SC	After work MC	One-stop path
The number of Individuals*	Osaka 1970	47	247	1472	1938	251	25375
	1980	[illegible]	263	1508	1591	290	20876
	Kyoto 1970	26	69	386	518	142	7133
	1980	59	133	594	671	256	9487
The average work trip duration(min)	Osaka 1970	35	49	47	51	33	49
	1980	28	47	50	52	35	52
	Kyoto 1970	26	37	41	40	26	42
	1980	19	41	39	41	26	43
The average travel time of additional trip (min)	Osaka 1970	32	53	42	43	26	-
	1980	33	48	41	39	21	-
	Kyoto 1970	26	38	41	41	24	-
	1980	19	40	38	37	18	-
The average working hours in office(hrs)	Osaka 1970	6.70	5.95	6.15	7.38	7.73	9.52
	1980	5.90	6.42	5.97	7.37	7.37	9.63
	Kyoto 1970	5.48	5.98	6.15	7.40	8.23	9.55
	1980	6.47	7.08	5.87	7.53	7.92	9.63
The average sojourn duration(min)	Osaka 1970	110	172	144	164	115	-
	1980	130	146	164	146	91	-
	Kyoto 1970	121	163	158	163	110	-
	1980	114	135	181	140	89	-
The average starting time of the first trip(hour:min)	Osaka 1970	9:01	8:34	7:49	7:50	7:58	7:51
	1980	9:17	8:35	7:45	7:49	7:47	7:50
	Kyoto 1970	9:18	9:05	7:52	7:54	7:52	7:56
	1980	9:15	8:26	7:45	7:49	7:50	7:53
The average ending time of the last trip(hour:min)	Osaka 1970	21:15	19:54	19:24	20:21	20:42	19:05
	1980	21:15	19:54	19:31	19:58	19:54	19:14
	Kyoto 1970	20:02	19:52	19:21	20:06	20:46	18:56
	1980	21:01	19:49	19:49	19:38	19:53	19:00

* Individuals who made walk trips are excluded from this tabulation because trip duration data are not available in the 1970 file. This is the main reason for the difference in sample size between Table 3 and Table 4.

Those who did not make trips and those who were not employed are excluded from the sample. In addition, the individuals are screened to include only those who made at least one trip to either Osaka City or Kyoto City. The subsamples of these two groups of individuals are referred to as the Osaka sample and the Kyoto sample.

It is important to note that the additional activity is not limited to nonwork activities (e.g., shopping, personal business, and social recreation) but includes business-related activities, such as attending a conference held outside the office base. In this respect the data sets of this study are quite different from those used in the previous analyses of two-stop chains, in which the additional activity is limited to nonwork activities. The main reason for including business-related activities is that business trips account for a large percentage of total trips made by workers in the metropolitan areas of Japan. For example, business trips in Osaka in 1980 accounted for 31.8 percent of total person trips, while work trips accounted for 26.9 percent, nonwork trips 9.9 percent, and home trips 31.4 percent. Also, workers tend to integrate business activities into their daily travel. For example, in Osaka in 1980, 21.4 percent of business trips were made in home-business-work or work-business-home chains, 57.0 percent in office-based chains, and 21.6 percent in separate home-based chains. It is likely that business-related activities influence the formation of workers' trip chains significantly.

Temporal and Spatial Stability of the Distribution of Path Types

The distributions of path types involving two stops are presented in Table 3 for the Osaka and Kyoto subsamples in 1970 and 1980, respectively. The table indicates that the additional activity was pursued most frequently in office-based chains during work and in single-chain paths after work. The percentage of before-work engagement and multichain after-work engagement is less than 30 percent and shows a decrease in 1980.

Office-based engagement increased substantially between 1970 to 1980 in both Osaka and Kyoto. At the same time the fraction of after-work engagement in Osaka decreased from 57.5 percent in 1970 to 40.9 percent in 1980; that in Kyoto decreased from 61.8 percent to 49.9 percent during the same period. This tabulation thus offers ample evidence that the distribution of path types is not stable over time. Comparing the Osaka and Kyoto samples, it can be found that the percentage of multichain paths is smaller in Osaka than in Kyoto, for both before-work and after-work engagements. These differences are presumably due to longer commuting durations in the much larger Osaka metropolitan area.

The increase in during-work engagement may be a factor leading to the increase in trip rates between 1970 and 1980 shown in Table 2. Figure 3 shows the distribution of the mode used for the additional trip by path type (the rail mode includes bus). Notable is the considerable decrease in walk trips and the increase in trips by "other" modes found in multichain paths, both before work and after work. This reflects the surge in bicycle use that took place in Japan between 1970 and 1980.

In case of during-work engagement, the differences between 1970 and 1980 are due to the increase in walk trips and the decrease in car trips. Although statistical data do not offer a direct explanation for this sharp increase, it presumably is due to the increase in nonwork activities outside the office base, especially eating meals, during the lunch break. Trip generation for shopping, eating meals, and social recreation increased by 25 percent in 1980, and the frequency of restaurants as a destination land use type more than doubled.

Temporal Stability in Time Use

Several indicators of time use in two-stop paths are presented in Table 4. They include the averages of commuting travel time, hours spent in the office, travel time for the additional activity, duration of the additional activity, departure time of the first trip, and arrival time of the last trip. Figure 4 presents the average daily time use pattern by path type for the Osaka sample in 1970 and 1980,

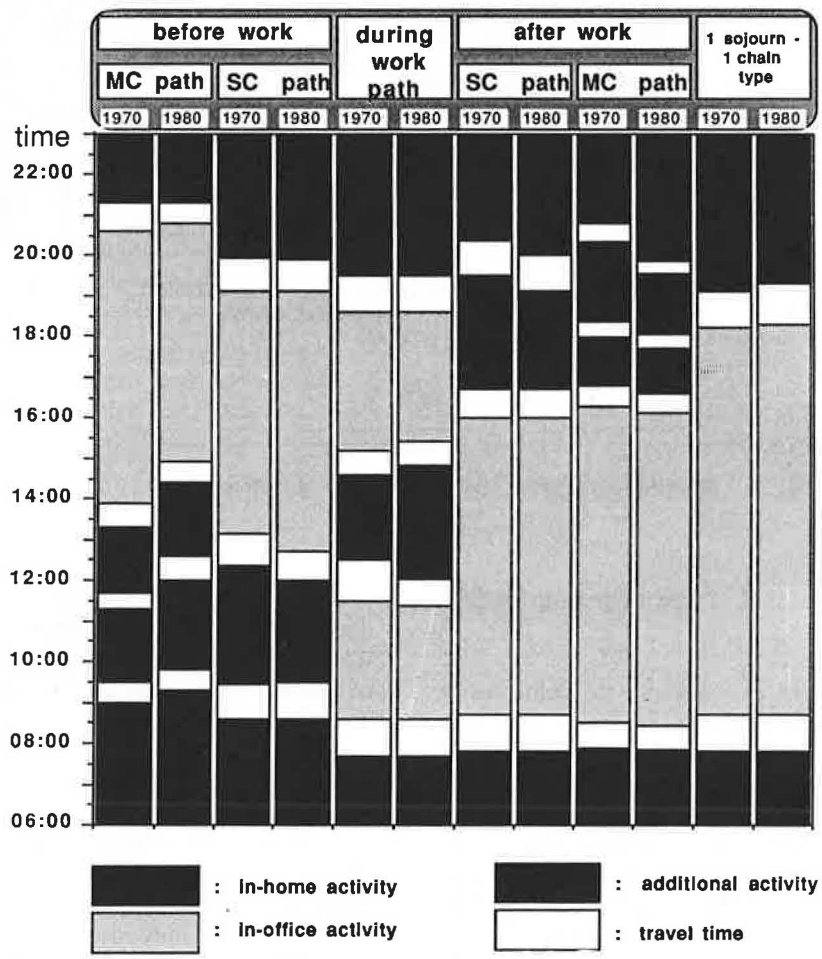


FIGURE 4 Time use patterns by path type (Osaka).

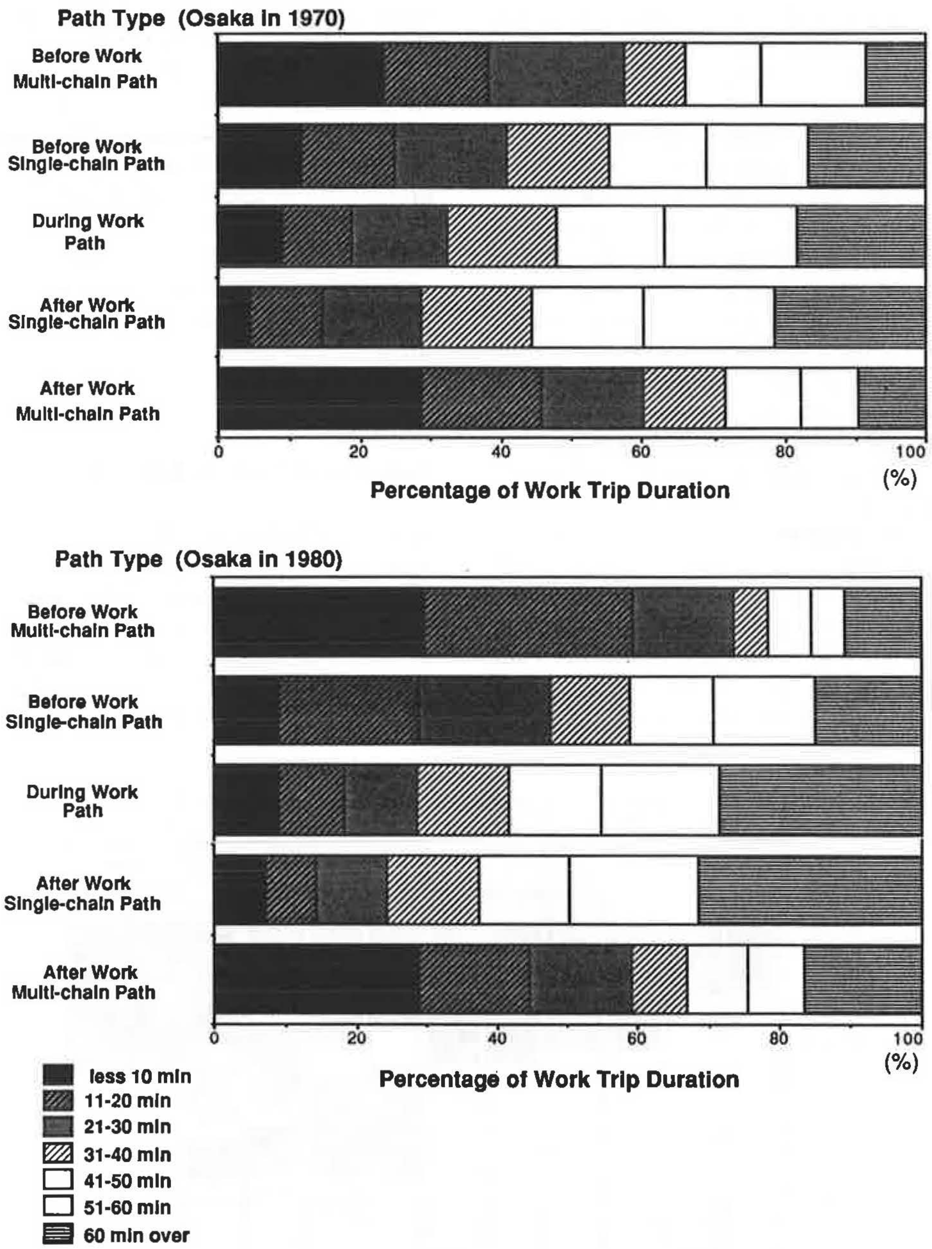


FIGURE 5 Distribution of work trip durations by path type.

respectively. Observations that can be made from Table 4 and Figure 4 include:

1. The time use pattern of workers with before-work activity engagement is different from those of workers with during-work and after-work engagements. The difference is in the starting time of the first trip and the ending time of the last trip. The average starting time for before-work engagement is approximately one hour later than in the other path types (including home-work-home, one-stop paths), where the average starting times are between 7:45 and 8:00 a.m. This result is consistent with those obtained by Damm (*14*) and Adiv (*27*), and demonstrates the impact of the work schedule upon trip chaining.

2. The durations of the additional activity show marked decreases in 1980 in after-work engagement and before-work, single-chain engagement. On the other hand, that of during-work engagement became significantly larger. Together with the increase in the frequency of during-work engagement, the result suggests that workers' engagement in out-of-office activities in office-based chains intensified in 1980.

3. Most of the additional activities pursued after work can be classified into nonwork activities, such as shopping, social recreation, and personal business. It is quite remarkable that those individuals who pursued activities after work tended to return home earlier in 1980 than in 1970. Similar results were obtained in a temporal stability analy-

sis using data from the Southeast Michigan area in 1965 and 1980 (*32*).

In summary, the statistical analysis thus far has indicated that there were considerable changes in the travel and out-of-home activity patterns of workers between 1970 and 1980 in the Osaka and Kyoto areas.

ANALYSIS OF THE RELATIONSHIPS DERIVED FROM THE LINEAR-CITY MODEL

Figure 5 presents the distribution of path types in the Osaka data for 1970 and 1980 against commuting trip duration. With a χ^2 value of 226.4 (df = 24) in 1970 and 301.9 (df = 24) in 1980, the contingency tables underlying the figure indicate an extremely significant association between the two factors. Commuting travel time corresponds to L/v and, given v, it is proportional to L. An inspection of the figure indicates that the likelihood of multichain paths decreases as commuting trip duration increases. The tendency is consistent with the theoretical result that $\partial DU/\partial L < 0$.

The distribution of path types is shown by the travel mode used for the additional activity in Figure 6 for the Osaka data set. Contingency tables with high values of χ^2 in both 1970 and 1980 again indicate a significant association between the two factors. This is due in part to the correlation between the mode choice and the home-to-work distance. The results show that after-work, single-chain paths are significantly overrepresented among the users of public transit, who often engage in long-distance commuting by rail. The dense commercial development surrounding railway stations in Japan is definitely a contributing factor. On the other hand, it appears that the walk mode in 1970 is associated with multichain paths, with the additional activity pursued in a separate home-based chain, both before and after work. The same tendency is not found in 1980; the walk mode is associated with the during-work activity engagement.

As an indicator of the propensity to link trips, consider the ratio of multichain paths to all paths (*MC* ratio). This

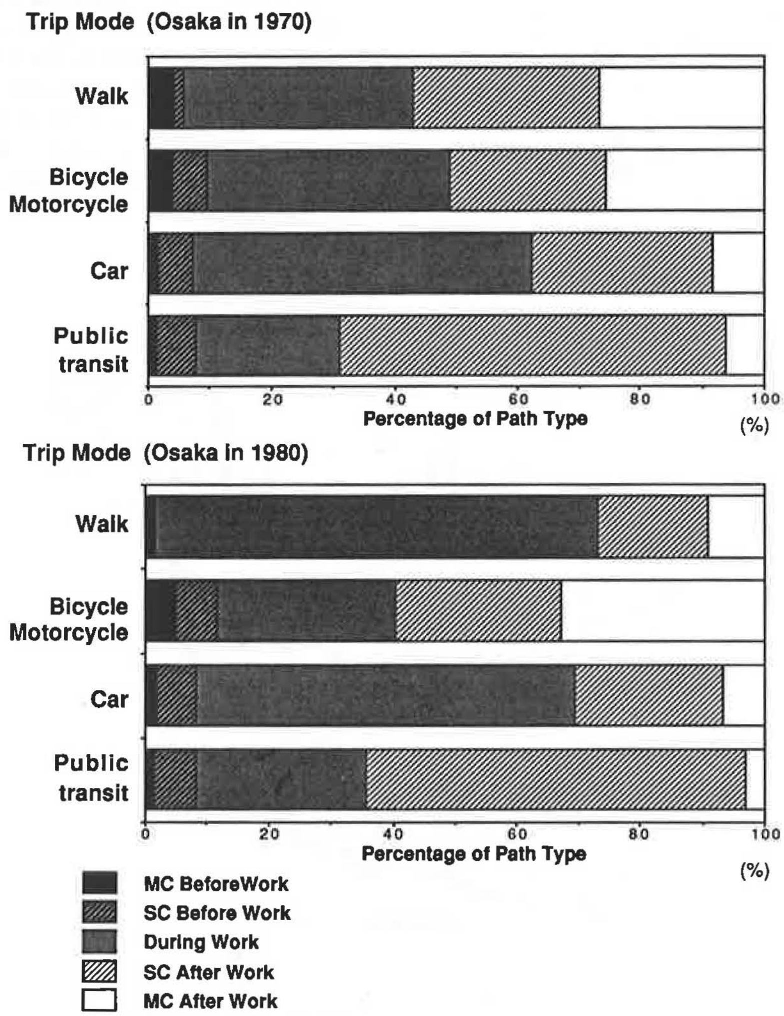


FIGURE 6 Distribution of path types by trip mode for additional activity.

