TRANSPORTATION RESEARCH RECORD 944

Transportation Forecasting: Analysis and Quantitative Methods

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1983

Transportation Research Record 944
Price $18.20
Edited for TRB by Scott C. Herman

modes
1 highway transportation
2 public transit
4 air transportation

subject areas
12 planning
13 forecasting

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

(Transportation research record; 944)
1. Transportation–Forecasting–Mathematical models–Congresses. 2. Transportation–Planning–Mathematical models–Congresses. I. National Research Council (U.S.) Transportation Research Board. II. National Academy of Sciences (U.S.) III. Series.

TE7.H5	no. 944	380.5 s	84-18963
[HE11]		[380.5]	
ISBN 0-309-03665-8		ISSN 0361-1981	

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Kenneth W. Heathington, University of Tennessee, chairman

Transportation Forecasting Section
George V. Wickstrom, Metropolitan Washington Council of Governments, chairman

Committee on Passenger Travel Demand Forecasting
Frank S. Koppelman, Northwestern University, chairman
Moshe E. Ben-Akiva, Werner Brog, William A. Davidson, Christopher R. Fleet, Susan Hanson, Joel L. Horowitz, Stephen M. Howe, Terry Kraft, Steven Richard Lerman, Eugene J. Lessieu, Eric J. Miller, Eric Ivan Pas, Martin G. Richards, Aad Ruhl, Earl R. Ruiter, James M. Ryan, Bruce D. Spear, Timothy J. Tardiff

Committee on Traveler Behavior and Values
David T. Hartgen, New York State Department of Transportation, chairman
Peter M. Allaman, Julian M. Benjamin, Katherine Patricia Burnett, Melvyn Cheslow, David Damm, Michael A. Florian, Thomas F. Golob, David A. Hensher, Joel L. Horowitz, Frank S. Koppelman, Steven Richard Lerman, Jordan J. Louviere, Arnim H. Meyburg, Richard M. Michaels, Alfred J. Neveu, Robert E. Paaswell, Earl R. Ruiter, Yosef Sheffi, Peter R. Stopher, Antti Talvitie, Mary Lynn Tischer

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units, officers, and members are as of December 31, 1982.

Contents

Authors of the Papers in This Record

Allen, William G., Jr., Volvo of America, 1 Volvo Drive, Building B, Rockleigh, N.J. 07647; formerly with Barton-Aschman Associates, Inc.
Ben-Akiva, Moshe, Department of Civil Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-178, Cambridge, Mass. 02139
Boyce, D.E., Transportation Planning Group, Department of Civil Engineering, University of Illinois at Urbana-Champaign, 208 N. Romine Street, Urbana, Ill. 61801
Brög, Werner, Socialdata, GmbH, Hans-Grässel-Weg 1, 8000 München 70, West Germany
Chon, K.S., Department of Civil Engineering, State University of New York at Buffalo, Buffalo, N.Y. 14260
Chu, Chaushie, Department of Civil Engineering, University of Petroleum and Minerals, Dhahran, Saudi Arabia
Daskin, Mark S., Department of Civil Engineering, Northwestern University, Evanston, Ill. 60201
DeJohn, Anthony, Bureau of Statewide Planning, New Jersey Department of Transportation, 1035 Parkway Avenue, Trenton, N.J. 08625
Doherty, Michael J., Institute for Interdisciplinary Engineering Studies, Room 304 Potter, Purdue University, West Lafayette, Ind. 47906
Doxsey, Lawrence B., Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, Mass. 02142
Eash, R.W., Chicago Area Transportation Study, 300 W. Adams Street, Chicago, Ill. 60606
Geva, Itzhak, Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4
Gur, Yehuda J., Transportation Research Institute, Technion–Israel Institute of Technology, Haifa, Israel
Haghani, Ali E., Department of Civil Engineering, Northwestern University, Evanston, Ill. 60201
Han, Anthony F., Institute of Transportation Studies, 109 McLaughlin Hall, University of California, Berkeley, Calif. 94720
Hauer, Ezra, Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4
Hocherman, Irit, Transportation Research Institute, Technion–Israel Institute of Technology, Haifa, Israel
Jacobson, Jesse, 1382 Beacon Street, Brookline, Mass. 02146; formerly with Transportation Systems Center, U.S. Department of Transportation
Kermanshah, Mohammad, Department of Civil Engineering, University of California at Davis, Davis, Calif. 95616
Kitamura, Ryuichi, Department of Civil Engineering, University of California at Davis, Davis, Calif. 95616
Koppelman, Frank S., Department of Civil Engineering, Northwestern University, Evanston, Ill. 60201
Landau, Uzi, Department of Civil Engineering, Technion–Israel Institute of Technology, Haifa, Israel
Lee, Y.J., Transportation Planning Group, Department of Civil Engineering, University of Illinois at Urbana-Champaign, 208 N. Romine Street, Urbana, Ill. 61801
Lisco, Thomas E., Central Transportation Planning Staff, 27 School Street, Boston, Mass. 02108
McDonald, Kathie G., Research and Analysis, Planning Division, Metropolitan Dade County Planning Department, 44 West Flagler Street, Suite 1100, Miami, Fla. 33130; formerly with Schimpeler-Corradino
Meyburg, Arnim H., Department of Civil and Environmental Engineering, Cornell University, Hollister Hall, Ithaca, N.Y. 14853
Neveu, Alfred J., Transportation Data and Analysis Section, Planning Department, New York State Department of Transportation, 1220 Washington Avenue, State Campus, Albany, N.Y. 12232
Prashker, Joseph N., Transportation Research Institute, Technion–Israel Institute of Technology, Haifa, Israel
Schultz, Gordon W., Barton-Aschman Associates, Inc., 1400 K Street, N.W., Washington, D.C. 20005
Sheffi, Yosef, Center for Transportation Studies, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-131, Cambridge, Mass. 02139
Sparrow, F.T., Institute for Interdisciplinary Engineering Studies, Room 304 Potter, Purdue University, West Lafayette, Ind. 47906
Stopher, Peter R., Schimpeler-Corradino Associates, 300 Palermo Avenue, Coral Gables, Fla. 33134
Sullivan, Edward C., Institute of Transportation Studies, 109 McLaughlin Hall, University of California, Berkeley, Calif. 94720
Supernak, Janusz, Department of Civil Engineering, San Diego State University, San Diego, Calif. 92182; formerly with Drexel University

Talvitie, Antti, Department of Civil Engineering, State University of New York at Buffalo, Buffalo, N.Y. 14260
Tarem, Zvi, Center for Transportation Studies, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-050, Cambridge, Mass. 02139
Wermuth, Manfred J., Institute fur Stadtbauwesen, Technische Universität Braunschweig, Postfach 3329, 3300 Braunschweig, West Germany

Development of Survey Instruments Suitable for Determining Nonhome Activity Patterns

WERNER BRÖG, ARNIM H. MEYBURG, AND MANFRED J. WERMUTH

Generation of travel behavior data by means of empirical surveys is an important element of transportation planning. At the same time, relatively little attention has been paid to the rules for collecting and determining the methodological quality of the data. The methodological design of such surveys is relatively complicated because of a number of influence factors that may ultimately be reflected in the validity of the results. The issue of survey instrument design is discussed in detail. A number of methodological tests are examined that were intended to improve one of the weak points in surveys of travel behavior—the design of such instruments. Initially, it was concluded that a diary-type instrument would have to be used to ensure proper recording of trip details. An ideal diary was developed that was used in several surveys. But it became evident that this instrument design, in spite of its high methodological quality, was unsuitable for large-scale surveys, such as those frequently used in transportation planning, because of organizational and cost problems. Therefore, an additional series of tests was developed to simplify these diaries and to transform them into a form suitable for large-scale mail-back surveys. Each test series was tested empirically with detailed documentation of reporting deficiencies. Thus it was possible to present in an understandable manner the development of a survey instrument of desirable quality. The final version of the instrument design, which was the outgrowth of the empirical tests, has been used subsequently in numerous large-scale applications in several countries. In the course of these applications the methodological quality of the design was confirmed, which ultimately justified the development costs.

The influence of measurement procedures and measurement (survey) instruments on measurement results has to be recognized at the outset of any empirical survey. Therefore, the survey procedure has to be included as part of the overall research approach (1). Typically, a measurement process (i.e., survey procedure) is composed of a number of elements that can be subsumed under the following categories (2,3):

1. Problem formulation, theoretical reference frame, analysis concept;
2. Base population, sampling unit, sampling procedure, weighting, population values;
3. Survey method and instrument(s);
4. Survey implementation, response rates; and
5. Data preparation, evaluation, and analysis.

The third and fourth categories are the subjects of this paper. The development and use of survey instruments designed to measure actual nonhome activity patterns are described in this paper.

Empirically measured travel behavior is the most important input to transportation planning decisions because it constitutes the basis for explanation and prediction of future travel activities. Methodological deficiencies of this measure have direct consequences for all subsequent phases of the transportation planning process.

Meanwhile, the mail-back household survey, which measures nonhome activity patterns, has become a standard component of transportation planning. Generally, the survey instruments used in this process are the result of years of developmental work. In this paper such a developmental process is retraced in terms of content and chronology on the basis of the KONTIV design (4).

Two aspects will be emphasized. First, the laborious path of such developmental work, including its accompanying setbacks, is illustrated. Second, it will be shown that basic methodological research also can produce, as by-products, fundamental and substantive analytical and theoretical insights.

EARLY DEVELOPMENTS

When preliminary developmental work toward the improvement of methods for measuring nonhome activity patterns started in Germany in 1972, the generally accepted method for empirical surveys was the personal interview. For example, in an intensive personal interview survey (5), the course of the daily trips to work or school was investigated in addition to various other aspects. Three main bases for criticism arose out of such survey efforts:

1. The survey measured average rather than actual travel behavior;
2. Information (e.g., about travel time) was estimated by the interviewee; and
3. Only a segment of the individual's mobility was investigated.

Consequently, the results of such interview information were unsatisfactory when validated on the basis of objectively measured values for travel time, distance, and cost. For example, only three-quarters of automobile drivers estimated their travel time within a tolerance level of ±25 percent. (Admittedly, the generation of objective comparative data is difficult in this instance.) On average, travel time was underestimated by 11 percent (1).

For the transit user the situation was quite different. Although the share of respondents with reports of travel time within the tolerance level of ±25 percent was greater (namely, 79 percent), the average error was substantially higher and in the opposite direction, namely an average overestimation of 36 percent [see Table 1 (4)].

The strong distortions caused by these misestimates are described in Table 2 (4), which gives a breakdown of trips into their access, egress, and travel-time components. Automobile drivers claim to have spent, on average, only a total of 6 min on access and egress, including the search for parking spaces, whereas transit users recorded 62 min for access, egress, waiting, and transfer times.

The methodologically oriented reader of such results could draw two significant conclusions. First, the reported travel behavior and characteristics deviated substantially from reality even though these respondents experienced the real values of these trip elements twice during each working (school) day. Second, the biases are of a systematic nature and apparently are related to the user's attitude toward the respective travel mode. Hence, in the case of public transit, the particularly disturbing access, egress, waiting, and transfer times are overestimated drastically.

From a conceptual point of view, these results [which were substantiated in several other studies (6)] indicated that the subjective perception of such measures constitutes an important determinant of travel modal choice. This concept has found entry into the relevant models under the terms perception and perceived values (7). The methodological analysis of these findings leads to two conclusions. First, data about travel behavior must not be collected (inquired about) in a general form

Table 1. Accuracy of travel-time estimates for automobiles and transit (4).

Item	Reported (interview) Travel Time for	
	Automobile	Public Transit
Sample size	800	520
Correct estimates (within ±25 percent error) (%)	72	79
Incorrect estimates (> 25 percent error) (%)	28	21
Index of average deviation from the correct travel time (objective time = 100)	89	136

Table 2. Reported estimates of travel-time components for automobile and transit users (4).

Item	Travel Time (min)
Automobile users (n = 800)	
Walk from residence to parking; from parking to destination	6
In-vehicle travel time	41
Search for parking at destination	1
Total	48
Transit users (n = 520)	
Walk from residence to boarding stop; from alighting stop to destination	28
In-vehicle travel time	22
Total waiting and transfer time	34
Total	84

(i.e., not in terms of average values); they need to have a concrete temporal reference. Second, activities cannot be viewed in isolation. Instead, complete daily activity patterns are needed to constitute the basis of analysis.

It could be shown, for example, that the recording of beginning and termination times of a trip is more accurate than the direct reporting of trip lengths. The implications of this for further methodological considerations are as follows. First, the data about travel behavior need to be collected for specific survey days. Second, a diary-type survey instrument should be used, which requires entries about complete daily activity sequences. Third, a written survey form is preferable to the personal interview. However, this does not indicate by what means the survey instrument should be delivered to the respondents, i.e., by mail or by means of an interviewer.

DEVELOPMENT OF AN ACTIVITIES DIARY

Based on the recognition that surveys about general (or average) travel behavior and of estimated information lead to invalid results, an activity diary (8) was developed in 1972, in which the target population (sample) was asked to record in writing its complete daily activity set for specific survey dates.

This diary (see Figures 1-4) was a brochure of about 8 x 6 in. in size, the cover of which listed the name of the target person, the day of the week, and the date of the respective survey day. On the inside cover were 12 numbered lines for trip entries, where the odd-numbered trips were designated by a different color in order to make this page of the diary visually clearer and more appealing. On this page the respondents were supposed to enter the

Figure 1. Cover of trip diary for en route use.

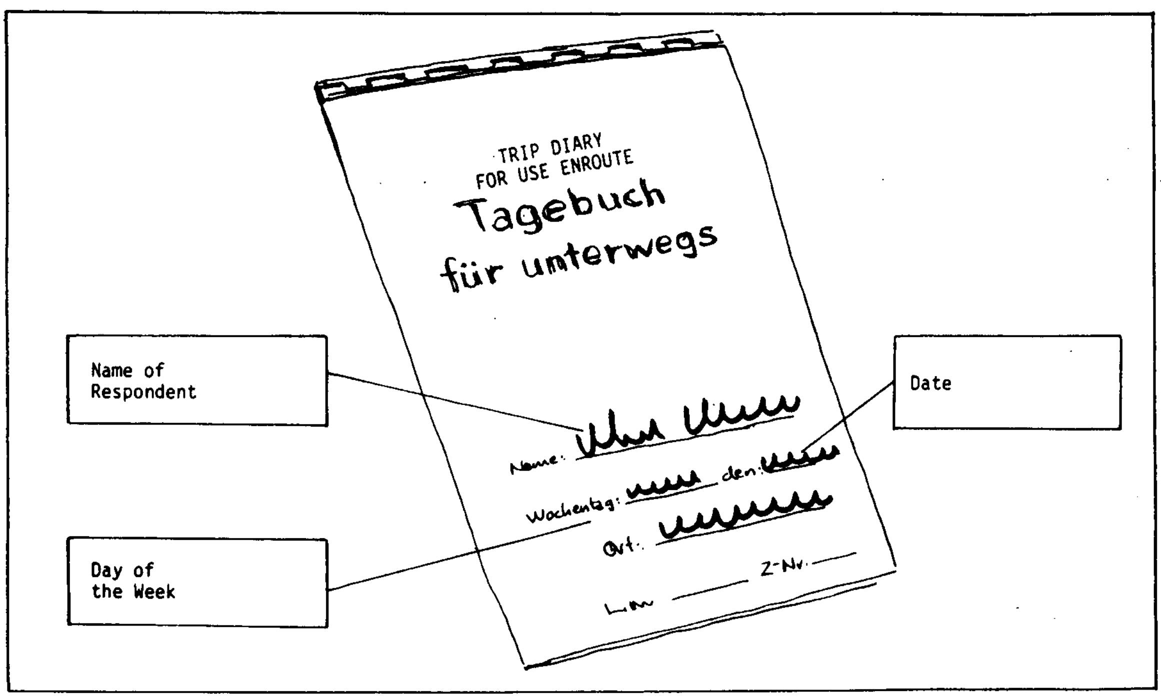


Figure 2. Inside of trip diary.

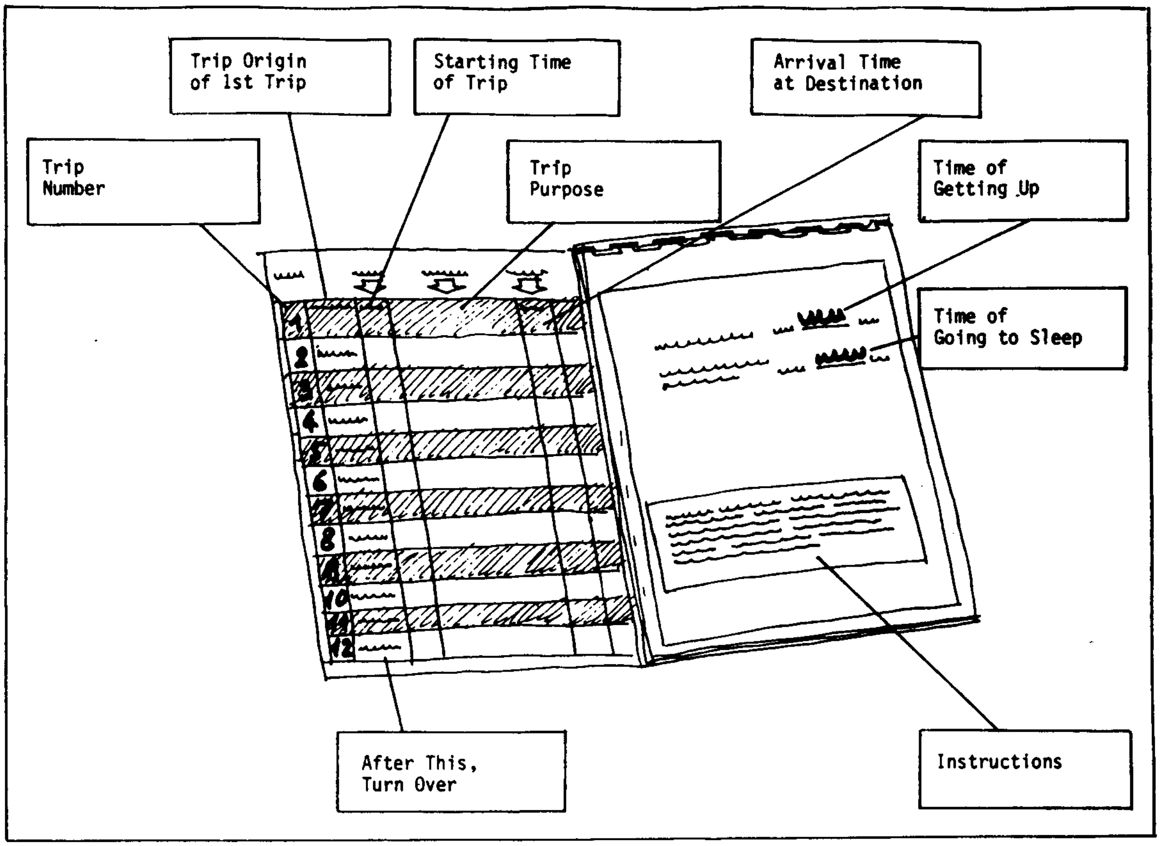


Figure 3. Trip register of diary for respondent.

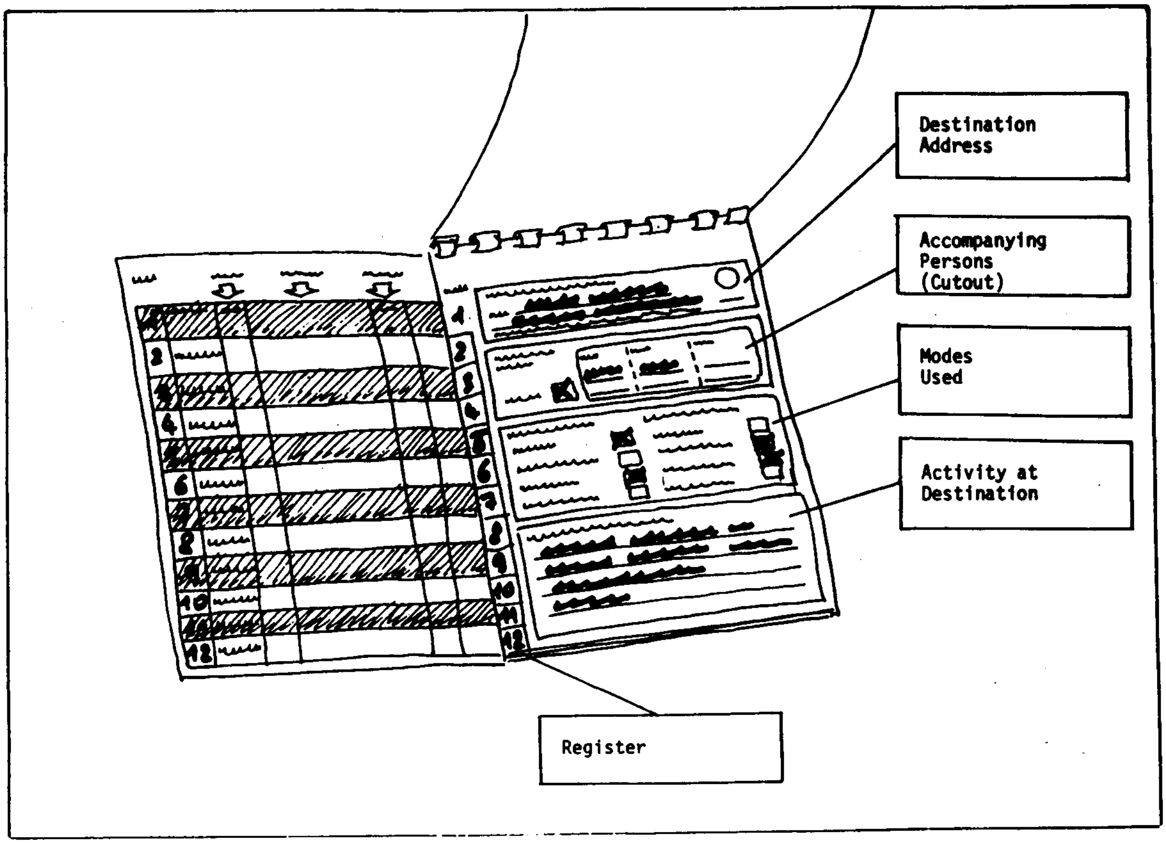
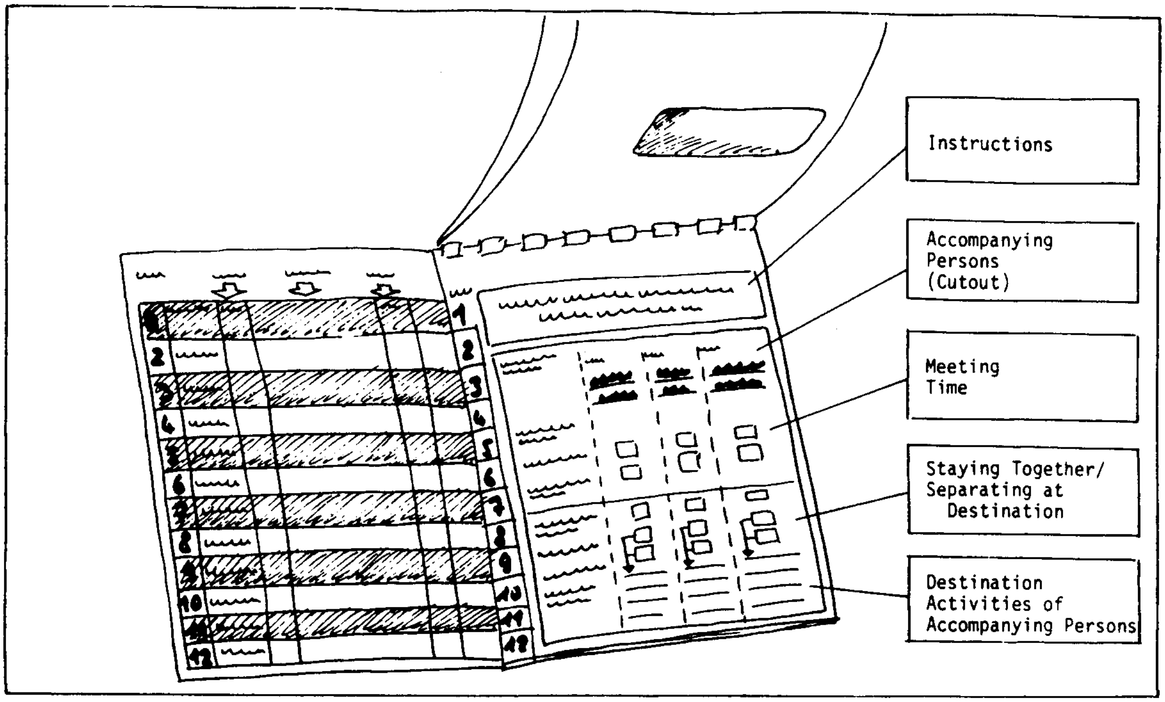


Figure 4. Trip register for accompanying person(s).



most important aspects of their sequence of activities during that day; i.e., location of the day's first activity (usually home), starting time of the first trip, activity associated with that trip (e.g., work), and time of arrival at destination.

All subsequent trips for that day were recorded according to the same pattern on the inside cover. Thus the temporal sequence of activities and the reasons (trip purposes) for the diverse nonhome activities were determined. At the same time, the format and layout of the instrument ensured that this rough record of daily activities could be outlined in the course of the day (i.e., en route, close to the time of the occurrence of any particular activity). This constituted the basis for the additional questions in the activities diary.

Separate survey sheets for each trip were afixed to the top of the inside right cover. There were two sheets for each trip; the first was to be used by the target person who was completing the diary. The second trip sheet referred to any possible accompanying traveler. These individual survey sheets were equipped with a register that made it simple to locate quickly the two sheets that belonged to any one trip. A color code was used for each trip that corresponded to the color scheme of even- versus odd-numbered trips recorded on the left inside cover.

The survey form for a specific trip performed by the respondent contained the following information:

1. Accurate address of destination,
2. Specification of up to three accompanying persons (e.g., neighbor, son, uncle),
3. All travel modes used on a particular trip, and
4. Detailed description of the destination activity.

A window was cut in the space where the specification of the accompanying person was recorded so that this specification appeared on both sheets (for the respondent and the accompanying person) without the need to record the same information twice. The form for the accompanying person contained information as to whether that person had accompanied the respondent from the start of the trip, whether the person stayed with the respondent at the destination, and, if applicable, what the person did subsequently.

ORGANIZATIONAL PROCESS FOR USE OF ACTIVITY DIARY

The diary was intended to be completed by the respondents, but the demands on the respondents both in terms of time and contents comprehension were substantial, especially for first-time use. The necessary instructions could not be transmitted easily in writing to the respondent. Hence the use of interviewers was necessary, but they played the role of advisors rather than interviewers.

The procedure went as follows. First, the interviewer conducted a preinterview with the respondent, collecting the relevant sociodemographic data. The interviewer explained the structure of the diary and helped fill in the sequence of activities for the day before the interview. Then the diaries were handed to the respondent for subsequent unassisted reporting of activities on the specified survey days.

Finally, a postinterview was arranged to discuss the respondents' experiences with the diaries, to review the completed diaries, to make any corrections or additions that came to light at that time, and to collect the completed diaries. By this technique it was possible to determine how well respondents had fared with the diaries and how complete the recorded information was.

The technique of a personal trip diary represented significant progress both in terms of content and method. With respect to content, the diary, which required the reporting of entire activity sequences, by necessity also provided information for the transportation planner about walk and bicycle trips that had been ignored typically up to that time. The high share of nonmotorized travel in total individual mobility was registered with some

Figure 5. Timetable for interview work plan.

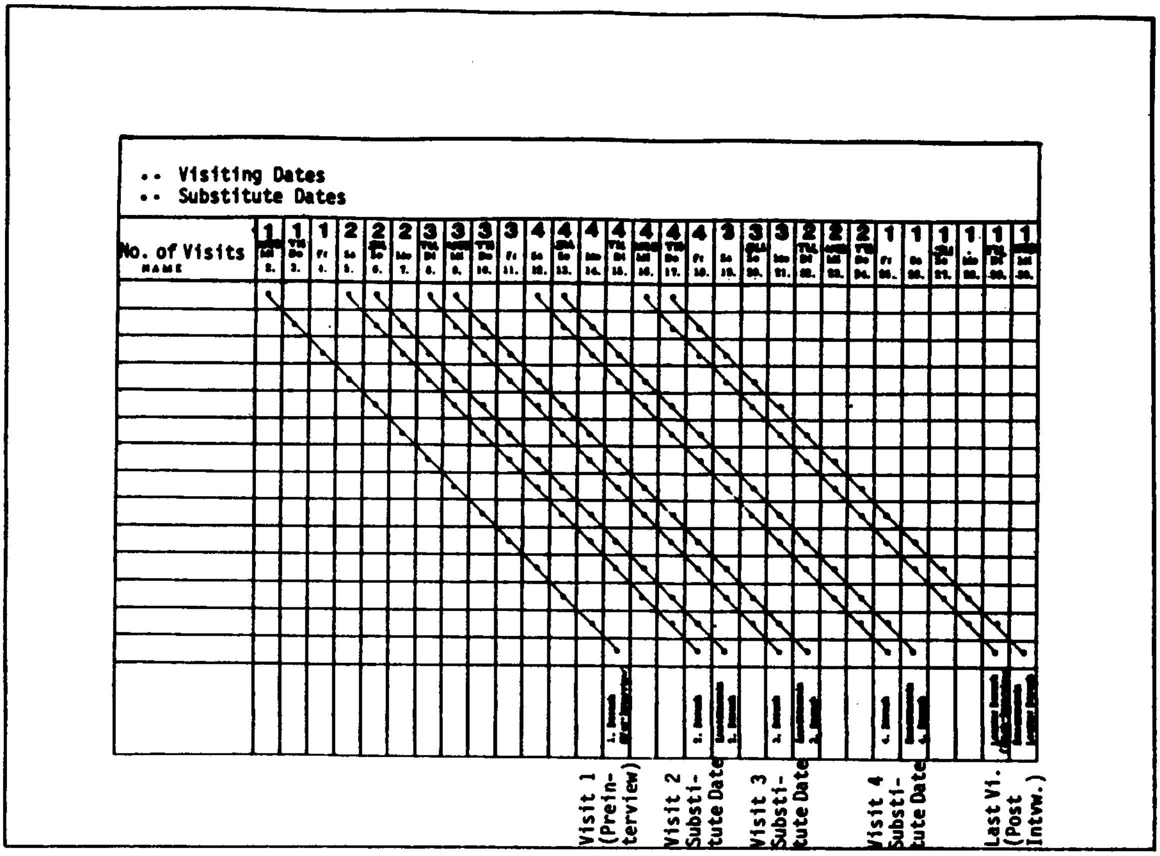


Table 3. Influence of interviewer on reported number of trips (8).