TABLE 5 RELATIONSHIP BETWEEN TRIP MODE AND *MC* RATIO BY HOME LOCATION (OSAKA)

Year	Home location	Mode	No. of paths MC	No. of paths Total	MC ratio	Mean work trip duration
1970	North-Osaka	Rail	55	768	7.3(%)	46.7 (min)
		Car	22	163	13.5 *	37.4 *
	East -Osaka	Rail	53	640	8.3	47.1
		Car	14	150	9.3 *	41.8 *
	South East Osaka	Rail	19	244	7.8	52.7
		Car	5	51	9.8 *	50.3 *
	South-Osaka	Rail	37	454	8.1	52.6
		Car	14	78	17.9 *	46.2 *
	Osaka city	Rail	200	1084	18.5	33.1
		Car	82	387	21.2 *	22.5 *
1980	North-Osaka	Rail	32	788	4.1(%)	47.1 (min)
		Car	7	166	4.2 *	47.0 *
	East -Osaka	Rail	33	698	4.7	47.8 *
		Car	15	140	10.7 *	49.7
	South East Osaka	Rail	12	228	5.3	55.1
		Car	3	44	6.8 *	53.5 *
	South-Osaka	Rail	23	454	5.1 *	55.2 *
		Car	3	71	4.2	58.1
	Osaka city	Rail	80	775	10.3	34.3
		Car	37	280	13.2 *	24.9 *

ratio is shown in Table 5 by trip mode (private car vs. rail, including bus). To control for commuting distance, the tabulation is prepared by residential zones as shown in Figure 7. Except for one case, the faster mode with a smaller average trip duration has a larger *MC* ratio. The table thus offers empirical evidence in support of the theoretical result that $\partial DU/\partial v > 0$.

The tendencies just discussed are found in general in both the 1970 and 1980 data. The results indicate that time and space constraints do influence workers' trip chaining behavior in important ways. The results also point to the important effects the commuting distance and speed of travel have upon trip chaining.

CONCLUSION

A model of trip chaining behavior under time-space constraints was developed in this study; the likelihood of combining stops into multistop chains is related to parameters that characterize an urban area. The analysis was concerned with workers' trip chaining behavior where a nonwork activity is introduced to the basic home-work-home chain. The theoretical results indicated that the likelihood of pursuing an additional activity in a separate home-based chain will increase with the speed of travel, and that of multistop chains will increase with commuting distance, travel cost, and the density of opportunities.

The empirical results using data sets from two areas and two points in time support the relation between commuting distance and trip chaining. In addition the results

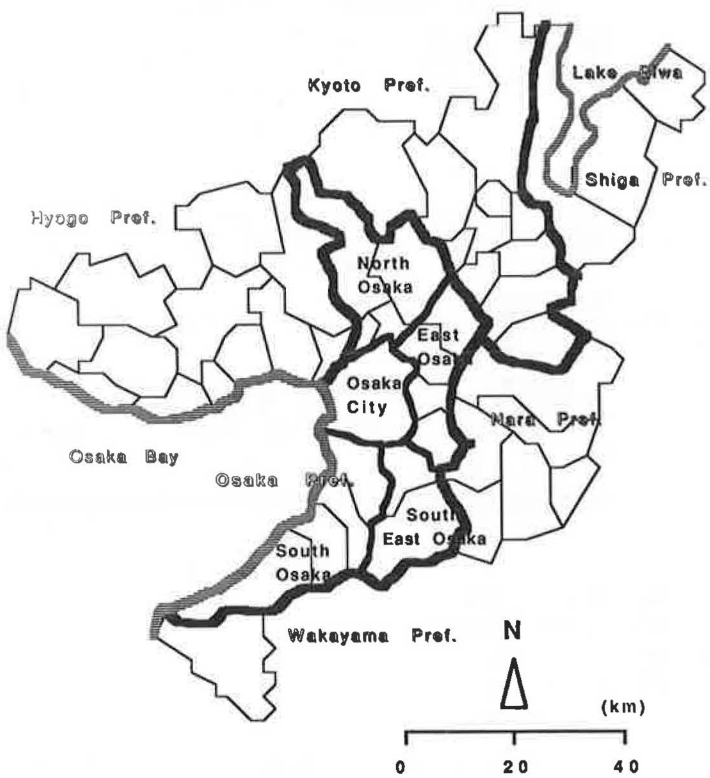


FIGURE 7 Study area and zoning.

support the theoretically obtained relation that faster travel speed encourages activity participation in one-stop chains. The empirical analyses also showed that trip chaining does not remain stable over time despite the fact that trip rates are very stable. In particular, a drastic increase in office-based, nonwork activity engagement on foot was found between 1970 and 1980.

ACKNOWLEDGMENT

Funding for this study was provided through a grant from the National Science Foundation to the University of California, Davis.

REFERENCES

1. T. Hagerstrand. What About People in Regional Science? *Papers of Regional Science Association*, 24, 1970, pp. 7–21.
2. P. M. Jones, M. C. Dix, M. I. Clarke, and I. G. Heggie. *Understanding Travel Behavior*. Gower, Aldershot, England, 1983.
3. K. Kondo and R. Kitamura. Time-Space Constraints and the Formation of Trip Chains. *Regional Science and Urban Economics*, 17, 1987, pp. 49–65.
4. K. G. Goulias and R. Kitamura. On the Spatial Distribution of Linked Trips in a Uniform Two-Dimensional Space. Unpublished paper. University of California, Department of Civil Engineering, Davis, 1988.
5. T. Sasaki. Estimation of Person Trip Patterns Through Markov Chains. In *Traffic Flow and Transportation* (G. F. Newell, ed.), American Elsevier, New York, 1972, pp. 119–130.
6. K. Kondo. Estimation of Person Trip Patterns and Modal Split. In *Transportation and Traffic Theory* (D. Buckley, ed.), Elsevier, New York, 1974, pp. 715–742.
7. K. Nishii and T. Sasaki. An Analysis of Travel Patterns and Modal Split in a Business Trip Chaining. *Proceedings of World Conference on Transport Research*, forthcoming (1986).
8. S. R. Lerman. The Use of Disaggregate Choice Models in Semi-Markov Process Models of Trip Chaining Behavior. *Transportation Science*, 13, 1979, pp. 273–291.
9. R. Kitamura and T. N. Lam. A Time Dependent Markov Renewal Model of Trip Chaining. In *Proceedings of the Eighth International Symposium on Transportation and Traffic Theory* (V. F. Hurdle, E. Hauer, and G. N. Steuart, eds.), University of Toronto Press, Toronto, Canada, 1983, pp. 376–402.
10. M. E. O'Kelly. A Model of the Demand for Retail Facilities, Incorporating Multistop, Multipurpose Trips. *Geographical Analysis*, 13, 1981, pp. 134–148.
11. A. Borgers and H. Timmermans. A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas. *Geographical Analysis*, 18, 1986, pp. 115–128.
12. S. Hanson. Urban-Travel Linkages: A Review. In *Behavioral Travel Modelling* (D. Hensher and P. Stopher, eds.), Croom Helm, London, 1979, pp. 81–100.
13. F. S. Chapin. *Human Activity Patterns in the City*. John Wiley and Sons, New York, 1974.
14. D. Damm. Parameters of Activity Behavior for Use in Travel Analysis. *Transportation Research*, Vol. 16A, No. 2, 1982, pp. 135–148.
15. W. W. Recker, M. G. McNally, and G. S. Root. A Model of Complex Travel Behavior: Part I—Theoretical Development, and Part II—An Operational Model. *Transportation Research*, 20A, 1986, pp. 307–318 and 319–330.
16. J. Horowitz. A Utility Maximizing Model of the Demand for Multidestination Non-Work Travel. *Transportation Research*, 14B, 1980, pp. 369–386.
17. T. Adler and M. Ben-Akiva. A Theoretical and Empirical Model of Trip Chaining Behavior. *Transportation Research*, 13B, 1979, pp. 243–257.
18. T. van der Hoorn. Experiments with an Activity-Based Travel Model. *Transportation*, 12, 1983, pp. 61–77.
19. T. van der Hoorn. Development of an Activity Model Using a One-Week Activity-Diary Data Base. In *Recent Advances in Travel Demand Analysis* (S. Carpenter and P. Jones, eds.), Gower, Aldershot, England, 1983, pp. 335–349.
20. R. Kitamura and M. Kermanshah. Identifying Time and History Dependencies of Activity Choice. *Transportation Research Record 944*, TRB, National Research Council, Washington, D.C., 1983, pp. 22–30.
21. R. Kitamura and M. Kermanshah. A Sequential Model of Interdependent Activity and Destination Choice. *Transportation Research Record 987*, TRB, National Research Council, Washington, D.C., 1984, pp. 81–89.
22. F. Southworth. Multi-Destination, Multi-Purpose Trip Chaining and Its Implications for Locational Accessibility: A Simulation Approach. *Papers of the Regional Science Association*, 57, 1985, pp. 107–123.
23. D. Swiderski. A Model for Simulating Spatially and Temporally Coordinated Activity Sequences on the Basis of Mental Maps. In *Recent Advances in Travel Demand Analysis* (S. Carpenter and P. Jones, eds.), Gower, Aldershot, England, 1983, pp. 313–334.
24. R. Kitamura, T. N. Lam, and N. Hioki. Stochastic Simulation of the Temporal and Spatial Interactions in Urban Travel. In *1983 Winter Simulation Conference Proceedings*, Vol. 1. Institute of Electrical and Electronics Engineers, New York, 1983, pp. 289–299.
25. C. V. Oster. The Second Role of the Work Trip: Visiting Non-work Destination. *Transportation Research Record 728*, TRB, National Research Council, Washington, D.C., 1979, pp. 79–81.
26. S. Hanson. The Importance of Multipurpose Journey to Work in Urban Travel Behavior. *Transportation*, 9, 1980, pp. 229–248.
27. A. Adiv. The Structure of the Work-Trip Based on Analysis of Trip Diaries in the San Francisco Bay Area. In *Recent Advances in Travel Demand Analysis* (S. Carpenter and P. Jones, eds.), Gower, Aldershot, England, 1983, pp. 117–136.
28. T. F. Golob. A Descriptive Analysis of Trip-Chaining Behaviour in the Netherlands. In *Proceedings, Seminar H, PTRC Summer Annual Meeting*, PTRC Education and Research Service, London, 1984.
29. J. C. Thill. Demand in Space and Multipurpose Shopping: A Theoretical Approach. *Geographical Analysis*, 17, 1985, pp. 114–129.
30. R. Kitamura. Recent Developments in Trip Chaining Analysis. Paper Presented at the PTRC Summer Annual Meeting, Bath, England, 1987.
31. L. D. Burns. *Transportation, Temporal, and Spatial Components of Accessibility*. D.C. Heath, Lexington, Mass., 1979.
32. L. P. Kostyniuk and R. Kitamura. Temporal Stability of Urban Travel Patterns. *Transportation Policy and Transportation Decision Making*, 2, 1984, pp. 481–500.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Formulation of Trip Generation Models Using Panel Data

RYUICHI KITAMURA

This study addresses the question of whether conventional models based on cross-sectional data alone account for trip generation adequately. Alternative model formulations are examined using a panel data set to determine whether elements associated with past time points should be considered in trip generation analysis and, if so, which elements should be introduced into the model. The analysis shows that estimated model coefficients and *t*-statistics differ substantially depending on model specification, and that allowing for serial correlation and incorporating a lagged dependent variable both significantly improve the model's fit. The results indicate that serial correlation should be incorporated whenever feasible for efficient estimation and improved fit and that it is safer to ignore state dependence (dependence of observed trip generation on that from a previous time point) and incorporate serial correlation, than to ignore serial correlation and incorporate state dependence. This significance of serial correlation, which presumably is due to omitted variables that are longitudinally correlated, suggests that important determinants of trip generation lie outside the set of variables that have traditionally been considered in travel behavior analysis.

Suppose a panel data set consisting of observations of the same behavioral units at several points in time is available for developing models of travel behavior. The modeling effort in this case will be more complex than with a cross-sectional data set because of the wider range of variables available—that is, measurements of behavior and contributing factors from previous observation periods. In addition, one must select a model formulation that best represents the behavior from several classes of models that can be developed with panel data.

Choosing a particular model formulation quite often implies adopting a particular behavioral hypothesis. One focus of model specification effort using panel data is how to represent dynamic characteristics of observed behavior. Two major hypotheses are (1) that behavior is static (or contemporaneous): namely, behavior at time t can be explained by factors observed at time t; and (2) that behavior is dynamic and cannot be fully explained by contemporaneously observed factors. If the former is the case, cross-sectional data suffice. If the latter is the case one may hypothesize further: (2a) behavior can be described adequately by introducing explanatory variables from past observation points into the model; (2b) past measurements of the behavior itself must be introduced into the model; and (2c) serial correlation in unobserved elements needs to be taken into account.

As is often the case in modeling travel behavior, no theory exists that dictates a priori which model formulation should be used. The selection of a model formulation is empirical, at least until such a theory can be constructed. In this context, the implications of recent observations of dynamic aspects of travel behavior, such as response lags, habit persistence, and history dependence, are important (*1–4*). These results suggest that dynamic models represent travel behavior more accurately and meaningfully. Then how best can the dynamism be captured in a quantitative model of travel behavior?

This study addresses the question of whether incorporating elements observed in past time points is worthy in trip generation analysis and, if so, which elements should be introduced into the model. The models examined include conventional cross-sectional model, lagged dependent variable model (which is formulated using a dependent variable from a previous time point as an explanatory variable), lagged independent variable model, and cross-sectional and lagged dependent variable models with serially correlated errors. The objective of this study is to examine these alternative model formulations and to determine the most relevant representation of trip generation behavior.

The weekly total numbers of person trips, travel time expenditures, and number of social-recreation trips, as reported by household members, are used as the dependent variables of this study. Models for these variables should ideally be formulated within a simultaneous equations framework together with car ownership models. This is not done in this study. Rather, this analysis is conducted to determine the model formulation to be used in such a modeling effort, which is to follow the present study as an extension of the model system reported elsewhere (*3, 5, 6*). This study considers only linear models.

The rest of this paper is organized as follows. The alternative model formulations considered are described in the next section. Then the data set and estimation procedures used in the study are described briefly, and the results of a preliminary analysis of trip generation using a cross-sectional model are presented. The relative importance of lagged independent variables and lagged dependent variables is next evaluated. Following this, the com-

Department of Civil Engineering, University of California at Davis, Davis, Calif. 95616.

peting hypotheses of serial correlation and state dependence are examined. The latter hypothesis underlies the use of the lagged dependent variable model. A summary and conclusions of the study are last.

ALTERNATIVE MODEL FORMULATIONS

This section describes the alternative model formulations considered in this study. The *cross-sectional model* has the form

$$Y(t) = \beta' X(t) + \varepsilon(t) \quad t = 1, \ldots, T \tag{1}$$

where $Y(t)$ is a scalar measure of observed behavior at time t, β is a vector of coefficients, $X(t)$ is a vector of explanatory variables at time t that are uncorrelated with $\varepsilon(t)$, $\varepsilon(t)$ is a random error term, and T is the number of equi-spaced observation points. In this formulation, observed behavior at time t is related to contributing factors also observed at t. Obviously, this model can be estimated using cross-sectional data. The assumption underlying the formulation is that Y changes immediately in response to a change in X and that the value of Y does not depend on the past history of either X or Y itself.

Alternatively, the behavioral relation of interest may be formulated as

$$\begin{aligned} Y(t) &= \beta' X(t) + \tau' X(t-1) + \cdots \\ &\quad + \mu' X(t-s) + \varepsilon(t) \\ t &= s+1, \ldots, T \end{aligned} \tag{2}$$

where β, τ, and μ are coefficient vectors and s is a positive integer. In this formulation behavior is assumed to be history-dependent; that is, behavior at time t is assumed to be a function of contributing variables measured at $t - s$ through t. A change in X is not fully reflected in Y immediately; Y will change gradually over time and reaches a new equilibrium after s periods, provided that no further change takes place in X. Accordingly, the model depicts behavior with response lags. This model is called the *lagged independent variables model* in this study.

Another possible formulation is

$$Y(t) = \theta Y(t-1) + \beta' X(t) + \varepsilon(t) \quad t = 2, \ldots, T \tag{3}$$

where θ is a scalar coefficient and $|\theta| < 1$. This model is known as the *lagged dependent variable model* and is often referred to as the *dynamic model.*

The assumption that $Y(t)$ depends on $Y(t - 1)$ implies that $Y(t)$ is a function of the entire history of X and ε. Applying the above recursive relation repeatedly, the following is obtained;

$$\begin{aligned} Y(t) &= \beta' X(t) + \theta\beta' X(t-1) + \theta^2\beta' X(t-2) + \cdots \\ &\quad + \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2\varepsilon(t-2) + \cdots \end{aligned} \tag{4}$$

Accordingly, the lagged dependent variable model exhibits behavior similar to that of the lagged independent variables model. The former model is advantageous when only a short panel (with small T) is available; if the specification of Equation 3 is correct the model can be estimated with observations from as few as two time points, even though $Y(t)$ depends on the entire history. The model of Equation 4 is a special case of the model of Equation 2 with $s = -\infty$, except for the difference in the error structure.

The lagged dependent and lagged independent models are both special cases of the general formulation

$$\begin{aligned} Y(t) &- \theta_1 Y(t-1) - \theta_2 Y(t-2) - \cdots \\ &= \beta' X(t) + \beta_1' X(t-1) + \beta_2' X(t-2) + \cdots + \varepsilon(t) \end{aligned} \tag{5}$$

which is referred to as a transfer function model in time-series analysis (*7*). Discussions of this class of models can be found, for example, in Griliches (*8*).

The error term of each model may be assumed to be serially correlated; that is,

$$\varepsilon(t) = r\varepsilon(t-1) + u(t) \tag{6}$$

where r is the coefficient of serial correlation and $u(t)$ is independent of ε. This is a convenient scheme to adopt when unobserved variables do not change their values frequently and are longitudinally correlated.