Day	Avg No. of Trips	Mobility Index (first day = 100.0)
1	5.14	100.0
2	4.90	95.3
3	4.66	90.7
Visit by interviewer		
4	5.02	97.7
5	4.66	90.7
6	4.76	92.6
7	4.43	86.2
Visit by interviewer		
8	4.82	93.8
9	4.45	86.6
10	4.67	90.9
11	4.74	92.2
Visit by interviewer		
12	4.83	94.0
13	4.52	87.9
14	4.48	87.2

surprise, at least in the Federal Republic of Germany.

From a methodological point of view, progress was achieved because travel behavior had not been recorded in general and average terms, but rather according to actual activities, and estimates had been replaced by methodologically superior techniques. Nevertheless, the problems remained that one survey day provided only a segment of an individual's mobility behavior, and that travel behavior could vary from day to day.

Based on these problems it was decided to investigate the travel activities of a population for two consecutive weeks, with each day requiring the completion of a separate diary. Because it could be expected that the motivation for completing these diaries would decrease with time, the interviewers took on the additional task of visiting the sample households and providing the respondents with renewed encouragement. Also, respondents were handed diaries for only 3 to 4 days at a time, which were then checked and exchanged against new ones for the next set of days. Only highly qualified and sensitive interviewers could be used for this difficult task. Therefore the sample was divided into several subsamples for which the survey weeks were staggered. Hence the interviewers did not have to conduct all preinterviews and postinterviews on the same days. Instead, they received a rather complicated work plan (see Figure 5) according to which they had to conduct the preinterviews, the repeat visits, and the postinterviews on specific days for specific households.

This form of survey organization permits a time-series investigation with diaries. It is clear, however, that such surveys have to be limited in terms of sample size because of organizational and financial constraints.

The evaluation of the data collected by means of these diaries indicates that the expensive advisory function performed by the interviewers was absolutely necessary. As indicated by the data in Table 3 (8), the number of trips recorded for the first day was highest, with all subsequent days showing a decline. This continuity was interrupted only for the days following a visit by an interviewer, i.e., the number of reported trips increased only to decrease again until the next visit.

FURTHER DEVELOPMENTS OF ACTIVITY DIARY

It became clear that, from a methodological point of view, this diary approach constituted the best instrument in the early 1970s. However, it was not suitable for use in large-scale surveys that cover large geographical or time dimensions. The objectives for further developmental work were the elimination of the interviewer (advisor) and the simplification of the diary to such a degree that self-administered, mail-back surveys would become feasible.

In the course of a new pretest series, the diaries were still delivered by interviewers. But the interviewers would only hand out an instruction sheet to the sample households, rather than providing detailed verbal explanations. The completed diaries were returned by mail, thus eliminating the possibility of checking the diaries for accuracy and completeness.

The these reasons, this pretest was subjected to a systematic error analysis of each diary, which revealed the following results:

1. About one-third of the diaries did not contain any recognizable errors,
2. About one-fifth contained mistakes that could be corrected subsequently by means of careful data preparation (e.g., missing return trips home, inaccurate destination address), and
3. Another fifth showed mistakes of such severity that the diary was unusable or only partly usable [see Table 4 (8), Version 1].

A more detailed analysis of the mistakes indicated that

1. Forty percent of the errors pertained to the trip destination address, most of which could be corrected subsequently;
2. Approximately 25 percent of the errors occurred in the trip-purpose specification, most of which could be corrected; and
3. A little less than one-quarter of the deficiencies pertained to incomplete information, mostly missing trips; only 14 percent of these could be reconstructed in the data preparation phase [see Table 5 (8), Version 1].

Overall, about three-quarters of the recognizable errors could be corrected (Table 5). This result was considered satisfactory. In principle, it appeared feasible to conduct such surveys with purely written instructions accompanying the survey instrument. The relatively high number of unusable or only partly usable survey responses were attributable to the complexity of the required recording procedure that had not been altered up to this stage in the development of the survey instrument.

Before tackling this particular issue, another problem had to be addressed, which pertained to the content of the survey instrument, namely the definition of the term trip and the recording of travel modes. Up to this version of the diary, a trip was understood as the activity that links two geographically separate places where the respondent pursued activities. Therefore, it was necessary to record all modes of travel that were necessary to overcome the spatial separation. This aspect resulted in the following issues:

1. It was possible that respondents did not record walk trips that were necessary in conjunction with the use of individual or public transportation modes;
2. If a travel mode had to be used repeatedly (e.g., different subway, bus, or street car lines), this mode could only be recorded once; and
3. The sequence of use for the different modes was not immediately discernable from the diary entries.

The methodological solution that eliminated these issues completely could only lie in the definition of trip as comprising each individual mode used on a specific travel segment. This meant that a separate survey sheet would have to be used for each change of mode. The obvious disadvantage was the increased reporting effort required of the respondent.

The results of a test with a diary that used the trip definition just outlined were as follows.

1. The number of usable diaries did not change.
2. The number of diaries with correctable minor errors increased slightly (see Table 4, Version 2).
3. The number of recorded trips per diary increased from 4.21 to 4.79, as was to be expected. Of course, this increase was directly related to the change in trip definition. In fact, when the number of trips were compared on the basis of the same trip definition, the second, more work-intensive version of the diary led to a reduction in the number of trips by about 10 percent.
4. The total number of errors per diary decreased from 3.41 in the pretest to 3.05, which was attributable mainly to improvements in the reporting of destination addresses. This is plausible because this address now was the parking garage, the bus stop, and so forth.

Table 4. Response quality for activity diary (8).

Item	Version 1[a]	Version 2[b]
Sample size	118	133
Usable diaries (%)		
Without mistakes	62	60
With small mistakes	18	20
Total	80	80
Unusable or only partly usable diaries (%)	20	20

[a]Every activity represents a trip. [b]Every mode used constitutes a trip.

Table 5. Reporting errors for activity diary (8).

	Total (n = 2,522)		Version 1[a] (n = 402)		Version 2[b] (n = 405)	
Item	Percent	Correctable Errors (%)	Percent	Correctable Errors (%)	Percent	Correctable Errors
Error in destination address	60	46	40	36	26	19
Error in trip purpose	20	16	28	24	52	45
Error in mode used	4	2	3	1	1	-
Error in specification of time	5	2	5	2	1	-
Incomplete reporting	11	6	23	14	21	8
Total	100	72	99	77	101	72

[a]Every activity represents a trip. [b]Every mode used constitutes a trip.

5. The number of incorrectable errors increased from 0.78 to 0.85 per daily diary. More than half of the errors pertained to trip-purpose information (see Table 5, Version 2).

These results suggested a return to the former trip definition because the problems that gave rise to a change in trip definition could be overcome by other means:

1. Walk trips as access and egress elements could be supplemented at the time of data preparation (verification);
2. The sequence of travel mode used and multiple use of a mode on a single trip could be constructed easily on the basis of origin-destination information, in case this is important information for a specific study; and
3. The majority of investigations that deal with explanation and prediction of, and the ability to influence, travel behavior are mainly directed toward the main mode used on a trip.

FROM ACTIVITY DIARY TO PERSONAL SURVEY FORM

From a methodological and theoretical point of view, it can be concluded that the diary met the requirements of methodological quality extremely well. Nevertheless, as stated previously, the use of a diary becomes problematic for large, possibly widely dispersed, populations. The financial and organizational costs for the necessary interviewer advice and for the instrument layout make it somewhat questionable.

This implied that a survey instrument had to be developed for large-scale surveys that maintained high methodological quality while at the same time was technically simpler and more suitable for self-administration by the respondents. With the survey content given (namely measurement of all trips during a day characterized by times, purpose, destination, and travel mode used), the following aspects gained importance in the further development of the survey instrument: formulation of questions, arrangement of questions, layout, and communications between respondents and survey administrators.

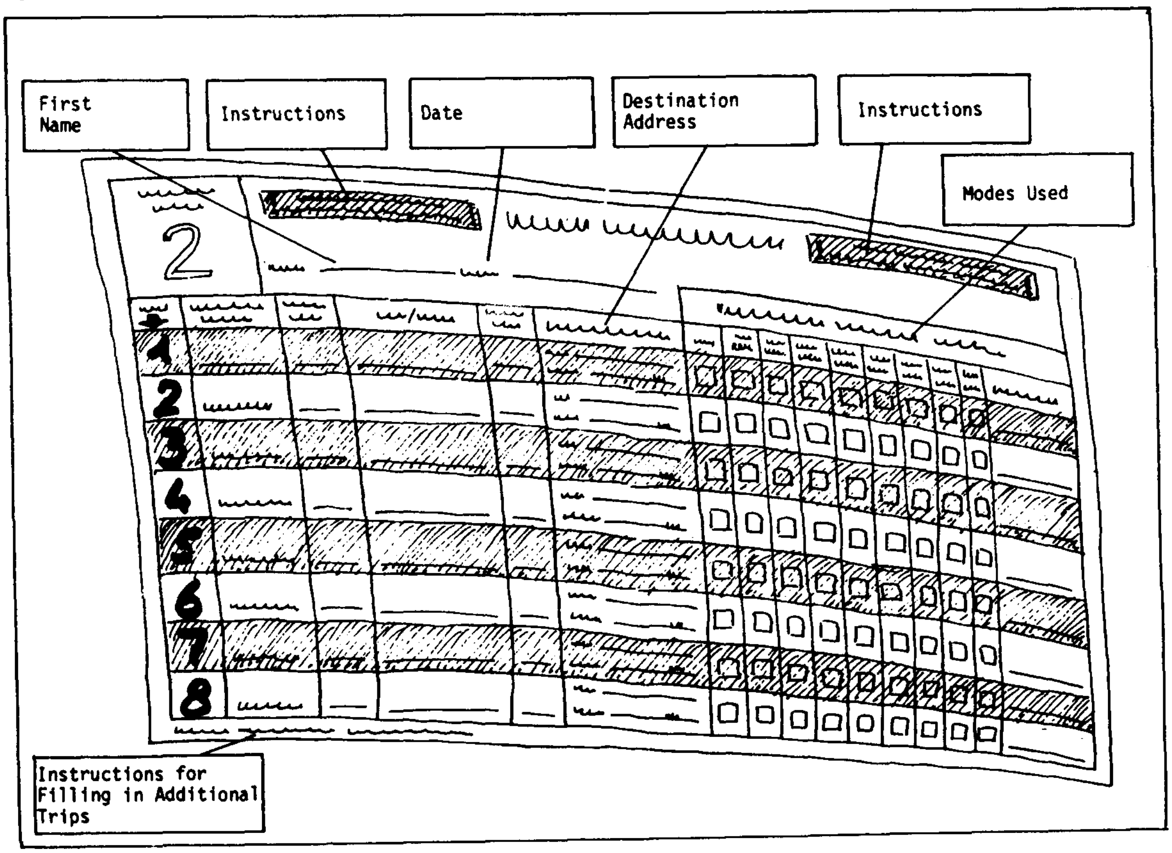
First Pretest Phase for Questionnaire Development

A multiphase pretest series was performed in order to transform the activity diary to a survey instrument suitable for large-scale surveys (9). The main effort during the first pretest concentrated on generating preferably a single-sheet questionnaire out of an extensive diary, while still being able to record all trips of a survey day. This requirement had several consequences: (a) the number of recorded trips had to be more limited, (b) the brief summary of the sequence of the day's activities (inside front cover of diary) had to be deleted, and (c) space for comments and open questions was to become limited.

Two questionnaires were developed for this first pretest that differed with respect to the formulation and arrangement of the questions and the layout. In the first questionnaire trips had to be recorded in rows. Trip purpose had to be entered in longhand rather than checked off on a preprinted listing. All trip characteristics could only be listed once. Each trip row contained fields for making longhand entries and squares for checkoff marks (Figure 6).

In the second version of this questionnaire trips had to be recorded in columns. Trip purpose had to be recorded in longhand. For each block the most frequent and obvious categories of answers were

Figure 6. Row version of questionnaire.



given for easy checkoff; all other answers had to be provided in longhand (Figure 7).

The results of this first pretest stage can be summarized as follows.

1. The percentage of usable forms for the column version of the questionnaire was higher (97 percent) than the row version (92 percent) [see Table 6 (10)].
2. Sixty-two percent of the reported trips contained incorrect or incomplete information; 46.4 percent were correctable [see Table 7 (8,10), First Pretest Phase].
3. Most deficiencies in reporting pertain to the destination address (41.9 percent of all trips), but most of them are minor problems because the majority of the addresses can be located, given the geographical aggregation level typically used in transportation planning (see Table 7, First Pretest Phase).
4. In the row version an increasing number of errors occurred with respect to trip purpose for the return trip home. This is attributable to the open form of the question used in this version.
5. The average number of daily trips measured in this pretest was 3.59 trips per person compared with

Figure 7. Column version of questionnaire.

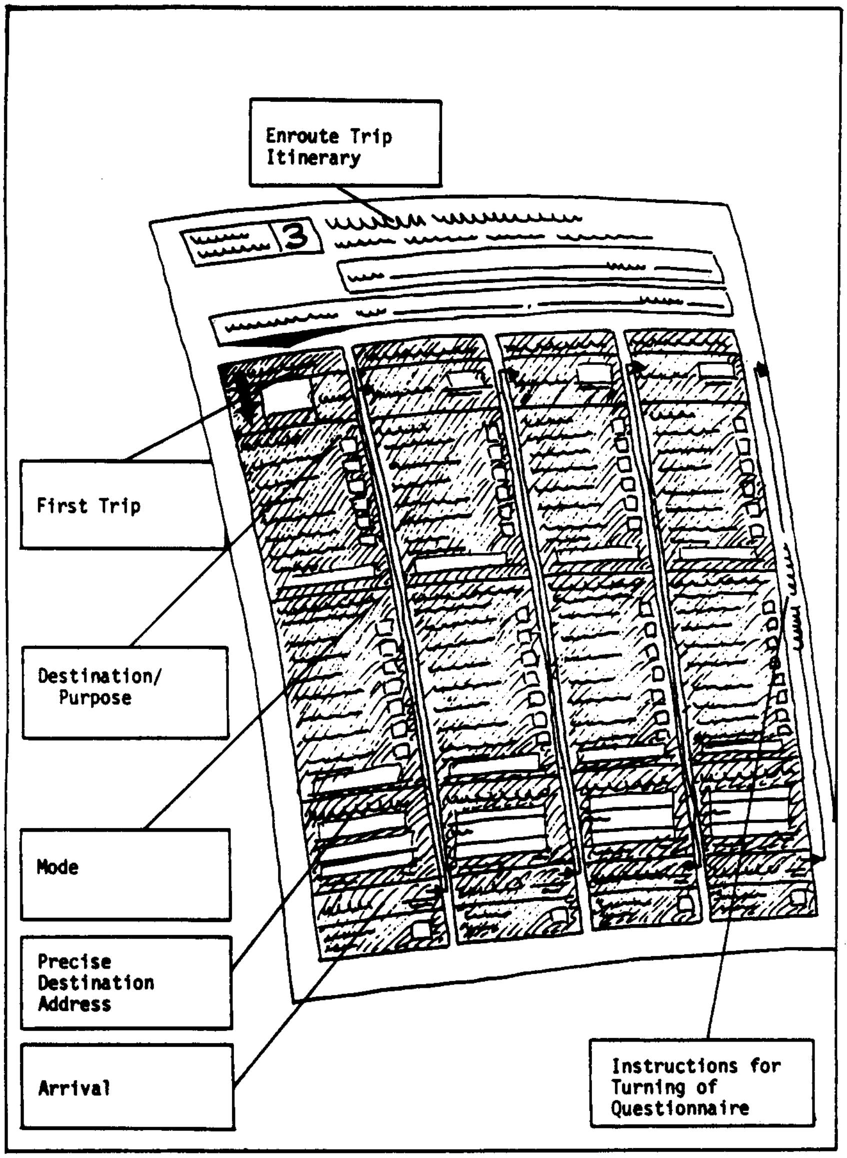


Table 6. Reporting quality for column and row versions of questionnaire (10).

Questionnaire Version	Sample Size	Usable Questionnaires (%) Without Error	Usable Questionnaires (%) Correctable Questionnaire	Usable Questionnaires (%) Total	Unusable or Partly Usable Questionnaire (%)
Column layout	59	88	9	97	3
Row layout	58	89	3	92	8

Table 7. Incorrect and incomplete reporting (8,10).

Pretest Phase	Incorrect Reports per 100 Trips by Trip Characteristic								Incomplete Reports per 100 Trips	
	Destination Address		Purpose		Mode		Timing of Departure and Arrival			
	Total	Noncorrectable	Total	Noncorrectable	Total	Noncorrectable	Total	Noncorrectable	Total	Noncorrectable
First	41.9	2.9	4.4	2.6	1.6	1.0	5.5	3.6	8.6	5.5
Second	29.6	8.1	4.7	1.6	1.9	0.8	0.8	0.4	2.0	0.4

4.21 trips reported in the diary. The reasons for this lie in the absence of an interviewer providing additional motivation for responding and in the layout of the questionnaire.

Second Pretest Phase

The second pretest again made use of the row and column versions of the questionnaire (see Figures 6 and 7). However, this time the layout was improved substantially. Dual color printing made the questionnaire more readable and visually more appealing. In the column version the fields and squares for recording answers and checkmarks, and in the row version all odd-numbered trips, appeared in a different color from that used on the rest of the form. Also, emphasis of certain important information was achieved through varying letter size and thickness.

These changes in layout were supposed to improve the results of the first pretest phase in two respects. First, the clearer distinction between individual trips impresses more on the respondent that all trips for a day were to be recorded. Second, the visual emphasis was supposed to reduce the share of unanswered questions because the respondent could see immediately where entries were expected to be made.

The second pretest phase is distinguishable from the first one mainly because the questionnaires were to be tested under the conditions of a mail-back survey; i.e., respondents had to master the questionnaire responses exclusively on the basis of the written instructions provided, and the respondents had to be motivated in writing to participate in the survey.

Two variations of the column version, distinguished by their different spatial arrangements, were developed for purposes of a mail-back survey. Both variations were printed on one sheet, one of them a folded version where all trips could be recorded across that page. The other version was printed on both sides of a smaller sheet, with the implication that the sheet had to be turned over after the first four trips had been recorded on the front. This last version, of course, had a postage cost advantage.

The results of this second pretest phase were as follows.

1. The number of reported trips increased from 3.59 during the first phase to 3.97 trips, which can be attributed to the improved layout. The remaining discrepancy with respect to the 4.20 trips per person per day obtained in the diary is explainable because no control and immediate corrections function can be provided in the mail-back questionnaires.
2. The row version contained the largest number of incomplete answers (39.9 percent of all trips), whereas the front and back column version contained the fewest (37.9 percent). These differences are not dramatic, but it should be emphasized that the number of errors was successfully reduced for all questionnaire versions compared with the first pretest phase [see Table 8 (9,10)].
3. The number of mistakes with respect to the destination address decreased from 41.9 to 29.6 percent. Unfortunately, the share of noncorrectable errors increased from 2.9 to 8.1 percent (see Table 7, Second Pretest Phase). It is worth mentioning that the first pretest phase was conducted in Munich, where a greater amount of professional deciphering of address information could be provided by the administering agencies (Socialdata GmbH and Technical University Munich) than in the case of the second pretest phase, which took place in other German cities. Of course, the three questionnaire versions used were identical; i.e., the destination address had to be provided in longhand [see Table 9 (9,10)].
4. The row version had more errors in the trip purposes, as was the case in the first pretest phase. Again, the reason was because of the open answer format (Table 9).
5. The number of unusable questionnaires and noncorrectable entries increased with the age of the respondent. Older people had particular difficulties with the accurate reporting of trip purposes.
6. For complicated trip sequences (i.e., those that involve more than travel to and from a single destination or involve several intermediate activities), the number of unusable responses was high. Trip purpose and destination address appeared to cause the most difficulties.

Table 8. Incorrect and incomplete reporting of trips in relation to different questionnaire versions (9,10).

Questionnaire Version	Reported Trips	Incorrect and Incomplete Trip Reports		Incorrect and Incomplete Reports per 100 Trips	
		Total	Noncorrectable	Total	Noncorrectable
Column version with foldout	1,384	540	146	39.0	10.6
Column version with front-to-back printing	1,148	436	138	37.9	12.1
Row version	1,253	500	144	39.9	11.5
Total	3,785	1,476	428	39.0	11.3

Table 9. Incorrect and incomplete trip reports by trip characteristic and questionnaire version (9,10).

Questionnaire Version	Incorrect Reports per 100 Trips by Trip Characteristic								Incomplete Reports per 100 Trips	
	Destination Address		Purpose		Mode		Timing of Departure and Arrival			
	Total	Noncorrectable	Total	Noncorrectable	Total	Noncorrectable	Total	Noncorrectable	Total	Noncorrectable
Column version with foldout	29.7	7.8	4.0	0.9	1.9	1.1	1.4	0.8	2.0	0.0
Column version with front-to-back printing	29.7	9.7	4.0	1.0	1.1	0.4	0.6	0.1	2.5	0.9
Row version	29.5	7.3	6.1	2.9	2.7	0.7	0.2	0.2	1.4	0.4
Total	29.6	8.1	4.7	1.6	1.9	0.8	0.8	0.4	2.0	0.4

In summary it can be concluded that the column version resulted in higher reporting accuracy. The decisive impetus to use this version in future surveys, however, was provided by a second criterion that was investigated in this pretest phase--willingness to respond.

The front-to-back variation on the column version led to a better response rate: approximately 80 percent as compared with the row version of about 70 percent.

Communication Between Survey Agency and Respondent

In the previous sections a distinction was made between two forms of communication: personal delivery and pickup of the survey forms (first pretest) and self-administered mail-back surveys (second pretest). The impact of these two methods on response accuracy was investigated. However, communication still has two additional important implications: response rate and survey cost per respondent.

These two aspects were investigated in another pretest series. Eight different forms of communication were tested, including a mix of personal and postal delivery and pickup. For the case of postal service use, additional distinctions were made as to whether prior notification by postcard was provided, and whether the recipients of the survey instrument received reminders by telephone on the actual prescribed survey day.

The results of these methodological tests were clear [Table 10 (8)]. Even the simplest postal service method (method 1) resulted in a better response rate (73 percent) than the most costly personal attention method (method 7) by means of interviewers (70 percent response rate). A response rate of 81 percent was achieved by means of the most expensive postal method [i.e., including notification and reminder by telephone (method 4)]. Even this method is less expensive than the least-expensive personal method (method 5).

On the basis of these results it was decided to conduct such surveys in writing by the mail-back process and to ensure as good a response rate as possible by written notification and reminder notices (8).

Further Aspects of Survey Instrument Design

Three additional aspects of questionnaire design that often are relevant in specific practical applications are as follows: (a) ease of coding for computer analysis, (b) consistency of questionnaire contents, and (c) surveys for foreign nationals.

Questionnaire Design for Computer Processing

Frequently, questionnaires were and are designed such that they meet the demands of researchers in the best possible manner. These demands and standards, however, often run counter to the needs of the survey respondent. Outstanding examples for this are the attempts to design the survey questionnaires in machine-readable form. A comparison of two substantially identical questionnaires, one in machine-readable format and the other with a normal layout, produced the following results (11):

1. The machine-readable form produced 10 percent fewer activities,
2. The number of deficient questionnaires was almost 3 times as high,
3. The number of unusable questionnaires was almost 4 times as high, and
4. With identical strategies for increasing the response rate, the machine-readable form produced a 66 percent rate and the normal layout a 79 percent rate.

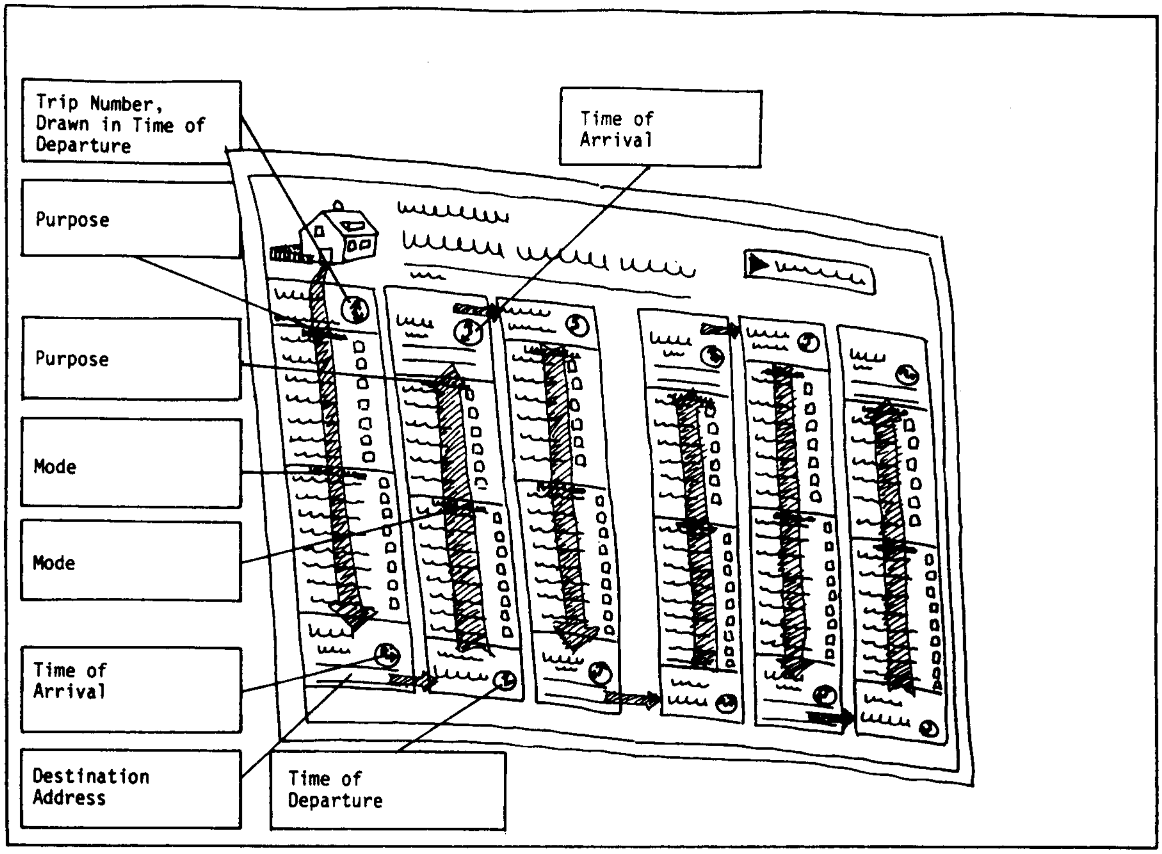
Consistency of Questionnaire Content

In addition to the design and layout, the questionnaire content has a significant effect on the willingness to respond. The logic of the questionnaire content (as perceived by the respondent) rather than the length is important. In this context it can be shown that it is feasible to transmit to the respondent the necessity of answering related and internally consistent sets of questions, but that the respondent's comprehension and willingness to respond is reduced markedly when this rule is violated.

Table 10. Response rates and survey cost as a function of questionnaire distribution and collection methods (8).

Distribution and Collection Method	Response Rate (%)	Cost-Index per Response	Sample Size (households)
Method 1-postal distribution and return	73	100	1,188
Method 2-notification, postal distribution and return	78	101	1,196
Method 3-postal distribution, reminder on survey day, postal return	77	104	1,193
Method 4-notification, postal distribution, reminder on survey day, postal return	81	113	1,191
Method 5-postal distribution, personal pickup	64	188	544
Method 6-personal delivery, postal return	63	215	517
Method 7-personal delivery, personal pickup	70	278	1,071

Figure 8. Column version of questionnaire for foreigners.



This point is illustrated in the following table on the basis of three surveys (4) with different degrees of internally logical sets of questions:

Item	Version 1: Complete Internal Logic	2: Partial Internal Logic	3: No Internal Logic
Sample size	55,107	19,380	12,091
Response rate (%)	81	77	67

Version 1 contained questions about demographics and nonhome activities (i.e., the internal logic was fully recognizable). Version 2 included additional, somewhat related questions (i.e., a logical unit was present, in part). Finally, in version 3 sets of questions of entirely different content were added (i.e., the logical unity was lost). The data in the table indicate that the response rate was affected quite substantially.

Surveys for Foreign Nationals

In several countries with sizable groups of foreign nationals it is sometimes necessary to survey this population segment of a specific study area. Typically, one of the following survey techniques is used. Either the foreigners receive the standard local-language form as it is distributed to the domestic population sample in the hopes that they have acquired sufficient local-language facility, or they receive a version prepared in their native language.

The second method obviously is the better approach, but it is not sufficient to generate adequate response in terms of numbers and quality. Because foreigners do not only differ in their native language but also in terms of mentality (e.g., perception of time), forms of expressions, and communications, a straight technical translation of the survey instrument cannot suffice to provide them with a survey form adequate for their needs (see Figure 8). In order to generate a questionnaire of equal content it was necessary to conduct similar types of pretest series as were described for the development of the local-language questionnaire in earlier sections of this paper. Different techniques and presentations had to be tested.

Such a questionnaire was developed for Turkish and Yugoslav residents of Berlin, Germany, and it was used in the context of a large-scale survey in that city (12). A meaningful comparison of the response quality between the German and foreign-language versions of the questionnaire can be made for the reporting of trip destinations because that aspect was probably most difficult for foreigners to answer accurately. The results indicated that the difference in response accuracy was insignificant, and it was certainly much better than had been observed in other surveys involving foreign residents [see Table 11 (12)].

Table 11. Example of response quality for German and Turkish and Yugoslav residents of Berlin, Germany (12).

Item	German Residents	Turkish and Yugoslav Residents
Sample size	19,000	2,000
Reporting quality of destination address (%)		
Directly usable	78	72
Usable with extra effort	20	18
Not usable	2	10
Response rate (%)	77	71

The questionnaire design for foreigners also has a direct impact on the response rate. A 13 percent difference in response rates could be observed between a straight technical translation and a specially designed survey form. According to the data in the following table (12), an additional increase of 9 percent was possible by means of special foreign-language telephone and written assistance and information:

Survey	Sample Size	Response Rate (%)
Straight technical translation	3,000	49
Specially designed survey form	1,084	62
Specially designed survey form with special assistance provided	2,712	71

CONCLUSIONS

The details of the developmental process involved in generating a survey instrument that meets criteria of high methodological quality, high expected response rates, suitability for large-scale surveys into travel behavior, and relatively low costs have been described. Through a number of real-world tests it was demonstrated that a variety of design aspects can have substantial influence on one or more of the preceding criteria. Each test series was tested empirically, with detailed documentation of reporting deficiencies.

The tests revealed how important methodological research into improved survey design can pay off in terms of better and more complete survey results and, hence, in terms of more reliable and valid inputs into travel modeling and transportation planning. Uncritical use of unproven survey instruments can have a profound influence on the efforts by transportation planners and policy decision makers.

In this paper the evolution of better travel survey instruments based on diary-generated information through research performed in Germany has been discussed. It should be made clear that many of the methodological insights gained in the course of these developments have been implemented in sophisticated travel data-collection efforts in the United States. Excellent examples of such efforts have been presented in two recent papers (13,14).