DATA SET AND INITIAL EXPLORATION

The observations from waves 1, 3, 5, and 7 of the Dutch National Mobility Panel survey are used in this study (these waves will be referred to hereafter as periods 1, 2, 3, and 4, respectively). The four surveys were conducted in the spring of 1984, 1985, 1986, and 1987, respectively. In each survey a weekly travel diary was collected from each member of the sample household who was at least 12 years old. The aim, format, and characteristics of this panel data set are discussed elsewhere (*9–12*).

The dependent variables of the analysis in this study are weekly measures of trip generation and travel time expenditure. These variables are used without any transformation because a weekly total is likely to lead to residuals with desirable properties because of the law of large numbers. Previous model estimation results using the same data set and weekly measurements agree with this expectation.

Heteroscedasticity is accounted for by applying the weight,

$$W_i = a(1 + |\hat{Y}_i|)^b \tag{7}$$

where i refers to the household. Parameters a and b are obtained by regressing the squared residual from ordinary least squares (OLS) estimation on the OLS prediction. This weight has been selected after examining several other formulations. Models that assume serial correlation are estimated by regressing $Y(t) - rY(t - 1)$ on $X(t) - rX(t - 1)$ and iterating on r. A convergence is obtained quickly, usually within five iterations, with the relative change in r being reduced to less than 1 percent of its absolute value.

A set of explanatory variables was selected after an examination of a wide range of variables, including the number of household members (≥12 years old) who kept a travel diary, number of workers, number of drivers, number of children by age group, household income, car ownership, education, a rough indicator of transit service levels, and their transformations. The variables that appear in the models presented in this paper are defined in Table 1.

Table 2 shows weighted least squares (WLS) estimates of the coefficients of a cross-sectional total person trip generation model obtained from wave-5 observations ($t = 3$). The sample of this model estimation consists of those 829 households that participated in waves 1, 3, and 5 surveys. Note that some variables are multiplied by the number of diary keepers (NDIARIES) so that their coefficients represent the difference in trip generation per diary keeper.

The most dominant factor is the number of diary keepers in the household (NDIARIES). The estimated coefficient value indicates that a diary keeper on average will add approximately 20 trips to weekly household trip generation. The number of added trips will on average be 3.71 trips fewer if this person is not a driver and 1.85 trips more if the person is employed. The three variables representing the number of children by age group are all significant and indicate the increased household travel needs resulting from the presence of children.

The education level of a household is defined in this study as that of the person who has the highest education in the household. The two dummy variables used in the model to represent household education are both significant and show that the number of reported person trips

TABLE 2 CROSS-SECTIONAL MODEL OF WEEKLY HOUSEHOLD PERSON TRIP GENERATION (WLS, $t = 3$)

R^2	.586	
F	89.2	
df	(13,815)	
	β	t
NDIARIES	20.150	9.63
NONDRIVERS	-3.714	-3.46
NWORKERS	1.853	1.86
CHLD0-6	3.537	3.48
CHLD7-11	3.588	3.35
CHLD12-17	5.796	3.71
HIEDUC*NDIARIES	1.275	2.08
LOEDUC*NDIARIES	-2.646	-3.37
√INCOME*NDIARIES	-.026	-.12
ONECAR*NDIARIES	.861	.94
TWOCAR*NDIARIES	-.074	-.06
BOV-HIGH*NDIARIES	-.940	-1.32
FOR-LOW*NDIARIES	-1.125	-1.06
Constant	3.006	
N	829	

TABLE 1 DEFINITION OF THE VARIABLES USED IN THE MODELS OF THIS STUDY

Variable	Definition
NTRIPS	Total no. of person trips reported by household members during the survey week
NTRNTRIPS	Total no. of transit trip segments reported by household members during the survey week
NDIARIES	No. of household members (≥ 12 years old) who filled out diaries
NDRIVERS	No. of licensed drivers in household
NONDRIVERS	No. of individuals of at least 12 years old who did not hold a driver's license
NWORKERS	No. of employed household members
CHLD0-6	No. of children between 0 and 6 years old
CHLD7-11	No. of children between 7 and 11 years old
CHLD12-17	No. of children between 12 and 17 years old
√INCOME	Square-root of annual household income divided by 100
LOEDUC	1 if the household member with the highest education had completed only primary school; 0 otherwise
HIEDUC	If at least one person in household has a college degree; 0 otherwise
ONECAR	1 if the household owns one car; 0 otherwise
TWOCAR	1 if the household owns two or more cars; 0 otherwise
BOV-HIGH	1 if the household resides in a metropolitan area with highly developed transit systems; 0 otherwise
BOS-MEDIUM	1 if the household resides in a community which is served by rail; 0 otherwise
FOR-LOW	1 if the household resides in a community which is not served by rail; 0 otherwise

increases with household education. The result, however, may be a reflection of reporting errors as well as of genuine variations in trip making. An earlier analysis of panel attrition using the same data set (*11*) suggests that households with lower education tended to underreport their trip making and tended to drop out of the panel. It is likely that the coefficients of these education variables represent in part the magnitude of underreporting.

Household income (√ INCOME) and transit service level indicators (BOV-HIGH and FOR-LOW) are all insignificant. Car ownership can be considered to be more closely associated with household mobility than is income itself, because it reflects long-term mobility choice. The estimation result shown in Table 2 indicates that car ownership, which is represented by two dummy variables (ONECAR, TWOCAR), is insignificant. This result is consistent with the earlier indication (*5*) that household person trip generation in the Dutch Panel data set is statistically independent of household car ownership. This is due in part to the inclusion of walking and bicycle trips. Nonetheless, its contradiction with the commonly held belief that car ownership is a major determinant of person trip generation is noteworthy.

In summary, the analysis of this section indicates that trip generation as depicted by this cross-sectional model is primarily a function of demographic characteristics of the household. Car ownership and transit service levels have little influence on the total number of person trips generated by household members over the period of a week.

SIGNIFICANCE OF LAGGED VARIABLES

The model development effort has considered a wide range variables and their transformations that extend beyond those typically used in trip generation models. The resulting model has a relatively high R^2 value of 0.586. Nevertheless, a question remains whether the set of variables used in the model adequately captures systematic variations in trip generation, and whether all pertinent elements are included in the model. This section is concerned with the possibility that observed variables from past time points, or lagged variables, influence trip generation.

If response lags, habit persistence, history dependence, and other longitudinal relationships significantly affect trip generation behavior, cross-sectional models such as the one discussed in the previous section would be a misspecification of the behavioral relationship. Response lags can be represented by introducing lagged independent variables—that is, explanatory variables observed at previous time points—into the model. Habit persistence and history dependence may be captured by the use of lagged dependent variables. The relative contributions of lagged dependent variables and lagged independent variables to the model's explanatory power are the focus of this section.

The main questions of this section are whether lagged variables significantly contribute to the model's explanatory power and, if so, whether a lagged dependent variable contributes more than do lagged independent variables. Suppose that the effect of a lagged dependent variable is due to the effect of independent variables from previous time points reflected in the lagged dependent variable. A special case of this is shown by Equation 4 derived from Equation 3 through Koych transformation. If this in fact is the case, then the fit of a model with lagged independent variables, but without a lagged dependent variable, should be approximately equal to that of a lagged dependent variable model. This question is examined by estimating both lagged independent and lagged dependent variable models.

The same set of 829 households that are in the data files for periods 1 through 3 is used to estimate models for period 3 trip generation ($Y(3)$), and the set of 645 households that are in the data files of periods 1 through 4 is used to estimate models of $Y(4)$. The results of WLS estimation are summarized in Table 3.

The same set of explanatory variables as in the cross-sectional model of Table 2, but from a previous period, is sequentially added to the cross-sectional model. The R^2 values shown in the table clearly indicate that the introduction of the X vector from each previous period does contribute to the model's goodness-of-fit. It is also clear, however, that the marginal improvement in R^2 is small and tends to decline as more sets are added. Based on the pseudo F-statistics presented in the table, the lagged independent variable vectors are mostly not significant (at $\alpha = 0.05$); the only exception is $X(t - 1)$ for $t = 3$.

The goodness-of-fit of lagged dependent variable models is presented at the bottom of Table 3. As noted earlier, the models are estimated through the WLS method iterated on r. The difference in the goodness-of-fit is substantial between the models with lagged X vectors and those with a lagged Y. For example, the lagged independent variable model estimated for the third period ($t = 3$) with $X(1)$ through $X(3)$ has an R^2 of 0.605, while the lagged dependent variable model with $X(3)$ and $Y(2)$ has an R^2 of 0.770. Similar differences between the two sets of models can be observed for $t = 4$. A further inspection of the estimation results indicated that the lagged Y is the most significant explanatory variable of the lagged dependent variable models estimated here. The results of this analysis offer an indication that the improvement in the model's fit realized by a lagged dependent variable is not merely a reflection of the effects of independent variables from previous periods.

This point is further examined by repeating the analysis for weekly total travel time expenditure and number of social-recreation trips using data from the third period ($t = 3$). The same estimation procedure was applied to these dependent variables. The results summarized in Table 4 agree with those obtained for total person trip generation; the X vectors from past periods are generally insignificant, and lagged dependent variables account for much larger portions of the total variations than do lagged independent variables.

Importantly, the difference in the goodness-of-fit between the lagged independent model and lagged dependent model is more pronounced with social trip generation. The

TABLE 3 IMPROVEMENT IN THE MODEL'S FIT DUE TO LAGGED INDEPENDENT AND LAGGED DEPENDENT VARIABLES IN TOTAL TRIP GENERATION MODELS

Explanatory Variables		t = 3	t = 4
X(t)	R^2	.586	.613
	F	89.2	77.4
	df	(13,815)	(13,631)
X(t), X(t-1)	R^2	.599	.622
	F	46.4	39.4
	df	(26,802)	(26,618)
Significance of X(t-1)	F	2.09	1.14
	df	(13,802)	(13,618)
X(t), X(t-1), X(t-2)	R^2	.605	.633
	F	31.1	26.9
	df	(39,789)	(39,605)
Significance of X(t-2)	F	.83	1.24
	df	(13,789)	(13,605)
X(t), X(t-1), X(t-2) X(t-3)	R^2		.643
	F		20.7
	df		(52,592)
Significance of X(t-3)	F		1.28
	df		(13,592)
Y(t-1), X(t)	R^2	.770	.776
	F	191.93	160.42
	df	(14,814)	(14,630)
Significance of Y(t-1)	t	-6.5	-9.8
Serial Correlation	r	.780	.805
N		829	645

Note: Each X vector contains the 13 explanatory variables shown in Table 2. All models are estimated using the weighted least squares (WLS) method with the weight shown in Eqn (7). To enable comparison across the models, the R^2 and F values shown have been recomputed using predictions by the respective models. The F values are therefore approximate and do not exactly follow F distributions. The prediction by the lagged dependent variable models was obtained using the residual from the previous period.

lagged independent variable model with $X(1)$ through $X(3)$ but without a lagged Y has a small R^2 of 0.299. The lagged dependent variable model with $X(3)$ and $Y(2)$, on the other hand, has an R^2 of 0.487. The difference in the R^2 values is extremely large, with the lagged dependent variable model accounting for 63 percent more variation than does the lagged independent variable model. The corresponding percentages are only 27 percent for total trip generation ($t = 3$) and 33 percent for travel time expenditure. The relative explanatory power of lagged dependent variables varies from variable to variable.

The social-recreation trip generation models, for which the effect of the lagged dependent variable is most substantial, have much smaller R^2's. If this observation can be generalized, the relative contribution of a lagged dependent variable increases when the other explanatory variables do not capture much of the variation in the dependent variable. In other words, as the fraction of unaccounted variations increases, so does the contribution of the lagged dependent variable, presumably because this variable reflects idiosyncrasy.

The findings thus far obtained offer valuable insight, but are subject to a limitation that arises from the nature of the variables used. Some of the independent variables in this analysis are longitudinally multicollinear because their values change only infrequently. Consequently, vectors of lagged independent variables do not offer much additional information and therefore do not substantially improve the model's fit. Unfortunately, it is not possible to determine, on the basis of the available data, whether this in fact is the major reason for the significance of the lagged dependent variables and the insignificance of the lagged independent variables, as evidenced in Tables 3 and 4. For this reason, lagged independent variable models are excluded from consideration in the analyses presented in the rest of this paper.

TABLE 4 IMPROVEMENT IN THE MODEL'S FIT DUE TO LAGGED INDEPENDENT AND LAGGED DEPENDENT VARIABLES IN TOTAL TRIP TIME EXPENDITURE AND SOCIAL-RECREATIONAL TRIP GENERATION MODELS

Explanatory Variables		Travel Time	Social Recreation
X(t)	R^2	.424	.283
	F	50.1	36.0
	df	(12,816)	(9,819)
X(t), X(t-1)	R^2	.441	.293
	F	26.6	18.7
	df	(24,804)	(18,810)
Significance of X(t-1)	F	2.15	1.28
	df	(12,804)	(9,810)
X(t), X(t-1), X(t-2)	R^2	.448	.299
	F	17.9	12.7
	df	(36,792)	(27,801)
Significance of X(t-2)	F	.73	.74
	df	(12,792)	(9,801)
Y(t-1), X(t)	R^2	.598	.487
	F	93.63	78.032
	df	(13,815)	(10,818)
Significance of Y(t-1)	t	-6.8	-9.3
Serial Correlation	r	.676	.697
N		829	829

STATE DEPENDENCE VERSUS SERIAL CORRELATION

The analysis of this section focuses on the possibility that the significance of the lagged dependent variables shown in the previous section is due to omitted variables that are longitudinally correlated. A competing hypothesis is that observed travel behavior is dependent on the past behavior itself and a lagged dependent variable is its true determinant. If the effect of longitudinally correlated, omitted variables can be represented by serially correlated errors, then these two hypotheses constitute a panel analysis version of serial correlation versus "state dependence."

To address the question, the following two additional models for total person trip generation are estimated: cross-sectional model with serially correlated errors (Model 2) and lagged dependent variable model assuming no serial correlation (Model 3). These models are summarized in Table 5 together with the cross-sectional model with no serial correlation (Model 1), which was given in Table 2, and the lagged dependent variable model with serially correlated errors (Model 4) whose summary statistics were given in Table 3.

It is evident from Table 5 that coefficient estimates and estimated *t*-statistics vary substantially depending on the model specification. For example, the coefficient of number of workers (NWORKERS) is positive in both cross-sectional and lagged dependent variable models without serial correlation (Models 1 and 3), while it is negative in the models where serial correlation is assumed (Models 2 and 4). The coefficient of number of children between 7 and 11 years old (CHLD7-11) is not significant, while the coefficient of income ($\sqrt{\text{INCOME}}$) is positive and more significant in Models 2 and 4 with serial correlation. The significance of the household education variables (HIEDUC and LOEDUC) differs drastically between the two cross-sectional models (Models 1 and 2).

In general the coefficients of the lagged dependent variable model without serial correlation (Model 3) are smaller in their absolute values than those of the cross-sectional model without serial correlation (Model 1). Such regularity, however, cannot be found between the two lagged dependent variable models (Models 2 and 4). Many of the explanatory variables in the cross-sectional model (Model 1) become insignificant when a lagged independent variable is introduced (Model 3), but this is not observed between Models 2 and 4 that assume serial correlation.

Most importantly, the fraction of variance explained is virtually the same between the two models with serial correlation with R^2's of 0.763 (Model 2) and 0.770 (Model 4). The coefficient estimates also show similarity. The R^2 of Model 3, on the other hand, shows that the lagged dependent variable by itself does not account for as much variation as does the serially correlated error (Model 2). The results suggest that trip generation is not state dependent, and the apparent significance of the lagged dependent variable is due to serially correlated errors, which in turn are presumably due to longitudinally correlated, omitted variables.

TABLE 5 COMPARISON OF DIFFERENT MODEL FORMULATIONS FOR TOTAL NUMBER OF PERSON TRIPS PER WEEK

	Cross-Sectional				Lagged Dependent			
	Model 1 (r = 0)		Model 2 (r > 0)		Model 3 (r = 0)		Model 4 (r > 0)	
R^2	.586		.763		.718		.770	
F	89.2		189.0		150.7		182.05	
df	(13,815)		(14,814)		(14,814)		(15,813)	
	ß	t	ß	t	ß	t	ß	t
r			.695				.780	
NTRIPS(t-1)					.525	19.41	-.171	-6.49
NDIARIES(t)	20.150	9.63	23.470	11.68	9.137	5.08	24.536	11.95
NONDRIVERS(t)	-3.714	-3.46	-2.316	-4.04	-2.574	-2.87	-1.823	-3.25
NWORKERS(t)	1.853	1.86	-2.428	-2.22	.254	.31	-3.139	-2.87
CHLD0-6(t)	3.537	3.48	1.961	1.19	2.342	2.77	1.529	.85
CHLD7-11(t)	3.588	3.35	.219	.13	1.696	1.91	-.152	-.08
CHLD12-17(t)	5.796	3.71	2.216	1.31	4.601	3.62	1.883	1.09
HIEDUC*NDIARIES(t)	1.275	2.08	-.471	-.59	.159	.32	-.685	-.81
LOEDUC*NDIARIES(t)	-2.646	-3.37	-.172	-.20	-.854	-1.30	.086	.10
√INCOME*NDIARIES(t)	-.026	-.12	.277	1.59	.132	.77	.271	1.65
ONECAR*NDIARIES(t)	.861	.94	.534	.72	.617	.81	.791	1.12
TWOCAR*NDIARIES(t)	-.074	-.06	-.690	-1.23	-.197	-.21	-.462	-.88
GOV-HIGH*NDIARIES(t)	-.940	-1.32	-2.408	-1.73	-.155	-.27	-3.470	-2.14
FOR-LOW*NDIARIES(t)	-1.125	-1.06	-1.868	-.90	-.736	-.84	-2.335	-.96
Constant	3.006		.436		.583		1.314	
N	829		829		829		829	

Note: In order to facilitate comparison across the models, the R^2 and F values have been reevaluated by computing predicted values of Y using WLS coefficient estimates. Residuals from t = 2 are used to compute predicted values of the models with serially correlated errors.