REFERENCES

1. A.H. Meyburg and W. Brög. Validity Problems in Empirical Analyses of Nonhome Activity Patterns. TRB, Transportation Research Record 807, 1981, pp. 46-50.
2. W. Brög. Mobility and Lifestyle--Sociological Aspects. Report of the 8th International European Conference of Ministers of Transport Symposium on Theory and Practice in Transport Economics (ECMT Publications Department of the Organization for Economic Cooperation and Development), Istanbul, Sept. 1979.
3. M.J. Wermuth. Effects of Survey Methods and Measurement Techniques on the Accuracy of Household Travel Behavior Surveys. In New Horizons in Travel Behavior Research (P.R. Stopher, A.H. Meyburg, and W. Brög, eds.), Lexington Books, Lexington, Mass., 1981, Chapter 28.
4. Sozialforschung Brög. KONTIV 75-77: Continuous Survey of Travel Behavior (Kontinuierliche Erhebung zum Verkehrsverhalten), Summary Vols. 1-3. Ministry of Transport, Munich, Germany, 1977 (in German).
5. W. Brög and W. Schwerdtfeger. Considerations on the Design of Behavioral Oriented Models from the Point of View of Empirical Social Research. In Transport Decisions in an Age of Uncertainty (E.J. Visser, ed.), Kluwer Academic Publishers, Hingham, Mass., 1977.
6. W. Brög. Latest Empirical Findings of Individual Travel Behavior as a Tool for Establishing Better Policy-Sensitive Planning Models. Presented at World Conference on Transport Research, London, England, April 1980.
7. W. Brög. Application of Situational Approach to Depict a Model of Personal Long-Distance Travel. TRB, Transportation Research Record 890, 1982, pp. 24-33.
8. Sozialforschung Brög. Travel Behavior in Connection with the Munich Commuter Train System (Verkehrsverhalten im S-Bahn-Bereich). Urban Development Office, Munich, Germany, Jan. 1973 (in German).
9. Sozialforschung Brög. Comprehensive Transportation Plan for Nuremberg (Gesamtverkehrsplan Nürnberg). Government of Middle Franconia, Munich, Germany, Preparatory Studies, Vol. 3, 1974 (in German).
10. M.J. Wermuth and G. Meerschalk. Representativeness of Written Mail-Back Household Surveys (Zur Repräsentanz Schriftlicher Haushaltsbefragungen). German Ministry of Transport, Munich, Germany, 1981 (in German).
11. Sozialforschung Brög. Investigation of Work and School Trips (Studie zum Berufs-und Ausbildungsverkehr). German Ministry of Transport, Munich, Germany, 1976 (in German).
12. Sozialforschung Brög. Transport Development Plan for Berlin [Verkehrsentwicklungsplan (VEP) Berlin]. Senator for Construction and Housing, Berlin, Germany, Final Rept., 1977 (in German).
13. P.R. Stopher and I.M. Sheskin. Toward Improved Collection of 24-H Travel Record. TRB, Transportation Research Record 891, 1982, pp. 10-17.
14. P. Rai, P.R. Stopher, and C. Brown. Origin-Destination Travel Survey for Southeast Michigan. TRB, Transportation Research Record 886, 1982, pp. 1-8.

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

Sequential, History-Dependent Approach to Trip-Chaining Behavior

RYUICHI KITAMURA

The characteristics of trip-purpose chains are examined, and a sequential model of trip chaining, which consists of history-dependent probabilities of activity choice, is developed. Statistical analyses of the study indicate that there is a consistent hierarchical order in sequencing activities in a chain where less-flexible activities tend to be pursued first. The analyses also indicate that the set of activities pursued in a chain tends to be homogeneous. Thus activity transitions are more organized and systematic than what a Markovian model would depict. Based on these findings, a sequential model of activity choice is formulated that, in spite of its simplified representation of the history of a chain, satisfactorily represents the observed behavior. Although the focus of the model is on direct linkages between activities, the model is capable of representing those characteristics associated with the entire set of activities in a chain. The results of the study strongly support the sequential modeling approach and indicate its practical usefulness in the analysis of trip-chaining behavior.

The importance of understanding trip-chaining behavior has long been recognized in connection with nonresidential trip generation (1) or with urban land use development (2). Underlying this is the dissatisfaction with the way tripmaking has been dealt with in the conventional transportation planning process or in location theory (3). As planning emphases in transportation shifted from infrastructure construction toward systems management and policy development, it was recognized that there was an increased need for a more fundamental understanding of travel behavior (4-6). The responses of urban residents to the recent oil crises (7,8) have made evident the importance of investigating trip-chaining behavior. Its importance is clearly seen when considering how the temporal and spatial distribution of trips in an urban area is affected by the way people organize their daily schedule of activities and combine trips. Statistical analyses have been accumulated to form a substantial body of empirical evidence [reviews of previous works on related subjects can be found in Hanson (5) and Damm (9)]. Yet many questions that have arisen in model-building efforts of trip-chaining behavior remain to be answered.

In this study one of the critical issues in trip chain modeling is addressed: representation of the decision structure involved in trip chaining. From the viewpoint that people plan and schedule beforehand a set of activities to be pursued in a trip chain, the decision process can be best represented as a simultaneous one that concerns the entire set. However, only few studies (10) have taken this approach in the past because of enormous difficulties involved in developing a practical simultaneous model of trip chaining. Most previous studies took sequential modeling approaches, which include the Markovian approach that has been traditionally used in trip chain analysis (1,3,11-14). The validity of the Markovian models, however, has not been thoroughly examined in the past, although several empirical observations (15-18) have indicated that trip chaining is not Markovian.

The objective of this study is to demonstrate that the inadequacy of previous sequential models is caused by their failure to represent patterns of activity sequencing and activity set formation in trip chaining, and further to demonstrate that trip-chaining behavior can be adequately described by sequential probabilities of activity choice that incorporate the history dependence of the behavior in a simple manner. The sequential approach has an obvious advantage because it represents the behavior by a simple model structure while avoiding combinatorial and other problems that may otherwise arise. At the same time, the approach may appear to be inconsistent with the viewpoint that trip chains are planned and scheduled beforehand while considering the entire set of activities, and not the transitions between activities. Whether a sequential analysis can adequately describe the behavior is, therefore, a critical question to be examined, because if the sequential approach is proven to be valid, it will lead to practical models of trip chaining that can be developed for a wide range of study objectives. This study is an effort to establish a basis for such development.

In examining the adequacy of the sequential approach, two aspects are discussed: sequencing of activities in a trip chain, and tendencies or preferences in formation of the set of activities to be pursued in a trip chain. (This study is concerned with types and sequences of activities in a chain, but not with their spatial or temporal attributes. A modeling effort that extends the present study into the temporal dimension can be found in a paper by Kitamura and Kermanshah presented elsewhere in this Record.) How these two aspects affect sequential probabilities of activity choice is demonstrated. Following this, empirical observations are made, and the nature of trip-chaining behavior is characterized.

BACKGROUND

The equivalence of the sequential and simultaneous approaches can be found in the following identity. By letting X_n be the nth activity in a trip chain for the case of three activities,

$$\Pr(X_1 = A, X_2 = B, X_3 = C) = \Pr(X_3 = C|X_1 = A, X_2 = B)\Pr(X_2 = B|X_1 = A)\Pr(X_1 = A) \quad (1)$$

The probability that a given set of activities is chosen and pursued in a given order can be represented by a set of sequential and conditional probabilities. (The same identity has been used in relating simultaneous and sequential formulations of discrete choice.) When the conditionality in Equation 1 is appropriately represented in sequential probabilities, then the sequential approach is equivalent to the simultaneous approach to trip chaining.

It may be argued that activity choice cannot be adequately described by probabilities that are conditioned only on the past; activity choice may also be dependent on future activities because a set of activities to be pursued may have been planned beforehand. Nevertheless, it can be seen that the backward dependency on the past implies forward dependency on the future as well. By using Bayes's rule,

$$\Pr(X_1|X_2) = [\Pr(X_2|X_1)\Pr(X_1)] / \left[\sum_{X_1} \Pr(X_2|X_1)\Pr(X_1)\right] \quad (2)$$

$$\Pr(X_1|X_2,X_3) = [\Pr(X_3|X_1,X_2)\Pr(X_2|X_1)\Pr(X_1)] \div \left[\sum_{X_1} \Pr(X_3|X_1,X_2)\Pr(X_2|X_1)\Pr(X_1)\right] \quad (3)$$

and so forth. A forward dependent probability can be always expressed as a function of backward dependent probabilities. That a choice is conditioned on the past implies that it is also conditioned on the future.

The preceding discussion indicates that the problems of previous sequential analyses, many of which used Markov chains, do not lie in their sequential structure, but rather they lie in their inadequate representation of the conditionality. In the following discussion it is assumed that there are patterns in sequencing activities in a set, and also that the choice probability of a given activity set is predetermined. The intensity of direct linkages between activities, or transition probabilities, which have been the main focus of previous studies, is viewed as a consequence of the patterns and preferences in choosing activity sets and sequencing activities. It is then shown that these patterns and preferences can be represented by the conditional transition probability, whereas the two Markovian assumptions--history dependence and stationarity (or time homogeneity)--are inadequate.

Suppose the number of activities in the set (denoted by k) is fixed and the individual is completely indifferent to the sequence of activities. Consider an activity set (ω) and two activity types (A and B). Because the sequencing is completely random, all the sequences obtained by permutating the activities in ω have the identical probability. Accordingly, for all ω,

$$\Pr(X_n = A, X_{n+1} = B|\omega) = \Pr(X_m = A, X_{m+1} = B|\omega) \quad (4)$$

and

$$\Pr(X_n = A|\omega) = \Pr(X_m = A|\omega) \qquad m,n = 1,2,\ldots,k-1 \quad (5)$$

Then, if A is included in at least one activity set,

$$\begin{aligned}\Pr(X_{n+1} = B|X_n = A) &= \left[\sum_\omega \Pr(X_n = A, X_{n+1} = B|\omega)\Pr(\omega)\right] \\ &\div \left[\sum_\omega \Pr(X_n = A|\omega)\Pr(\omega)\right] \\ &= \Pr[X_{m+1} = B|X_m = A]\end{aligned} \quad (6)$$

Namely, the pairwise activity transition probabilities are stationary. Note that this conclusion is not affected by the probability with which ω is chosen [$\Pr(\omega)$], i.e., it does not depend on the preferences in activity set choice.

Although the pairwise activity transition probabilities are stationary, they are not history independent even in this simplified case of random activity sequencing. Suppose that the choice probabilities of sets that include activities A, B, and D are zero, while those of other sets are positive. Then

$$\Pr(X_{n+1} = B| \ldots, X_\ell = C, \ldots, X_n = A) > 0 \quad (7)$$

and

$$\Pr(X_{n+1} = B| \ldots, X_\ell = D, \ldots, X_n = A) = 0 \quad (8)$$

Therefore, $\Pr(X_{n+1}|X_1, X_2, \ldots, X_n) \neq \Pr(X_{n+1}|X_n)$. For the activity transitions to be Markovian, the probabilities with which respective activity sets are chosen must conform with those depicted by the transition matrix of a Markov chain, a condition rather groundless to assume.

The pattern of sequencing activities in a trip chain is another source of history dependence, which also yields nonstationarity. Suppose activity A tends to be pursued before B, but the individual is indifferent to the sequencing of activity C. Then for ω that involves A, B, and C, the probability $\Pr(X_{n+1} = B| \ldots, X_n = C, \omega)$ varies depending on whether A has been pursued before C. Now suppose both A and B tend to be pursued earlier in a chain, but again C is equally likely to be pursued in any order. Then, $\Pr(X_{m+1} = A|X_m = C, \omega) > \Pr(X_{n+1} = A|X_n = C, \omega)$ if $m < n$. The first example indicates that sequencing patterns cause history dependence, and the latter indicates that pairwise transitions become nonstationary.

Any Markov chain exhibits certain patterns of activity set formation and sequencing. But the reversal is not always true; i.e., given patterns of set formation and sequencing cannot always be represented by a Markov chain. The discussion in this section also implies that sequencing and activity set formation can be represented when the conditional probabilities of activity transitions are appropriately specified. The failure of Markov chain models is caused by their invalid representation of the conditionality. In the following sections characteristics of trip chaining are first observed, and then a sequential model is proposed.

DATA SETS

Empirical observations of this study are made by using the 1965 Detroit area transportation and land use study (TALUS) data set, the 1977 Baltimore travel demand data set, and published transition frequency matrices from Chicago, Buffalo, and Pittsburgh [reported by Hemmens (19)]. The TALUS data set is most extensively analyzed, whereas the other sets are used to examine the generality of the results obtained. A significant advantage of the TALUS data set--a conventional origin-destination survey result--is its ample sample size, which is crucial for the analysis carried out in this study.

The original TALUS data file, which contained records of 320,090 trips made by 82,050 individuals, was screened to exclude those individuals who did not have a closed series of trips that originated and terminated at home (which may include intermediate returns to home), who had no car available to the household or did not hold a driver's license, who were younger than 18 years old, who used travel modes other than car, and those who made work trips on the survey day (walk trips are not recorded in the TALUS data unless they were work trips). The last criterion is introduced because of the substantial differences in travel and time use patterns between those who worked and those who did not on the survey day (20,21). As a result of this screening, the sample analyzed includes 76,025 trips and 27,901 trip chains made by 16,520 individuals [a geographical subsample of this was used in previous studies (16,20,21)].

All screening criteria are also applied to the Baltimore data set, and a sample of 1,789 trips and 697 trip chains made by 435 individuals is obtained. The transition frequency matrices from the other three metropolitan areas include all observations without comparable screening. As is clear from the screening criteria, the internal homogeneity of the sample is emphasized in this study, whereas some aspects of travel behavior are placed out of its scope, such as the effect of travel mode on trip chaining. Individuals with transit trips are eliminated for this reason, and they are not analyzed because their sample size is too small for statistical analysis.

The 27,901 trip chains in the sample from the TALUS data set contain 48,124 sojourns with an average chain length (average number of sojourns per chain) of 1.725. Although 62.2 percent of the total chains are single-sojourn chains, they account for only 36.0 percent of the total sojourns, and approximately two-thirds of the sojourns belong to multisojourn chains. The significance of multisojourn chains is evident. The average chain length of the Baltimore sample is 1.57, approximately 10 percent less than that of the TALUS sample. The average number of chains per person is 1.60, which compares with 1.689 of the TALUS sample.

The direct transitions between activities in trip chains in the TALUS and Baltimore samples were first analyzed by using a transition matrix, with the assumption that trip chaining can be represented by a stationary and history-independent Markov chain. These two samples are different from those of other studies in that the individuals who made work trips are excluded. Nevertheless, this preliminary analysis of the pooled transition matrices indicated that the present samples share many of the trip-purpose linkage patterns reported in the literature (1,11,19).

NONSTATIONARITY OF ACTIVITY TRANSITIONS

Although traditional Markov chain analysis (which uses the pooled transition matrix) offers a convenient means of data summarization, the implicit stationarity assumption that the same transition matrix applies to all transitions in a trip chain is too restrictive for rigorous analysis of the behavior. In this section the nature of trip chaining is explored by using a nonstationary Markov chain, where each step of transition has its own transition matrix that is not necessarily identical to those of other steps (the first step of transition refers to the transition from the first purpose to the second, the second step of transition is the one from the second purpose to the third, and so forth).

Nonstationarity in Trip-Purpose Chains

Nonstationarity in the observed trip-purpose transitions is statistically examined by applying the likelihood-ratio test (22). The results are summarized in Table 1. To eliminate empty cells in the frequency matrices for as many steps as possible, two trip purposes with fewer observed frequencies are merged with others, as indicated in the table. The test is conducted for the first nine transition matrices, and also for the eight matrices from steps 2-9. The null hypothesis is strongly rejected in both cases.

Together with the overall chi-square values, the data in Table 1 present chi-square statistics for the row and column of each trip purpose, where the row total represents the nonstationarity in the transition probabilities from the trip purpose, and the column total represents that to the trip purpose. For the first case (steps = 1, 2, . . ., 9), all rows and columns have significant statistics, except the column total for shopping, which indicates that shopping is pursued with a relatively stationary probability throughout a chain. The large chi-square value associated with the transitions to serve-passenger trips and that from social-recreation trips are also noted.

The second test excludes the transition matrix of the first step. The drastic reduction in the overall chi-square value from the first test indicates the extreme distinctiveness of the matrix for the first transition. Note that the first transition determines whether the individual pursues only one or more than one sojourn in a trip chain. The data in Table 1 also indicate that the variation in linkages with serve-passenger trips is a major source of nonstationarity in the second step and thereafter.

The pairwise distinctiveness of two successive transition matrices was also tested, and the first four matrices were found to be significantly different from each other (with chi-square values of 783.3 between the first and second steps, 92.1 between second and third, and 38.9 between third and fourth, all with df = 16). No significant difference was found after the fourth step. This is at least partly caused by the reduced sample size in the transition frequency matrices of later steps. At the same time, the implication of the result that the transition probabilities are stabilized in later steps of a trip chain is intuitively agreeable.

Variations in Linkage Patterns

The nonstationarity in trip-purpose transitions implies that a pair of activities may have strengthening or weakening linkages with each other, and that some activities tend to be pursued earlier or later in a chain. The data in Table 2 indicate by step of transition those trip-purpose pairs for which more than expected transitions are observed in respective steps. Many of the diagonal cells are significant in all steps, which indicates that activities of the same type continue to have strong linkages among themselves throughout the chain. There are also several trip-purpose combinations that are significant only in the first few steps or in later steps.

Table 1. Likelihood-ratio test of stationarity in trip-purpose transitions.

Trip Purpose	For s = 1, . . . , 9 Row Total[a]	For s = 1, . . . , 9 Column Total[b]	For s = 2, . . . , 9 Row Total[c]	For s = 2, . . . , 9 Column Total[d]
Home	–	357.6[e]	–	77.9[e]
Personal business[f]	337.6[e]	155.2[e]	46.0[g]	39.9
Social-recreation[h]	603.7[e]	157.4[e]	38.8	65.7[e]
Shopping	374.1[e]	36.5	93.0[e]	26.9
Serve passengers	270.1[e]	878.7[e]	132.2[e]	99.6[e]
Total[i]	1,585.4[e,j]	1,585.4[e,j]	310.0[e,k]	310.0[e,k]

Note: In places where degrees of freedom are indicated, the df for the column total cannot be defined in the conventional manner; therefore the ratio [(total df) ÷ (no. of columns)] is presented here.

[a] df = 32.
[b] df = 25.6.
[c] df = 28.
[d] df = 22.4.
[e] Significant at $\alpha = 0.005$.
[f] Includes school.
[g] Significant at $\alpha = 0.05$.
[h] Includes eat-meal trips.
[i] A definition of the log-likelihood ratio statistic is given in Anderson and Goodwin (22).
[j] df = 128.
[k] df = 112.

Table 2. Salient trip-purpose linkages in nonstationary transition matrices for steps 1-4.

Category	HOME	PBNS	SREC	MEAL	SHOP	SCHL	SVPS
PBNS		1,2,3,4			1		
SREC	1		1,2,3,4	1,2,3			
MEAL			1,2,3				3
SHOP	1,2				1,2,3,4		
SCHL	1						2
SVPS	3		1			1	1,2,3,4

Note: Steps 1 through 4 indicate the step of transition for which the cell has a chi-square value of 7.879 or greater with an expected frequency of 5 or greater. Abbreviations for trip-purpose categories are as follows: PBNS = personal business, SREC = social-recreation, MEAL = eat meal, SHOP = shopping, SCHL = school, and SVPS = serve passengers.

Especially noted is the transition from serve-passenger trips to home in the third step. The significance of this trip-purpose combination in this particular step is caused by the dominance of the trip-purpose sequence: serve passengers to other activity to serve passengers to home (a later section indicates that this is a typical sequence when a trip chain involves serve-passenger trips). Thus the result suggests that the observed nonstationarity is partly caused by the sequencing by the tripmaker of the activities within a trip chain.

The variations in trip-purpose linkages were further characterized by evaluating for respective steps the mean first passage times (MFPTs); that is, the expected number of transitions from an origin state until a destination state is visited for the first time (23). The result indicated that the linkages to personal business become weaker in later steps of a chain. On the other hand, the MFPTs to serve-passenger and social-recreation trips revealed strengthening linkages between these activities and others in later steps.

This analysis of nonstationarity in trip-purpose chains strongly suggests the existence of patterns in sequencing activities. An earlier section indicated that another possible source of the observed nonstationarity is the dependence of activity choice on the set of activities already pursued, which is closely related with the preferences in the choice of activity set. In the following sections these two aspects are discussed, and the reasons why such nonstationarity exists in trip-chaining behavior are illuminated.

ACTIVITY SEQUENCING IN A TRIP CHAIN

Consider the transition frequency matrix presented in Table 3, which gives direct transitions between activities in 10,555 multisojourn trip chains in the TALUS sample. The matrix is obviously not symmetric, i.e., the frequency of (i,j) transitions is not always similar to that of (j,i) transitions. Examination of this asymmetric nature of the matrix leads to inferences as to the sequencing of activities within a trip chain. Suppose that three activities (A, B, and C) are to be pursued in a chain. If the tripmaker is completely indifferent to the sequence of these activities, all of the 3! possible sequences would have the equal likelihood of occurrence, and the occurrence of each one of the 6 (= ${}_3C_2$ 2) possible direct transitions would have the identical probability. Accordingly, the observed transition frequency matrix must be symmetric. The asymmetric matrix of Table 3, therefore, suggests that certain activities tend to precede others in multisojourn chains.

Table 3. Asymmetry of pooled transition frequency matrix.

Category	PBNS	SREC	MEAL	SHOP	SCHL	SVPS
PBNS	1,527[a]	815[b]	212[b]	1,820[b]	27[c]	515[c]
SREC	462[d]	1,563[a]	285	1,091	15[d]	722
MEAL	114[d]	277	16[a]	191	17	158
SHOP	844[d]	1,122	188	3,109[a]	8[d]	687[d]
SCHL	46[e]	43[b]	27	58[b]	23[a]	59[f]
SVPS	618[e]	737	140	1,030[b]	93[g]	1,564[a]

Note: Abbreviations are defined in Table 2. The footnotes in the table, except a, give the significance of the asymmetry between (i, j) and (j, i) cells.

[a]Not part of the examination of asymmetry.
[b]Observation greater than expectation; significant at $\alpha = 0.005$.
[c]Observation less than expectation; significant at $\alpha = 0.05$.
[d]Observation less than expectation; significant at $\alpha = 0.005$.
[e]Observation greater than expectation; significant at $\alpha = 0.05$.
[f]Observation less than expectation; significant at $\alpha = 0.01$.
[g]Observation greater than expectation; significant at $\alpha = 0.01$.

Tendencies in Activity Sequencing

Examination of the pooled transition frequency matrix of Table 3 indicates that the transition frequencies that are statistically most asymmetric involve personal business; for example, 815 transitions from personal business to social-recreation versus 462 transitions from social-recreation to personal business; 212 transitions from personal business to eating meal versus 114 from eating meal to personal business; and so forth (the differences are significant at $\alpha = 0.005$). Obviously, personal business tends to be pursued in a chain before the other activities. School trips have a similar tendency, and they precede personal business trips more frequently, and serve-passenger trips have a tendency to precede school and personal business trips.

There are also several pairs of trip purposes of whose sequences the tripmaker is apparently indifferent: 285 transitions from social-recreation to eating meal versus 277 from eating meal to social-recreation; 1,091 from social-recreation to shopping versus 1,122 from shopping to social-recreation; and so forth. None of these differences is statistically significant at any appropriate level.

Nine of the 15 (= ${}_6C_2$) pairs of different trip purposes have statistically significant asymmetry ($\alpha = 0.05$). Based on these relationships, a hierarchy diagram is constructed to show the tendencies in sequencing activities within a trip chain (Figure 1a). The perfect consistency in the hierarchical relationship among the trip purposes is shown in the figure; for example, serving passengers, which precedes school, also precedes those trip purposes that school precedes. These consistent tendencies in the observed direct transitions are quite noteworthy.

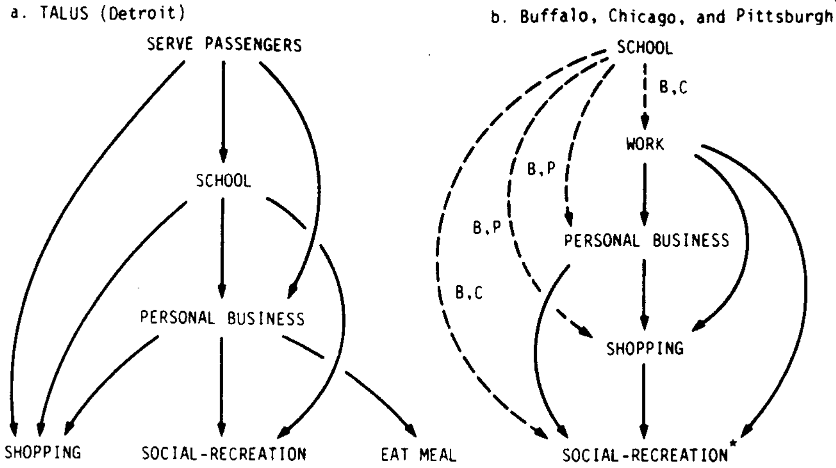
Hierarchical relationships among activities are evaluated in the same manner by using transition frequency matrices from Chicago, Buffalo, and Pittsburgh; these are summarized in Figure 1b. The result is in satisfactory agreement with the TALUS result. This is also the case for Baltimore, but the sample size is insufficient to be conclusive.

Activity Sequencing and Uncertainty

The hierarchical order of activities presented in Figure 1a,b indicates that activities in the higher order tend to be accompanied with spatial or temporal fixity, or both. For example, serving a passenger quite often implies that a person must be chauffeured to a given location by a given time, personal business such as banking must be pursued at a predetermined location, and so forth. The result indicates that activities of less flexibility tend to be pursued in a trip chain before more flexible activities, such as social-recreation and shopping. Cullen and Godson (24) argued that an individual's itinerary for a day is formed by articulating activities with less fixity around those activities with high spatial or temporal fixity or both, which act as pegs in daily activity scheduling. A previous study (20) revealed that serve-passenger trips largely prescribe an individual's daily travel pattern because of their fixity. The present study reveals another tendency in urban travel behavior: a relationship between sequencing of activities and their fixities.

The information available from the data set does not allow statistical determination of the reason why this sequencing pattern is observed. Nevertheless, the consistent observations from the four metropolitan areas offer the basis for constructing behavioral inferences on the subject. A rather

Figure 1. Hierarchy in activity sequencing in trip chains: TALUS (Detroit), Buffalo, Chicago, and Pittsburgh.



[1] A solid arrow indicates that the hierarchical relationship is observed in all three metropolitan areas. A broken arrow indicates that the relationship is observed in the area indicated by the initial (e.g., "B" for Buffalo).

*Includes eating meals. Serving passengers is excluded from the original tabulation

Note: Transit frequency tables are reported in Hemmens (19).

straightforward conjecture postulated here is that the sequencing pattern observed in this study is a result of individuals' consideration of uncertainty in activity planning.

Consider the case where an individual is combining both fixed and flexible activities into a chain. Quite typically, the exact amount of time required to accomplish an activity is not known to the individual beforehand. If a flexible activity is pursued first, and if it takes longer than initially thought, then the individual may not be able to attend the fixed activity in time. Note that an activity with spatial and temporal fixity by definition demands the individual to be at a certain location by a certain time. On the other hand, if the flexible activity takes less time, an unexpected block of time must be somehow spent. In either case, if the individual recognizes this uncertainty, it appears logical for him to pursue the fixed activity first. The observed activity sequencing pattern thus suggests that uncertainty plays a significant role in the activity planning of an individual. The pattern is perhaps a result of an individual's effort to minimize risks because of the uncertainty and to pursue a set of activities efficiently in a trip chain.

HISTORY DEPENDENCE IN ACTIVITY CHOICE

An earlier section indicated that preferences in activity set choice in general make activity transitions history dependent. The sequencing pattern observed in the previous section implies that activity choice depends on the series of activities already pursued in a trip chain. The strong direct linkages among activities of the same type also suggest history dependence. However, little exploration of the nature of history dependence in trip chaining has been made in the past, and most analyses were concerned only with direct linkages between pairs of activities. The analysis of this section, which focuses on the entire series of activities in trip chains, reveals additional characteristics of activity set formation and activity sequencing.

Although there are many possible ways of statistically examining the history dependence in trip chaining [e.g., triples used by Parkes and Wallies (25); also see Anderson and Goodman (22)], most of them encounter problems with sample size because of the scarcity in the sample of chains with a large number of sojourns. Accordingly, this study takes on an approach of tabulating the frequency of chains by the trip-purpose sequence and directly examining the history-independence assumption by using a contingency table analysis technique.

History Dependence of Three-Sojourn Chains

Consider those trip chains with three sojourns, namely, X_1, X_2, and $X_3 \neq$ home, and X_4 = home. The history-independence assumption can be stated for these chains as

$$\Pr(X_3 = k|X_1 = i, X_2 = j) = \Pr(X_3 = k|X_2 = j) \qquad (9)$$

for all i, j, and k $\neq$ home. Namely, the conditional probability that the third activity is k given the second activity (= j) is independent of the first activity (= i). This null hypothesis can be tested by tabulating, for given X_2, the frequencies of the third activity categories by the first categories, then by examining the independence of the resulting two-way contingency table. This contingency analysis is equivalent to applying a nonstationary Markov chain of the first order to test the history independence of three-sojourn chains. The results for 2,760 three-sojourn chains found in the TALUS sample are given in Table 4. To ensure a sufficient number of observations for each sequence of trip purposes, the original six trip-purpose categories are collapsed into four, as in Table 1.

In part A of Table 4 the results for those three-sojourn chains whose second trip purpose is personal business (including school) are presented. The row represents the first trip purpose, and the column represents the third trip purpose. If the history-independent assumption holds, then every row should have the same distribution of cell frequencies. The expected cell frequencies under this independence assumption are shown in parentheses.

As expected, the four contingency tables (parts A-D, Table 4) are all highly significant, which indicates that the conditional probability that a certain activity is pursued as the third activity, given the second one, does depend on the first activity pursued in the chain. Especially notable are the much higher-than-expected frequencies of the diagonal cells; tripmakers tend to repeat the same type of activity as the first and third activities in a trip chain. This recurrence of the same activity type is particularly noticeable for serve-passenger trips when the second purpose is not serving passengers (see parts A-C of Table 4). The diagonal cell for serving passengers alone accounts

Table 4. Frequencies of three-sojourn chains by sequence of trip purposes.

A X_2 = PBNS

X_1	X_3: PBNS	SREC	SHOP	SVPS	Total
PBNS	107 (66.2)	38 (42.5)	87 (84.6)	13 (51.7)	245 [54.7]
SREC	11 (20.3)	36 (13.0)	22 (25.9)	6 (15.8)	75 [51.5]
SHOP	24 (30.2)	16 (19.4)	65 (38.7)	7 (23.6)	112 [31.5]
SVPS	23 (48.3)	16 (31.1)	37 (61.8)	103 (37.8)	179 [143.1]
Total	165 [44.0]	106 [49.0]	211 [28.5]	129 [159.3]	611 [280.9]

B X_2 = SREC

X_1	X_3: PBNS	SREC	SHOP	SVPS	Total
PBNS	50 (17.6)	45 (65.8)	50 (35.9)	22 (47.7)	167 [85.3]
SREC	19 (29.4)	183 (109.5)	47 (59.8)	29 (79.4)	278 [87.8]
SHOP	9 (12.5)	51 (46.5)	50 (25.4)	8 (33.7)	118 [44.9]
SVPS	3 (21.5)	23 (80.3)	18 (43.9)	160 (58.2)	204 [249.9]
Total	81 [80.0]	302 [97.3]	165 [47.4]	219 [243.1]	767 [467.8]

C X_2 = SHOP

X_1	X_3: PBNS	SREC	SHOP	SVPS	Total
PBNS	82 (42.1)	50 (66.1)	163 (153.2)	16 (49.7)	311 [65.2]
SREC	7 (24.9)	85 (39.1)	66 (90.6)	26 (29.4)	184 [73.9]
SHOP	46 (51.6)	67 (80.9)	254 (187.6)	14 (60.9)	381 [62.6]
SVPS	9 (25.4)	24 (39.9)	41 (92.6)	114 (30.0)	188 [280.4]
Total	144 [61.9]	226 [66.6]	524 [59.5]	170 [294.0]	1064 [482.1]

D X_2 = SVPS

X_1	X_3: PBNS	SREC	SHOP	SVPS	Total
PBNS	11 (6.8)	10 (8.5)	10 (10.3)	17 (22.5)	48 [4.2]
SREC	6 (6.4)	15 (7.9)	8 (9.6)	16 (21.1)	45 [7.8]
SHOP	7 (6.4)	14 (7.9)	15 (9.6)	9 (21.1)	45 [14.7]
SVPS	21 (25.5)	17 (31.7)	35 (38.5)	107 (84.3)	180 [14.0]
Total	45 [3.5]	56 [18.1]	68 [3.6]	149 [15.6]	318 [40.7]

Personal business (PBNS) includes school, and social-recreation (SREC) includes eating meal. For other abbreviations, see Table 2.

( ): Expected cell frequency
[ ]: Row, column, or grand total of chi-square values.

for 40.0 percent of the total chi-square value of part A where the second purpose is personal business. The corresponding values are 38.1 percent for part B (X_2 = social-recreation), and 48.8 percent for part C (X_2 = shopping). The sequence of serve passengers to other activity to serve passengers is observed much more frequently than the expectation under the history-independence assumption, and it is found in 12.2 percent of the all three-sojourn chains, or in 36.6 percent of those three-sojourn chains that involve serve-passenger trips at all. The corresponding statistics from the Baltimore sample are 14 and 36 percent, respectively. This sequence pattern is obviously caused by the typical requirement that a person chauffeured and dropped off at a place has to be picked up later. The examination of individual cells of parts A-C also indicates that the probability that the third purpose is serving passengers is significantly smaller than the expectation when the first and second purposes are not serving passengers.

The data in Table 4 also indicate that the activities pursued in a chain quite often all fall within one trip-purpose category. For example, the sequences shopping to shopping to shopping and social-recreation to social-recreation to social-recreation are the most frequently observed sequences. This, together with the recurring tendency previously discussed, indicates that the activities pursued in a chain tend to be homogeneous. Of the 2,760 chains, 61.9 percent involve only one trip-purpose category, 23.6 percent involve two, and only 14.5 percent involve three different trip-purpose categories as defined here. These observations differ substantially from the expected values obtained by assuming complete independence in trip-purpose transitions (i.e., Markov chain of the 0th order): 6.9, 56.8, and 36.3 percent, respectively. In the Baltimore sample 72 percent of chains with three or more sojourns involve only one or two trip-purpose categories.

The sequences of activities in these three-sojourn chains showed exactly the same hierarchical order as in Figure 1. Note that this analysis takes into consideration the sequences of indirectly linked activities. This can be seen in part by examining the asymmetry of the matrices presented in Table 4.

Similar tabulations and analyses were done for 1,164 four-sojourn chains in the TALUS sample with the same classification of trip purposes into four categories. However, of the 256 (= 4^4) possible sequences of trip purposes, 150 had observed frequencies of 3 or less, which warranted only limited statistical examination of these chains. Even a data set of 76,025 trip records appears insufficient for rigorous statistical investigation of history dependence in trip chains. Nevertheless, available statistics indicate that the inferences made for the three-sojourn chains are likely to apply to the four-sojourn chains. For example, 547 (47 percent) of the all four-sojourn chains involved only one or two trip-purpose categories. Again, tripmakers tend to pursue only a few types of activities in a chian. Of the 292 chains that contain two or three serve-passenger trips, 220 (75.3 percent) involve the sequences of serve passengers to other activity to serve passengers, or serve passengers to other activity to other activity to serve passengers.

Possible Explanation of Homogeneity

Obviously, the temporal and spatial distributions of opportunities are among the factors that contribute to the homogeneity of activity types pursued in a trip chain. For example, pursuing personal business is not likely in the evening because businesses or shops are typically closed, and chains made in the evening tend to be social-recreation oriented. Commercial corridor development provides many shopping opportunities in close proximity, thus making shopping trip chains convenient and economical.

It may also be hypothesized that the individual has clear perception as to the compatibility of different types of activities in a chain [closely related is the viewpoint that individuals perceive different time periods of a day as suitable for different types of activities (15,24)]. In some extreme cases a set of activities may be viewed as a full course of activities and pursued as such; for example, movie to late dinner to home, or visit a friend to bowling lane to a pizza house to drive the friend home to home. The latter forms a four-sojourn chain that involves social-recreation and serve-passenger trips. If a bundle of activities is perceived as one integrated activity, introduction of a heterogeneous activity into the chain may not be acceptable to the individual. Even the way the individual is dressed may affect trip chaining. If the individual leaves home to pursue a homogeneous set of activities and is dressed suitably for this set, he may feel uncomfortable to visit locations for heterogeneous activities where he may be overdressed or underdressed. This may be especially the case in the TALUS sample because of its survey date and the conservative nature of the region.

SEQUENTIAL MODEL OF TRIP CHAINING

Based on the results of the previous sections, which examined the nature of trip-chaining behavior from various viewpoints, a simple model of the behavior is developed and tested. The key issue in the modeling effort is how to represent the past history of a chain in a simple and practical form.

Model Framework

The model development effort is based on the premise that the observed characteristics of trip chaining can be adequately represented by making the transition probabilities dependent on the past history of the chain. The findings that activities in a chain tend to be homogeneous and that inflexible activities tend to be pursued first suggest a rather simple and systematic structure of history dependence. The probability of a given activity transition strongly depends on the types of activities already pursued, but may not depend on the number of times these activities were engaged in or on the exact order in which they were pursued. For example, the sequencing pattern implies that once flexible activities have been pursued, the probability of an inflexible activity is small, but the number of the previous flexible activities and their sequence may have only a negligible effect on the probability. The exact representation of the history, as shown in Equation 1, may not be necessary, and a simpler representation may be adequate.

The conditional probability of activity choice is formulated as follows:

$$\Pr(X_{n+1}|X_1,X_2,\ldots,X_n)=\Pr(X_{n+1}|X_n;D_{1n},D_{2n},\ldots D_{Kn}) \qquad (10)$$

where D_{jn} is a binary (0-1) variable, which indicates whether activity type j has been pursued in the chain by the nth transition, and K is the number of activity categories used to represent the history of the chain. The model assumes that activity transition probabilities depend on direct linkages; thus the probability of the next activity (X_{n+1}) is conditioned on the current activity (X_n). History of a chain, however, is not represented by the entire series of activities pursued, but by a set of binary variables ($D_{1n}, \ldots, D_{Kn}$).

For example, suppose that activities are classified into four categories (serve passengers, personal business, social-recreation, and shopping), and the D_{jn}'s are defined for j = 1, 2, 3 as

D_{1n} = 1 if serving passenger has been pursued by n, 0 otherwise;

D_{2n} = 1 if personal business has been pursued by n, 0 otherwise; and

D_{3n} = 1 if social-recreation or shopping has been pursued by n, 0 otherwise.

Two activity categories (social-recreation and shopping) are grouped together in defining the D_{jn}'s (note how they are tied in the activity sequencing in Figure 1a). There are eight possible values of vector $D_n = (D_{1n}, D_{2n}, D_{3n})$. Recall that some transitions are likely to occur with the sequencing pattern found earlier, whereas others are less likely to occur. Naturally, the probabilities of the first group will be larger than those of the second group for given X_{n+1} and X_n. The activity choice probability of the model captures this variation because it is conditioned on D_n and thus replicates the sequencing pattern. The preferences in activity set choice are represented by the conditional probability in a similar manner.

The process ($X_1, X_2, \ldots$) depicted by the history-dependent probability $\Pr(X_{n+1}|X_n; D_{1n}, D_{2n}, D_{3n})$ can be represented as a stationary, history-independent Markov chain process if the states are redefined and the state space is expanded appropriately. The set of states and the structure of the transition matrix for this example are shown in Figure 2. The states are now defined in terms of both

Figure 2. Stationary transition matrix of history-dependent trip-chaining model.

"TO" STATE

"FROM" STATE	D_n		SVPS	PBNS	SREC	SHOP	SVPS	PBNS	SREC	SHOP	SVPS	PBNS	SVPS	SREC	SHOP	PBNS	SREC	SHOP	SVPS	PBNS	SREC	SHOP	HOME
1.	(1,0,0)	SVPS					*					*		*	*								*
2.	(0,1,0)	PBNS						*			*						*	*					*
3.	(0,0,1)	SREC							*	*			*			*							*
4.	(0,0,1)	SHOP							*	*			*			*							*
5.	(1,0,0)	SVPS					*					*		*	*								*
6.	(0,1,0)	PBNS						*			*						*	*					*
7.	(0,0,1)	SREC							*	*			*			*							*
8.	(0,0,1)	SHOP							*	*			*			*							*
9.	(1,1,0)	SVPS									*	*									*	*	*
10.	(1,1,0)	PBNS									*	*									*	*	*
11.	(1,0,1)	SVPS											*	*	*					*			*
12.	(1,0,1)	SREC											*	*	*					*			*
13.	(1,0,1)	SHOP											*	*	*					*			*
14.	(0,1,1)	PBNS														*	*	*	*				*
15.	(0,1,1)	SREC														*	*	*	*				*
16.	(0,1,1)	SHOP														*	*	*	*				*
17.	(1,1,1)	SVPS																	*	*	*	*	*
18.	(1,1,1)	PBNS																	*	*	*	*	*
19.	(1,1,1)	SREC																	*	*	*	*	*
20.	(1,1,1)	SHOP																	*	*	*	*	*

Note: States 1 through 4 are only for the first transition. A "*" indicates that the transition probability between the states are positive; otherwise the probability is zero.

PBNS includes personal business and school, SREC includes social-recreation and eating meals, SHOP represents shopping, and SVPS represents serving passengers.

activity type and past history of the chain (i.e., vector D_n). Accordingly, the past history, as expressed by D_n, is automatically specified when the state of the process is designated, which implies that

$$\Pr(X'_{n+1}=j|X'_n=i;D_{1n},D_{2n},D_{3n})=\Pr(X'_{n+1}=j|X'_n=i)=p_{ij} \qquad (11)$$

where X'_n is the redefined nth state. Namely, the process ($X'_1, X'_2, \ldots$) is a Markov chain process. This expansion makes the analysis quite straightforward, and statistical evaluation of the model can be done as in a standard Markov chain analysis.

Estimation Result

Sequential models of activity linkage are estimated by using the 27,901 trip chains in the TALUS sample with the classification of activities into four types, as in the preceding example. Five models with different transition structures are examined:

1. Stationary, history-independent model;
2. Nonstationary, history-independent model;
3. History-dependent model with three elements in D_n;
4. History-dependent model with four elements in D_n; and
5. History-dependent model, a hybrid of models 3 and 4.

Models 1 and 2 are studied here as references against which the history-dependent models are compared. Based on the results presented earlier, model 2 assumes a stationary transition matrix after the fourth transition. Model 3 is the one described in the previous example. The history indicator D_n of model 4 is defined for the four activity types without grouping shopping and social-recreation together, as in model 3. Model 5 uses the same D_n as model 4. However, no further difference is assumed in model 5 as to the history dependence of activity transitions after serving passengers, personal business, and either one of social-recreation or shopping are all pursued in a chain.

Transition probabilities of each model are estimated by the maximum likelihood method. The goodness of fit in terms of the log-likelihood value and square sum of errors in predicting the frequency of each activity sequence is given in Table 5. The latter statistic excludes chains with five or more sojourns (about 4 percent of the entire sample) for computational reasons. The improving goodness of fit of the model found in the table as the number of parameters increases is not surprising. More important, however, is that systematic prediction errors diminish as more thorough treatment of history dependence is made. The agreement between the observed and predicted frequencies of respective activity sequences is shown in Figure 3 for models 1 and 5.

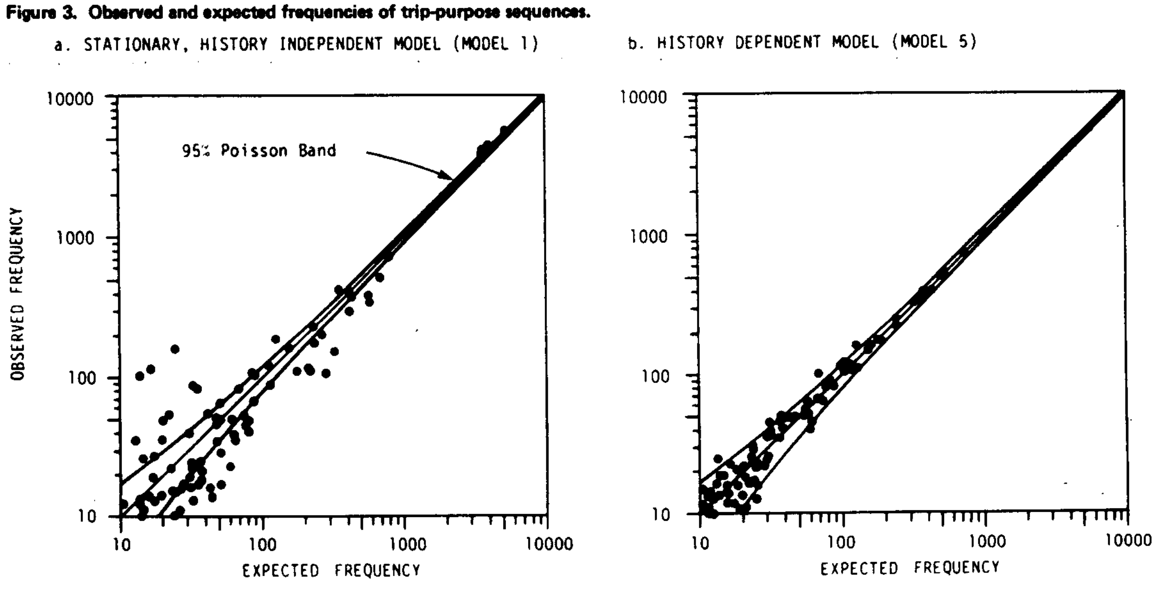
Model 1 (Figure 3a), a standard Markov chain model, significantly underestimates the frequencies of single-sojourn chains, overestimates most of two-sojourn sequences, and makes extremely large errors in evaluating the frequencies of chains that involve recurrence of activities, especially those involving the following sequence: serve passengers to other activities to serve passengers. The nonstationary model (model 2) almost perfectly replicates the distribution of chain lengths. Nevertheless, chains starting with shopping are mostly underestimated, and sequences that involve serve-passenger trips are estimated with large errors.

Table 5. Performance of alternative models of activity transition.

No.	Model	No. of Parameters	L	$-2(\Delta L)$ Chi-Square	$-2(\Delta L)$ df	SSE
1	Stationary, history independent	20	-58,403	–	-	723,648
2	Nonstationary, history independent	100	-57,668	1,471.2	80	117,969
3	History dependent, three-element D_n	100	-56,642	3,522.0	80	34,330
4	History dependent, four-element D_n	180	-56,336	4,133.4	160	8,689
5	History dependent, hybrid	150	-56,362	4,083.0	130	8,737

Note: L = log likelihood; $-2(\Delta L) = -2[(L \text{ of model 1}) - (L \text{ of the model})]$; and SSE = square sum of errors.

Figure 3. Observed and expected frequencies of trip-purpose sequences.



The history-dependent models (models 3-5) largely improve these defects. Model 3, however, still shows significant errors for chains that involve social-recreation or shopping trips, which suggests that grouping these two activity types when representing the history of a chain is not adequate. Examination of the log-likelihood value between this model and models 4 and 5 also indicates this. The performances of models 4 and 5 (Figure 3b) are satisfactory, and only few activity sequences are predicted with significant errors. Note that model 5 performs almost as well as model 4, even though it has 30 less parameters. The satisfactory agreement between the observed and predicted frequencies implies that the patterns in activity sequencing and activity set formation are well represented by the model, and also that the model adequately captures the history of a chain. A simple representation of the history of a trip chain by means of a set of binary variables makes possible a satisfactory replication of trip-chaining behavior.

CONCLUSIONS

The statistical analysis of this study found that there is a consistent hierarchical order in sequencing activities where less-flexible activities tend to be pursued first. It was also found that the set of activities pursued in a trip chain tends to be homogeneous. Thus activity transitions are more organized and systematic than what a Markovian process would depict. The homogeneity of activity types, patterns in sequencing activities, history dependence, and nonstationarity in activity transitions are all closely interrelated. Accordingly, it was possible to develop a sequential, history-dependent model of activity transition that, in spite of its simplified representation of the history of a chian, well replicated the observation. Although the focus of the model was on direct transitions of activities, the model was capable of representing those characteristics found for the entire chain (e.g., homogeneity and recurrence of activities and patterns in indirect transitions). The result strongly supports the sequential modeling approach adopted in this study. The usefulness of the model can be enhanced when the history-dependent probabilities are related to exogeneous factors. This is another step that must be taken before the sequential model can be applied to practical problems.

Although the focus of this study was on the basic characteristics of trip chaining and its representation by sequential probabilities, the study results have some practical implications. The strong regularity implied by the homogeneity of trip chains suggests that people's responses to changes in the travel environment may be limited, as far as trip chaining is concerned. People organize their trip chains while considering the types of activities, but they may not necessarily minimize travel distance or cost. The importance of uncertainty in activity scheduling suggested by the observed sequencing pattern also implies this. Thus travel patterns may be less sensitive to travel cost than what was expected. The rather surprising result that the post-energy crisis Baltimore sample has a mean chain length that is 10 percent shorter than that of the 1965 Detroit sample also supports this claim. This conjecture, however, is subject to further investigation. Additional subjects that can be suggested for future investigation include examination of hierarchical relationships in time allocations and spatial choices for activities in a trip chain, extension of the analysis to incorporate temporal and spatial aspects and verifying the present findings in that context, and investigation of the characteristics of all trip chains made by an individual within the study period and of the interdependence among these chains.

ACKNOWLEDGMENT

This study was partly supported by a grant from the National Science Foundation to the University of California, Davis.

REFERENCES

1. F.E. Horton and P.W. Shuldiner. The Analysis of Land-Use Linkages. HRB, Highway Research Record 165, 1967, pp. 96-107.
2. J.D. Nystuen. A Theory and Simulation of Intra-urban Travel. In Quantitative Geography--Part I: Economic and Cultural Topics (W.L. Garrison and D.F. Marble, eds.), Department of Geography, Northwestern Univ., Evanston, Ill., Studies in Geography 13, 1967, pp. 54-83.
3. S. Hanson and D.F. Marble. A Preliminary Typology of Urban Trip Linkages. The East Lakes Geographer, Vol. 7, 1971, pp. 49-59.
4. P.M. Jones. Methodology for Assessing Transportation Policy Impacts. TRB, Transportation Research Record 723, 1979, pp. 52-58.
5. S. Hanson. Urban-Travel Linkages: A Review. In Behavioral Travel Modelling (D.A. Hensher and P.R. Stopher, eds.), Groom Helm, London, England, 1979, pp. 81-100.
6. W. Brög and E. Erl. Application of a Model of Individual Behavior (Situational Approach) to Explain Household Activity Patterns in an Urban Area and to Forecast Behavioral Changes. In Recent Advances in Travel Demand Analysis (S. Carpenter and P.M. Jones, eds.), Gower, Aldershot, England, 1983, pp. 350-370.
7. D.T. Hartgen, A.J. Neveu, J.M. Brunso, J.S. Banas, and J. Miller. Changes in Travel During the 1979 Energy Crisis. Planning Research Unit, New York State Department of Transportation, Albany, Prelim. Res. Rept. 171, 1980.
8. R.L. Peskin. Policy Implications of Urban Travel Response to Recent Gasoline Shortages. Presented at Energy Users Conference, San Antonio, Tex., 1980.
9. D. Damm. Theory and Results of Estimation: A Comparison of Recent Activity-Based Research. In Recent Advances in Travel Demand Analysis (S. Carpenter and P.M. Jones, eds.), Gower, Aldershot, England, 1983, pp. 3-33.
10. T. Adler and M.E. Ben-Akiva. A Theoretical and Empirical Model of Trip Chaining Behavior. Transportation Research, Vol. 13B, 1979, pp. 243-257.
11. J.O. Wheeler. Trip Purposes and Urban Activity Linkages. Annals of the Association of American Geographers, Vol. 62, 1972, pp. 641-654.
12. S.R. Lerman. The Use of Disaggregate Choice Models in Semi-Markov Process Models of Trip Chaining Behavior. Transportation Science, Vol. 13, 1979, pp. 273-291.
13. 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.M. Jones, eds.), Gower, Aldershot, England, 1983, pp. 313-334.
14. 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.), To-

ronto University Press, Toronto, Ontario, Canada, 1983, pp. 376-402.
15. D. Damm. Interdependencies in Activity Behavior. TRB, Transportation Research Record 750, 1980, pp. 33-40.
16. R. Kitamura, L.P. Kostyniuk, and M.J. Uyeno. Basic Properties of Urban Time-Space Paths: Empirical Tests. TRB, Transportation Research Record 794, 1981, pp. 8-19.
17. V. Vidakovic. A Study of Individual Journey Series: An Integrated Interpretation of the Transportation Process. In Traffic Flow and Transportation (G.F. Newell, ed.), Elsevier, New York, 1972, pp. 57-70.
18. K. Kondo. Estimation of Person Trip Patterns and Modal Split. In Transportation and Traffic Theory (D.J. Buckley, ed.), Elsevier, New York, 1974, pp. 715-742.
19. G.C. Hemmens. The Structures of Urban Activity Linkages. Center for Urban and Regional Studies, Univ. of North Carolina, Chapel Hill, Urban Studies Res. Monograph, 1966.
20. R. Kitamura. "Serve-Passenger" Trips as a Determinant of Travel Behavior. In Recent Advances in Travel Demand Analysis (S. Carpenter and P.M. Jones, eds.), Gower, Aldershot, England, 1983, pp. 137-162.
21. L.P. Kostyniuk and R. Kitamura. Life Cycle and Household Time-Space Paths: Empirical Investigation. TRB, Transportation Research Record 879, 1982, pp. 28-37.
22. T.W. Anderson and L.A. Goodman. Statistical Inferences About Markov Chains. Annals of Mathematical Statistics, Vol. 28, 1957, pp. 89-110.
23. J.C. Kemeny and J.L. Snell. Finite Markov Chains. Van Nostrand, Princeton, N.J., 1960.
24. I. Cullen and V. Godson. Urban Networks: The Structure of Activity Patterns. Pergamon Press, London, England, 1975.
25. D. Parks and W.D. Wallis. Graph Theory and the Study of Activity Structure. In Timing Space and Spacing Time--Volume 2: Human Activity and Time Geography (T. Carlstein et al., eds.), Edward Arnold, London, England, 1978, pp. 75-99.

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

Identifying Time and History Dependencies of Activity Choice

RYUICHI KITAMURA AND MOHAMMAD KERMANSHAH

In this study a sequential model of activity patterns is formulated that consists of time- and history-dependent models of activity choice. This analytical framework is used to identify time-of-day and history-dependent characteristics of activity choice by statistically testing a series of hypotheses. The results indicate that the simplest expression of the history of activity engagements is an adequate descriptor, and also that non-home-based activity choice is conditionally independent of the activities in the previous chains, given the activities pursued in the current trip chain. Interdependencies of activity types across trip chains are also characterized by estimated model coefficients. The results of the study indicate that the decisions associated with the entire activity pattern can be decomposed into interrelated activity choices whose conditional dependencies can be statistically evaluated.

The way individuals schedule their daily activities and organize their itineraries has immediate impacts on the spatial and temporal distribution of trips, or needs for trips, in an urban area. Therefore, representing how the choice and scheduling of activities are done and how travel patterns are formed are critical elements in travel-demand forecasting as well as in basic travel-behavior research (1-3). This is especially so when attempting to forecast the impacts of novel changes in the travel environment or when seeking a transportation policy that will accomplish given objectives most effectively.

The mechanism by which trips as induced demand are generated is complex. Even when only scheduling is considered (i.e., when and in what order a given set of locations is visited and how these visits are arranged into trip chains), there are numerous scheduling possibilities. Choice of activities and their locations further complicates the problem. Constraints that govern the behavior are not limited to monetary and time budgets as in the classical utility maximization framework in economics, but include spatial and temporal fixity constraints associated with the respective activities (4), interpersonal linkage constraints (5), and other types of constraints that portray the travel environment of each individual (6). The interrelated activity choices underlying an activity-travel pattern are dependent on the time of day, as many previous studies on time use have indicated (7,8). Previous empirical evidence (9, and paper by Kitamura elsewhere in this Record) at the same time indicates that the choices are dependent on history, i.e., the set of activities already pursued on that day.

These aspects of daily activity and travel behavior are all of particular importance for the understanding and forecasting of the behavior. In particular, the time-of-day and history dependencies of activity choice may be viewed as the most fundamental elements, whose adequate representation will lead to representation of other important aspects of the behavior as well. For example, the preferences in forming a set of activities in a trip chain can be described by sequential probabilities of activity choice when their history dependencies are appropriately incorporated (see paper by Kitamura elsewhere in this Record). By specifying the structure of the time-of-day and history dependencies and estimating the model statistically, an important objective can be accomplished: characterization of activity and travel patterns along the time dimension. When the model includes exogenous factors that are related to changes in the travel environment or in the population characteristics, then the model serves as a tool for forecasting possible changes in activity and travel behavior.

An extension of a previous sequential analysis of activity linkages is described in this study (see paper by Kitamura elsewhere in this Record), and an attempt is made to identify the structure of time and history dependencies of activity choice. The objective is to demonstrate that the conditional dependency of activity choice can be properly represented by a simple model structure that can be statistically estimated and conveniently applied to practical problems. The dependency is examined by testing a set of hypotheses and by inferring its characteristics. Alternative model specifications are examined, and home-based and non-home-based activity-choice models are estimated.

The results of hypothesis testing and model estimation indicate that a simple indicator of the history of the behavior--a set of binary variables each representing whether an activity of a given type has been pursued--best explains the activity choice. Home-based choice that determines the first activity in a trip chain is shown to be dependent on the past activity engagement, but non-home-based choice is conditionally independent of the activities in the previous chains, given the activities pursued in the present chain. Strong time-of-day dependencies in activity choice, whose temporal variations are well captured by the model, are also shown in the study. The results of the study consistently indicate that the time and history dependencies of the behavior can be represented by a simple model structure, and suggest that a set of sequential activity-choice models can be developed to represent and forecast the characteristics of daily activity and travel behavior.

BACKGROUND

Because individuals develop their daily itineraries while considering the set of activities to be pursued during a certain period, activity choices (or travel choices) cannot be analyzed individually, but the interdependencies among them must be adequately accounted for. Such interdependencies have been noted across different time periods of a day (9), or among activity choices in a trip chain (see paper by Kitamura elsewhere in this Record). Another aspect of activity and travel behavior is the existence of various types of constraints that govern behavior (5,6,10). Many constraints are unobservable if typical survey data are the only information sources. All these characteristics of tripmaking make causal representation of the behavior quite complex.

A possible representation of activity- and travel-choice behavior uses the concept of optimization together with the assumption that the observed activity-travel pattern is the one preferred the most by the individual (11). Let a_n be the type of nth activity, t_n be its starting time, d_n be its duration, and ℓ_n be the location where the activity is pursued. For simplicity, only these four aspects are considered here. By letting $a = (a_0,\ldots,a_{N+1})$, and so forth, the activity- and travel-scheduling behavior can be formulated as follows:

$$\text{Maximize } U = U(a,t,d,\ell) \tag{1}$$

$$\begin{aligned}\text{Subject to}\quad & t_{n+1} - (t_n + d_n) = s(\ell_n, \ell_{n+1}, t_n + d_n)\\ & 0 < t_0,\ldots,t_{n+1} < T;\quad \ell_0 = \ell_{N+1} = \text{home}\\ & 0 < d_n\\ & a_n \in C, \ell_n \in E \quad n = 1,\ldots,N\\ & g_i(a,t,d,\ell) = 0 \quad i = 1,\ldots,G\end{aligned}$$

where

$s(i,j,s)$ = travel time between locations i and j when the trip begins at time s,
N = total number of sojourns (including intermediate sojourns at home),
C = set of activity types, and
E = set of opportunity locations.

The first constraint simply represents the temporal continuity condition, the second represents the condition where the individual's path must originate and terminate at home within time T, and the third condition is the nonnegativity of activity durations. Additional constraints are represented in a general form by function g in this formulation. Function U, which may be called a utility function, includes not only the type and duration of each activity but also its starting time. This is because the regularity and rhythms in time use patterns strongly suggest that the utility of an activity of a given type is a function of the time when it is pursued.

Not quite obvious from this formulation is the discrete nature of the optimization problem, i.e., resources are not always allocated to all activities and some activities simply may not be pursued at all during a given period. Accordingly, the classical constrained optimization approach (12,13) is not applicable to this problem if this formulation is to be applied to disaggregate data where behavior during a relatively short period (e.g., 1 day) is recorded. The problem is also much more complex than that of a traveling salesman. Not only the order of visits, but also the number of visits, their locations, the way these visits are organized into trip chains, and their timing must be endogenously determined. When this complexity as a mathematical programming problem is combined with the additional constraints, the task involved in representing the behavior as an optimization problem and obtaining its solution appears to be prohibitive. Perhaps the number of possible activity-travel patterns recognized by the individual is relatively small (2) because of the constraints and limited information the individual has, but this is not the case for the observer who attempts to analyze and predict the behavior without comparable knowledge on microscopic factors that influence each individual.

[The approach taken by Adler and Ben-Akiva (14) avoids these difficulties and at the same time retains the simultaneous structure of analysis by modeling the behavior as a discrete choice among alternative activity-travel patterns. The approach is quite effective in analyzing characteristics of activity-travel choice. Determining the probability with which a given pattern will be chosen, however, requires that all feasible patterns be enumerated.]