This point is further examined by estimating another model of the form,

$$Y(t) = \theta Y(t-1) + \beta_0' X(t) + \beta_1' X(t-1) + u(t) \quad (8)$$

which is shown below to be equivalent to Model 2 when certain conditions are met. Suppose Model 2 of Table 5 represents the true relationship; that is,

$$Y(t) = \beta' X(t) + \varepsilon(t) \qquad \varepsilon(t) = r\varepsilon(t-1) + u(t) \quad (9)$$

where r is the coefficient of serial correlation. Then

$$Y(t) = rY(t-1) + \beta' X(t) - r\beta' X(t-1) + u(t) \quad (10)$$

Therefore, the coefficient vectors of Equations 8 and 10 are related as

$$r = \theta, \quad \beta_0 = \beta, \quad \text{and} \quad \beta_1 = -r\beta_0 = -\theta\beta_0 \quad (11a\text{–}c)$$

if Model 2, or Equation 9, holds true. Hendry and Mizon (*13*) suggest the use of Equation 11c to discriminate between serial correlation and state dependence (also see *14*). The results of estimation are summarized in Table 6.

The coefficient of the lagged dependent variable (θ) of Equation 8 is 0.649, which is very close to the estimated serial correlation coefficient (r) of 0.695 (Model 2). The R^2 values vary only slightly between the two models (0.763 and 0.771). The results again favor the conclusion that trip generation is not state dependent and that the apparent state dependency is due to serially correlated errors. Nevertheless, it is important to note that some elements of β are substantially different from the corresponding elements of β_0, and not all elements of $\beta_1 + r\beta$ and $\beta_1 + \theta\beta_0$ diminish, which should be the case if Equations 11a through 11c hold. In addition, the lagged dependent variable of Model 4 is significant, as the t-statistic shown in the table indicates.

In conclusion, the analysis of this section has shown that the model's explanatory power increases substantially by incorporating serially correlated errors. It also has been shown that a lagged dependent variable also adds to the model's explanatory power but to a much lesser extent. The analysis has indicated that trip generation is both serially correlated and state dependent, but that the former plays a more dominant role.

A model with serial correlation should be used whenever possible. This can be concluded from the discrepancy in

TABLE 6 COMPARISON OF THE COEFFICIENT ESTIMATES OF TWO FORMULATIONS OF CROSS-SECTIONAL MODEL WITH SERIALLY CORRELATED ERRORS

	Model 2	Eqn (8)					
R^2	.763	.771					
F	189.0	106.5					
d.f.	(14,814)	(26,802)					
	β	β_0	β_1	$-\beta\mu$	$\beta_1+\mu\beta$	$-\theta\beta_0$	$\beta_1+\theta\beta_0$
NDIARIES	23.470	22.50	-17.00	-16.31	-.69	-15.24	-1.76
NONDRIVERS	-2.316	-1.47	.24	1.61	-1.37	1.50	-1.27
NWORKERS	-2.428	-3.45	4.44	1.69	2.75	1.58	2.87
CHLD0-6	1.961	-.70	2.66	-1.36	4.02	-1.27	3.93
CHLD7-11	.219	-3.47	4.33	-.15	4.49	-.14	4.48
CHLD12-17	2.216	-.79	1.58	-1.54	3.12	-1.44	3.01
HIEDUC*NDIARIES	-.471	-1.10	1.09	.33	.77	.31	.79
LOEDUC*NDIARIES	-.172	-.43	-.28	.12	-.40	.11	-.39
√INCOME*NDIARIES	.277	.25	-.24	-.19	-.05	-.18	-.06
ONECAR*NDIARIES	.534	2.03	-.61	-.37	-.24	-.35	-.26
TWOCAR*NDIARIES	-.690	1.26	-.89	.48	-1.37	.45	-1.34
BOV-HIGH*NDIARIES	-2.408	-4.61	4.18	1.67	2.51	1.56	2.62
FOR-LOW*NDIARIES	-1.868	-1.03	-.16	1.30	-1.46	1.21	-1.37
NTRIPS(t-1) (θ)		.649					
r	.695						

the coefficient estimates as well as difference in the R^2 values between Model 1 and Model 2 of Table 5. The estimation results of this study also indicate that state dependence, although statistically significant, may be ignored as far as the model's fit is concerned. The similarity of the coefficient estimates between Model 2 and Model 4, but the distinct coefficient estimates of Model 3, and the R^2 values of these models, all suggest that it is safer to ignore state dependence and incorporate serial correlation, than to ignore serial correlation and incorporate state dependence.

The significant serial correlation is due presumably to omitted variables that are longitudinally correlated. Considering that an extensive set of demographic and socioeconomic variables was considered in the model development of this study, it may be concluded that the omitted variables are not likely among the variables typically collected in transportation surveys.

CONCLUSION

The trip generation analysis of this study has shown that allowing for serial correlation and incorporating a lagged dependent variable significantly improve the model's fit. The improvement that a lagged dependent variable offers cannot be attributed to the effects of independent variables from previous time points. Serial correlation, however, contributes more substantially to the model's fit than does a lagged dependent variable. Trip generation is both serially correlated and state dependent, but the former plays a more dominant role.

The exploration of alternative model specifications has shown that estimated model coefficients and *t*-statistics differ substantially depending on model specification. Serial correlation should be incorporated whenever feasible for efficient estimation and improved fit. The results of this study indicate that it is safer to ignore state dependence and incorporate serial correlation than to ignore serial correlation and incorporate state dependence.

The significant serial correlation is due presumably to omitted variables that are longitudinally correlated. Considering that an extensive set of variables was considered in the model development of this study, it is not likely that the omitted variables are among those typically used in travel demand models and conventionally contained in travel survey data. Some determinants of travel behavior appear to lie outside the set of variables that have traditionally been considered in travel behavior analysis.

Finally, the contribution of serially correlated errors and lagged dependent variables to the predictive performance of a model has not yet been examined. The work on this subject is in its early stages, and only limited results have been reported (*15*). The many waves of weekly trip diary data now available from the Dutch National Mobility Panel data set offer a unique opportunity for further examination of this important issue.

ACKNOWLEDGMENTS

This research was performed while the author was at Bureau Goudappel Coffeng (BGC), Deventer, the Netherlands, in July through September 1987. For the support he received, the author is grateful to the BGC staff—in particular, to Jacqueline Visser, who prepared the data file used in this study. The valuable comments by Toon van der Hoorn on an earlier version of this paper are also gratefully acknowledged. The funding was provided by Rijkswaterstaat, Dienst Verkeerskunde, Dutch Ministry of Transport and Public Works.

REFERENCES

1. P. B. Goodwin. A Panel Analysis of Changes in Car Ownership and Bus Use. *Traffic Engineering and Control*, 27, October 1986, pp. 519–525.
2. P. B. Goodwin. *Family Changes and Public Transport Use 1984–1987: A Dynamic Analysis Using Panel Data.* Report prepared for the Project Bureau of the Netherlands Ministry of Transport, Bureau Goudappel Coffeng, Deventer, the Netherlands, 1987.
3. R. Kitamura. Determinants of Household Car Ownership and Utilization: A Panel Analysis. Presented at the International Conference on Travel Behavior, Aix-en-Provence, France, October 1987.
4. R. Kitamura and T. van der Hoorn. Regularity and Irreversibility of Weekly Travel Behavior. *Transportation*, 14, 1987, pp. 227–251.
5. R. Kitamura. A Panel Analysis of Household Car Ownership and Mobility. *Proceedings of the Japan Society of Civil Engineers*, No. 383/IV-7, 1987, pp. 13–27.
6. R. Kitamura. A Dynamic Model System of Household Car Ownership Trip Generation and Modal Split: Model Development and Simulation Experiment. In *Proceedings of the 14th Conference of the Australian Road Research Board, Part 3*, ARRB, Vermont South, Victoria, Australia, 1989, pp. 96–111.
7. G. E. Box and G. M. Jenkins. *Time Series Analysis, Forecasting, and Control.* Holden-Day, San Francisco, 1976.
8. Z. Griliches. Distributed Lags: A Survey. *Econometrica*, 35, 1967, pp. 16–47.
9. J. M. Golob, L. J. M. Schreurs, and J. G. Smit. The Design and Policy Applications of a Panel for Studying Changes in Mobility Over Time. In *Behavioural Research for Transport Policy*, VNU Press, Utrecht, the Netherlands, 1986, pp. 81–95.
10. T. F. Golob and H. Meurs. Biases in Response Over Time in a Seven-Day Travel Diary. *Transportation*, 13, 1986, pp. 163–181.
11. R. Kitamura and P. H. L. Bovy. Analysis of Attrition Biases and Trip Reporting Errors for Panel Data. *Transportation Research A*, 21A, 1987, pp. 287–302.
12. H. Meurs, P. van de Mede, J. Visser, and L. van Wissen. *Analysis of Panel Data.* Bureau Goudappel Coffeng, Deventer, the Netherlands, 1987.
13. D. F. Hendry and G. E. Mizon. Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England. *Economic Journal*, 88, 1978, pp. 549–563.
14. G. S. Maddala. Recent Developments in the Econometrics of Panel Data Analysis. *Transportation Research A*, 21A, 1987, pp. 303–326.
15. T. Van der Hoorn and R. Kitamura. Evaluation of the Predictive Accuracy of Cross-sectional and Dynamic Trip Generation Models Using Panel Data. Paper presented at the 66th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1987.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Experimental Investigation of Route and Departure Time Choice Dynamics of Urban Commuters

HANI S. MAHMASSANI AND DOUGLAS G. STEPHAN

The interaction between departure time and route switching in response to experienced congestion in a traffic commuting system is investigated using an experimental approach in which actual commuters interact over a period of seven weeks in a simulated commuting system. This experiment extends previous observational work along this line by incorporating the route choice dimension, which is essential to an understanding of user choice dynamics in congested urban systems. Another important feature of this experiment is the consideration of two groups of commuters with different information availability levels interacting in the same system, thereby allowing the analysis of the effect of information on user behavior and performance. The implications of the findings for the design of information-related strategies for the relief of congestion in urban networks are also discussed.

Urban transportation planning methods have generally been concerned with a static description of link flows on a network, where flows and travel times are assumed to be invariant over the duration of the peak period, under presumed user equilibrium conditions. The primary dimension of choice thus available to tripmakers is route choice, as users are assumed to select routes in such a way as to achieve equilibrium conditions whereby no user can improve his or her travel time by unilaterally switching routes. An important dimension of choice that is missing in this framework is trip timing, which determines the time-varying flow patterns on the network. The variation of flows and travel times within the peak period has been documented in several studies (*1*, *2*). Furthermore, the interaction between trip timing and route selection is an important element in the analysis of time-varying flows in congested urban networks.

It is thus increasingly apparent that existing urban transportation planning methods are not adequate to address certain classes of operational problems where the time-dependent behavior of flows and the daily variation of these flows in response to congestion patterns and management strategies are of critical importance. The need for operational models that can support the design and evaluation of strategies for coping with congestion in urban networks provides an important motivation for the study of the dynamics of user decisions. This need is timely in light of the mounting national awareness of the rapidly deteriorating traffic conditions in urban areas, as well as the potential of strategies that rely on recent developments in information and communication technologies, including in-vehicle navigation and guidance systems. In addition, the preponderance of certain classes of practical problems, such as major freeway and arterial reconstruction activities that necessitate the loss over significant periods of critical capacity on vital arteries, calls for methods to predict the impact of such reductions and of possible mitigation strategies.

In the past five years, there have been several theoretical contributions to the solution of time-varying departure patterns that satisfy dynamic user equilibrium conditions; the vast majority of these have considered the highly idealized situation of a single origin-destination pair connected by a single route with an intervening bottleneck (*3*–*7*). Extensions that consider both departure time and route choices in similarly idealized situations have also been developed (*8*, *9*). All these formulations were, however, limited to the solution of time-varying departure patterns (on one or more routes) at some "final" equilibrium state; they were not particularly well developed on the behavioral side, for the sake of analytical tractability and for lack of an observational basis for more elaborate behavioral theories. Perhaps the most advanced on the user decisions side is the work of Ben-Akiva et al. (*9*, *10*), where a well-structured decision framework, though still lacking an observational basis, is articulated in the context of a stochastic user equilibrium formulation that can be solved numerically.

There have been several models of the static decisions made by individual tripmakers, particularly of trip timing, where the problem is viewed as one of selecting a discrete alternative among the possible set of discretized departure time alternatives (*11*–*14*). Relatively few studies have addressed the route choice decision from a behavioral perspective; some of these are reviewed in the recent contribution by Ben-Akiva et al. in an intercity context (*15*). An interesting application to bicycle path selection in an urban context, using a stated preference experimental

Department of Civil Engineering, University of Texas at Austin, Austin, TX 78712

approach, has also been presented (*16*). The joint departure time and route choice selection problem has recently been addressed by Abu-Eisheh and Mannering (*17*), using a small sample of university employees whose actual decisions are assumed to reflect prevailing equilibrium conditions in a relatively lightly congested system.

None of the preceding studies have addressed the day-to-day mechanisms by which individual tripmakers may adjust their decisions, in the short term, in response to perceived and experienced congestion in the traffic system. Simple learning rules by which users adjust their anticipated travel times on alternate routes, given their prior experiences with the system, were proposed by Horowitz in a theoretical contribution to the stability of stochastic user equilibrium in an idealized two-link network (*18*). Alternative rules for departure time adjustment were explored by Mahmassani and Chang (*19*) in the context of a simulation framework to study the dynamics of the interaction between commuter decisions and congestion in traffic systems. An essential element in the further development of knowledge and methods in this area, however, is the observation of actual user behavior as it interacts with the traffic system's performance. Recognizing that the complexity of this dynamic interaction seriously reduces the ability of conventional survey methods to obtain the data that would be necessary for such investigation within practical resource limitations, an experimental approach was recently proposed by Mahmassani, Chang, and Herman (*20*). It consists of observing, over several weeks, real commuters interacting through a computer-simulated traffic system. Decisions supplied daily by the participants, acting independently, form the input to a traffic simulation model that yields the respective consequences (arrival times) of these decisions. Feedback from the simulated results is supplied to the commuters who provide decisions on the next day, and so on. This nonprohibitive alternative to large-scale field experiments allows insights into the overall system's dynamic properties and provides a basis for theoretical development and model specification and testing.

In previous papers, two such experiments have been reported (*20–22*). Both consist of an identical commuting corridor with a single highly congested route, where users have a choice of departure time only. The two experiments differ in the information available to users on the system's prior performance. In the first case, all commuters are provided only with limited information on their own arrival time on the previous day. In the second, they are provided with complete information on the previous day's performance, in the form of arrival times for all possible departure times. In addition to the overall evolution of the system, including spatial and temporal convergence behavior under the two informational scenarios, exploratory analyses of the dynamics of individual user behavior have been reported in the previous papers (*20–22*). Formal econometric models of the departure time adjustment mechanisms and associated trip time prediction rules operating at the individual level have also been developed on the basis of these two experiments (*23–26*), and further model development and hypothesis testing is still ongoing with regard to the complex dynamic processes governing commuter decisions.

Having developed and refined the experimental methodology in a single route context, it is natural to extend its application to more complicated situations encountered in actual commuting, particularly those allowing users a choice of route in addition to that of departure time. A third experiment has therefore been conducted for this latter situation, thereby allowing the investigation of the dynamics of user choices along these two interacting dimensions in congested commuting systems. In this third experiment, participants are divided into two information availability groups (limited vs. complete) operating side by side in the same system, providing the opportunity to examine the effect of information when the latter is provided to only a fraction of the total population. The results of this experiment are presented in this paper, with regard to the system's overall evolution and convergence patterns, the dynamics of user decisions, particularly switching frequency and the interaction between route choice and departure time switching decisions, as well as the effect of information on these processes. The analysis presented in this paper is primarily exploratory, aimed at providing an overview and general insights into the questions of interest and suggesting hypotheses for more extensive formal model building and testing.

In the next section, the experimental procedure is described, focusing primarily on those details that are different from the previous two experiments and that have a direct bearing on the interpretation of the results. This is followed by a discussion of the system's overall evolution and convergence patterns. The effect of information is then examined, followed by a comparative assessment of system performance by route from the perspective of user equilibrium principles. After that, the main focus is on the analysis of the interaction of route choice and departure time switching decisions. Concluding comments are presented in the final section, along with questions that will be addressed in subsequent modeling work.

COMMUTING CONTEXT AND EXPERIMENTAL PROCEDURE

The procedure followed in this experiment is essentially similar to that followed in the previous two, which are described in detail elsewhere (*20–22*). The commuting context now consists of two parallel roadway facilities, each nine miles long (with access limited to a finite number of entry points), used by adjoining residents in their home-to-work morning commute to a common destination, such as a city's central business district (CBD) or major suburban industrial park. The two facilities serving the corridor consist of (1) Route 1, a four-lane highway (two lanes in each direction) with a speed limit of 50 mph, and (2) Route 2, a two-lane arterial street (one lane in each direction) with a speed limit of 40 mph. Both facilities are assumed to be of uniform width and geometry throughout their nine-mile length. The two facilities do not intersect, thus no crossing over from one to the other is possible in the same trip. As before, the corridor is divided into nine identical sectors, one mile in length, with the work destination located at the end of sector 9 and sector 1 being the farthest outbound. The commuters' residences are located in sectors 1 through 5 only, with no trips generated from the remaining sectors. For simplicity, it is assumed that the access time from a given residence to either facility is the same.

To achieve a meaningful level of congestion in the system, the total number of trips generated in each of the five residential sectors is doubled from the 400 used in the two previous experiments to 800 tripmakers per sector, or a total of 4,000 trips during the peak period. It should be noted, however, that while two routes exist in this system compared to only one route in the two previous experiments, the overall level of congestion in terms of the ratio of overall volume to the total number of available lanes has increased by one-third. The number of participating commuters was accordingly increased to 200, or double the 100 used in each of the previous experiments. The participants were assigned equally and randomly to the five residential sectors, and the decisions of each participant were treated as those of 20 identical tripmakers for traffic simulation purposes.

All participants were staff members at the University of Texas at Austin. To avoid possible bias due to professional knowledge, none of the participants were recruited from the Civil Engineering Department or the Community and Regional Planning Program. Furthermore, individuals who had participated in either of the two previous experiments were excluded from this one to control for initial bias and learning effects. Participants were selected and instructions given so as to avoid cooperative behavior in the experimental responses, and to prevent access to uncontrolled sources of information. Careful monitoring and subsequent analysis do not suggest any violations of the intended individual, noncooperative character of the decision process.

In keeping with the assumptions of the previous experiments, participants were instructed that they needed to be at work by 8:00 A.M., with no late arrival tolerated at the workplace, which is quite similar to their actual working conditions. The purpose of imposing the identical work start time and no-lateness conditions in this and earlier experiments is to limit nonessential complication in interpreting the results, and to keep the number of participants at a manageable level while allowing a meaningful level of interaction to develop in the traffic system.

In the first two experiments, all users were assumed to have the same level of information availability; thus, no comparative advantage existed in this regard among participants in a given experiment. For instance, in the first experiment, one's own experience in the commuting system was the only source of information available. Thus each participant was provided with the time at which she or he arrived at the work destination given that individual's chosen departure time on the previous day (*20, 21*). In the second experiment, information was also available from exogenous sources, and participants were supplied with arrival times corresponding to an array of possible departure times (ranging from 7:00 to 7:50 A.M. in five-minute increments) for their residential sector, presented in the form; "If you had left at 7:00 you would have arrived at 7:15" (*22*). The new feature in the third experiment is that both limited (or myopic) and full information availability levels are included. Specifically, in each sector, participants are assigned equally to one or the other information availability group and, accordingly, receive either myopic or full information throughout the experiment. This allows the effect of information availability on performance to be studied and compared for unequally informed individuals operating simultaneously in the same commuting system.

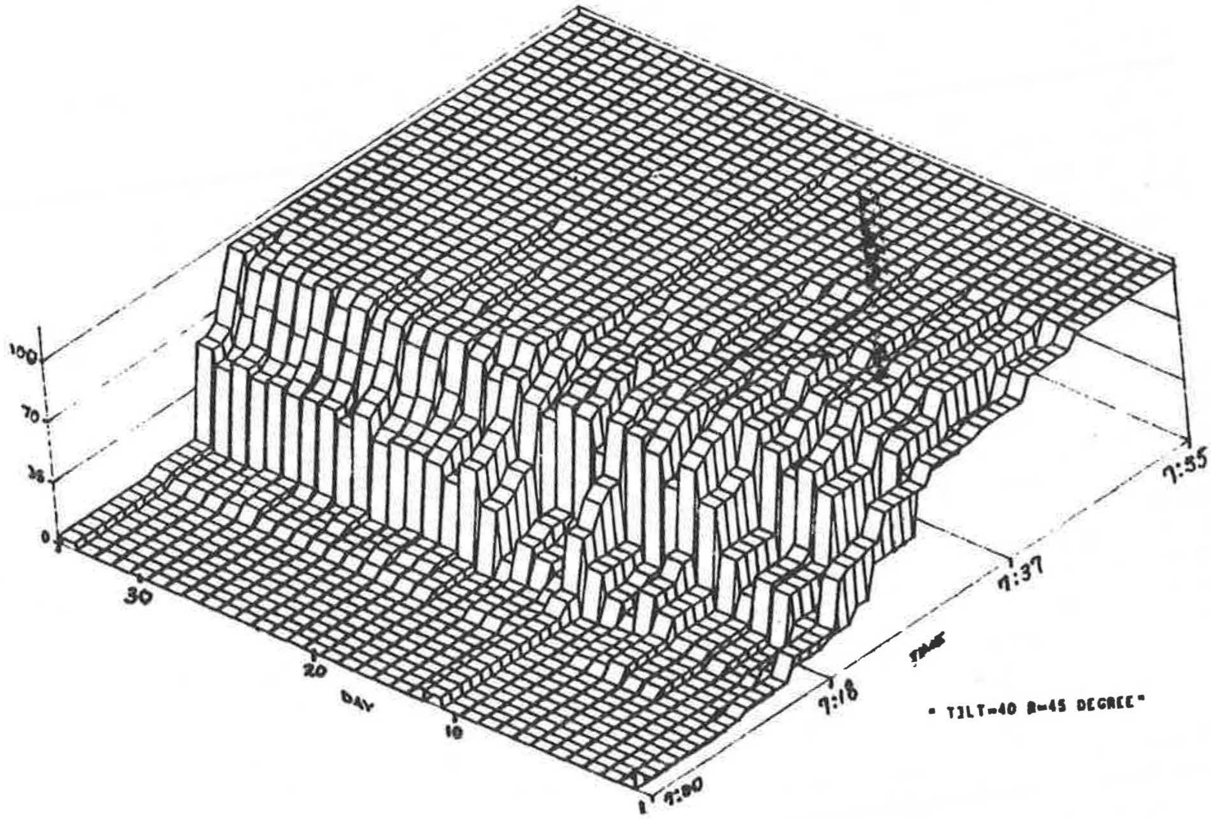
Finally, for completeness, the experimental procedure, beyond the selection of participants and their assignment to the various residential sectors and information groups, is summarized as follows:

1. Supply each participant i, $i = 1, \ldots, 200$, with initial information and instructions.
2. On day t, all participants supply their departure time and route decisions DT_{it} and RT_{it}; these are aggregated by sector into time-dependent departure functions for each route, $D_{krt}(T)$, where T is the time of day, $k = 1, \ldots, 5$ is the subscript for the sector of origin, and $r = 1, 2$ is that for the route.
3. The departure functions are input to a special-purpose, macroparticle traffic simulation model (or MPSM) (*27*), which yields the respective arrival times AT_{it}, travel times TT_{it}, and other pertinent traffic performance measures. Note that because the two routes do not overlap, each can be simulated independently from the other, given the respective input functions on day t.
4. If the maximum experiment duration is reached or steady state is achieved, stop. Otherwise, set $t = t + 1$, supply each participant with information on actual performance on the preceding day, according to her or his information availability group, and go to step 2 for updated departure time and route decisions from the participants.

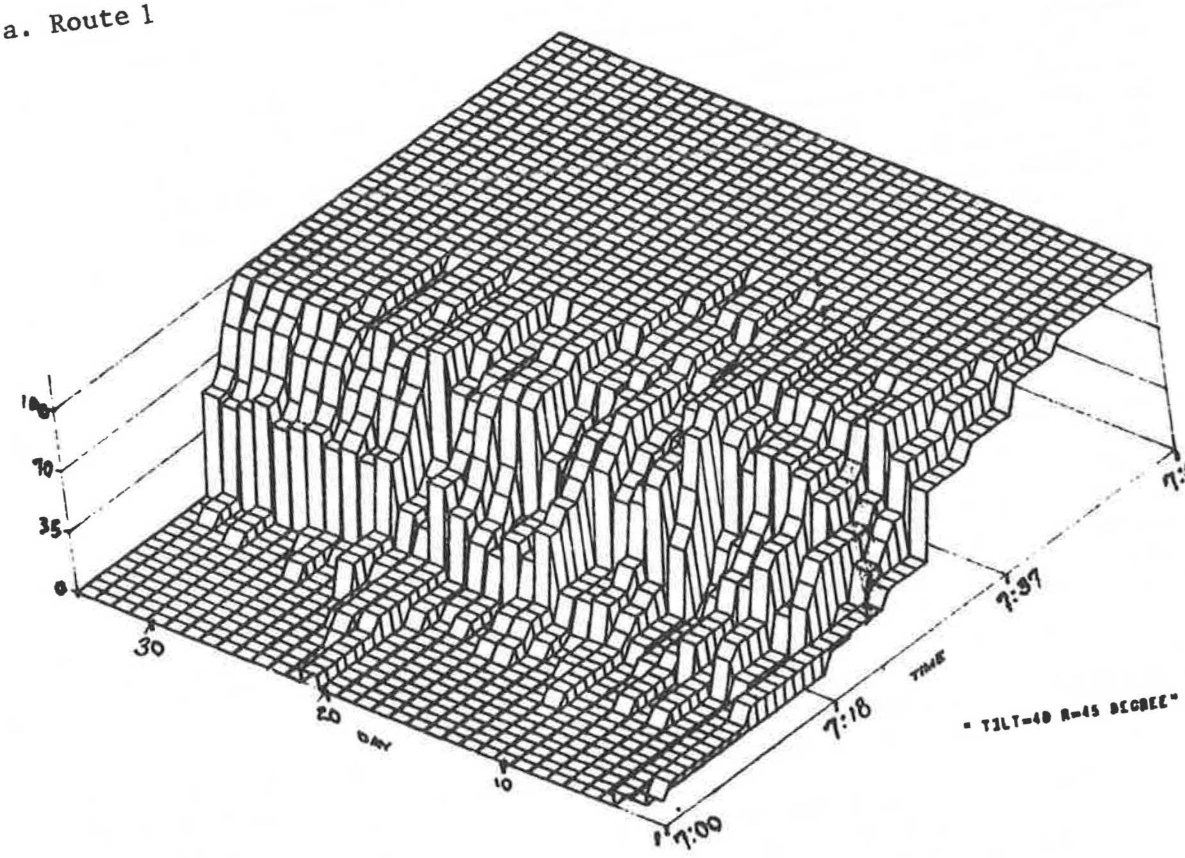
The experimental results are presented next.

SYSTEM EVOLUTION AND CONVERGENCE

Convergence in this system is achieved when the commuters are no longer adjusting their route or departure time choices and the resulting departure time distributions for each route and sector remain the same from one day to the next. In both previous experiments, convergence to a steady state was achieved as of day 20 for the limited information case, and day 29 for the full information case (*20, 22*). Complete convergence, however, was still not reached in this third experiment after seven weeks (34 days) of administering the experiment. Figures 1 (a, b) and 2 (a, b) show the evolution of the daily cumulative departure time distributions for routes 1 and 2, for sectors 1 and

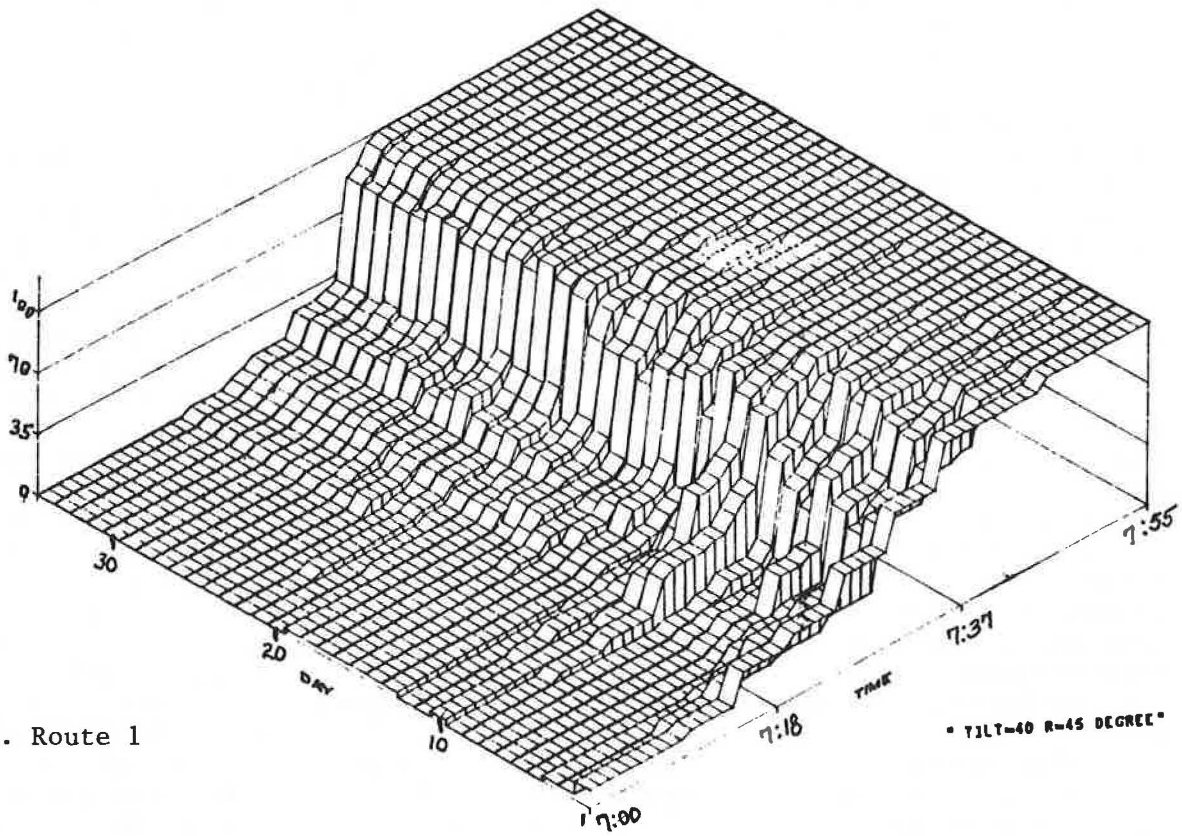


1.a. Route 1

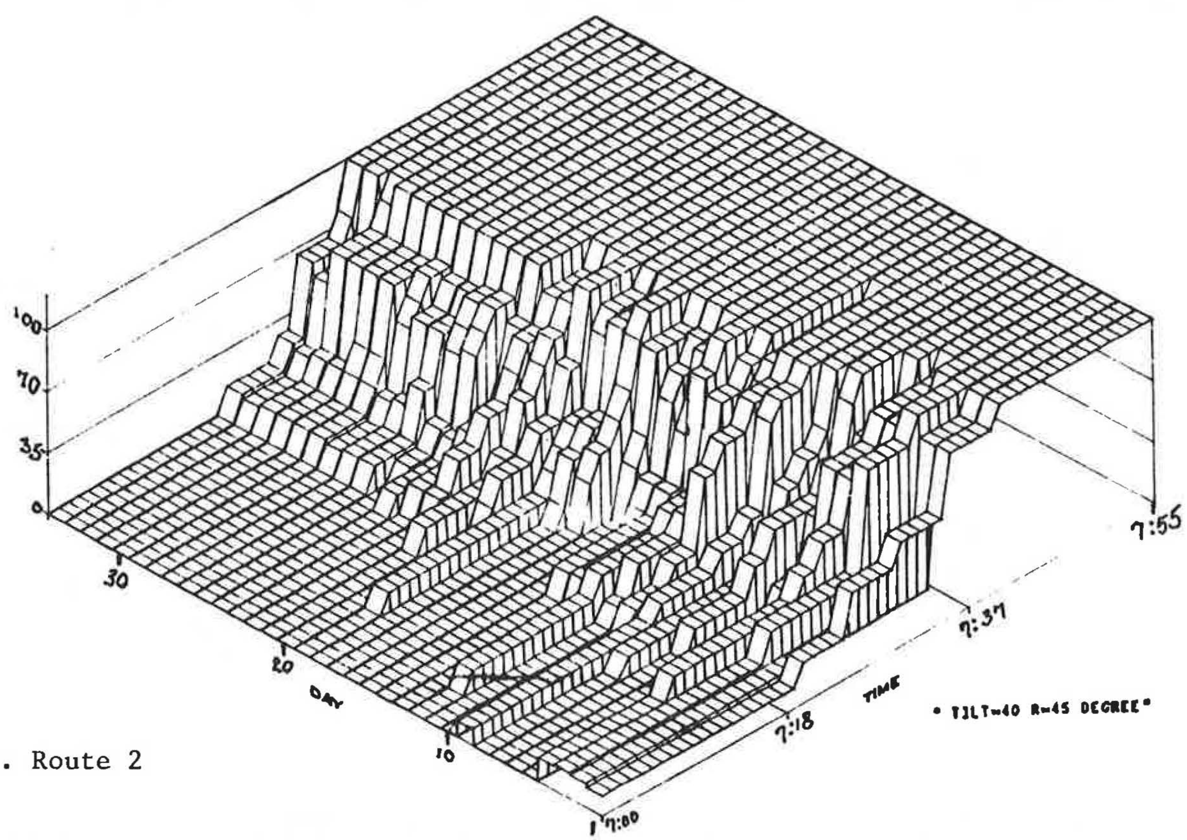


1.b. Route 2

FIGURE 1 Evolution of daily cumulative departure time distribution for Sector 1.