An alternative approach to the analysis of activity and travel patterns is a sequential one, which is based on the following identity:

$$\Pr(a,t,d,\ell) = \prod_{n=0}^{N} \Pr[a_{n+1}, t_{n+1}, d_{n+1}, \ell_{n+1} | a_{(n)}, t_{(n)}, d_{(n)}, \ell_{(n)}] \tag{2}$$

where $a_{(n)}$ is a vector of the first (n + 1) elements of a, i.e., $a_{(n)} = (a_0,a_1,\ldots,a_n)$; and $t_{(n)}$, $d_{(n)}$, and $\ell_{(n)}$ are similarly defined. This approach, where choices are analyzed one by one in a sequence, represents the preferences in choosing patterns given that, if $U(a,t,d,\ell) \le U(a',t',d',\ell')$, then $\Pr(a,t,d,\ell) \le \Pr(a',t',d',\ell')$. The approach has an advantage in that it reduces the size of the problem to a manageable one, and the preferences to the entire pattern can be correctly represented if the conditional dependencies of the sequential probabilities are properly incorporated. A recent study indicated that the sequence of activities in a trip chain can be adequately represented by a simple sequential model, whereas failure to capture the conditional dependency leads to erroneous results (see paper by Kitamura elsewhere in this Record).

An interesting example of a sequential approach can be found in Horowitz (15), where the concept of time-dependent utility is used. A similar concept is used in the present study, but emphasis is on the identification of time and history dependencies of the behavior. The works by Damm (9), Damm and Lerman (16), and Jacobson (17) are noted here because certain facets of the complex behavior are carefully selected in these studies so that the size of the problem can be reduced and the analysis can be carried out meaningfully and effectively by using econometric methods.

There are two tasks involved in developing a sequential model of activity and travel for forecasting purposes. The first is the identification of the structure of the conditional dependency, which is a prerequisite for proper functioning of the model. Because representing the history as in Equation 2 will not serve practical purposes because of its excessive information requirements, some simple yet accurate forms must be sought. The second task is to relate the sequential probabilities to exogenous factors, especially those that closely represent planning options and policies.

The time factor is of critical importance in developing such a probabilistic model of activity choice because of the strong correlation between time of day and activity, as noted earlier. Incorporating the time variable is also important because it will make probabilistic representation of the constraints that affect the behavior more meaningful and accurate. In particular, the effect of time constraints cannot be appropriately represented without the time variable [e.g., Hagerstrand's prism is approximated by time-dependent probabilities of spatial choice (18)]. A previous study (19) indicated that married women who are not employed and who are in the childbearing stage tend to return home early in the evening; this can be viewed as being a result of the constraints imposed by family responsibilities. The sequential probabilities can depict such constraints when they are specified as time-of-day dependent and when they include appropriate variables that represent individuals' attributes.

APPROACH

In this study the activity choice along the time dimension is analyzed, and the main focus of the study is on the identification of the time- and history-dependent nature of the choice. The spatial aspect is suppressed in this study. The model specification and estimation effort is based on the following formulation of the sequential probability:

$$\begin{aligned} d\Pr[a_{n+1}, t_{n+1} | a_{(n)}, t_{(n)}] &= \Pr[a_{n+1} | t_{n+1}; a_{(n)}, t_{(n)}] \, d\Pr[t_{n+1} | a_{(n)}, t_{(n)}] \\ &= \Pr[a_{n+1} | t_{n+1}; a_{(n)}, t_{(n)}] \, d\Pr[t_{n+1} - t_n | a_n, t_n; a_{(n-1)}, t_{(n-1)}] \end{aligned} \tag{3}$$

where $a_{(n)} = (a_0, a_1, \ldots, a_n)$ as before, and $t_{n+1} - t_n$ is called the sojourn duration in the nth state that, in this formulation, includes the duration of the nth activity and trip time to its location (the activity duration and trip time are treated separately in the empirical analysis presented in later sections). The sequential probability is expressed as a product of activity-choice probability given the time of the choice and the probability density of the duration of the nth sojourn. The focus of this study is on the first element: time- and history-dependent activity-choice probability.

The activity-choice probability is formulated as a function of time, history, and other factors by using the multinomial logit structure, i.e.,

$$\begin{aligned} \Pr[a_{n+1} = j | t_{n+1} = t; a_{(n)}, t_{(n)}, y] \\ = \exp\{V_j[t, a_{(n)}, t_{(n)}, y]\} / \sum_k \exp\{V_k[t, a_{(n)}, t_{(n)}, y]\} \end{aligned} \tag{4}$$

where y is a vector of socioeconomic attributes of the individual and t is the time of day. The conditional dependence in Equation 3 is now represented in the model by its explanatory variables that represent the history of the behavior and the time of day. It is therefore assumed that the random error terms of the model possess all the desirable properties, including their statistical independence across the choices in the sequence. Although it is possible to use more elaborate formulations of the random elements (20,21), which may lead to an interesting examination of history dependence, this study does not extend its scope to analysis of the dependence structure of the unobservables. [Note that the validity of the error term specification depends on model specification, and it is an empirical issue in that sense (22).]

The time dependency of activity choice is represented by introducing time variables into the logit function. For example, suppose that the effect of time of day on relative activity-choice odds can be expressed by gamma functions, i.e.,

$$\begin{aligned} \exp[V_j(t, \ldots)] / \exp[V_i(t, \ldots)] = [K t^a \exp(-bt)] \\ \div [t^c \exp(-dt)] \qquad a,b,c,d,K > 0 \end{aligned} \tag{5a}$$

(Note that neither the numerator nor the denominator is necessarily a distribution function.) Then,

$$V_j(t, \ldots) - V_i(t, \ldots) = \ell nK + (a - c)\ell nt - (b - d)t \tag{5b}$$

Although it is not possible to determine these parameter values uniquely, the time effects can be represented simply by introducing t and $\ell n(t)$ into function V. The model specification effort in the following sections also considers polynomial and exponential functions of t.

By using this framework, various hypotheses regarding the nature of the conditional dependencies can be examined statistically and the model can be specified subsequently. This study rejects without examination the null hypothesis that activity choice is independent of time of day. The critical hypotheses that need to be examined statistically include the following:

1. Activity choice is independent of the set of activities pursued in the past;
2. Activity choice is conditionally independent of the set of activities pursued in previous trip chains, given the activities pursued in the current chain;
3. Given whether activities of respective types have been pursued or not, activity choice is conditionally independent of the number of times the activities were pursued;
4. Given whether activities of respective types have been pursued or not, activity choice is conditionally independent of the amount of time spent in the past for each type of activity;
5. Activity choice is independent of the number of trip chains made in the past; and
6. Activity choice does not depend on the time spent since the individual left home.

An appropriate representation of the history of an activity pattern is sought through the examination of these hypotheses, and the nature of history dependency is inferred from the results.

DATA SET AND VARIABLES

In this study the statistical analysis of a sample from the 1977 Baltimore travel demand data set is used. Analysis of nonwork activities is the main subject of this study, and only those individuals who did not make work trips on the survey day are analyzed. The records in the data set are screened, and individuals who were younger than 18 years old, who did not hold a driver's license, and whose households did not have a car available are eliminated. A detailed description of the screening criteria used can be found in Kitamura (see paper elsewhere in this Record). The screened sample used in this study includes 927 activity choices in 356 trip chains made by 217 individuals.

Activities are defined in terms of the trip-purpose categories in the data set, which are grouped into four types: personal business, social-recreation, shopping, and serve passengers. Home-based activity-choice models are estimated with these activity types as alternatives. Two additional categories enter models of non-home-based choice: temporary return to home and permanent return to home for the day [similar binary classification of the home state can be found in Lerman (23)]. Accordingly, the non-home-based models are estimated with six alternatives.

As variables representing individuals' attributes, the age, sex, education, employment status, household income, household size, number of children, family life cycle, household role, and car ownership are examined in this study. The household-role variable is defined in terms of the sex and employment status of the individual. The life-cycle-stage variable is defined in terms of the marital status of the adult members, their ages, and the age of the youngest child. The definitions of those variables that appear in the models presented in this paper are given in Table 1.

Table 1. Definition of explanatory variables in activity-choice models.

Variable and Abbreviation	Definition
School-aged children (SCHLAG)	Binary variable: 1 if the age of the youngest child in the household is between 5 and 12, 0 otherwise
Household role (ROLE)	Binary variable: 1 if the individual is a female and not employed, 0 otherwise
No. of children (CHLDRN)	No. of household members who are 17 years old or younger and not married
Household income (INCOME)	Median value of the household's annual gross income category ($)
No. of cars (CARS)	No. of cars available to the household
Time of day (t)	Time of days in hours; the study period begins at 4:00 a.m. when t = 4.0, and ends at 4:00 a.m. the next day when t = 28.0
Activity engagement in previous chains in	
Personal business (PBNS01H)	Binary variable: 1 if activities of the indicated type were pursued in the trip chains previously made
Social-recreation (SREC01H)	
Shopping (SHOP01H)	
Serve passengers (SVPS01H)	
Activity engagement in the current chain in	
Personal business (PBNS01C)	Binary variable: 1 if activities of the indicated type have been pursued in the current trip chain
Social-recreation (SREC01C)	
Shopping (SHOP01C)	
Serve passengers (SVPS01C)	
Out-of-home time (OHTIME)	Cumulative amount of time spent so far outside home for both trips and activities
No. of chains (CHAINS)	Cumulative number of home-based trip chains made so far
Current activity	
Personal business (PBNS)	Binary variable: 1 if the current activity is of the indicated type
Social-recreation (SREC)	
Shopping (SHOP)	
Serve passengers (SVPS)	

HOME-BASED ACTIVITY-CHOICE MODEL

Because the examination of alternative hypotheses regarding the structure of time and history dependencies is an important concern of the study, a series of models, each being developed to test a specific hypothesis, is presented in this section. The first in the series involves only socioeconomic attributes of the individual as its explanatory variables (model 1 of Table 2). The model as a whole is significant with $\alpha = 0.005$, but the amount of variation explained by the model is relatively small ($\rho^2 = 0.0256$). Nevertheless, meaningful relationships are found from the estimation result. The coefficient of the variable that represents the presence in the household of children aged between 5 and 12 (SCHLAG) indicates a positive contribution of this variable to the engagement of serve-passenger trips. The role variable (ROLE), which has a value of 1 when the individual is female and not employed, indicates that these individuals carry out shopping and serve-passenger trips more often than do the others. The coefficient of the number of children (CHLDRN) indicates the negative effect that the presence of children has on the engagement in social-recreation by the adult members.

The fit of the model improves when time variables are introduced into the model with six additional coefficients (model 2). The log-likelihood ratio statistic has a value of $\chi^2 = 46.14$, with degrees of freedom (df) = 6 for the six new coefficients. Clearly the time of day has a substantial influence on activity engagement. The nature of the time dependency of activity choice is presented later in this section by using a history-dependent model.

Examination of the history dependence of home-based choice uses the following variables to represent the past history of activities: 0-1 binary variables, each representing whether an activity of a given type has been pursued in the past; the number of sojourns made for each activity type; and the cumulative amount of time spent for each activity type. These variables are used because of their conciseness as summary variables of the history. The possible effects on activity choice of the exact sequence of the past activities, their respective durations, and their occurrence times are considered to be negligible.

Each set of history variables is tested, and on the basis of its significance the nature of history dependence is inferred. The results indicate that the simplest representation of the history--the set of binary indicators of activity engagement--explains the choice better than any other sets examined here (model 3). Although the other sets of variables are all significant, they do not explain as large a portion of variations as does the set of binary variables. Whether the individual has pursued an activity of a given type or not does affect the home-based activity choice, but how many times and how long the activities were engaged in do not have as decisive an effect. This rather unexpected result is encouraging because of its implication that the history of behavior can be expressed in quite a simple manner in representing the condi-

Table 2. Home-based activity-choice models.

Variable	Personal Business Coefficient	Personal Business t-Statistic	Social-Recreation Coefficient	Social-Recreation t-Statistic	Shopping Coefficient	Shopping t-Statistic	Serve Passengers Coefficient	Serve Passengers t-Statistic
Model 1[a]								
Constant			-0.5797	-0.86	-0.0485	-0.08	0.0698	0.11
SCHLAG							0.8583	2.76
ROLE					0.4968	1.99	0.5702	1.92
CHLDRN			-0.1895	-2.02				
ℓn (INCOME)	0.2877	1.23	0.5893	2.37	0.2801	1.23		
Model 2[b]								
Constant			-3.3830	-5.55	-10.984	-2.01	5.0561	1.01
SCHLAG							0.8585	2.76
ROLE					0.4787	1.89	0.5743	1.93
CHLDRN			-0.2469	-2.52				
ℓn (INCOME)	0.3299	1.37	0.5063	2.02	0.2136	0.92		
t	-0.4240	-1.57	-0.2050	-0.76	-0.7928	-2.85		
ℓn (t)	4.1127	1.21	0.4227	1.18	10.498	2.90		
Model 3[c]								
Constant			-3.7499	-0.65	-11.876	-2.10	4.8241	5.25
SCHLAG							0.6289	1.89
ROLE					0.3895	1.44	0.6413	1.96
CHLDRN			-0.2086	-2.07				
ℓn (INCOME)	0.3542	1.42	0.5169	1.97	0.2093	0.87		
t	-0.3522	-1.22	-0.1310	-0.46	-0.7315	-2.56		
ℓn (t)	3.7260	1.04	4.0236	1.08	10.634	2.89		
PBNS01H			-0.8605	-1.98	-1.6576	-3.34	-1.0187	-1.90
SREC01H			0.6166	0.91	0.8793	1.25	1.5780	2.20
SHOP01H			-0.2300	-0.47	-0.5662	-1.09	-0.2014	-0.36
SVPS01H			0.3252	0.52	0.9638	1.67	1.8125	3.14

Note: Sample = 356 home-based activity choices. $L(\beta)$ = log-likelihood with the model coefficients, $L(C)$ = log-likelihood without explanatory variables (constant terms alone), $L(0)$ = log-likelihood without any coefficients, and $\rho^2 = 1 - L(\beta)/L(C)$. The chi-square values presented are defined as $-2[L(C) - L(\beta)]$.

[a] $L(0) = -493.52$, $L(C) = -490.27$, $L(\beta) = -477.70$, $\chi^2 = 25.14$ (df = 7), and $\rho^2 = 0.0256$.
[b] $L(\beta) = -454.63$, $\chi^2 = 71.29$ (df = 13), $\rho^2 = 0.0727$, and χ^2 for the set of time variables = 46.14 (df = 6).
[c] $L(\beta) = -434.36$, $\chi^2 = 111.81$ (df = 25), $\rho^2 = 0.114$, and χ^2 for the set of activity indicators = 40.54 (df = 12).

tional dependency of activity choice. Another history descriptor--the number of chains completed in the past--was found to be insignificant.

These models are developed primarily to examine alternative hypotheses; thus the selections of explanatory variables are not necessarily finalized as they are presented in Table 2. A similar model is estimated after eliminating some of the insignificant variables of model 3, and its coefficients for the binary variables are given in Table 3 to indicate how the past engagement in an activity of one type affects the choice of another activity type. In the table the estimated set of coefficients is adjusted by adding a constant to the coefficients for each activity type. The value of the constant is arbitrary, and that value that makes the row sum of the adjusted coefficients zero is used in developing the table.

The result indicates that engagement in personal business in the past has a positive influence on the choice of the same activity type later. The same tendency can be found for serving passengers; choices of personal business or serve-passenger trips are positively correlated across trip chains. The negative diagonal value for shopping indicates that people tend not to pursue shopping in two or more trip chains; it suggests that people have been consolidating their shopping trips into fewer trip chains. A negative coefficient of social-recreation on personal business indicates that there are patterns in sequencing activities across trip chains, and personal business tends not to be pursued if the previous chains included social-recreation trips. The pattern found here is quite similar to that found earlier as to the sequencing of activities within a trip chain (see paper by Kitamura elsewhere in this Record).

The time-dependent nature of home-based activity choice can be seen in Figure 1, which presents against the time axis both the observed relative frequencies of chosen activity types and the choice probabilities depicted by the model. The observed shopping frequency coincides naturally with the typical stores' hours, and it peaks in the early afternoon. Personal business tends to be pursued in the morning, whereas the relative frequency of social-recreation increases toward the end of the day. The serve-passenger activity has a rather irregular pattern with peaks in the early morning (chauffeuring children or workers, perhaps), early afternoon, and late evening.

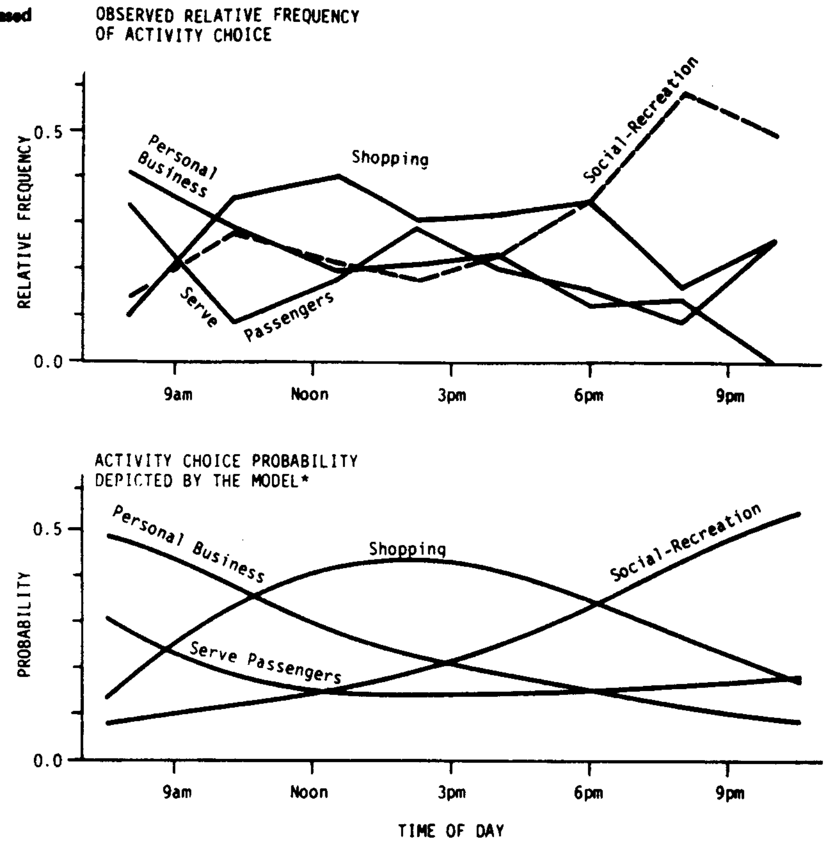
The data in the figure indicate that the observed tendencies are well replicated by the estimated

Table 3. Effects of activity engagements in previous chains on home-based activity choice.

Activity Engagement in Previous Chains[a]	First Activity of Current Chain: Personal Business	Social-Recreation	Shopping	Serve Passengers
Personal business	0.8278	0.0602	-0.7564	-0.1317
Social-recreation	-0.4109	-0.4109	0.0612	0.7606
Shopping	0.1024	0.1024	-0.3072	0.1024
Serve passengers	-0.6059	-0.6059	0.1782	1.0336

[a] 1 if engaged, 0 otherwise.

Figure 1. Observed and predicted probabilities of home-based activity choice.



*Independent variable values used are: INCOME = 20000, CHLDRN = 4, ROLE = 0, SCHLAG = 1, PBNSO1H = 0, SRECO1H = 0, SHOPO1H = 0, and SVPSO1H = 0.

activity-choice model. The activity-choice probabilities are evaluated by assuming the independent variable values, as shown in the figure; therefore, they are not readily comparable with the observed relative frequencies that represent the entire sample. Nevertheless, satisfactory agreement is shown in the figure between the observation and the prediction by the model. The irregularities in the probability of serve-passenger trips are not well represented by the model, although the overall tendency is captured. If it is shown that the observed irregularities are not caused by the small sample size, then the model specification must be altered to reflect them.

NON-HOME-BASED ACTIVITY CHOICE

Non-home-based activity choice is studied in a manner similar to home-based activity choice by examining hypotheses of history and time dependencies of the choice. Additional hypotheses that are included here are concerned with the relative magnitudes of the dependencies on the activities in the previous trip chains and on those in the current chain. Also of interest are the effects of elapsed time since the beginning of the chain and the total out-of-home time on the decision to return home. The variables used to represent the history of the behavior include 0-1 activity engagement indicators defined for the current chain and for the chains previously made, total activity time by activity type in the current chain and in the previous chains, number of sojourns made by activity type in the current chain and in the previous chains, number of chains made in the past, elapsed time since the individual left home, and the cumulative out-of-home time spent.

The models tested and their goodness of fit are given in Table 4 without presenting the estimated coefficients of the respective models. The conclusions of this hypothesis testing are summarized as follows:

1. Given the history of the current chain, activity choice is conditionally independent of the activity engagement in the previous chains;
2. The number of sojourns made and the time spent for each activity type in the current chain are correlated with the observed activity choice, but the 0-1 activity engagement indicators best explain the choice;
3. The elapsed time since the beginning of the chain is not a significant factor influencing the decision to return home;
4. The non-home-based choice is most closely correlated with the time of day, whereas activity history and socioeconomic attributes of the individual have less effects on choice; and
5. The choice of the next activity is affected by the type of current activity.

Perhaps the most significant finding is that non-home-based activity choice is conditionally independent of the history of activity engagement in the previous chains. (No sets of history variables for the previous chains were statistically significant when they were included in the model together with a set of history variables for the current chain.) This may appear to indicate that activity choice repeats itself and that all chains made by an individual are probabilistic replicas of each other. However, this is not the case because the home-based choice that determines the first activity of a chain

Table 4. Alternative specifications of non-home-based activity-choice models.

Model No.	Origin Indicator	Time of Day	Socioeconomic	History Indicators Set 1	History Indicators Set 2	L(β)	χ^2 (df = 4)
1[a]	X	X	X			-813.585	–[b]
2	X	X	X	X		-810.396	–[c]
3	X	X	X	X	No. of sojourns, past	-807.659	5.474
4	X	X	X	X	Activity time, past	-806.258	8.276
5	X	X	X	X	0-1 activity, past	-808.649	3.494
6	X	X	X	X	No. of sojourns, cumulative	-803.366	14.060
7	X	X	X	X	Activity time, cumulative	-807.694	5.404
8	X	X	X	X	No. of sojourns, present	-800.499	19.794
9	X	X	X	X	Activity time, present	-805.467	9.858
10	X	X	X	X	0-1 activity, present	-797.013	26.766
11[a]	X		X	X	0-1 activity, present	-842.981	–[d]
12[a]		X	X	X	0-1 activity, present	-808.820	–[e]

Note: The origin indicator includes three binary variables: PBNS, SREC, and SVPS. Time of day is represented by three independent variables: t, exp(t), and exp(-t). The set of socioeconomics includes four variables: number of children, 0-1 binary variable for presence of school-aged children, income, and number of cars. The history indicators include two sets of variables: set 1 consists of cumulative out-of-home time, elapsed time, and number of chains made previously; and set 2 includes the variables indicated in the table. X indicates that the variable is included in the model.

[a]These models are tested against model 10; the other models are tested against model 2.
[b]Effect of history, χ^2 = 33.14, df = 7.
[c]Reference model.
[d]Effect of time of day, χ^2 = 91.94, df = 5.
[e]Effect of direct linkages, χ^2 = 23.61, df = 10.

is dependent of the past history, as discussed in the previous section. Thus the history dependence of the non-home-based choice is indirectly represented through the history dependence of the home-based choice.

The strong time dependency of non-home-based choice must be noted. The contribution of the five time coefficients to the explanatory power of the model is represented by a chi-square statistic of 91.9 (df = 5), whereas that of the socioeconomic variables is 16.1 (df = 4), and that of the history variables is 33.1 (df = 7). Obviously, time of day is the most critical determinant of the non-home-based choice.

The final form of the non-home-based activity-choice model that was selected on the basis of the hypothesis testing results is given in Table 5. A set of three binary variables (PBNS, SREC, SHOP) is used to represent the type of current activity, i.e., the activity just completed at the time of the transition to the chosen activity. Many of the nine coefficients that apply to these variables are significant and indicate the strength of direct linkages between activity types. Compared with the home-based choice models, fewer socioeconomic attribute variables are used in the model. The number of children and the presence of school-aged children have the same effects on activity choice as in the home-based choice model.

The coefficients of the car-ownership variable are positive (but not significant) for the temporary return to home, and they are negative for the permanent return to home. The indication is that the individuals from households with more cars tend to make more trip chains, but the number of sojourns in a chain may tend to be fewer. A similar tendency was found in a previous study that analyzed a 1965 Detroit data set (24). The negative coefficients of the cumulative out-of-home time and the number of chains are quite noteworthy, although they are not statistically significant at α = 0.05. The coefficients apply to the permanent return to home and imply that the more time the individual has spent outside home and the more chains he has made, the less likely he is to terminate his out-of-home activity pursuit of the day. The result suggests that

Table 5. Non-home-based activity-choice model.

Variable	Personal Business Coefficient	Personal Business t-Statistic	Social-Recreation Coefficient	Social-Recreation t-Statistic	Shopping Coefficient	Shopping t-Statistic	Serve Passengers Coefficient	Serve Passengers t-Statistic	Home Coefficient	Home t-Statistic	Absorbing Home Coefficient	Absorbing Home t-Statistic
Constant	-0.5306	-1.01	-2.4408	-3.08	-2.7799	-3.19	-2.5840	-2.89	0.9237	1.17		
PBNS	1.4074	1.59	1.7789	2.17	2.7420	2.57			1.1598	1.44	1.1598	1.44
SREC	1.1186	1.96			1.8215	2.25			0.5169	1.09	0.5169	1.09
SHOP					0.9220	1.29			-0.3812	-1.10	-0.3812	-1.10
t			0.1809	3.23	0.0884	1.74	0.1490	2.52				
exp(-t/10)									1.3261	0.77		
exp(t/10)											0.0520	6.04
CHLDRN			-0.2116	-1.66								
SCHLAG							0.4998	1.31				
CARS									0.0580	0.59	-0.2478	-2.32
PBNS01C	1.2327	2.62			0.4349	1.16						
SREC01C			0.6284	1.51	0.6789	1.76						
SHOP01C					1.4726	3.53						
SVPS01C							1.2009	2.83				
OHTIME											-0.0009	-1.30
CHAINS											-0.1498	-1.26

Note: L(0) = -1023.090, L(C) = -905.722, L(β) = -794.778, χ^2 = 221.89 (df = 26), and ρ^2 = 0.123. Sample is based on 571 non-home-based activity choices.

individuals pursue either very few or very many activities on a given day. This may be a result of activity scheduling over a longer time span, e.g., a week.

In summary, the hypothesis testing and model specification efforts presented in these two sections have indicated that the activity choice is dependent on both the time of day and the history of the activity. But the structure of the history dependency is rather simple. The binary history indicators that represent whether activities of respective types have been pursued in the past or not are correlated with the activity choice more strongly than is the number of sojourns or the time spent for each activity type in the past. Furthermore, non-home-based activity choice is found to be conditionally independent of the activity history in the previous chains, given the history in the current chain. It appears that activity choice is dependent more strongly on more recent activities. The significance of the variables that represent the direct linkages also indicates this.

DISCUSSION OF RESULTS

Identifying the dependencies across a series of activity choices is critically important for the development of a practical tool for analyzing and forecasting daily activity and travel behavior. In this study the structural form of a sequential model of activity patterns was formulated, and conditional probabilities of activity choice that used the multinomial logit structure were specified. This framework was then used to examine the nature of time and history dependencies in activity choice with the assumption that time of day and the history of the behavior are the most fundamental factors that influence activity choice.

The examination of a series of hypotheses indicated that the simplest representation of the history of the behavior--a set of binary activity engagement indicators--is an adequate descriptor and best explains activity choice. Non-home-based activity choice is strongly affected by time of day and also by current activity type, but socioeconomic attributes of the individual and history variables have less influence on non-home-based choice than on home-based choice. Non-home-based activity choice was also found to be conditionally independent of the activity history in the previous chains, given the activity history in the current chain, whereas home-based activity choice had interdependencies in the activity types across trip chains. The results of the study are encouraging and indicate that a set of simple models that can be conveniently estimated is capable of representing individuals' daily activity and travel behavior together with the interdependencies across the choices involved. The study has indicated that the decisions associated with the entire activity pattern can be decomposed into interrelated activity choices whose conditional dependencies can be statistically evaluated.

The models presented in this study, however, are not immediately applicable to practical problems because the types of exogenous variables included are limited. This limitation is mainly caused by the aspatial nature of the study. The models must be extended to spatial activity-choice models with land use and transportation network variables introduced as explanatory variables. Note that the land use variables in this context must be defined in terms of both the spatial distribution of opportunities and their availabilities along the time dimension. When land use variables are defined in this manner, then the activity choice can be related to the availabilities of various opportunities in different time periods of a day.

Such effort of modeling the activity choice in the spatial dimension will encounter a new problem: representation of the attractiveness of an opportunity, or a group of opportunities such as a zone. This is not a trivial task when the assumption of the conventional approach that a travel choice can be separated from the rest and can be analyzed independently is discarded, and when the interdependencies across the choices are acknowledged. The interdependencies imply that a choice of an opportunity is influenced by both the past and intended future behavior. The conventional formulation of the attractiveness of a zone that uses the attributes of that zone alone is not adequate when the individual has in mind additional activities to be pursued elsewhere. In other words, when trip chaining is considered, the traditional definition of the attraction becomes inadequate, and the attractiveness of a zone as an origin from which the next activity site will be reached must be evaluated and incorporated into the attraction measure. This can be done by using the concept of expected utility, in which the attractiveness of a zone is a function of not only its own attributes but also those of other zones. Another aspect, which was not emphasized in this study, is the structural relationship between the activity duration and activity choice. It may be the case that the relationship varies depending on the time of day or on the past history of the behavior. Examination of the interdependence structure of the unobservables also remains as a subject of future research.

ACKNOWLEDGMENT

This study was partly supported by a grant from the National Science Foundation to the University of California, Davis.

REFERENCES

1. P.M. Jones, M.C. Dix, M.I. Clarke, and I.G. Heggie. Understanding Travel Behavior. Transport Studies Unit, Oxford Univ., Oxford, England, 1982.
2. K.P. Burnett and S. Hanson. Rationale for an Alternative Mathematical Approach to Movement as Complex Human Behavior. TRB, Transportation Research Record 723, 1979, pp. 11-24.
3. W. Brog and E. Erl. Application of a Model on Individual Behavior (Situational Approach) to Explain Household Activity Patterns in an Urban Area and to Forecast Behavioral Changes. In Recent Advances in Travel Demand Analysis (S. Carpenter and P.M. Jones, eds.), Gower, Aldershot, England, 1983, pp. 350-370.
4. I. Cullen and V. Godson. Urban Networks: The Structure of Activity Patterns. Pergamon Press, London, England, 1975.
5. I.G. Heggie and P.M. Jones. Defining Domains for Models of Travel Demand. Transportation, Vol. 7, 1978, pp. 119-135.
6. T. Hagerstrand. What About People in Regional Science? Papers, Regional Science Association, Vol. 24, 1970, pp. 7-21.
7. J. Tomlinson, N. Bullock, P. Dickens, P. Steadman, and E. Taylor. A Model of Students' Daily Activity Patterns. Environment and Planning, Vol. 5, 1973, pp. 231-266.
8. M. Shapcott and P. Steadman. Rhythms of Urban Activity. In Timing Space and Spacing Time--Volume 2: Human Activity and Time Geography (T. Carlstein et al., eds.), Edward Arnold, London, England, 1978, pp. 49-74.