2.a. Route 1



2.b. Route 2

FIGURE 2 Evolution of daily cumulative departure time distribution for Sector 3.

3, respectively, for the duration of the experiment. Although these distributions had not converged for these sectors by the last day of the experiment, it is clear that the departure patterns on each route were beginning to exhibit markedly reduced changes from day to day toward the latter part of the experiment. Likewise, none of the departure time distributions for any other route and sector combinations had converged as of day 34, though all had stabilized to varying degrees.

Figures 3 and 4 depict the day-to-day evolution of the average travel time and average (of the absolute value of) schedule delay experienced by users in each sector. The schedule delay is defined as the difference between actual and preferred arrival time, the latter having been supplied by each participant at the beginning of the experiment as the desired arrival time in the absence of congestion. These plots further illustrate the dampening of the fluctuation in the average performance of each sector toward the end of the experiment, especially in sectors 1 and 2. Figure 5 depicts the daily percentage of users switching either route or departure time (or both) in each sector, indicating that approximately 11 percent were still switching by the end of the experiment.

The fact that this experiment did not converge, despite the additional time relative to the previous two experiments, which did converge, is due to two primary factors. The first is the increase in overall system congestion, noted in the previous section; the addition of the alternate route and the doubling of the number of participants increased overall congestion in the commuting system on a trips per available-lane basis by 33 percent. Second, the availability of an additional choice dimension to the commuters can be expected to increase the overall level of switching activity and thus the time needed to converge. It is therefore not surprising that this more congested system would take longer to converge, in light of the simulation results presented by Mahmassani and Chang (*19*). Because the experiment was stopped, it is not possible to assert that the system would eventually have converged, although it appears to have been headed in that direction. It is not clear that continuation of the experiment would have added much to the general conclusions and analysis that can be performed on the results attained up to this point. Practicality and cost considerations precluded the possibility of going beyond the seven weeks that were undertaken, especially since serious concerns would arise regarding the goodwill and interest of the participants, thereby jeopardizing the quality of the experimental data.

In addition to the temporal aspects of system convergence, its spatial characteristics are also of interest. In the previous two experiments, it was observed that sectors in which users are facing greater day-to-day uncertainty and unpredictability, reflected in the day-to-day fluctuations of travel time, required a longer period to converge. Although none of the sectors attained a steady state by the end of this experiment, switching activity can be examined as an indicator of the probable order in which sectors might have converged and of the sector's relative closeness

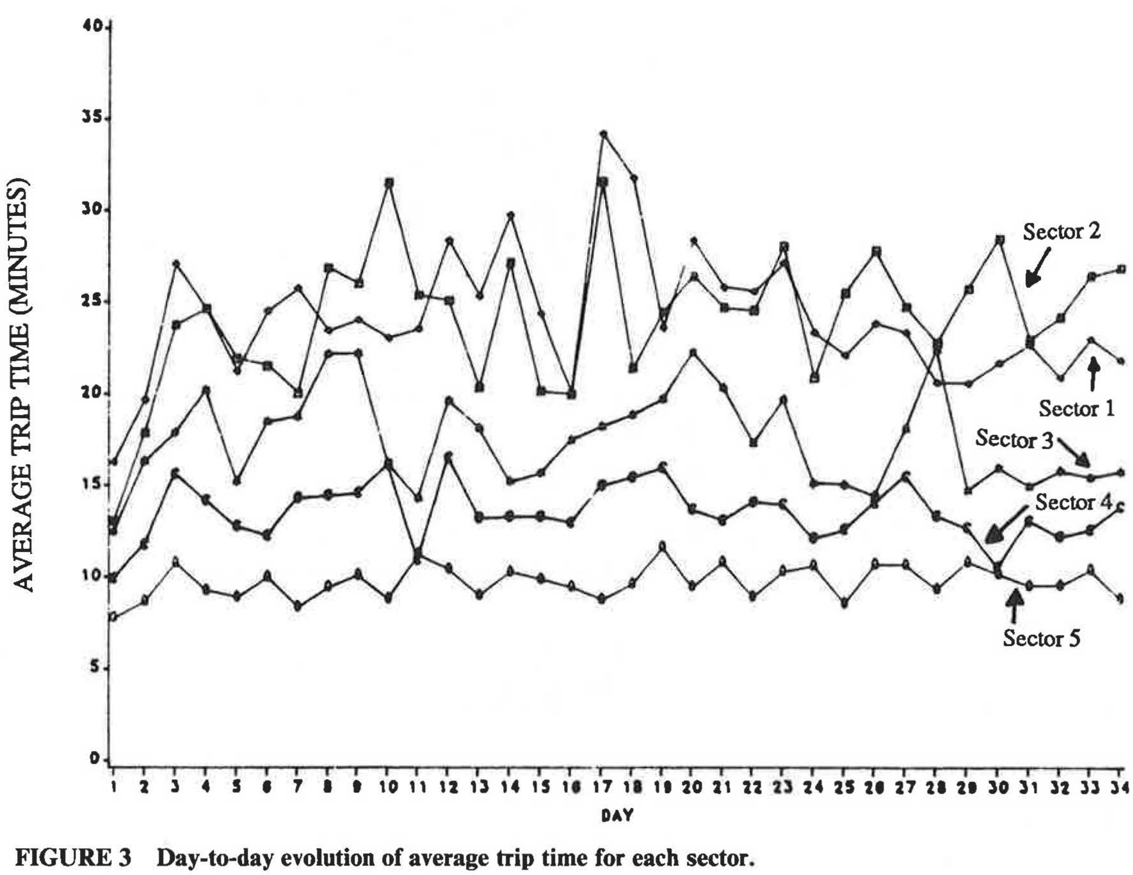


FIGURE 3 Day-to-day evolution of average trip time for each sector.

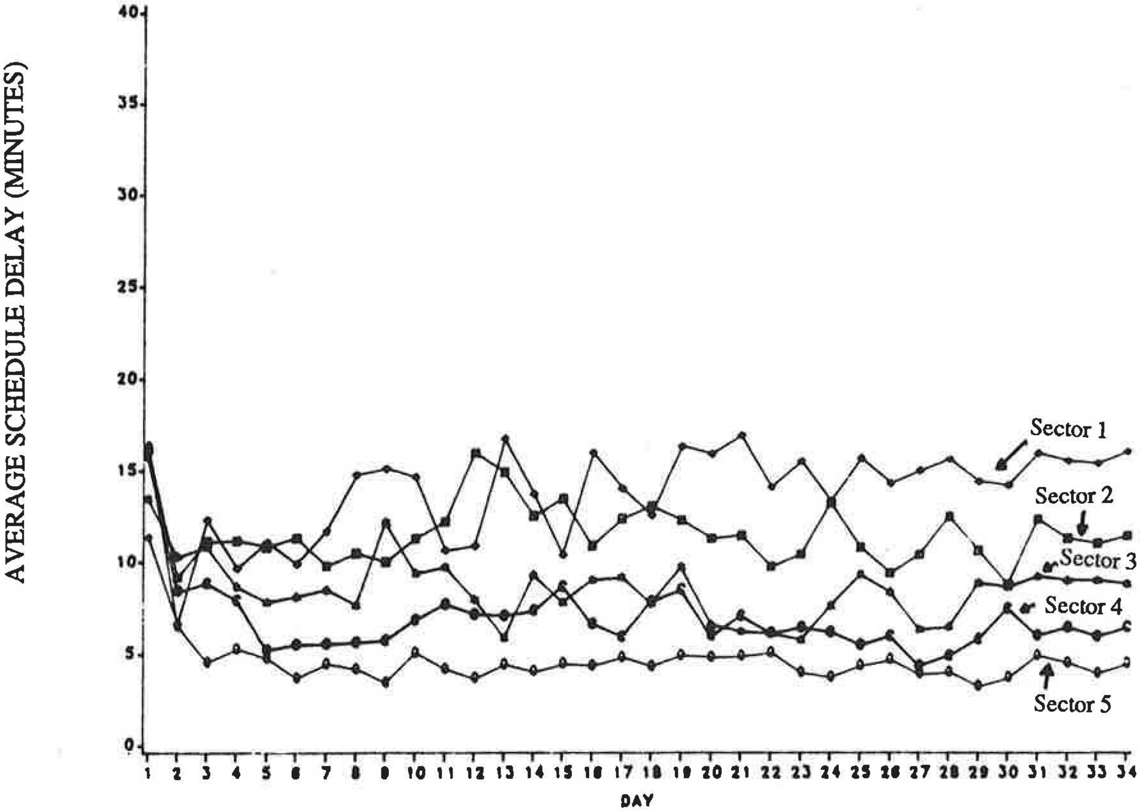


FIGURE 4 Day-to-day evolution of average absolute schedule delay for each sector.

to steady state. Figure 5 appears to indicate that sectors 2 and 1 experienced greater overall switching activity than did the other three sectors. A nonparametric test was used to address formally the dependence of switching activity on the sector. The null hypothesis is that switching activity is independent of sector or, more precisely, that each rank ordering of the sectors is equally likely (i.e., not the result of a common underlying order). The test measure is Friedman's statistic, related to Kendall's *W*-coefficient, which measures the agreement of several rank orderings of a given set of objects (*28*). In this case, the rank order of the sectors in terms of switching frequency is considered on each day of the experiment. Letting R_{ij} denote the rank of sector j, $j = 1, \ldots, 5$ on day i, $i = 2, \ldots, 34$, the calculation of the test statistic can be found in standard textbooks (*28*). Under the null hypothesis, this statistic is χ^2 distributed with 4 degrees of freedom. In this case, the calculated value is 42.2, while the 99.9th percentile value of the χ^2 distribution is 18.5, allowing the clear-cut rejection of the null hypothesis. Thus, as expected, the switching activity varies by sector. The sums of the ranks R_{ij}, namely, $R_{\cdot j} = \sum_{i=2}^{34} R_{ij}$, for $j = 1, \ldots, 5$, provide a maximum likelihood estimator of the common underlying rank order of the sectors (*29*). Accordingly, the calculated values are $R_{\cdot 1} = 84$, $R_{\cdot 2} = 57.5$, $R_{\cdot 3} = 134.5$, $R_{\cdot 4} = 114.5$, and $R_{\cdot 5} = 104.5$, indicating that sector 2 ranks highest in terms of the extent of switching frequency, followed by sector 1. The corresponding fluctuation in system performance is depicted in Figures 3 and 4, further corroborating the already mentioned findings of the previous experiments (*22*). The difference is that greater interaction seems to exist in this experiment among the sectors (particularly 3, 4, and 5), reflected in the less clear-cut conclusions, probably due to the higher level of overall congestion in the system.

In the next section, the effect of information availability on the system's evolution and the relative performance of the two user groups are examined.

EFFECT OF INFORMATION AVAILABILITY

As explained in the description of the experimental procedure, two user groups were defined on the basis of the level of information made available daily to each group. Differences between the limited information and complete information groups are examined in (1) how well each group performs (in terms of user costs) relative to the other and (2) the extent of the effort that members of each group have to exert in the process. Furthermore, by comparison with the previous experiments, it is possible to comment on the effect of the fraction of the population that has an informational advantage (i.e., is supplied with complete information). In the previous experiments, Mahmassani and Tong (*22*) found that when all users in a system (with a single route) were given complete information, the final

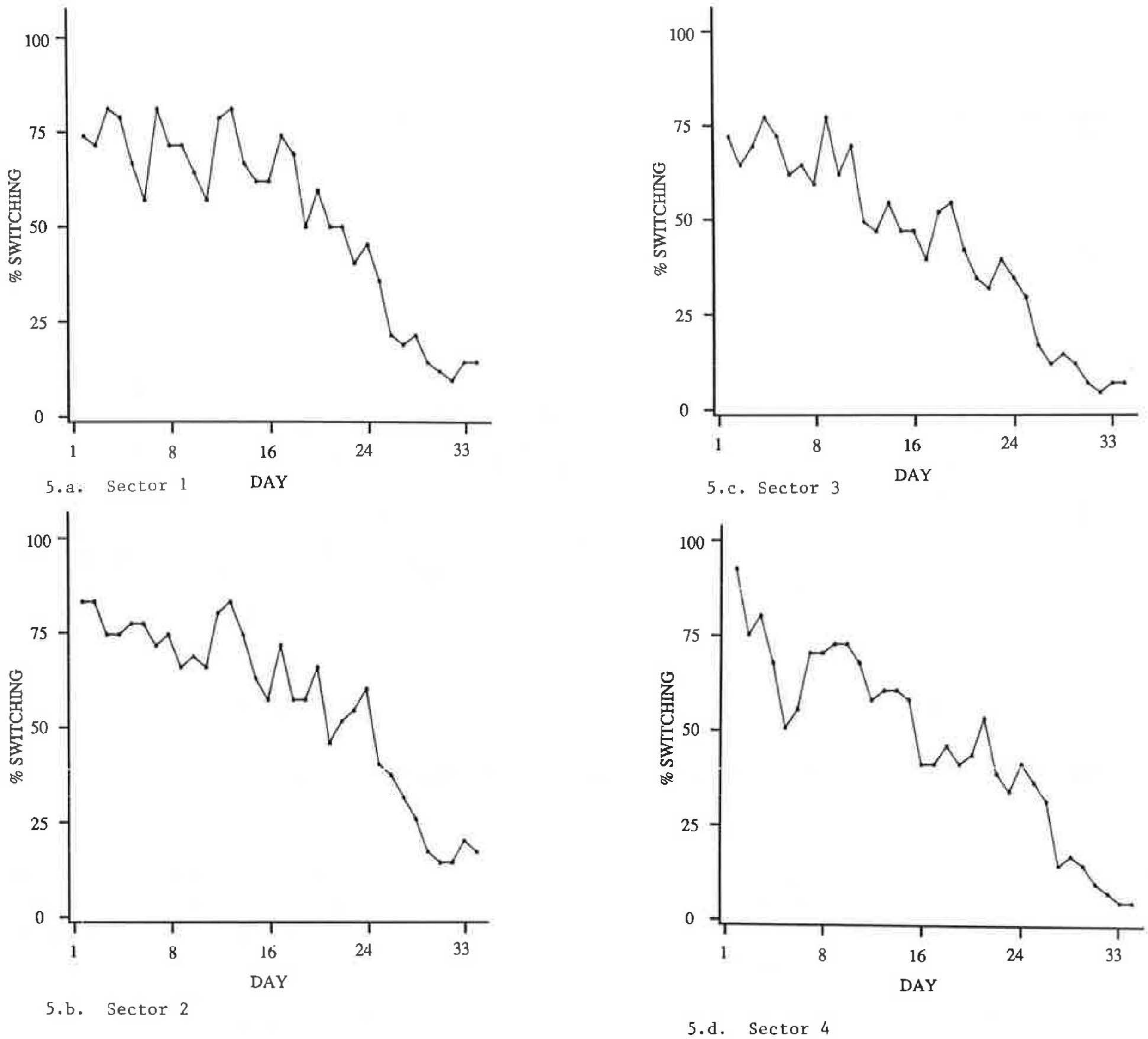


FIGURE 5 Day-to-day evolution of percentage of users in each sector switching or departure time (possibly both).

state reached was superior to that attained under the limited information situation. This superiority was established in terms of the average schedule delay and average travel time experienced by users in each sector at steady state. It was also found that this superior performance took a longer period to reach than in the limited information case, with users exhibiting greater overall daily switching activity in search of an acceptable solution.

Because this experiment did not converge completely, there is no "final state" at which to conduct the comparison of the performance of the two information availability groups. Since day-to-day fluctuations had subsided toward the end of the experiment, however, the averages of the performance measures of interest taken over the last four days are treated as representative of the "final state." The average schedule delay and average travel time experienced by the users in each sector (over the last four days) are calculated separately for each information availability group, and are constrasted in Figure 6 (the sector number is shown near each point in the figure). With the exception of sector 3, the performance of the full information group dominated the other. In sector 3, the average travel time is about 19 percent lower for the full information group, although the schedule delay is only slightly higher (by about 2 percent). These findings conform to intuition and are consistent with the earlier results of Mahmassani and Tong (*22*).

Regarding the relative effort exerted by the two groups to achieve the foregoing level of performance, the intensity of daily switching activity can be examined for each group. Figure 7 depicts the day-to-day evolution of the respective fractions of users in each information group that switch either route or departure time (possibly both). The plots suggest that the behavior of the complete information group was approaching convergence more rapidly than that of the limited information group. It is also useful to note that the complete information users in sector 3 constitute the only subgroup that had reached a steady state

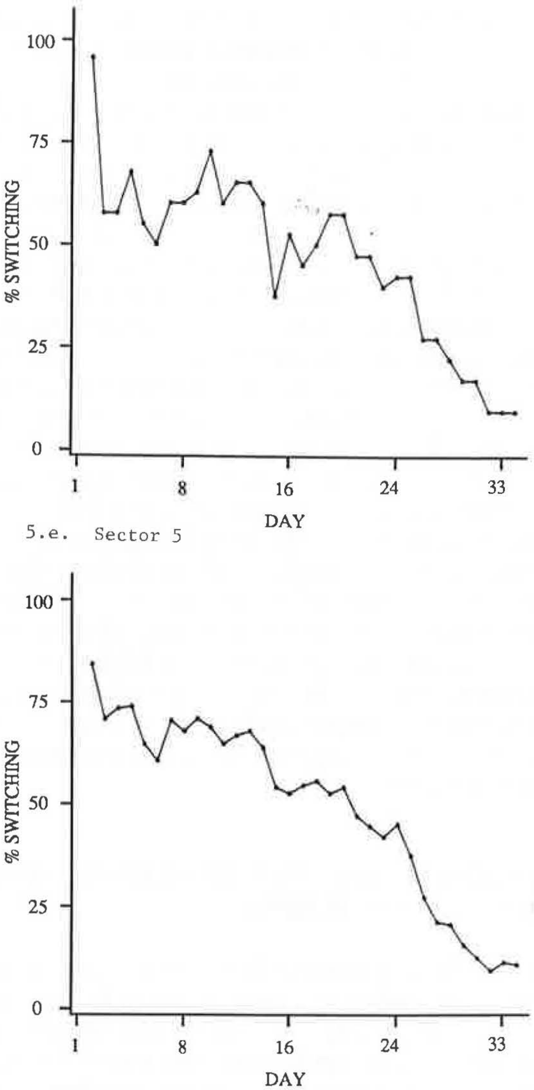


FIGURE 5 *(continued)*

(i.e., were no longer switching decisions) by the end of this experiment. These results would appear to contradict the findings of the two previous studies, in which providing information to *all* users in the single route system increased overall switching frequency and resulted in longer time to convergence. An important distinction in this experiment, however, is that there are two groups of users with two different levels of information availability "competing" in an interactive system. That the group with more information appears to reach a more desirable outcome at a smaller switching cost than the less fortunate group seems consistent with intuition. Actually, the earlier findings appeared counterintuitive but were explainable upon the realization that no user had an inherent advantage over the other in terms of availability of knowledge about the system's performance. Thus more information to all led to higher expectations and a more persistent search by all (*22*). Here, more information to only one group allowed that group to outperform the other, resulting in a more rapid decrease in switching activity. This then illustrates that the effect of information depends on the fraction of users with access to this information. Greater diffusion could reduce the competitive advantage that a given level of information might convey.

PERFORMANCE BY ROUTE

Although it is not possible to test the validity of the standard user equilibrium (UE) conditions in the context of this experiment, it is nevertheless helpful to examine differences in the performance experienced by users on the two available routes. Table 1 presents the average travel time experienced by users (from each sector) on each route, calculated over the four-day final state, as described previously. In all cases, the average travel time experienced on Route 1 is less than on Route 2, which has lower capacity and lower free flow speed. This does not seem to be in agreement with static UE conditions, which require

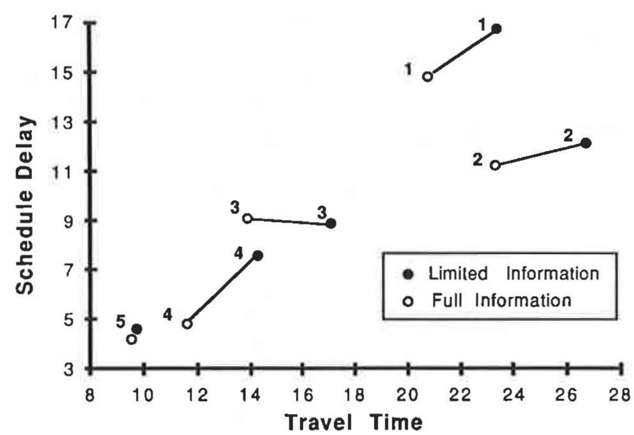


FIGURE 6 Comparison of average performance of two information availability groups at final state (average over last four days).

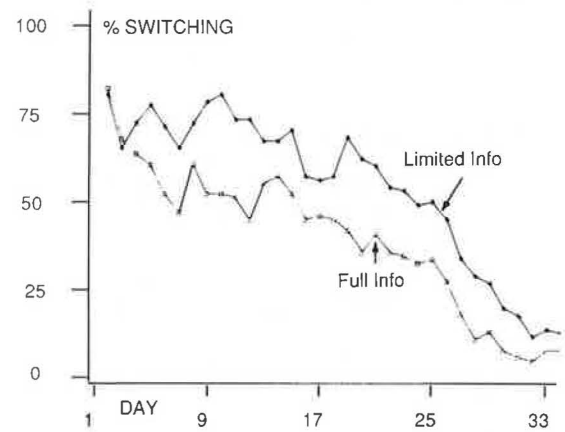


FIGURE 7 Day-to-day evolution of fraction of users in each information group who switch departure time or route (possibly both).

travel times on all used alternative routes between a given origin-destination pair to be equal at equilibrium.

Although equilibrium was not completely reached in this system, a more important consideration is that the preceding conditions are recognized to be inappropriate for time-dependent flows, where travel times vary within the peak period. Plots of the travel time by departure time on each route (excluded here for space considerations) clearly confirm that different travel times are experienced on each route for different departure times in this experiment. The extension of the UE principle to the time-dependent situation is predicated on the assumption that the decision process involves a trade-off between the travel time and schedule delay associated with a particular joint departure, time-route choice alternative (*3*, *8*). Table 2 presents the average of the absolute value of the schedule delay, also taken over the four-day final state described earlier. The values for sectors 1, 2, and 3 reflect performance that is consistent with such a trade-off, indicating that residents of these sectors who use Route 2 experience a lower average schedule delay than those who use Route 1. No such advantage exists for residents of sector 4, however, who incur an uncompensated higher travel time on Route 2 relative to Route 1. For sector 5, Route 1 simply dominates Route 2 on both cost components, yet the latter still captures a fraction of the sector 5 commuters.

If the utility function of a commuter can be represented as a linear weighted sum of travel time and schedule delay, then the relative magnitudes of the quantities reported in Tables 1 and 2 for sectors 1, 2, and 3 indicate that schedule delay must be valued considerably more highly than travel time in order to satisfy the dynamic user equilibrium conditions (for convenience, assume that users in a given sector have identical utility functions). Estimates for the marginal rates of substitution between the two cost components can be calculated as the ratio of the difference in average schedule delay between the two routes to the corresponding difference in average travel time, yielding −0.17, −0.47, −0.19, 0.0, and 1.52 for sectors 1 through 5, respectively. These values can be interpreted as the minutes of schedule delay one is willing to incur to reduce travel time by one minute. Alternatively, by taking the inverses of these numbers, it can be seen that one minute of schedule delay can be worth more than five minutes of travel time. Furthermore, in both sectors 4 and 5, travel time apparently cannot compensate for schedule delay, suggesting that users may effectively be indifferent to travel time relative to schedule delay. These findings therefore corroborate the results of the previous two experiments. Further insights can naturally be obtained through a more detailed econometric analysis of the individual-level data, which will be addressed in future work. In the next section, the exploratory aggregate analysis is continued, with the focus on the interrelation of route and departure time switching decisions.

TABLE 1 AVERAGE TRAVEL TIME EXPERIENCED BY SECTOR OF ORIGIN ON EACH ROUTE AT FINAL STATE (AVERAGE OVER LAST FOUR DAYS)

	Route 1	Route 2	Both Routes
Sector 1	20.2	25.5	22.1
Sector 2	24.5	26.2	25.1
Sector 3	12.5	26.9	15.5
Sector 4	11.4	15.8	12.9
Sector 5	8.9	11.2	9.6

TABLE 2 AVERAGE ABSOLUTE SCHEDULE DELAY EXPERIENCED BY SECTOR OF ORIGIN OF EACH ROUTE AT FINAL STATE (AVERAGE OVER LAST FOUR DAYS)

	Route 1	Route 2	Both Routes Combined
Sector 1	16.1	15.2	15.7
Sector 2	11.8	11.0	11.5
Sector 3	9.6	6.9	9.0
Sector 4	6.2	6.2	6.2
Sector 5	3.5	7.0	4.4

INTERACTION BETWEEN DEPARTURE TIME AND ROUTE SWITCHING

The interaction between the day-to-day adjustments of the departure time and route choice decisions is addressed in this section in an effort to suggest and explore some hypotheses about the mechanisms that govern this behavior. Of particular interest is the extent to which the switching of route and of departure time occur independent of one another. This has methodological implications for the formulation of an appropriate decision structure and model specification, as well as practical implications for the design of control and management strategies.

Table 3 presents the percentage of users in each sector who switched departure time at least n times, with $n = 1, \ldots, 33$. Similar information is presented in Table 4 for those who change route, indicating that more users switched departure time than route and that users switched departure time more frequently than route. Figure 8 depicts the day-to-day evolution of the fraction of participants who switched both departure time and route on a given day and the fraction of those who switched only one or the other (but not both). Clearly, the latter fraction is considerably less than the former on most days. Figure 9 shows the day-to-day patterns of the (marginal) fraction that switched route and of the fraction that switched departure time, further illustrating the considerably higher frequency of departure time changes on a daily basis. The average number of departure time switches made between consecutive route changes is presented in Table 5, indicating an initial reluctance to change routes and suggesting

TABLE 3 PERCENT OF PARTICIPANTS MAKING AT LEAST *n* DEPARTURE TIME SWITCHES BY SECTOR

n	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
1	100	100	98	100	98
2	100	97	95	100	95
3	100	97	95	95	92
4	100	97	92	93	88
5	100	94	92	90	85
6	95	94	90	88	82
7	90	91	82	80	82
8	88	86	80	78	70
9	83	83	72	71	68
10	79	83	70	68	68
11	79	83	65	68	65
12	71	83	60	58	60
13	69	80	55	56	58
14	67	77	55	56	55
15	57	69	50	51	52
16	43	63	40	51	48
17	40	60	35	44	45
18	40	54	30	36	42
19	36	46	25	34	35
20	31	40	22	29	35
21	26	37	20	29	28
22	19	29	18	17	22
23	17	20	18	15	20
24	12	20	15	15	20
25	12	17	10	15	18
26	10	17	8	12	12
27	7	17	8	10	12
29	5	9	2	5	8
30	2	9	2	2	8
31	2	9	0	0	2
32	0	6	0	0	2
33	0	3	0	0	2

TABLE 4 PERCENT OF PARTICIPANTS MAKING AT LEAST *n* ROUTE SWITCHES BY SECTOR

n	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
1	76	63	55	68	50
2	71	57	50	59	38
3	55	34	28	46	20
4	48	31	20	37	15
5	33	20	10	24	15
6	29	20	10	15	12
7	19	14	8	7	8
8	19	14	2	5	8
9	10	6	2	2	8
10	10	6	2	2	8
11	10	3	0	2	8
12	7	3	0	2	5
13	2	0	0	0	5
14	2	0	0	0	5
15	2	0	0	0	2
16	2	0	0	0	2
17	2	0	0	0	2
18	2	0	0	0	2
19	2	0	0	0	2
20	0	0	0	0	2
21	0	0	0	0	2
22	0	0	0	0	2

that commuters might first respond to experienced fluctuations in a given traffic system by first changing departure time, keeping the same route as usual. After the first route change, Table 5 reveals that users switched departure time approximately 1.5 to 2.5 times more often than route.

Further inspection of Figures 8 and 9 reveals a marked similarity in the patterns depicted in the two graphs, suggesting a rather high correlation among particular pairs of plotted variables. In particular, the joint switches of route and departure time closely parallel those of route,

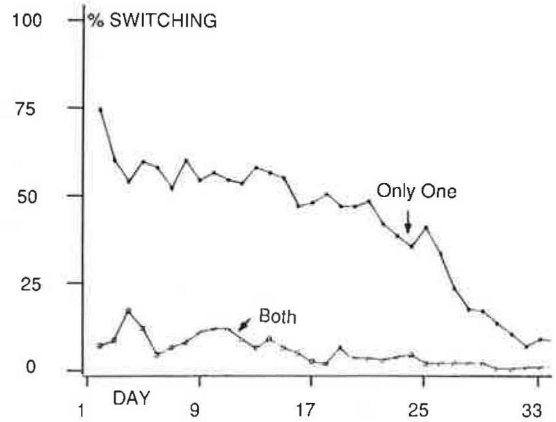


FIGURE 8 Day-to-day evolution of percent of users switching both route and departure time and fraction switching only one.

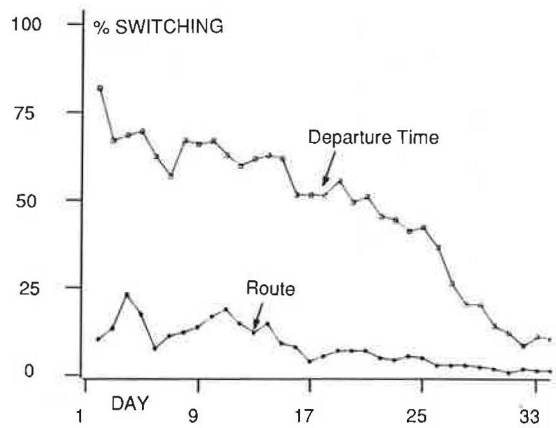


FIGURE 9 Day-to-day evolution of percent of users switching departure time and fraction switching route.

whereas changes of only one parallel those of departure time. This interdependence is further explored hereafter. A simple way to assess the validity of the independence hypothesis is to examine the compliance with the fundamental rule of probability that states that the product of the marginal probabilities of two independent events is equal to the probability of the joint occurrence of the two events. Using the sample fractions as estimates of the underlying probabilities, the product of the respective fractions of changing route and departure time is compared with that of changing both, on a daily basis. Figure 10 displays the results. In the figure, the "estimated" fraction is the above product, and the "observed" is the actual fraction in the sample of those switching both. Of course, if the independence hypothesis were to hold, then all the plotted points should lie on the straight line with slope 1. There is a clear, systematic violation of this hypothesis, however, since most points seem to lie above that line.

A more formal test of the aforementioned independence hypothesis can be performed using the chi-squared statistic. Let N_{ijt} denote the number of participants who, on day

TABLE 5 AVERAGE NUMBER OF DEPARTURE TIME SWITCHES BETWEEN $K - 1$TH AND KTH ROUTE SWITCH

K	Number of Partipants Making K Route Switches	Average Number of Departure Time Switches
1	124	3.65
2	109	1.81
3	73	2.58
4	60	1.68
5*	41	1.66

*Less than 20% of the participants had 6 or more route switches

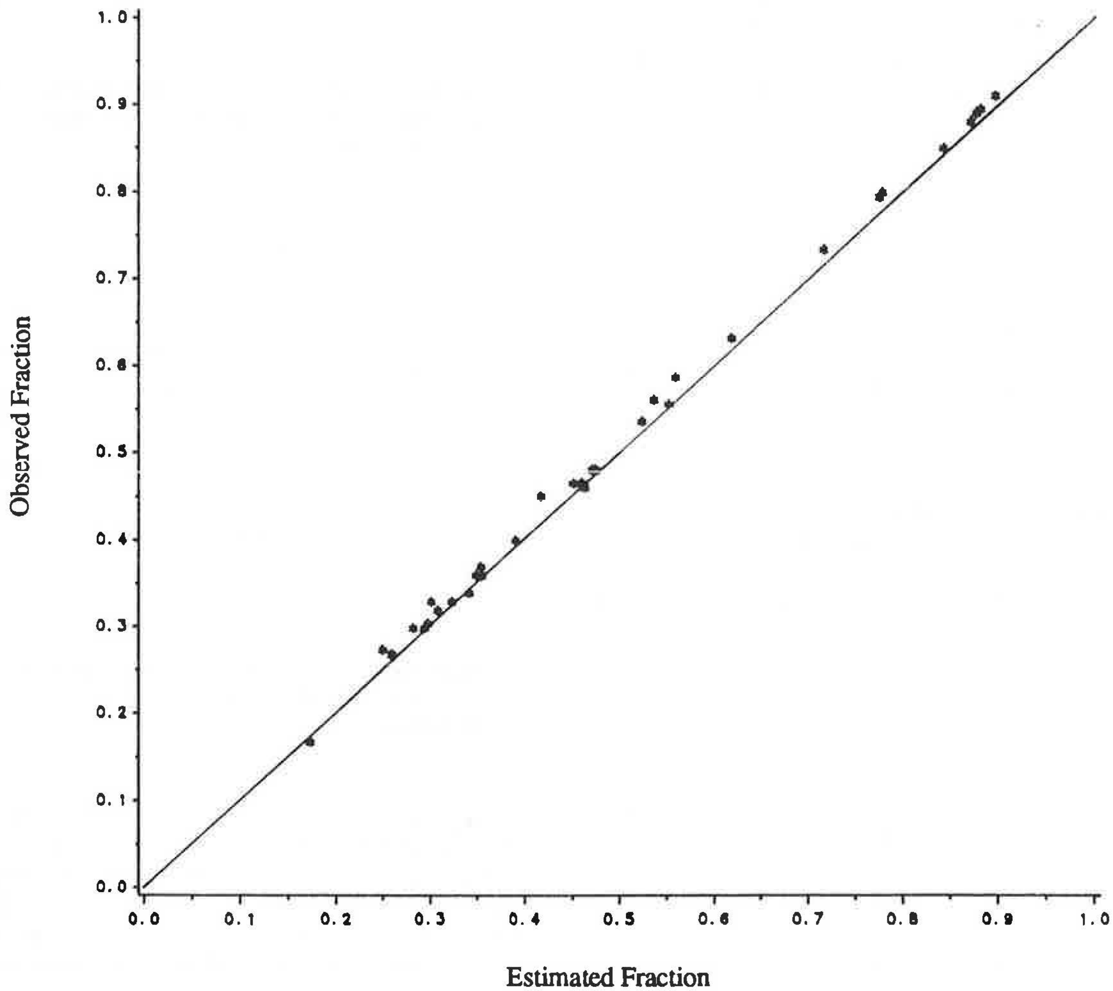


FIGURE 10 Observed versus estimated fractions in test of independence between departure time and route-switching behavior.

t, make decision i regarding departure time, where $i = 0$ indicates no change and $i = 1$ indicates a change, and decision j regarding route choice (same notational convention as i). Then, for a given day t (and dropping the subscript t for clarity of notation), the statistic

$$\chi^2 = \sum_i \sum_j \frac{\left[N_{ij} - \left(\sum_i N_{ij} \cdot \sum_j N_{ij} \Big/ \sum\sum_{ij} N_{ij}\right)\right]^2}{\sum_i N_{ij} \sum_j N_{ij} \Big/ \sum\sum_{ij} N_{ij}}$$

is χ^2 distributed (with one degree of freedom) under the null hypothesis that route switching is independent of departure time switching. Because daily data might result in cells with very few observations, it was necessary to combine the daily data into weekly groups. The results of the test are presented in Figure 11 in the form of a plot of the calculated values of the test statistic for each week of the experiment. Also shown on this plot are horizontal lines corresponding to the 95th and 90th percentiles of the χ^2 distribution (with one degree of freedom). As can be seen, the null hypothesis of independence can be rejected

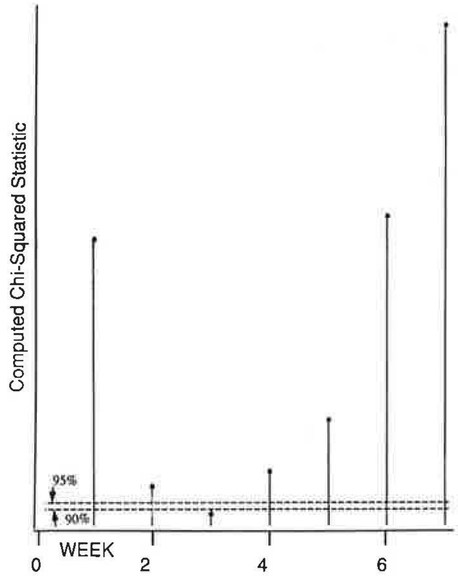


FIGURE 11 Results of chi-squared test of independence between departure time and route-switching behavior.

for six of the seven weeks with over 95 percent confidence; the only exception is in the third week, where that hypothesis can be rejected with slightly under 90 percent confidence.