9. D. Damm. Interdependencies in Activity Behavior. TRB, Transportation Research Record 750, 1980, pp. 33-40.
10. J. Supernak. Travel-Time Budget: A Critique. TRB, Transportation Research Record 879, 1982, pp. 15-28.
11. G.S. Root and W.W. Recker. Toward a Dynamic Model of Individual Activity Pattern Formation. In Recent Advances in Travel Demand Analysis (S. Carpenter and P.M. Jones, eds.), Gower, Aldershot, England, 1983, pp. 371-382.
12. M.J. Beckmann and T.F. Golob. A Critique of Entropy and Gravity in Travel Forecasting. In Traffic Flow and Transportation (G.E. Newell, ed.), Elsevier, New York, 1972, pp. 109-117.
13. J.H. Niedercorn and B.V. Bechdolt. An Economic Derivation of the Gravity Law of Spatial Interaction. Journal of Regional Science, Vol. 9, No. 2, 1969, pp. 273-282.
14. T. Adler and M. Ben-Akiva. A Theoretical and Empirical Model of Trip Chaining Behavior. Transportation Research, Vol. 13B, 1979, pp. 243-257.
15. J. Horowitz. A Disaggregate Demand Model of Non-Work Travel that Includes Multi-Destination Travel. TRB, Transportation Research Record 673, 1978, pp. 65-71.
16. D. Damm and S.R. Lerman. A Theory of Activity Scheduling Behavior. Environment and Planning A, Vol. 13, 1981, pp. 703-718.
17. J. Jacobson. Models of Non-Work Activity Duration. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Ph.D thesis, 1978.
18. R. Kitamura and M. Hioki. A Time Dependent Stochastic Model of Spatial Interaction: Model Framework and Properties. Presented at Annual Midcontinent Meeting of the Regional Science Association, Vail, Colo., June 1982.
19. L.P. Kostyniuk and R. Kitamura. Life Cycle and Household Time-Space Paths: Empirical Investigation. TRB, Transportation Research Record 879, 1982, pp. 28-37.
20. J.J. Heckman. Statistical Models for Discrete Panel Data. In Structural Analysis of Discrete Data with Econometric Applications (C.F. Manski and D. McFadden, eds.), MIT Press, Cambridge, Mass., 1981, pp. 114-178.
21. C.F. Daganzo and Y. Sheffi. Multinomial Probit with Time-Series Data: Unifying State Dependence and Serial Correlation Models. Environment and Planning A, Vol. 14, 1982, pp. 1377-1388.
22. D. McFadden, K. Train, and W.B. Tye. An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. TRB, Transportation Research Record 637, 1977, pp. 34-46.
23. S.R. Lerman. The Use of Disaggregate Choice Models in Semi-Markov Process Models of Trip Chaining Behavior. Transportation Science, Vol. 13, 1979, pp. 273-291.
24. R. Kitamura, L.P. Kostyniuk, and M.J. Uyeno. Basic Properties of Urban Time-Space Paths: Empirical Tests. TRB, Transportation Research Record 794, 1981, pp. 8-19.

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

Equilibrium Traffic Assignment on an Aggregated Highway Network for Sketch Planning

R.W. EASH, K.S. CHON, Y.J. LEE, AND D.E. BOYCE

An application of the equilibrium traffic assignment algorithm on a simplified highway network, such as might be used for sketch planning, is described. Analysis zones in the assignment are also substantially larger than in most conventional traffic assignments. The algorithm for equilibrium traffic assignment is introduced, followed by a discussion of the problems with equilibrium traffic assignment in a sketch-planning application. Next, the network coding procedures for the case study are examined. Results of the sketch-planning assignment are then evaluated against a comparable regional assignment of the same trips. Finally, there is a discussion of how this research fits into the programs of a transportation planning agency.

An application of equilibrium traffic assignment to sketch planning is presented in this paper. Trips are assigned onto an aggregated network with a limited number of links, nodes, and zone centroids. One arterial link in the sketch-planning network is equivalent to a number of links in a conventionally coded regional highway network, and one sketch-planning zone is substantially larger than a zone in the regional assignment at the same location. The traffic assignment algorithm used in the study converges to approximately equal path travel times for multiple paths between origin-destination zone pairs. The algorithm is available to most transportation planning agencies.

A major portion of the paper is spent on a comparison of this sketch-planning assignment with a regional traffic assignment of a large trip table onto a detailed coded highway network. This comparison is complicated by the different number of intrazonal trips in the two traffic assignments; therefore, a procedure was developed to determine the significance of the additional intrazonal trips in the sketch-planning assignment. Vehicle miles of capacity and travel, vehicle hours, and average speeds predicted by the two assignments are summarized at the regional and zonal levels.

In the introductory sections of the paper the equilibrium traffic assignment algorithm and the network coding procedures for the sketch-planning network are documented. A simple method for aggregating links and summing regional link capacities into sketch-planning link capacities is then described. The question of the best network aggregation procedure is not considered. Moreover, a solution of this network aggregation problem was not an objective of the research, but rather a data requirement. The principal concern of this paper is to demonstrate a satisfactory correspondence between traffic assignments on the sketch plan and regional networks. Finally, a few implications of this research for work programs of transportation planning agencies are discussed.

METHODOLOGICAL APPROACH

The equilibrium concept was first formulated for minimum time-path traffic assignment by Wardrop (1). Given that travel times on a network link increase with traffic, a highway network is in equilibrium if the travel times along all paths that are used between each origin-destination are equal, and no unused path has a lower time. In other words, no driver has an incentive to change paths.

Several algorithms were developed in the early 1970s to determine the equilibrium traffic flows, and one version of the algorithm is now available in the Urban Transportation Planning System (UTPS) computer programs for transportation planning supported by UMTA and FHWA (2). The formulation of the algorithm discussed here follows the work of Nguyen (3) and LeBlanc et al. (4) and is consistent with the algorithm available in the UTPS program UROAD.

For a given trip table, the equilibrium assignment of traffic may be found by solving a nonlinear mathematical programming problem. The solution to this problem is that set of traffic flows on network links that minimizes a nonlinear convex mathematical function (called an objective function), the value of which depends on the traffic flows. These flows must also satisfy a second set of linear equations called constraints. In general terms, the constraints on the objective function ensure that all solutions are feasible trip assignments; that is, all trips in the trip table are assigned to the network, and negative link flows are prohibited.

The objective function is to minimize the sum of the areas under each link's travel-time and traffic volume congestion function from zero to the assigned flow. To understand the interest in minimizing the sum of these areas requires some mathematical analysis beyond the scope of this paper. It is only important to understand that the link flows that correspond to the minimum value of this objective function are those that satisfy the equilibrium conditions.

Summary of Equilibrium Traffic Assignment Algorithm

The algorithm to solve the equilibrium traffic assignment problem is based on a nonlinear optimization technique developed by Frank and Wolfe (5). Theirs is an iterative approach that starts with an initial feasible solution that satisfies the constraints, determines a feasible direction to move that improves the objective function, and then calculates how far to move in this direction. This results in a new feasible solution, and the procedure iterates until the objective function cannot be improved.

A network composed of links with congestion functions, a trip table for assignment, and a first solution that is a feasible assignment of trips to the network are given. The equilibrium conditions are normally not met by this first trip assignment. Application of the method by Frank and Wolfe then involves the following steps.

1. Compute the travel time on each link by using volumes in the current solution.
2. Trace minimum time-path trees from each origin to all destinations by using the link times computed in step 1.
3. Assign all trips for each origin to each destination to the minimum paths computed in step 2 (this produces an all-or-nothing trip assignment).
4. Linearly combine the current link volumes (v_a) of the solution and the new all-or-nothing link volumes (w_a) of the assignment to obtain a new current solution (v'_a) that minimizes the objective function:

$$\sum_a \int_0^{v'_a} S_a(x)dx \tag{1}$$

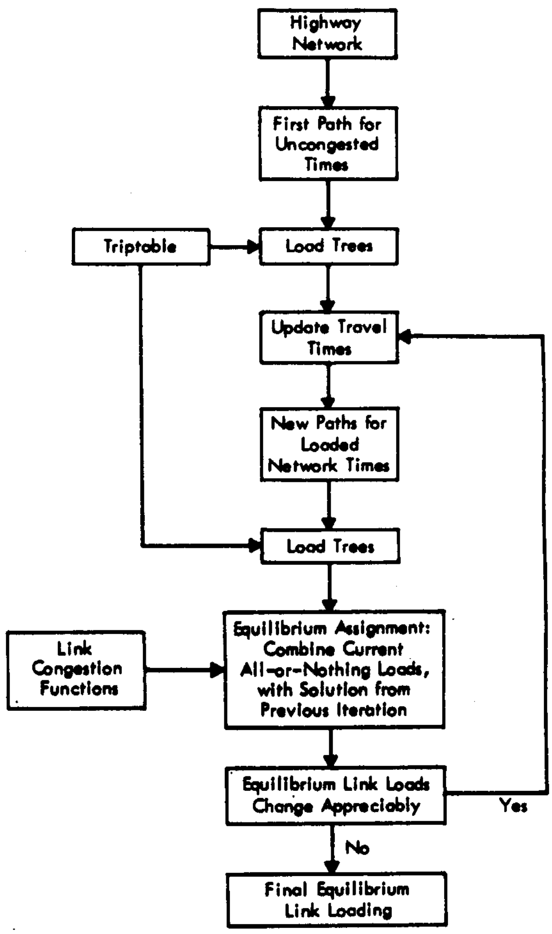
where

$v'_a = (1-\lambda)v_a + \lambda w_a$ = new current solution volume on link a,
$S_a(x)$ = link congestion function for link a, and
λ = constant between 0 and 1.

5. If the solution has converged sufficiently, stop; otherwise return to step 1.

The sequence of program steps is shown in the flowchart in Figure 1.

Figure 1. Equilibrium algorithm program steps.



Equilibrium Traffic Assignment and Sketch Planning

The obvious problems in applying equilibrium traffic assignment to sketch planning are how to simplify the traffic assignment network and analysis zones and the nature of the travel-time and traffic volume congestion function for such a network. Previously, researchers have constructed sketch-planning networks either by eliminating minor and lightly traveled links (6) or by aggregating links in a detailed network into summary links (7). The sketch-planning network for this project combines these two approaches and includes all freeway and expressway links with a grid network of aggregate links for arterial streets.

A number of time and volume relationships have been developed for traffic assignment when the coded network resembles an actual highway network (8). The most widely used is the Bureau of Public Roads (BPR) formula available in the program UROAD:

$$T = T_o \, [1 + 0.15\,(v/c)^4] \quad (2)$$

where

T_o = uncongested (zero traffic flow) travel time on the link,
T = estimated link travel time, and
v/c = ratio of link traffic volume to link capacity.

In a conventionally coded highway assignment network, each link is a street or highway segment, the attributes of which can be observed. To illustrate the detail coded into these networks, only local streets and rural roads used principally for land access are omitted in the regional network used by the Chicago Area Transportation Study (CATS). It is reasonable, therefore, to assert that the coded network built from all these individual links reflects the supply characteristics of the regional highway network.

If the regional network links to be combined in a sketch-planning link can be identified and acceptable regional network link congestion functions exist, then two methods for developing aggregate congestion functions appear plausible. First, the regional time and volume relationships can be mathematically combined to form an aggregate congestion function. Alternatively, a general link congestion function can be applied to a summary link, the attributes of which are aggregate quantities. Both methods were attempted in this project.

Further problems in using equilibrium assignment for sketch planning are caused by the larger analysis zones and the corresponding smaller trip table. For a conventional regional assignment, an analyst might have a trip table with a thousand or more zones. By comparison, no more than a few hundred zones can be used in a sketch-planning application.

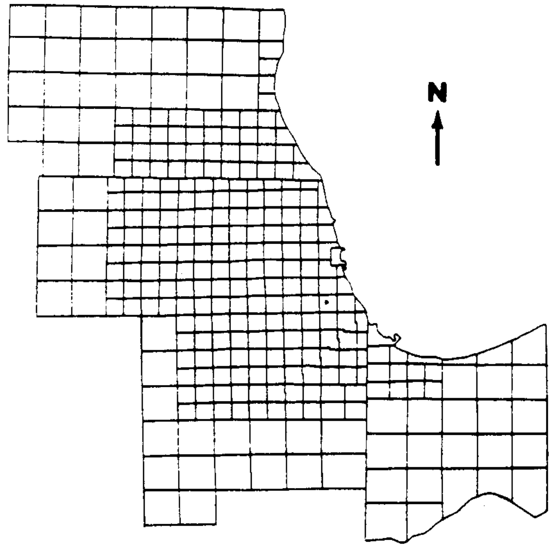
More trips occur within a zone when larger zones are used in a traffic assignment. Because intrazonal trips are not assigned to the highway network, this means that estimated traffic is reduced. This under-assignment of trips, in turn, affects congestion in the highway network and the travel times predicted by the link congestion functions. The effect of this larger number of intrazonal trips on the sketch-planning assignment was evaluated in this project.

The smaller trip table causes cell values to increase, and more trips are loaded onto the network at zone centroids. The links immediately adjacent to centroids are then loaded with all the traffic from the larger area covered by the sketch-planning zone. These links tend to be overassigned, which also affects the travel times predicted by the link congestion functions. Fortunately, this problem is mitigated by running more iterations of the equilibrium algorithm to load more paths in the sketch-planning network.

Coding the Sketch-Planning Network

The first step in coding the sketch-planning network was selection of the system of analysis zones. The zone system used in the project was developed by combining the CATS regional zones into a suitable number of areal units. Each sketch-planning zone usually includes four to nine regional zones. The resulting zone system covers the eight-county northeastern Illinois (six counties) and northwestern Indiana (two counties) region; it is shown in Figure 2. There are 317 sketch-planning zones compared to the 1,797 zones used in the CATS regional traffic assignments.

Figure 2. Sketch-planning zone system.



The basic areal unit in the region is the survey township, a roughly 6-mile[2] land unit originally surveyed in the mid-1800s. All of the CATS zone systems, including the sketch-planning zones, make use of these survey townships. A majority of the sketch-planning zones are quarter-townships (approximately 9 miles[2]), with full townships as the next largest group of zones. At the state line between Indiana and Illinois, a few zones are slightly larger than full townships, and several smaller odd-sized zones are along the lakefront.

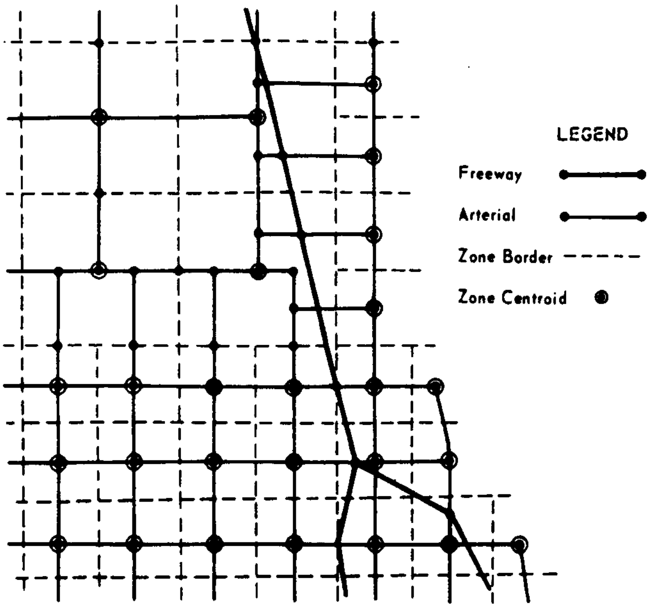
Zones are covered by a network of bidirectional arterial and freeway links (9). Each zone's centroid is located at the center of a zone and is connected by two to four arterial street links to produce a fairly regular grid network over the region. Freeway and expressway links are then coded on top of this regular grid of arterial street links, with interchanges placed approximately at their actual locations. A portion of the sketch-planning network is shown in Figure 3.

Links are coded as either arterials or freeways (expressways). Attributes coded for each sketch-planning link include beginning and ending node numbers, link length, type of area where link is located, link free speed, and link capacity. The type of area where the link is located is coded by using municipal boundaries and zone populations. Link free speed is then estimated for each link by using the area and facility types. All coding was done in the usual UTPS format, except that the traffic count field was used for link capacity and UROAD was altered to accept link capacities in this field.

Arterial Link Capacity and Congestion Functions

The original approach in the project to develop

Figure 3. Example of sketch-planning network coding.



sketch-planning arterial street network congestion functions was to aggregate mathematically the BPR formula congestion functions used in the regional network, as described by Morlok (10). This approach can straightforwardly be applied for two or more consecutive links, or for two parallel links between the same two nodes. The intent was to construct the sketch plan arterial network congestion functions from the regional network congestion functions by repeated aggregation using these two relationships. This approach proved far too time consuming to be completed manually, and it was believed that preparing suitable software to accomplish the work required substantial efforts beyond the scope of the project.

Given the geometry of the sketch-planning network and the arrangement of zones (two regular grid patterns offset so that each zone boundary is usually crossed by only one summary arterial street link), an obvious method for estimating sketch-planning arterial link capacities was to sum regional arterial network capacities along the edge of a sketch-planning zone. This was accomplished by first overlaying the sketch-planning zones on the regional highway network to identify the regional arterial street links crossing a zone boundary, and then summing the appropriate regional link capacities. This procedure is shown in Figure 4. Note that in this example the capacity of the first regional link is shared with the adjacent zone.

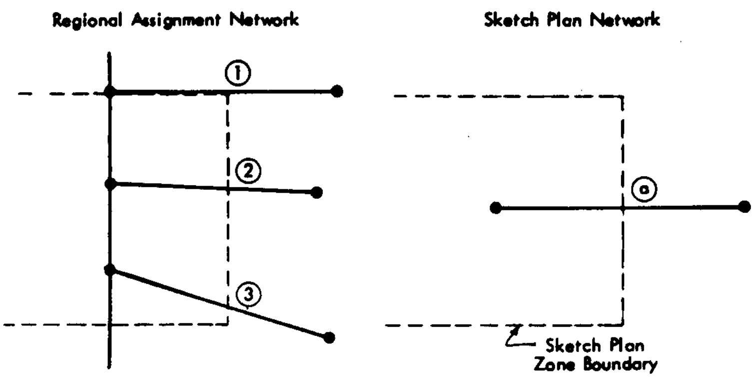
EVALUATION OF SKETCH-PLANNING ASSIGNMENT

Because there are separate zone systems in the regional traffic assignment and in the sketch-planning traffic assignment, intrazonal trips in the two assignments are different. This makes it difficult to compare the two assignments because fewer trips are assigned onto the sketch-planning network and fewer vehicle miles of travel are produced. To remove this bias from the comparison of the regional and sketch-planning assignments, an estimate of these added intrazonal trips and missing vehicle miles was needed.

A second assignment of trips onto the regional highway network was performed with a trip table that contained only the additional intrazonal trips in the sketch-planning assignment, i.e., the trips that became intrazonal when the regional zones were aggregated. This partial trip table was created by scanning the regional trip table and eliminating all entries that would be intrazonal in the sketch-planning zone system. The resulting intrazonal trip table was then assigned onto the same minimum time paths used in the regional traffic assignment. The proportion of the intrazonal trip table assigned to each minimum time path was the same as the proportion of the full trip table assigned to that minimum time path. Link volumes from the intrazonal trip assignment were then subtracted from the original link volumes of the regional assignment to produce a revised vehicle mile estimate.

Another difference between the two assignments is the number of iterations of the equilibrium algorithm. In the regional traffic assignment, five separate all-or-nothing assignments are completed,

Figure 4. Estimation of sketch-planning arterial link capacities.



Regional Link	Capacity	Sketch Plan Link	Capacity
1	C_1	a	$\frac{1}{2} C_1 + C_2 + C_3$
2	C_2		
3	C_3		

which correspond to four iterations of the equilibrium algorithm. For the sketch-planning assignment, 10 all-or-nothing assignments are prepared (9 iterations of the equilibrium algorithm); therefore, each interchange has the opportunity to travel 5 added paths. However, the sketch-planning assignment is still less expensive in computer costs. This points out the trade-off between detail in the assignment network and the number of paths that can be practically loaded in the equilibrium algorithm. As the network becomes more detailed, the cost of building minimum time paths increases, thereby restricting the number of iterations of the equilibrium algorithm that can be completed.

The CATS regional and sketch-planning assignments in the project are given in Table 1. Both assignments are for a 1-hr 1975 trip table in the morning peak period. The sketch-planning network is less than one-tenth the size of the regional network, even allowing for the fact that the regional network extends slightly beyond the eight-county area covered by the sketch-planning network. The data in Table 1 indicate that the number of trips in the two assignments is slightly different because of rounding during the allocation of the regional trip table into sketch-planning zones. There are an additional 137,000 intrazonal trips in the sketch-planning assignment.

Table 1. Summary of regional and sketch-planning network assignments.

Item	Sketch-Planning Network	Regional Network[a]
Analysis zones	317	1,797
Network nodes	820	12,040
One-way links	2,422	37,065
Assigned interzonal trips	1,016,900	1,140,400
Unassigned intrazonal trips	192,800	55,800
Number of iterations (all-or-nothing assignments)	10	5
Computing time[b] (CPU)	3 min, 45 sec	163 min, 7 sec
Memory required[b] (maximum bytes)	600K	540K

[a]The regional network covers a slightly larger area than the eight-county Chicago region.
[b]IBM 3033, Operating System VS2.

The last two items in Table 1 give the relative computer costs of the two assignments. The sketch-planning assignment was accomplished with the UTPS program UROAD (slightly modified to use link capacities from the network link file and an efficient line-search procedure), whereas the regional assignment made use of the PLANPAC programs originally prepared in the mid-1960s by the FHWA (11), with a separate program for the equilibrium algorithm (12). Different programs are required because of the size of the regional network, which is too large for the version of the UROAD program used in the project. Identical functions (path building, assignment, line search between all-or-nothing assignments, and calculation of link times) are carried out in both cases.

Computer memory requirements for the two assignments are about equal because UROAD allocates memory space according to the largest node number (more than 6,000 in this case) instead of the number of nodes in the network, and also because the individual PLANPAC programs can be written more efficiently for memory use because each program performs only a single function. The computer time required to run the sketch-planning assignment is almost insignificant compared with the regional assignment, even though twice as many iterations are performed for the sketch-planning assignment.

Regional Travel Comparison

The results of the regional and sketch-planning assignments within the eight counties are given in Tables 2-5. Because the sketch-planning network does not include any ramps between freeways or between freeways and arterials, ramps are not included in the regional vehicle miles of capacity (Table 2). Even without ramps, slightly more capacity is available in the regional network than in the sketch-planning network. Total arterial capacity in the two networks is surprisingly close, however, considering the crude method used to estimate the capacity of the sketch-planning arterial street links.

Several reasons can be cited for the discrepancy between the freeway capacities in the two networks. A few short freeway segments, most only a mile or so in length, are omitted from the sketch-planning network. Another reason is that the sketch-planning freeway links are coded somewhat abstractly as straight links between freeway interchanges. This tends to understate the actual length of these links.

The data in Table 3 give vehicle miles of travel for the two networks. Vehicle miles on ramps are included in the regional assignments, even though their capacity was omitted. Although ramps are not coded in the sketch-planning network, the vehicle miles of travel that would occur on ramps are approximated by additional travel to reach the single interchange node. Vehicle miles on ramps that connect freeways are included in the freeway category, and vehicle miles on ramps between freeways and arterials are split evenly between both route types in the regional assignment figures. Only vehicle miles within the eight counties are tabulated.

The data in Table 3 also describe the impact of the intrazonal trips in a comparison of the two assignments. When the sketch-planning and regional assignments are first compared, there is a difference of 4 percent in the division of vehicle miles between freeways and arterials. Twenty-nine percent of the unadjusted regional vehicle miles are assigned to freeways, whereas 31 percent of the sketch-planning vehicle miles occur on freeways. When the regional assignment is adjusted for the different number of intrazonal trips, part of this difference is explained. The great majority of trips in the intrazonal trip table is assigned onto arterials because these trips are short and are not likely to use a freeway.

The difference between the total vehicle miles in the sketch-planning assignment and the adjusted regional assignment is about 5 percent, and the extra vehicle miles on sketch-planning network freeways account for nearly all of the difference. After reviewing the coding of the two networks, it is clear that freeways in the sketch-planning network have some advantages that freeways in the regional network do not have. In addition to the slight undercoding of distance along sketch-planning freeway links noted earlier, the only radial links included in the sketch-planning network are freeway links, so paths made up only of arterial links must be longer than comparable paths in the regional assignment network.

Vehicle hours of travel for the assignments are given in Table 4. These estimates follow the pattern established in Table 3. Although total vehicle hours in the sketch-planning assignment and in the adjusted regional assignment are nearly equal, the distribution of the vehicle hours between freeways and arterials is somewhat different. In the sketch-

Table 2. Vehicle miles of capacity for eight-county region.

Highway	Vehicle Miles of Capacity (000s) Sketch Plan	Regional[a]	Intrazonal	Regional Less Intrazonal	Sketch Plan/ Regional
Freeway	3,870	4,241	NA	NA	0.91
Arterial	14,286	14,584	NA	NA	0.98
Total	18,156	18,825	NA	NA	0.96

Note: NA = not applicable.
[a]Ramp capacities not included.

Table 3. Vehicle miles of travel for eight-county region.

Highway	Vehicle Miles of Travel (000s) Sketch Plan	Regional	Intrazonal	Regional Less Intrazonal	Sketch Plan/ Regional Less Intrazonal
Freeway	3,476	2,985	3	2,982	1.17
Arterial	7,001	7,315	289	7,026	1.00
Total	10,477	10,300	292	10,008	1.05

Table 4. Vehicle hours of travel for eight-county region.

Highway	Vehicle Hours of Travel Sketch Plan	Regional	Intrazonal	Regional Less Intrazonal	Sketch Plan/ Regional Less Intrazonal
Freeway	104,962	81,549	103	81,446	1.29
Arterial	261,048	298,704	12,040	286,664	0.91
Total	366,010	380,253	12,143	368,110	0.99

Table 5. Average travel speed for eight-county region.

Highway	Avg Travel Speed (mph) Sketch Plan	Regional	Intrazonal	Regional Less Intrazonal	Sketch Plan/ Regional Less Intrazonal
Freeway	33.1	36.6	33.3	36.6	0.90
Arterial	26.8	24.5	24.0	24.5	1.09
Overall	28.6	27.1	24.1	27.2	1.05

planning assignment 29 percent of the vehicle hours are on freeways, whereas in the adjusted regional assignment only 22 percent of the vehicle hours are on freeways.

The data in Table 5 give the average network speeds computed as the ratio of vehicle miles to vehicle hours. Arterial links have higher average speeds in the sketch-planning assignment than in the regional assignment. Freeway average speeds in the sketch-planning assignment are slower than freeway speeds in the regional assignment because of the added freeway travel.

Travel at the Zone Level

Vehicle miles and average speeds from the two assignments were summarized and compared at the level of sketch-planning zones. Standard statistics were calculated for the distribution of these quantities among zones as well as the correlation between regional and sketch-planning values. All regional quantities used in this phase of the evaluation are actually adjusted quantities without the intrazonal trips added in the sketch-planning assignment.

Figures 5 and 6 are scattergram plots of the vehicle miles per sketch-planning zone and average sketch-planning zone speeds produced by the two assignments. Means and standard deviations for the vehicle mile and speed variables, and the square of the correlation coefficient between sketch-planning and regional variables, are also shown in each figure.

IMPLICATIONS FOR PLANNING AGENCIES

The question arises whether the work described in this paper is relevant for other transportation planning agencies. To a large extent, the sketch-planning zones and the geometry of the sketch-planning network used for this project are the result of the geography and township survey of the northeastern Illinois region. Because other metropolitan areas are spatially organized quite differently, it would be inappropriate to use the gridlike pattern of zones and arterial street links described here.

Figure 5. Scattergram of vehicle miles per zone.

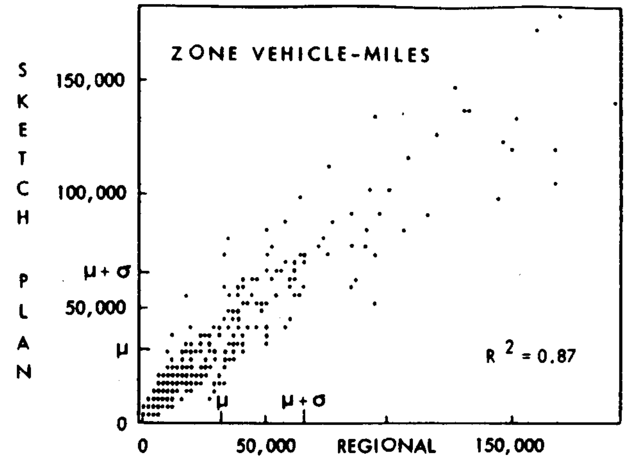
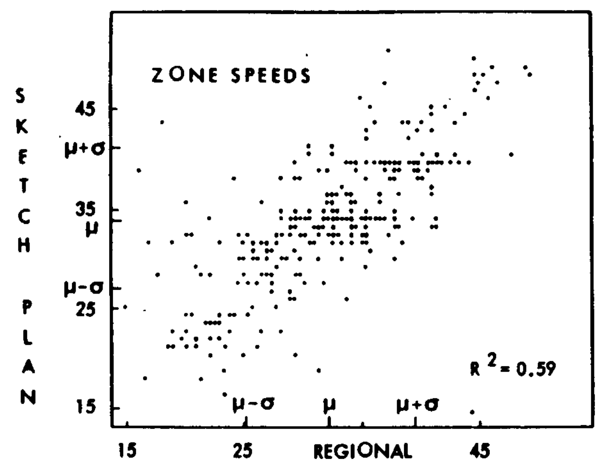


Figure 6. Scattergram of average zone speed.



Sketch-Planning Capabilities

In spite of the parochial nature of the zone pattern and network geometry of the example, some general conclusions can be drawn concerning the characteristics of equilibrium traffic assignment by using larger zones and simpler network coding. The most surprising result was that the different intrazonal trips in the two assignments did not significantly affect assignment results. For example, even though zones were 4 to 9 times larger in the sketch-planning assignment, the error introduced in the regional vehicle miles was less than 3 percent. It is apparent that even larger zones could be used without seriously biasing the traffic estimates.

The assignment of traffic is more seriously affected by the coding of the underlying arterial street network. In this case study the grid network of arterial streets increased arterial travel distances; as a result, the loadings on sketch-plan freeway links exceed the regional assignment values. The method used to estimate capacity in the sketch-plan arterial street network appears adequate, given the vehicle miles and average speeds that resulted.