The computed values of the test statistic also provide a useful measure of the magnitude of the discrepancy between the actual (observed) number of switches in each cell and the corresponding theoretical ones that would hold under the independence assumption. Thereby an informal index of the "extent" of the dependence between the two decisions is provided. Viewed from this angle, the evolution over time of the values shown in Figure 11 reveals a striking pattern whereby route and departure time switching decisions exhibit increasing deviation from independence as the system evolves beyond the third week. In the first week, a large deviation from independence can be detected, followed by two weeks of relatively weaker dependence. Having established the dependence between these two decisions, it is important to attempt to describe how this dependence arises in terms of the underlying behavioral mechanisms. An explanation within an extension of the boundedly rational decision framework proposed by Mahmassani and Chang (*21*, *30*) for the day-to-day dynamics of departure time decision making is presented next.

As noted earlier, the pattern of switching of both route and departure time closely parallels that of route, while the switching of only one of the two choice dimensions appears to parallel that of departure time. These patterns suggest that when users switch route, they also switch departure time. This can be verified by the conditional probability of switching departure time given that a route switch is taking place. These conditional probabilities are presented in Table 6, revealing that approximately 80 percent of the participants also switched departure

TABLE 6 CONDITIONAL PROBABILITY OF SWITCHING DEPARTURE TIME GIVEN *K*TH ROUTE CHANGE

K	Number of Participants Switching Routes for Kth Time	Conditional Probability
1	124	0.81
2	109	0.68
3	73	0.68
4	60	0.70
5*	41	0.66

*Less than 20% of the participants had 6 or more route switches

time the first time they switched route and approximately 70 percent did so on the succeeding route switches. Table 7 further presents the conditional probability of switching departure time given that only one of the two is changed. This probability is well over 90 percent on most days, confirming that route switches are in most cases accompanied by departure time switches.

The behavioral implications of the preceding observations are essentially as follows. When users switch only one of the two available decisions, it is most frequently departure time; when they switch route, they generally also switch departure time. This suggests that user behavior in this context might be governed by a hierarchical structure, with departure time switching taking precedence over the route switching decision. Thus the user's initial response to performance failures is likely to be a switch of departure time; continued failure to achieve satisfactory performance, or a particularly "large" failure, would trigger a more drastic response involving the switching of both route and departure time.

This behavior can be explained within the boundedly rational decision framework presented earlier for departure time decisions (*21*, *30*). It relies on Simon's well-known satisficing decision process proposed as a behaviorally realistic alternative to strict rational behavior (*31*). In its application to departure time decisions, user behavior in the commuting system is viewed as a dynamic boundedly rational search for an acceptable departure time (*21*). The acceptability of a given departure time is determined relative to an aspiration level, operationalized in the form of a dynamically varying indifference band of schedule delay (corresponding to acceptable arrival times).

The extension to the decision context where both route and departure time choices are available to commuters can explain the phenomena discussed earlier. Essentially, user behavior can now be viewed as if governed by two indifference bands $IBD \leq IBR$: one associated with the switching of departure time, denoted by IBD, which would be a subset of another interval IBR associated with changing route (and departure time in most cases). Since these would vary across users and over time, $IBD_{i,t}$ and $IBR_{i,t}$ are used to refer to user i's values on day t. Thus, letting $PM_{i,t}$ denote the performance measure for a given user i on day t, if $PM_{i,t} \leq IBD_{i,t} \leq IBR_{i,t}$, then user i will maintain the same departure time and route on day t. If $IBD_{i,t} < PM_{i,t} \leq IBR_{i,t}$, then the user will switch departure time

TABLE 7 CONDITIONAL PROBABILITY OF SWITCHING DEPARTURE TIME GIVEN THAT ONE IS SWITCHING ONLY ROUTE OR ONLY DEPARTURE TIME FOR *K*TH TIME

K	Number Switching Exactly one for Kth Time	Number Also Switching Departure Time	Conditional Probability
1	195	186	0.95
2	192	169	0.88
3	189	172	0.91
4	186	170	0.91
5	180	167	0.93
6	174	161	0.92
7	170	153	0.90
8	157	145	0.92
9	151	146	0.97
10	141	134	0.95
11	134	129	0.96
12	118	115	0.97
13	111	111	1.00
14	109	107	0.98
15	97	97	1.00
16	87	86	0.99
17	79	78	0.99
18	68	68	0.99
19	54	53	0.98
20	48	48	1.00

NOTE: Less than 20% of the participants switched exactly one more than 20 times

only; however, when $IBD_{i,t} \leq IBR_{i,t} < PM_{i,t}$, then both route and departure time will be changed. The performance measure has not been specified in the preceding, although the schedule delay is the primary candidate given the previous results on the departure time problem (where travel time was virtually insignificant) and the preliminary results of modeling work on this new data set. It is not necessary to preclude at this point, however, the possibility of using some weighted combination of schedule delay and travel time. Furthermore, note that the effect of repeated failures can be captured in the dynamic equations of the indifference bands, as shown in Mahmassani and Chang (*23*).

The analysis just presented has been repeated separately for each group of users sharing the same information availability status. No particular differences were found between the two groups in the nature of the interrelation between departure time and route switching decisions. These results are not presented here for space considerations, and can be found in a separate report (*32*). As noted earlier, the preceding results constitute a set of hypotheses that will be tested in the context of formal econometric modeling work.

CONCLUSION

The experiment reported in this paper extends the work conducted to date in this research program in two important directions: (1) the inclusion of the route choice dimension in addition to that of departure time and (2) the consideration of two user groups with different information availability levels interacting in the same simulated commuting system. Naturally, the addition of these features greatly adds to the complexity of the system and to the subsequent analysis of the results. On the other hand, it yields important insights into some of the least studied and least understood aspects of tripmaker behavior in congested urban networks.

This added complexity and the higher congestion level in this commuting system relative to that prevailing in the two previous experiments have resulted in the inability of the system to converge to a steady state where all users are satisfied with the consequences of their departure time and route decisions over an extended experimental period of 34 days. However, a general trend of decreased switching activity over time was exhibited by the users' behavior. Taken with the results of previous experiments, a clearer picture of a traffic system's day-to-day evolution in response to major supply-side changes is beginning to emerge. The effect of information availability on the behavior and performance of given user groups is of particular interest. In this regard, the results of this experiment are to a large extent consistent with a priori expectations based on intuition; that is, users with more information clearly outperform those with limited information when both are competing in the same system. The fraction of total users with particular information levels, however, appears to be a significant determinant of the effect of this information. In the limit, if all users share the same "complete" or "limited" information, a system with complete information may experience greater turbulence in its evolution (*22*). These findings have important implications for ongoing efforts and interest to develop information-related strategies involving communication and information technologies, for the relief of traffic congestion in urban networks. In particular, the type and quantity of information, as well as its distribution across the user population, can greatly influence the resulting impact on the system.

The interdependence between route choice and departure time decisions is another important aspect of user behavior addressed in this paper. The exploratory aggregate analysis considered here points to the precedence of departure time shifts over route shifting in dealing with experienced unpredicted congestion in the system. The explanation for the observed behavior within the previously articulated boundedly rational decision framework appears to be plausible, and preliminary results of model estimation work are confirming this explanation. The insights based on this exploratory analysis form the principal hypotheses guiding the development and calibration of dynamic discrete choice model of user behavior.

Naturally, there are important limitations associated with the general experimental approach followed here, as well as with the particulars of the experiment described in the paper. Regarding the general approach of using real commuters in a simulated system, extensive discussion can be found elsewhere (*20*, *21*, *22*) and is not repeated here. Suffice it to say that this approach can play an important role in the development of theory and behavioral insights that can guide eventual complete or partial field validation. Regarding the specific elements of this experiment, the main new concerns (relative to the previous two experiments) relate to the omission of potentially important attributes that affect route choice. In particular, the visual dimension is missing, thereby excluding the influence of aesthetics. Similarly, pavement texture and condition and the resulting ride quality are other accessory attributes. For all practical purposes, the two routes must be considered as identical in all respects other than their operational characteristics.

ACKNOWLEDGMENTS

This paper is based on research funded by a National Science Foundation grant to the University of Texas at Austin. The contribution of Chee-Chun Tong and R. Jayakrishnan in the design and administration of the experiment, in preliminary data preparation as well as in substantive related discussions, is gratefully acknowledged and much appreciated. The cooperation of the participants and the numerous research assistants during the conduct of the experiment is particularly appreciated. The authors have benefited from the encouragement of and fruitful interaction with Robert Herman and Gang-len Chang in the course of their collaboration on this general problem area.

REFERENCES

1. R. Herman and T. Lam. Trip Time Characteristics for Journeys to and From Work. *Transportation and Traffic Theory: Proceedings of the Sixth International Symposium* (D. J. Buckley, ed.). A. H. & A. W. Reed, Sydney, Australia, 1974, pp. 57–86.
2. C. Hendrickson, D. Nagin, and E. Plank. Characteristics of Travel Time and Dynamic User Equilibrium for Travel-to-Work. *Proceedings of the Eighth International Symposium on Transportation and Traffic Theory* (V. F. Hurdle et al., eds.). University of Toronto Press, Toronto, Ontario, Canada, 1983, pp. 321–347.
3. C. Hendrickson and G. Kocur. Schedule Delay and Departure Time Decisions in a Deterministic Model. *Transportation Science*, Vol. 15, 1981, pp. 62–77.
4. P. H. Fargier. Influence du Mecanisme de Choix de l'Heure de Depart sur la Congestion du Trafic Routier. *Proceedings of the Eighth International Symposium on Transportation and Traffic Theory* (V. F. Hurdle et al., eds.). University of Toronto Press, Toronto, Ontario, Canada, 1983, pp. 223–263.
5. M. Kuwahara and G. F. Newell. Queue Evolution on Freeways Leading to a Single Core City During the Morning Peak. *Transportation and Traffic Theory: Proceedings of the Tenth International Symposium* (N. H. Gartner and H. H. M. Wilson, eds.). Elsevier, New York, 1987, pp. 21–40.
6. C. F. Daganzo. The Uniqueness of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck. *Transportation Science*, Vol. 19, 1985, pp. 29–37.
7. M. J. Smith. The Existence of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck. *Transportation Science*, Vol. 18, 1984, pp. 385–394.
8. H. Mahmassani and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science*, Vol. 18, 1984, pp. 362–384.
9. M. Ben-Akiva, A. de Palma, and P. Kanaroglou. Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates. *Transportation Science*, Vol. 20, 1986, pp. 164–181.
10. M. Ben-Akiva, A. de Palma, and P. Kanaroglou. Effects of Capacity Constraints on Peak-Period Traffic Congestion. *Transportation Research Record* 1085, TRB, National Research Council, Washington, D.C., 1986, pp. 16–26.
11. S. Cosslett. The Trip Timing Decision for Travel to Work by Automobile. *Demand Model Estimation and Validation.* The Urban Travel Demand Forecasting Project Phase I Final Report, Vol. V. Institute for Transportation Studies, University of California at Berkeley, Berkeley, Calif., 1977.
12. K. Small. The Scheduling of Consumer Activities: Work Trips. Working Paper, Econometric Research Program, Department of Economics, Princeton University, Princeton, N.J., 1978.
13. M. Abkowitz. Understanding the Effect of Transit Service Reliability on Work Travel Behavior. *Transportation Research Record* 794, TRB, National Research Council, Washington, D.C., 1981, pp. 33–41.
14. C. Hendrickson and E. Plank. The Flexibility of Departure Times for Work Trips. *Transportation Research A*, Vol. 18A, 1984, pp. 25–36.
15. M. Ben-Akiva, M. J. Bergman, A. J. Daly, and R. Ramaswamy. Modelling Inter Urban Route Choice Behavior. *Ninth International Symposium on Transportation and Traffic Theory* (J. Volmuller and R. Hamerslag, eds.). VNU Science Press, Utrecht, The Netherlands, 1984, pp. 299–330.
16. P. H. L. Bovy and M. A. Bradley. Route Choice Analyzed with Stated-Preference Approaches. *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 11–20.
17. S. Abu-Eisheh and F. L. Mannering. A Discrete/Continuous Analysis of Commuters' Route and Departure Time. Presented at the 66th Annual Meeting of the Transportation Research Board, Washington, D.C., 1987.
18. J. L. Horowitz. The Stability of Stochastic Equilibrium in a Two-Link Transportation Network. *Transportation Research B*, Vol. 18B, 1984, pp. 13–28.
19. H. S. Mahmassani and G.-L. Chang. Experiments with Departure Time Choice Dynamics of Urban Commuters. *Transportation Research B*, Vol. 20B, 1986, pp. 297–320.
20. H. S. Mahmassani, G.-L. Chang, and R. Herman. Individual Decisions and Collective Effects in a Simulated Traffic System. *Transportation Science*, Vol. 20, 1986.
21. H. S. Mahmassani and G.-L. Chang. Dynamic Aspects of Departure Time Choice Behavior in a Commuting System: Theoretical Framework and Experimental Analysis. *Trans-*

portation Research Record 1037, TRB, National Research Council, Washington, D.C., 1985, pp. 88–101.
22. H. S. Mahmassani and C.-C. Tong. Availability of Information and Dynamics of Departure Time Choice: Experimental Investigation. *Transportation Research Record 1085*, TRB, National Research Council, Washington, D.C., 1986, pp. 33–47.
23. H. S. Mahmassani and G.-L. Chang. Specification Estimation of a Dynamic Departure Acceptability Mechanism. Presented at the 65th Annual Meeting of the Transportation Research Board, Washington, D.C., 1986.
24. G.-L. Chang and H. S. Mahmassani. Travel Time Prediction and Departure Time Adjustment Behavior Dynamics in a Congested Traffic System. *Transportation Research B* (in press).
25. G.-L. Chang. *Departure Time Dynamics in the Urban Transportation Network*. Ph.D. dissertation. Department of Civil Engineering, The University of Texas at Austin, Austin, Tex., 1985.
26. C.-C. Tong, H. S. Mahmassani, and G.-L. Chang. Travel Time Prediction and Information Availability in Commuter Behavior Dynamics. *Transportation Research Record 1138*, TRB, National Research Council, 1987, pp. 1–7.
27. G.-L. Chang, H. S. Mahmassani and R. Herman. A Macroparticle Traffic Simulation Model to Investigate Peak-Period Commuter Decision Dynamics. *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 107–121.
28. W. J. Conover. *Practical Non Parametric Statistics*, 2nd ed. Wiley, New York, 1980.
29. P. D. Feigin and A. Cohen. On a Model of Concordance Between Judges. *Journal of the Royal Statistical Society B*, Vol. 40, 1978, pp. 203–213.
30. H. S. Mahmassani and G.-L. Chang. On Boundedly-Rational User Equilibrium in Transportation Systems. *Transportation Science*, Vol. 21, 1987, pp. 89–99.
31. H. Simon. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, Vol. 69, 1955, pp. 99–118.
32. D. G. Stephan. Route Choice and Departure Time Decision Dynamics for Urban Commuters. M.S. thesis. Department of Civil Engineering, The University of Texas at Austin, Austin, Tex., 1987.

Publication of this paper sponsored by Committee on Traveler Values and Behavior.