Overall results from the sketch-planning assignment compared reasonably well with the regional assignment, and zone level assignment quantities were well correlated with regional assignment counterparts. Results from the sketch-planning assignment are, therefore, probably adequate for estimating most highway travel characteristics, including operating costs, emissions, and gasoline consumption, at regional and subregional levels.

Sketch Planning in Work Programs of Planning Agencies

Given these sketch-planning attributes relative to those of a conventional regional assignment, the sketch-planning methodology appears most applicable to long-range systems planning and strategic planning that deals with dramatic changes in transportation supply or demand characteristics. Project-level and corridor planning will almost always require more detailed network coding and smaller analysis zones. Nevertheless, the zone system and network in this sketch-planning example may be used to represent the balance of a region outside the corridor of interest.

Long-range systems planning concentrates on projected traffic or patronage for evaluation of alternative regional networks with different combinations of new major highway and transit investments. Unfortunately, the number of alternatives investigated is often limited because of the resources needed to support the conventional forecasting procedures. Less expensive approaches, such as the sketch-planning methods discussed here, will allow more alternatives to be tested and still provide reasonable estimates of traffic on major highway facilities.

There is also a trend in long-range transportation planning away from the evaluation of alternative networks of major facilities. In many metropolitan areas prospects for new major investments are limited, and future planning will emphasize more general transportation investment strategies for different energy, demographic, social, and economic resource scenarios. Sketch-planning approaches appear more suited for strategic planning than conventional techniques because more scenarios can be investigated and enough detail remains to accurately predict regional and subarea transportation impacts.

ACKNOWLEDGMENT

The research described in this paper is part of a larger project to develop a family of sketch-planning models that analyze general urban transportation and location issues. This project is a cooperative effort between the University of Illinois at Urbana-Champaign and CATS, with university staff primarily responsible for theoretical and software development. The resulting prototype versions of the sketch-planning models are being used in several projects in the CATS work program. The work at the university is supported by grants from the National Science Foundation and UMTA.

REFERENCES

1. J.G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineers, Part 2, Vol. 1, 1952, pp. 352-378.
2. UTPS Reference Manual. UMTA and FHWA, 1979.
3. S. Nguyen. An Algorithm for the Traffic Assignment Problem. Transportation Research, Vol. 4, 1974, pp. 391-394.
4. L.J. LeBlanc, E.K. Morlok, and W.P. Pierskalla. An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. Transportation Research, Vol. 9, 1975, pp. 309-318.
5. M. Frank and P. Wolfe. An Algorithm for Quadratic Programming. Navel Research Logistics Quarterly, Vol. 3, 1956, pp. 95-110.

6. G.R.M. Jansen and P.H.L. Bovy. The Effect of Zone Size and Network Detail on All-or-Nothing and Equilibrium Assignment Outcomes. Traffic Engineering and Control, Vol. 23, 1982, pp. 311-317, 328.
7. Y. Chan. A Method to Simplify Network Representation in Transportation Planning. Transportation Research, Vol. 10, 1976, pp. 179-191.
8. D.E. Boyce, B.N. Janson, and R.W. Eash. The Effect on Equilibrium Trip Assignment of Different Link Congestion Functions. Transportation Research, Vol. 15A, 1981, pp. 223-232.
9. S. Pathak, A. Rosenbluh, and D. Michals. Sketch Planning System. Chicago Area Transportation Study, Chicago, Tech. Memorandum, 1975.
10. E.K. Morlok. Introduction to Transportation Engineering and Planning. McGraw-Hill, New York, 1978, pp. 504-508.
11. Computer Programs for Urban Transportation Planning: PLANPAC/BACKPAC General Information. FHWA, 1977.
12. R.W. Eash, B.N. Janson, and D.E. Boyce. Equilibrium Trip Assignment: Advantages and Implications for Practice. Transportation Research Record 728, 1979, pp. 1-8.

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

Network Design Application of an Extraction Algorithm for Network Aggregation

ALI E. HAGHANI AND MARK S. DASKIN

The performance of a network extraction algorithm is described, and the algorithm is tested by using the network design problem. A network is chosen as the original network and is aggregated at different levels. The results of the optimal decision making under a common set of alternative actions are then compared against the original and the aggregated networks. The results suggest that the network aggregation algorithm is a useful tool in simplifying networks to reduce the computational burden associated with the network design problem, and to allow a broader range of policy options to be tested in a fixed amount of computer time than would be allowed by using the original disaggregated network.

Network aggregation is the art and science of condensing a given network into another one that (a) is small enough to be managed efficiently and effectively, and (b) preserves some desired characteristics or satisfies certain objectives or both (1). The usefulness of network aggregation schemes is particularly evident in instances when similar problems are to be solved on a network, or sensitivity analyses of various types are to be performed. Dealing with the detailed network in solving such problems entails high costs in terms of computer storage and time.

There are two main approaches to the network aggregation problem: network element (link or node) extraction and network element abstraction. Extraction of network elements means deletion of the elements of the network that are identified as being insignificant based on a prespecified criterion. Abstraction of the elements collapses the insignificant ones into pseudo or dummy elements. Network element extraction has the disadvantage of causing network disconnection (because of the removal of links). As a result, the remaining links of the network will be overloaded if the origin-destination (O-D) trip matrix is not adjusted appropriately. Network element abstraction is more difficult to perform. It is hard to transform the original network into an aggregate one, and moreover, it is even harder to translate the actions taken on the aggregate network into actions on the detailed network because of drastic changes in the topology of the network that occur during the aggregation process.

The primary objective in developing a network aggregation scheme should be to find an aggregation process that, when applied to a detailed network, results in an aggregate network that retains the physical appearance of the original one as much as possible. Thus when solving a decision-making problem, such as the network design problem on the aggregate network, the results should be easily transferable to the original one. With this in mind, and because the abstraction process changes the topology of the network and cannot effectively serve the process, it is proposed that an aggregation algorithm, which focuses primarily on link extraction, be used. Node extraction is a process that follows link extraction; when all links incident to a node are extracted, the node will be extracted. The algorithm is presented in the next section.

NETWORK EXTRACTION ALGORITHM

Let N(V,A) be a network, where V is the set of vertices or nodes and A is the set of arcs or links. Let T be the set of destinations and S be the set of origins, S and T $\subset$ V. Let x_i^t be the flow over link i destined to t, x_i^s be the flow over link i originated from s, and x_i^{st} be the flow over link i that originates from s and is destined to t, $i \epsilon A$, $s \epsilon S$, and $t \epsilon T$. Also, let x_i denote the flow over link i,

$$x_i = \sum_{s\epsilon S} x_i^s = \sum_{t\epsilon T} x_i^t = \sum_{s\epsilon S} \sum_{t\epsilon T} x_i^{st},\ i\epsilon A,\ s\epsilon S, \text{ and } t\epsilon T \quad (1)$$

Moreover, let D = (d_s^t) be the O-D trip matrix. Finally, let $C_i(x_i)$ represent the average cost of travel on link i at flow x_i that is continuous, differentiable, Riemann integrable, convex, and strictly increasing.

It is assumed that the distribution of flow over a transportation network is based on Wardrop's first principle (2)--user equilibrium (3). There are some links in the network that, after the distribution of the flow has taken place, will not carry a significant amount of traffic. These links are the ones that will be focused on in the aggregation process to be described. The criterion used for the identification of insignificant links is defined as follows.

A link in a network is insignificant if the corresponding equilibrium flow is below α percent of the maximum equilibrium link flow in the network. The level of network aggregation changes, depending on the value of α; as α increases, the network becomes more aggregated and vice versa.

The reason for choosing the level of flow in the links as a criterion for identifying the insignificant links is that many transportation problems deal with the equilibrium flow levels in the network links. It has already been proved (1) that the equilibrium flow level in the significant or nonextracted links remains unchanged when the aggregation scheme is used. By preserving the level of equilibrium flow in the nonextracted links, the aggregation scheme should produce an aggregate network that is more representative of the detailed network than is an aggregation process that failed to preserve these flow levels. This should be particularly important in solving problems in which the objective function is based on the level of equilibrium flow in the links. The network design problem is one such problem.

Thus the network extraction algorithms as it has been coded, is presented. A more rigorous presentation is included elsewhere (1). The inputs to the algorithm are the specifications of the original network N(V,A), the average link cost functions $C_i(x_i)$, $i \epsilon A$, and the O-D trip matrix D. Either the maximum number of links to be extracted or the maximum α percentile denoting the cutoff point between the insignificant and significant link flows should also be given. Through this process certain prespecified links in the aggregate network can be maintained; also, specific links can be extracted. Furthermore, the algorithm extracts the links one by one and provides the results after each iteration. As a result, several different aggregate networks are obtained. The principle is to extract insignificant links and to update the trip matrix such that the flow level in the remaining links of the aggregate network remains unchanged. The algorithm is as follows.

Step 1: Specify α or the maximum number of links that may be extracted (M). Solve the equilibrium flow problem. Let x_i^*, $i \epsilon A$ be the equilibrium flow on link i.

Step 2: Identify the unextracted link k with the minimum flow. Let t(k) and h(k) denote the tail and head nodes of link k. Compute

$$\alpha_k = x_k^*/\mathrm{Max}\,(x_i^*) \tag{2}$$

If $\alpha_k > \alpha$ as specified in step 1, or if the number of extracted links is greater than the maximum number of links that may be extracted (specified in step 1), stop. Otherwise disaggregate the flow on link k by specifying the origin and destination of all flow on the link, which is done by solving the equilibrium flow problem on the most aggregate network generated. [An outline of how the O-D specific link flow (x_k^{st}) may be obtained from the solution procedure to the equilibrium flow problem, and the problems associated with the nonuniqueness of this quantity, are discussed elsewhere (1).] Go to step 3.

Step 3: Discard link k. Declare t(k) a destination (if it is not already such a node) and h(k) an origin (if it is not already such a node).

Step 4: Update the trip matrix as follows: (a) type I entry, where t(k) is a destination, i.e.,

$$\hat{d}_s^{t(k)} = d_s^{t(k)} + x_k^t \tag{3}$$

where $d_s^{t(k)}$ is the original O-D matrix element and is taken to be zero if the t(k) is a new destination, and $\hat{d}_s^{t(k)}$ is the updated O-D trip matrix element; (b) type II entry, where h(k) is an origin, i.e.,

$$\hat{d}_{h(k)}^t = d_{h(k)}^t + x_k^t \tag{4}$$

where $d_{h(k)}^t$ is the original O-D trip matrix element and is taken to be zero if h(k) is a new origin, and $\hat{d}_{h(k)}^t$ is the updated O-D trip matrix element; and (c) type III entries, where all remaining entries of the O-D trip matrix ($\hat{d}_s^t$) are substituted by $d_s^t - x_k^{st}$ (subtracting from d_s^t the part of the demand from s to t that is now destined to a new destination and that will reoriginate from a new origin).

Certain properties of the algorithm are worth mentioning. First, as previously noted, the algorithm preserves the level of equilbrium flow in the links of the network that are not extracted. Second, in cases in which all of the nodes of the network are not both origins and destinations, the algorithm will increase the number of origins and destinations in the aggregate network. This in turn might have adverse effects on the computation time of the network design problem by increasing the number of origins and the associated time for computation of the minimum paths in the network. This situation has not been examined in this paper. However, this increase in computation time should be offset through other means.

Third, the result of the extraction process may be a set of disconnected subnetworks. If this occurs, the analysis of the aggregate network, now a set of subnetworks, will be much easier to undertake. In fact, in cases in which link extraction will increase the number of origins and destinations (and thereby increase the computation time for the network design problem), specification of the links to be extracted can force the aggregate network to be a set of disconnected subnetworks. In this way the computational savings obtainable by having disconnected subnetworks may be used to offset the increased time that results from additional origins and destinations.

Finally, in the network design problem it is shown that for a given budget level, the total cost to the users of the network, as measured on the detailed network, is overestimated by the solution to the network design problem that uses the aggregate network (1).

In the next section this algorithm is applied to an original network, and the network design problem is solved on the original detailed network and on a series of aggregate networks.

APPLICATIONS OF NETWORK EXTRACTION ALGORITHM TO NETWORK DESIGN PROBLEM

Problem Description

The network design problem is that of finding a set of feasible actions or projects from among a collection of such actions that, when implemented, optimize the objective function(s) being considered. The feasibility of a set of actions is determined by resource, physical, and environmental constraints (4). Traditionally, the objective function in the network design problem has been formulated as the minimization of the total number of vehicle hours of travel on the network, with flows and travel times computed based on user equilibrium. This is represented as

$$\underset{p\epsilon P}{\text{Minimize}} \sum_{i\epsilon A} x_i^* \, C_i \, (x_i^*) \tag{5}$$

Subject to budget constraint on the cost of implemented projects p

where x_i^* is the user equilibrium flow on link i, and P is a set of projects (p) under consideration for implementation. In solving the network design problem, a modified objective function, which was suggested by Poorzahedy (4) in his algorithm I, has been used. This form is as follows:

$$\underset{p\epsilon P}{\text{Minimize}} \sum_{i\epsilon A} \int_0^{x_i^*} C_i \, (v_i) dv_i \tag{6}$$

Subject to budget constraint on the cost of implemented projects p

where x_i^* is the user equilibrium flow on link i.

The modified form of the problem was selected because of the availability of a computer code to solve this problem. Also, solving this form of the problem has been found to be more efficient than solving the traditional formulation and generally results in similar actions being taken on the network (5).

Thus the results of a set of experiments designed to test the effectiveness of the proposed NA algorithm in solving the modified network design problem can be presented. For the detailed network, the Sioux Falls, South Dakota, network is used in the experiments because it is a well-documented network and has been used by other researchers (4,6,7) in analyzing network design problems.

The detailed network, which consists of 24 nodes and 76 links (or 38 link pairs, allowing two-way traffic movements), is shown in Figure 1 (4). The link travel costs $[C_i(x_i)]$ are given by functions of the form

$$C_i \, (x_i) = a_i + b_i \, (x_i)^4 \tag{7}$$

The constants a_i and b_i for each of the existing links in the network, as well as for the six candidate links, are given in Table 1 (4). Also provided in the table is the cost of implementing each of the candidate links. The first five projects represent improvements on existing links, whereas the sixth project is an entirely new link. Two different sets of experiments were considered. In the first, only the five improvement projects are used; in the second, all six candidate projects are used. The O-D matrix for this network is given in Poorzahedy (4).

Five aggregate networks are developed, which result from the extraction of 6, 12, 18, 24, and 30 links. The aggregate networks are shown in Figures

Figure 1. Original network (4).

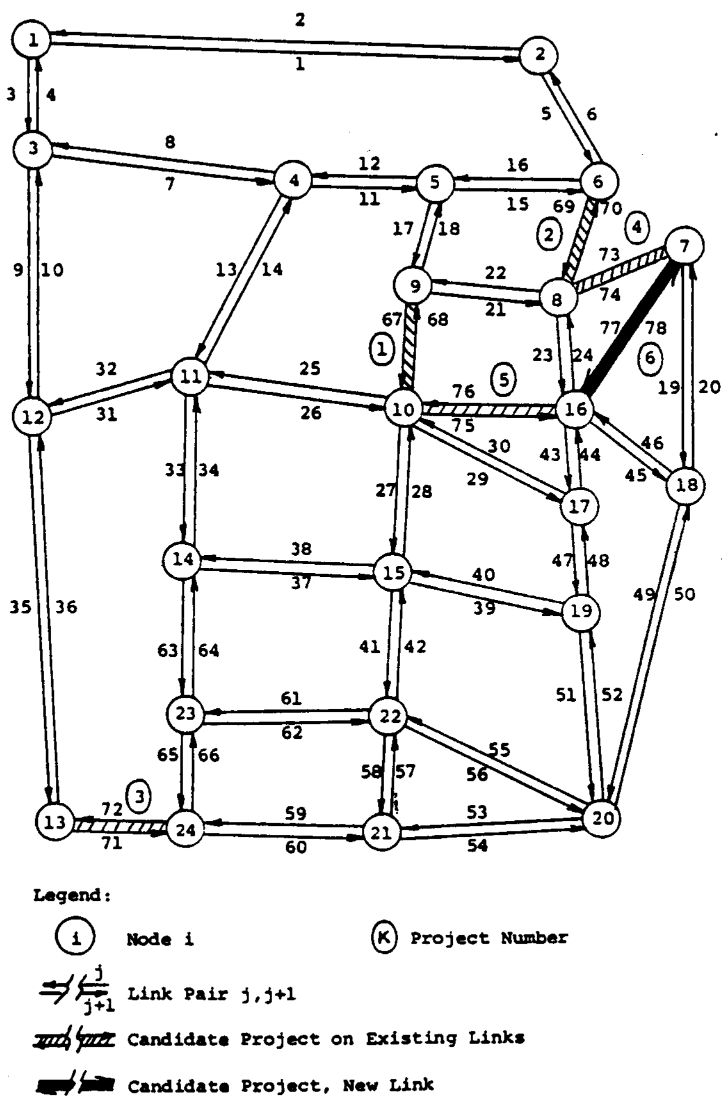


Table 1. Link parameters of test network III (4).

Link	a(x10^{-2})	b(x10^{-4})	Link	a(x10^{-2})	b(x10^{-4})	Link	a(x10^{-2})	b(x10^{-4})
1,2	5.96	0.00023	35,36	2.98	0.00011	69,70	2.17	0.05208
3,4	4.34	0.00017	37,38	4.52	0.10848	71,72	3.72	0.08928
5,6	5.17	0.12408	39,40	3.50	0.00104	73,74	2.50	0.01185
7,8	4.31	0.00069	41,42	3.50	0.00525	75,76	4.50	0.10800
9,10	4.14	0.00016	43,44	1.67	0.04008	77,78	-	-
11,12	2.16	0.00035	45,46	2.69	0.00025	New Data		
13,14	6.46	0.15504	47,48	2.31	0.05544	67,68	1.60	0.00037
15,16	4.17	0.10008	49,50	4.46	0.00017	cost	$625.x10^3$	
17,18	5.03	0.00755	51,52	3.99	0.09576	69,70	1.30	0.01562
19,20	2.18	0.00008	53,54	5.72	0.13728	cost	$650.x10^3$	
21,22	9.61	0.23064	55,56	4.71	0.11304	71,72	2.20	0.02678
23,24	4.82	0.11568	57,58	1.67	0.04008	cost	$850.x10^3$	
25,26	5.00	0.00750	59,60	3.29	0.07896	73,74	1.50	0.00355
27,28	5.87	0.00265	61,62	4.00	0.09600	cost	$1000.x10^3$	
29,30	8.04	0.19296	63,64	4.25	0.10200	75,76	2.70	0.03240
31,32	6.46	0.15504	65,66	1.88	0.04512	cost	$1200.x10^3$	
33,34	4.42	0.10608	67,68	2.75	0.00124	77,78	3.00	0.00321
						cost	$1500.x10^3$	

Note: a parameters are given in hours, and b parameters are given in hours ÷ (1,000 vehicles per day)^4.

2-6; the resulting O-D matrices are given in Haghani (1).

Results for the Aggregation Model

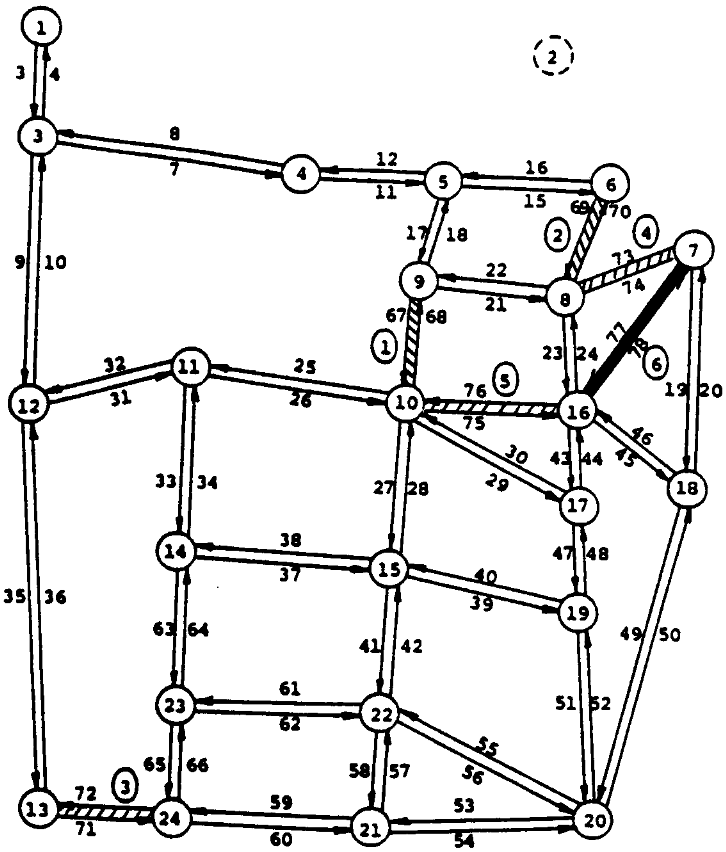
The five aggregate networks shown in Figures 2-6 and the detailed network shown in Figure 1, along with their corresponding O-D matrices, constitute the basis for the experiments. On each of these six networks, two network design problems were solved: one with the first five candidate projects, and the second with all six projects. The initial budget was set at $2,000,000 in all cases, and a complete sensitivity analysis (with respect to increases in the budget) was performed for all six networks and both design problems.

Results of the Five-Project Experiment

The results of the five-project experiment are summarized in Table 2. More detailed results are given in Haghani (1). The data in the table report (a) the percentage error in the total number of vehicle hours on the aggregate networks as compared with the detailed network, and (b) the number of projects that are selected differently when the network design problem is solved on the detailed and aggregate networks. Note that there are 18 unique budget levels that must be considered in performing the sensitivity analyses, beginning with a budget of $2,000,000 and ending with a budget of $4,325,000, which allows the implementation of all five projects. Of the 18 budget levels, 5 result in different solutions for the network design problem on the original and aggregate networks in the worst case.

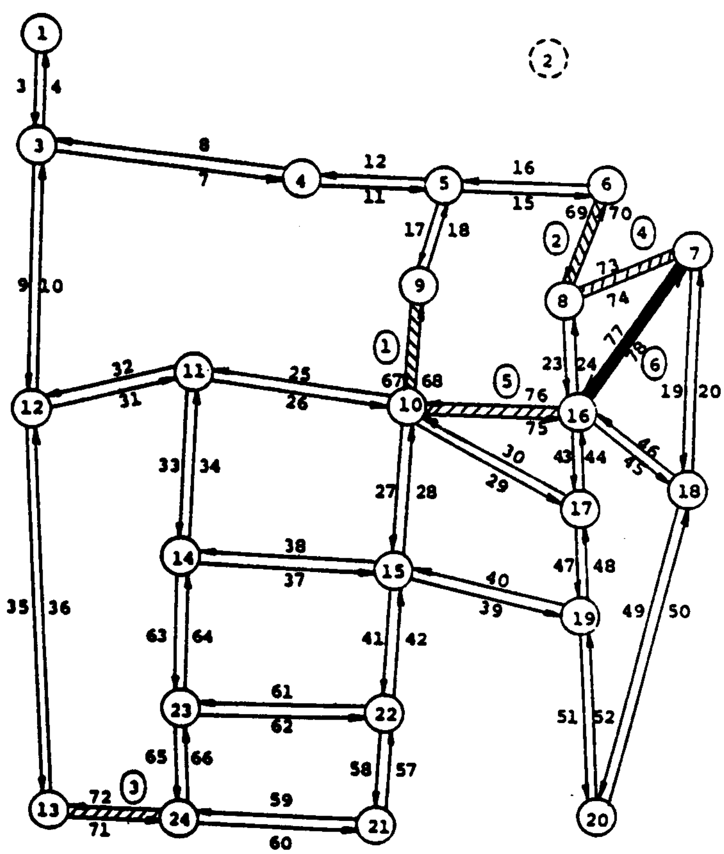
The data in Table 2 indicate that with six links deleted from the original network, the solutions to the network design problem on the original network and the aggregate network are identical for all budget levels. For higher levels of aggregation, discrepancies occur between the solution using the aggregate network and that found using the original network. Also note that most of the errors occur when the ratio of the budget level to the total cost

Figure 2. Aggregate network with 6 extracted links.



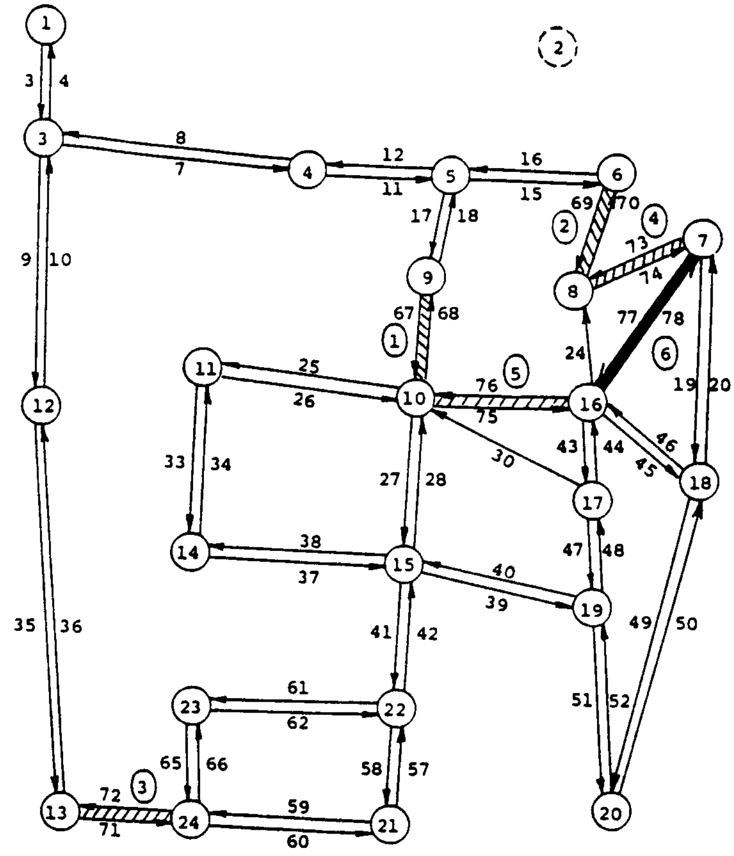
Note: Legend is given in Figure 1.

Figure 3. Aggregate network with 12 extracted links.



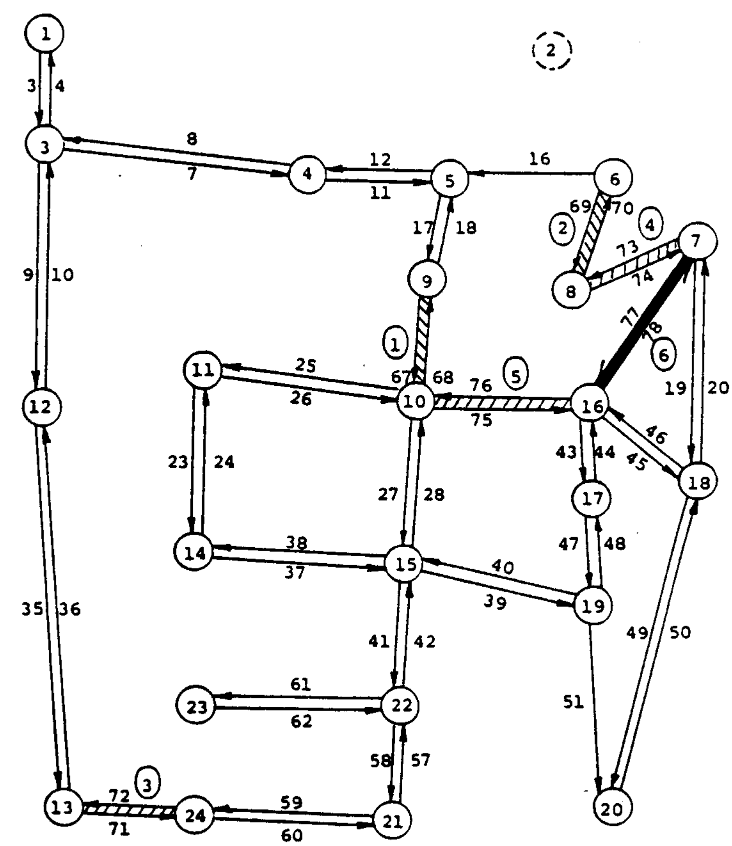
Note: Legend is given in Figure 1.

Figure 4. Aggregate network with 18 extracted links.



Note: Legend is given in Figure 1.

Figure 5. Aggregate network with 24 extracted links.



Note: Legend is given in Figure 1.

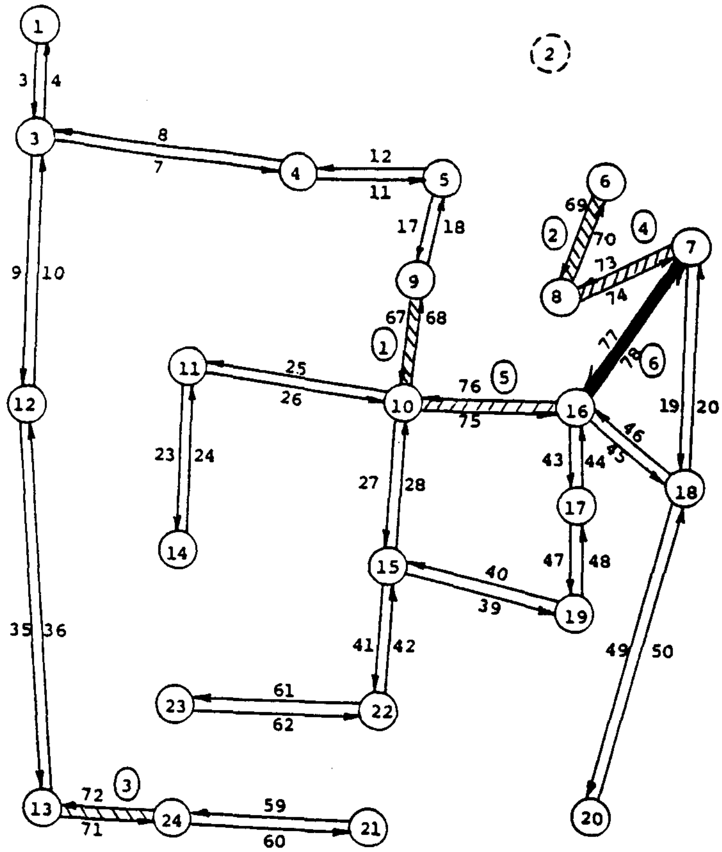


Figure 6. Aggregate network with 30 extracted links.

Note: Legend is given in Figure 1.

Table 2. Percentage of vehicle hour errors and number of misselected projects for five-project case.

		No. of Extracted Links									
		6		12		18		24		30	
Budget Levels ($000s)	No. of Budget Levels	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects
B = 2,000	1	0	0	3.11	1	3.11	1	0	0	3.11	1
B = 2,050	1	0	0	6.73	1	6.73	1	6.73	1	6.73	1
2,125 < B < 2,475	3	0	0	0	0	0	0	0	0	0	0
2,475 < B < 2,675	2	0	0	3.82	1	3.82	1	3.82	1	3.82	1
B = 2,675	1	0	0	7.95	1	7.95	1	7.95	1	7.95	1
2,700 < B < 3,325	5	0	0	0	0	0	0	0	0	0	0
3,325 < B < 4,325	4	0	0	0	0	0	0	0	0	0	0
B = 4,325	1	0	0	0	0	0	0	0	0	0	0

of all candidate links is low. In all cases in which the solution on the aggregate network differed from the solution on the detailed network, the number of misselected links was only one. By using vehicle hours as the measure of effectiveness, the maximum percentage error is 7.95 percent. The identity of the errors, the equality of their severity, and the similarity of their frequency across the various levels of aggregation suggest that the size of the network may be reduced significantly without increasing the magnitude of the errors. This phenomenon is also apparent in the case of six projects.

{Note that the maximum percentage error of 7.95 percent is computed as follows. At a given budget level, let $\underline{Y}_a$ and $\underline{Y}_o$ be the optimal solutions to the network design problem for the aggregate and original networks, respectively, where $\underline{Y} = (y_i)$ and $y_i = (0,1)$ if project i (is not, is) chosen to be in the optimal set. Also, let $V(\underline{Y})$ represent the decrease in the total number of vehicle hours in the original network that results from implementing project set $\underline{Y}$. The percentage error is defined as $[V(\underline{Y}_o) - V(\underline{Y}_a)/V(\underline{Y}_o)] \cdot 100$.}

Note also that the total travel time on the network is overestimated by the solution to the design problem on the aggregate networks as compared with the total time found when using the detailed network. This is also shown in the six-project experiment and, as noted at the end of the section on Network Extraction Algorithm, may be shown to be a general property of the aggregation process.

Results of the Six-Project Experiment

The results of the six-project experiment are given in Table 3. Again, the data in the table report two sets of statistics: (a) the percentage error in the total number of vehicle hours on the aggregate networks as compared with the number on the detailed network, and (b) the number of projects that are selected differently when the network design problem is solved on the detailed and aggregate networks. [More detailed results are given in Haghani (1).] There are now 40 unique budget levels that must be considered in performing the budget sensitivity analysis. Of the 40 budget levels, 9 result in different solutions for the network design problem on the detailed network and the aggregate networks in the worst case.

Again, when only six links are extracted from the network, the results on the detailed network and on the aggregate network are identical. As the level of network aggregation is increased, the number of budget levels at which discrepancies occur between the solutions on the two networks increases: when 12 or 18 links are extracted, errors occur at 5 of the 40 budget levels, or 12.5 percent of the time; with 24 links extracted, errors occur 20 percent of the time (in 8 cases); and with 30 links extracted, errors occur 22.5 percent of the time (in 9 cases). Again, the maximum number of misselected projects is 1, and the maximum difference in vehicle hours is 7.95 percent.

The relatively sharp increase in the number of budget levels at which errors occur, as the level of aggregation is increased from 18 extracted links to 24, suggests that a trade-off must be made between the expected accuracy of the results and the level of detail preserved in the network. The desired point on this trade-off curve depends on the use to which the analysis will be put; this is a decision best left to the analyst in each case.

Finally, there appear to be patterns to the misselected links. In both sets of experiments the aggregate network solution replaced project 2 or project 5 with project 3; in the six-project experiment, the aggregate solution also replaced project 6 with project 4. These replacements appear to be related to the degree of network aggregation in the neighborhood of the candidate projects. Also, because the errors occur at the same budget levels in all three cases, there appears to be a relationship between the errors and the specified budget levels. A more complete exploration of the relationship between these errors and (a) the degree of network aggregation in the neighborhood of the candidate projects and (b) the specific budget levels would be an interesting area for future research. The nature of the errors found in this analysis are discussed in greater detail later in the paper.

Computation Times

One of the major reasons for implementing a network aggregation process is, as already noted, the savings in computer time that should result from solving problems on smaller, less-detailed networks. The central processing unit (CPU) times required for the solution of the two network design problems, with all the budget sensitivity analyses on each of the six networks, are given in Table 4. All CPU times are for the Univac 1100, unless otherwise noted.

The network design problem for the five-project case may be solved 6 times faster on the most aggregate network considered in this analysis than on the original detailed network. In the six-project case, the ratio of CPU time on the detailed network to that on the most aggregate network is 4.4. However, these savings must be offset by the CPU time required to perform the aggregation process before the network design problem is solved. In some cases this may be of the same order of magnitude as the saving that results from using the aggregate network to solve the network design problem.

Nevertheless, several points are worth mentioning. First, the aggregation algorithm provides the analyst with one aggregate network for each link extracted from the detailed network. Thus the aggregate network that best suits the analysis purposes can be selected. Second, it is not likely that only a single network design problem will be solved; rather, a series of problems will be solved, each of which might be solved on a different aggregate network. For example, the analyst might elect to screen a large number of candidate projects by using a highly aggregated network. The most promising projects, along with others that might be of interest for nontechnical reasons, would be retained in more detailed analyses conducted on more detailed networks.

Third, with a given aggregate network, a large number of sensitivity analyses can be performed. As already indicated, it is likely that budget sensi-

Table 3. Percentage of vehicle hour errors and number of misselected projects for six-project case.

		No. of Extracted Links									
		6		12		18		24		30	
Budget Levels ($000s)	No. of Budget Levels	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects	Vehicle Hour Error (%)	No. of Misselected Projects
B = 2,000	1	0	0	3.11	1	3.11	1	0	0	3.11	1
B = 2,050	1	0	0	6.73	1	6.73	1	6.73	1	6.73	1
2,125 < B < 2,475	5	0	0	0	0	0	0	0	0	0	0
2,475 < B < 2,675	2	0	0	3.82	1	3.82	1	3.82	1	3.82	1
B = 2,675	1	0	0	7.95	1	7.95	1	7.95	1	7.95	1
2,700 < B < 3,325	9	0	0	0	0	0	0	0	0	0	0
3,325 < B < 4,325	12	0	0	0	0	0	0	0	0	0	0
4,325 < B < 4,825	4	0	0	0	0	0	0	0	0	0	0
4,825 < B < 5,825	4	0	0	0	0	0	0	6.18	1	6.18	1
B = 5,825	1	0	0	0	0	0	0	0	0	0	0

Table 4. CPU times for network design problems with complete budget sensitivity analysis on the original network III and the corresponding aggregate networks.

Aggregation Level	No. of Extracted Links	CPU Time[a] (sec) Five Candidate Projects	Six Candidate Projects
0	0[b]	50[c]	79[c]
0	0[b]	314.384	514.598
0.2797	6	246.280	398.487
0.3352	12	249.677	409.988
0.3413	18	249.002	417.027
0.3564	24	207.315	378.481
0.3780	30	51.496	124.228

[a]All figures, except those noted in footnote c, are from a UNIVAC 1100.
[b]Original network.
[c]Figures are from a CDC 6600, as reported by Poorzahedy (5).

tivity analyses will be performed. In addition, the sensitivity of the solution to additional constraints that require selected projects to be included in (or excluded from) the optimal solution may need to be analyzed. In all of these cases the network aggregation algorithm needs to be solved only once. Thus the CPU time for the network aggregation algorithm is best viewed as a fixed cost that may be distributed over a large number of analyses.

Finally, note that the CPU time involved in solving the network design problem decreases significantly as a result of extracting six links from the network in both the five- and six-project experiments. The CPU times for the cases of 6, 12, and 18 extracted links are comparable. A slight decrease in CPU time is experienced as a result of extracting 24 links, and a significant decrease is found when 30 links are extracted. This result, combined with the results outlined in the section Results for the Aggregation Model, which describes the accuracy of the results at various levels of aggregation, clearly suggests that there is an important tradeoff to be made between decreased computation costs (and greater network aggregation) on one hand and improved solution accuracy on the other hand.

In the sample problems previously discussed, it appears that desirable aggregation levels would correspond to either the extraction of 6 links (resulting in a moderate decrease in computer time and a high level of accuracy) or the extraction of 30 links (resulting in a large decrease in computer time at the expense of decreased solution accuracy). Intermediate levels of aggregation appear to result in relatively large solution errors without large compensations in terms of solution times. An interesting area of future research would be to determine whether or not the network aggregation algorithm results in such identifiable choices between aggregation and solution accuracy in other network design problems, and more generally, in other network problems.

SOURCES OF DISCREPANCY BETWEEN AGGREGATE NETWORK AND DETAILED NETWORK RESULTS

The results presented in the previous sections on the application of the proposed network aggregation algorithm to the network design problem are generally promising. In no case is the difference in the improvement in vehicle hours between the solutions on the detailed and the aggregate networks greater than 7.95 percent. Also, the two solutions differ by at most one candidate link in all cases. Nevertheless, there are several differences that warrant further explanation. As indicated in the following paragraphs, the test case selected is likely to exaggerate the extent of the differences that are likely to result in a more realistic planning context.

Two characteristics of the test problem will tend to result in an overestimation of the errors that result from using the network aggregation scheme. First, the Sioux Falls network being used is already a highly aggregate representation of the actual road network. This is evident when the range in equilibrium flows on the original network under the do-nothing option is examined; i.e., the maximum link flow is less than 4 times the minimum link flow. The average link flow is 12,989 vehicles and the maximum flow is 24,901 vehicles. The flow in the 30th link extracted from the network is 9,839 vehicles, or almost 40 percent of the maximum link flow. A real network is likely to exhibit a much greater range in equilibrium flows. If the extracted links truly carry an insignificant level of flow compared with the flow on the maximum flow link, the solutions to design problems on the aggregate networks are likely to be much better than they were in the test problem in which the flow levels on the extracted links were actually quite large and significant.

Second, the number of candidate links in the design problem was large relative to the total number of links in the detailed network. There are 6 two-way candidate links on a network with only 38 links. In the aggregate networks the situation is even more dramatic. When 30 (one-way) links are extracted, 26 percent of the links are being considered as candidate links. Thus the changes under consideration for the network are quite radical when compared with more realistic situations in which only 1 or 2 percent of the links are likely to be considered candidate links. Again, if the ratio of the number of candidate links to the number of links in the detailed network is small, the solution to the design problem on an aggregate network is more likely to replicate the solution on the detailed network than was found in the test problem, in which almost 16 percent of the links in the detailed network were candidate links.

In summary, the test network chosen for study is already a highly aggregate network that exhibits a relatively small range in equilibrium link flows. Also, the number of candidate links is extremely large relative to the total number of links in the network. It is expected that, if a more realistic detailed network is used, the solution to the network design problem using an aggregate network will more closely approximate the solution using the detailed network than was found in the small test network.

Finally, note that the aggregation process extracts links sequentially, thereby propagating computational errors and accumulating them in the final aggregate network. The resulting O-D trip matrix carries these errors to the decision-making model--in this case the network design model. Had a simultaneous extraction process been developed, this source of error would have been eliminated. To date, however, a simultaneous extraction process that circumvents the multiple counting danger has not been implemented.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

A network extraction algorithm for the network aggregation problem has been presented. The algorithm is based on the extraction of those links in a detailed network whose equilibrium link flows are less than a user-specified fraction of the maximum equilibrium link flow. The algorithm is sufficiently flexible to allow the analyst either to force certain links out of the detailed network or to retain particular links in the resulting aggregate network. Links are sequentially extracted from the network and, after each extraction, a modified O-D matrix is derived. The revision in the O-D trip matrix preserves the level of equilibrium flow in the nonextracted links. By extracting links sequentially, the algorithm provides the analyst with multiple aggregate networks--one after each link extraction.

The network aggregation algorithm was tested by examining the performance of a network design algorithm (4) on both a detailed network and five aggregate networks derived from the detailed network. The results are quite encouraging. The maximum percentage error in the improvement in vehicle hours of travel between a solution using the detailed network and a solution using an aggregate network was 7.95 percent. In most cases the same projects were selected for implementation when using both the detailed and the aggregate networks; when the solutions differed, at most one link was misselected when using the aggregate network. As suggested in the previous section, it is anticipated that even better results will occur when the algorithm is applied to networks that are larger and more realistic than is the 76-link, 24-node test network presented here.

Several promising areas for future research are suggested by this study. First, links are extracted from the network in increasing order of the ratio of the equilibrium flow in the link to be extracted to the maximum equilibrium link flow. Other criteria should also be investigated. For example, in certain contexts it may be desirable to extract links based on the ratio of the equilibrium flow in the link to the capacity of the link. Alternatively, hybrid criteria might be developed. For example, in the traditional formulation of the network design problem, the objective function is

$$\text{Minimize } Z = \sum_{\substack{\text{all links}\\ i}} x_i^* C_i(x_i^*) \tag{8}$$

which may be rewritten as

$$\text{Minimize } Z = \sum_{\substack{\text{links } i \text{ in aggregate}\\ \text{network}}} x_i^* C_i(x_i^*) + \sum_{\substack{\text{deleted}\\ \text{links } i}} x_i^* C_i(x_i^*) \tag{9}$$

In solving a network design problem on an aggregate network, it is hoped that changes in the network caused by the actions taken will not significantly affect the second term of the objective function and that it may, therefore, be treated as a constant and omitted from the calculations. This suggests that the rate of change in the objective function from a change in the flow on link i can be computed, and that links for which changes in the flow will only marginally change the objective function can be deleted. Specifically, the rate of change in the objective function because of a change in the flow of link (which is denoted W_i) is

$$\partial Z/\partial x_i = W_i = C_i(x_i^*) + x_i^* C_i'(x_i^*) \tag{10}$$

where $C_i'(x_i^*)$ is the derivative of $C_i(x)$ evaluated at $x = x_i^*$. A hybrid strategy would be to compute W_i for all links and to delete those links for which W_i is less than $\alpha \operatorname{Max}_i(W_i)$.

Second, the O-D trip matrix that is derived after each link is extracted is not unique because it is based on the O-D specific flows in the extracted link, which are not unique. The effect of other O-D matrices on the aggregate networks and on the uses to which those aggregate networks are put is worthy of additional research.

Third, to avoid multiple counting problems, a sequential link extraction procedure has been implemented. Research should be devoted to the development of a simultaneous link extraction procedure. Such a procedure would probably be faster than the sequential procedure that has been used and would be less prone to accumulating and propagating round-off errors from one aggregate network to the next.

Fourth, based on the network design experiments, it it suspected that the quality of the network design solution that uses an aggregate network is related to the degree of network aggregation in the neighborhood of the candidate links and to the ratio of the available budget to the budget required to implement all candidate links. Additional research should explore these relationships.

Finally, the algorithm should be tested on networks that have a limited number of origins and destinations to determine whether or not the increase in the size of the O-D matrix that results from the extraction algorithm increases the computation time more than the time is reduced because of the deletion of links. Recall that this did not occur in the network used in the set of experiments because all nodes were origins and destinations. If this does occur, it might limit the usefulness of the proposed approach to cases in which the increase in the size of the O-D matrix can be predicted to be small.

ACKNOWLEDGMENT

We would like to acknowledge Houssain Poorzahedy, whose ideas and assistance contributed significantly to this research. We would also like to acknowledge Hedayat Aashtiani, who helped secure a computer account for the component of the research conducted in Iran, as well as the computer staff of the Isfahan University of Technology in Iran for their technical assistance. Funds for this research were partly provided by a National Science Foundation grant; this support is also greatly appreciated.

REFERENCES

1. A.E. Haghani. A Network Extraction Approach to the Network Aggregation Problem. Northwestern Univ., Evanston, Ill., M.S. thesis, June 1982.
2. J.G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineers, Part II, Vol. 1, 1952, pp. 352-378.
3. M.C. Beckmann, B. McGuire, and C. Winston. Studies in the Economics of Transportation. Yale Univ. Press, New Haven, Conn., 1956.
4. H. Poorzahedy. Efficient Algorithms for Solving the Network Design Problem. Northwestern Univ., Evanston, Ill., Ph.D. dissertation, June 1980.
5. H. Poorzahedy and M.A. Turnquist. Approximate Algorithms for the Discrete Network Design Problem. Transportation Research, Vol. 16B, No. 1, 1982, pp. 45-55.
6. H.Z. Aashtiani. The Multimodal Traffic Assignment Problem. Massachusetts Institute of Technology, Cambridge, Ph.D. dissertation, May 1979.
7. L.J. LeBlanc. An Algorithm for the Discrete Network Design Problem. Transportation Science, Vol. 9, No. 3, 1975, pp. 183-199.

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

Quick-Response Procedures to Forecast Rural Traffic

ALFRED J. NEVEU

The development of a quick-response method to forecast traffic volumes at project sites located on the rural highway network is discussed. By using travel data from New York State's continuous count stations in rural locations and various state-, county-, and town-level demographic data, a set of elasticity-based models is derived. These models can forecast future year annual average daily traffic (AADT) as a function of base year AADT modified by various demographic factors. These models are estimated based on the type of service the roadway carries: interurban, urban to rural, and rural to rural. Nomographs and a user's manual that describes a simple seven-step process to use the model were developed and distributed to regional offices throughout New York State.

For highway improvements, the gap between available funds and potential projects is becoming wider as revenues from various sources (including gasoline sales taxes, vehicle registration fees, and driver's license fees) fall because of economic pressure or government-enforced conservation (although the $0.05 gasoline tax increase will ease some of the pressure). Costs of labor and materials are escalating faster than the national rate of inflation. At the same time, compounding the problem, increasing travel demands are placing an even greater burden on the U.S. highway system than in the past, thus worsening an already difficult situation.

These trends mean that the need for construction, rehabilitation, and regular maintenance of the highway network is greater than ever. Each year a large number of such projects, ranging from simple intersection improvements to large-scale facility construction, are identified as candidates for the limited financial resources available. Even in the best of times, not all projects can be funded; now, with reduced monies to fund projects, it is even more imperative that programming decisions be made in the most effective and efficient manner possible.

The selection of projects to be implemented is generally based on some evaluation process in which the costs and benefits of each project are compared. The various evaluation processes consider many factors in weighing each alternative, including safety, noise, air pollution, and energy. Each of these factors is, in turn, based on an estimate of the traffic volume that will use the facility under consideration. Thus the volume estimate determines, to a significant degree, which of the many projects will be implemented.

Travel forecasting methodology is highly advanced at the urban area level. Most large metropolitan areas have developed and implemented a fairly sophisticated set of computer-based travel simulation models based on the traditional four-step process. In a nonurban context, however, this process is not nearly so advanced. With many of the projects competing for the scarce funds coming from nonurban areas, it is important to improve and streamline forecasting procedures for rural travel needs. In this way it would be possible to evaluate many rural projects quickly and accurately, thus providing government officials with better information on which to base their programming decisions.

To fulfill this need, research was initiated by the Transportation Statistics and Analysis Section of the New York State Department of Transportation (NYSDOT) to develop a quick-response procedure to forecast traffic volumes on rural roads. The primary focus of this effort was the design and testing of a simple, fast method to forecast rural traffic volumes. In this paper previous efforts aimed at forecasting rural traffic are examined, the chosen methodology is described, and the results of the analysis are presented. Finally, some of the limitations of the procedure are discussed, and some possible solutions to the limitations are provided.

PAST EXPERIENCE

Little attention has been focused on the topic of forecasting volumes on rural roads. Much of the research that deals with the rural highway system has been in the area of design and construction of low-volume roads, travel to recreation facilities, or rural public transportation. An extensive literature review uncovered only two studies specifically concerned with forecasts of rural traffic volumes.

In 1958 Morf and Houska (1) examined the variation of traffic growth patterns on rural highways. They hypothesized that four factors were responsible for the variations in growth patterns observed on the Illinois rural highway network: geographic location, type and width of pavement, proximity to an urban area, and type of service provided by the roadway. This last factor was subdivided into four categories: interurban, interregional, urban to rural, and rural to rural.

The authors noted that the growth trends in sites close to urban areas were primarily a function of the expansion of the city. Therefore, the remaining analysis focused on rural highways outside the influence of an urban area.

A comparison by geographic location indicated minor differences in growth patterns. Slightly greater traffic increases were noted in northern rather than southern Illinois. Roadways with wider widths had correspondingly greater increases in traffic, but the authors believed that the wider roads were an effect of the volume increases, not a cause of them.

The only factor that had an appreciable effect on traffic growth rates was the characteristic of type of service. Highways with the greatest percentage of interurban or interregional service generally had the largest increases in travel. Roads that served largely urban-to-rural or rural-to-urban travel had the smallest increases. Based on these results, the authors projected volume trends on the rural highway network in Illinois for the different road types separately.

The study by Tennant (2) used the land use and traffic generation principle to outline a procedure to estimate rural road traffic in developing counties. By using various economic, social, land use, and travel data from the Mount Kenya region in Kenya, several trip-generation equations were estimated for both urban and rural zones in the study region. The results are almost identical; in both cases employment is a better predictor of trip generation than is vehicle ownership. The correlation coefficients that use either variable in the equation are all in the range of 0.5 to 0.9. Thus even vehicle ownership does a fair job of predicting trips per person. Examining traffic generation from different land uses revealed that 75 percent of the trips were generated by one of three land use types: retail and commercial, government administration, and road transport. Agricultural and residential land use areas did not generate many trips in this region. The author concluded that, obviously, more detailed research was needed, and, as a first-cut analysis, either vehicle ownership or employment could be used to forecast future rural trip generation.

DEVELOPING THE METHODOLOGY

Current practice at NYSDOT to forecast travel on rural highway links assumes that travel, represented as vehicle miles of travel (VMT), is directly proportional to population (note that these data are from an internal memo from W.S. Caswell to J. Shafer, "VMT Growth Factors for Minor Civil Divisions," January 14, 1975). Travel forecasts for urbanized areas are obtained from the network assignments for each area. In areas outside those geographic boundaries, a different procedure was developed. By using VMT per capita estimates by area from the 1972 National Transportation Study and population estimates for each town and county in the state from the New York State Department of Commerce, annual VMT growth rates by town were derived for the years 1972-1990. These rates were developed by first estimating total VMT for each area by using the VMT per capita data and the population estimates, then calculating the annual growth rates for each area.

Several problems surfaced as these VMT growth rates were used by the Department. First, it was recognized in the beginning that there is not necessarily a correlation between VMT and population. Inaccurate estimates of VMT may result from large amounts of nonresident travel drawn into or through the area. This is especially true in popular recreation areas. Second, although the population in New York state may decline (and did so between 1970 and 1980), the number of households may (and did) rise; thus this procedure would forecast a decline in VMT from 1970 to 1980 when, in actuality, travel was still increasing. Finally, there was no sensitivity to energy price or supply in this method.

Because of these shortcomings, the NYSDOT Transportation Statistics and Analysis Section initiated a research project to develop a procedure sensitive to these factors to forecast rural traffic to be used in the development and evaluation of highway-related projects. This new methodology was designed to meet several criteria. First, the procedure must be simple enough to be used on simple desk-top or hand-held calculators, which are generally available in most planning organization offices. It was believed that a large, cumbersome computer model would be inappropriate in this study. Second, the data used in the procedure must be easily available to the local or regional planner. This includes both historical trends and future predictions. Finally, it was believed that to be of maximum use to the project development staff, the procedure would forecast annual average daily traffic (AADT) at the project site, rather than VMT as was done previously.

An elasticity model formulation was selected as the appropriate model. In this model future year AADT is related to present year AADT and modified by changes in any number of background factors. The general form of the model is as follows:

$$AADT_f = AADT_p \left\{1.0 + e_1 [(X_{1,f} - X_{1,p})/X_{1,p}] + \ldots\right\} \quad (1)$$

where

$AADT_f$ = AADT in the future year,
$AADT_p$ = AADT in the present year,
$X_{1,f}$ = value of variable X_1 in the future year,
$X_{1,p}$ = value of variable X_1 in the present year,
e_1 = elasticity of AADT with respect to X_1.

The elasticity model was selected for several reasons. Because it was believed that the range of volumes over which the model would be applied would be much greater than that available in the calibration data set, a simple linear regression model that relates AADT to the background factors directly was deemed 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. Finally, the elasticity portion of the model calculates a growth factor directly, so the procedure can be easily transformed into a set of nomographs, thus further simplifying the work required by the user.

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

$$Y = a + a_1 x_1 + a_2 x_2 + \ldots \quad (2)$$

elasticity measures can be estimated by

$$e_i = a_i(\overline{X}_i/\overline{Y}) \quad (3)$$

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

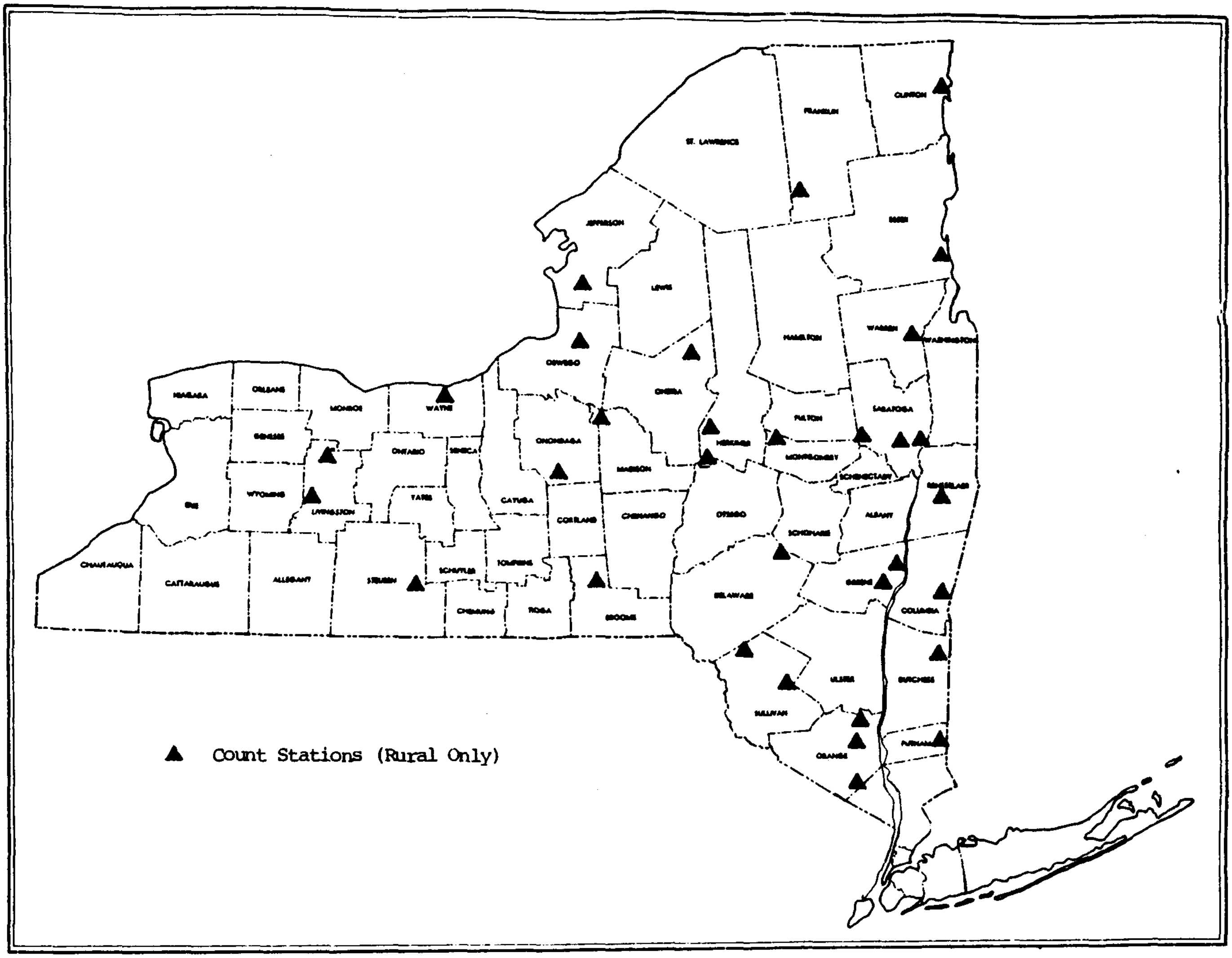
Data for the estimation of the background factors and elasticities came from a variety of sources. The AADT values were obtained from the continuous count program at NYSDOT. Only those stations classified as rural in nature were selected for use in the study. This yielded a total of 32 stations throughout the state (Figure 1). By using the town and county in which the station is located, the various background factors were collected. Information at the state, county, or town level was obtained from a variety of demographic factors, including population, households, automobile ownership, and employment. Some of these data were collected at more than one level of detail. A summary of the background factors collected at each level is as follows:

1. Town level--population, housing units, and households;
2. County level--population, housing units, households, automobile registrations, employment, labor force, personal income, and income per capita; and
3. State level--gasoline sales.

These data were collected for several years (1974-1978) and yielded a total of 5 observations for each station and 160 observations overall. These years were chosen to avoid any complications introduced by the energy emergency situations experienced during the past decade. Although the first energy crisis did encompass the early months of 1974, it was believed that the emergency had eased enough so that yearly totals for the variables would not be significantly affected.

The equations developed to uncover the most important background factors and to estimate their elasticities related each year's AADT for each station to the corresponding year's data for the background variables for that station's location. By using the results from the earlier study in Illinois (1), three different classes of roads were examined separately. These road classes were based on the type of service the road provides. By using functional class as the determinant, the three service types were Interstates (representing interurban and interregional service), principal arterials (repre-

Figure 1. Rural continuous count stations.



senting rural-to-urban service), and minor arterials and major collectors (representing rural-to-rural service). Thus three sets of elasticities and three forecasting models were derived.

Several regression analyses were performed by using a stepwise linear regression program. In the initial runs, one of the income variables was entered into the model. However, future values for either of those income variables are difficult to forecast, especially in an economy that is undergoing such rapid changes. Given the earlier criterion for using variables that are easily available and simple to forecast, all further analyses eliminated any income variables from consideration.

In addition, throughout the remainder of the analyses, town or county housing units appeared in many of the equations. In this case, although the relationship has statistical significance, the causal relationship to travel must be questioned. It was believed that households (sometimes defined as occupied housing units) were a better determinant of travel. Therefore, whenever housing units at any level entered the equations, the corresponding household value was substituted. This resulted in extremely small reductions in explanatory power of the models, but the models had a much better causal foundation.

The final regression equations, along with the R^2 values, t-statistics, and elasticities are as follows. For Interstates,

$$\text{AADT} = -1097.870 + 0.051 \text{ county automobiles} + 9.042 \text{ town households} \quad (4)$$

$$R^2 = 0.65 \quad t = 2.49 \quad t = 6.86$$
$$F = 25.13 \quad e = 0.228 \quad e = 0.832$$

For principal arterials,

$$\text{AADT} = -3013.145 + 0.125 \text{ county households} + 0.866 \text{ town population} \quad (5)$$

$$R^2 = 0.77 \quad t = 4.98 \quad t = 7.27$$
$$F = 45.75 \quad e = 0.572 \quad e = 0.760$$

For minor arterials and major collectors,

$$\text{AADT} = 2867.129 + 0.619 \text{ town households} \quad (6)$$

$$R^2 = 0.20 \quad t = 4.95$$
$$F = 24.52 \quad e = 0.314$$

Each of the models are relatively simple, with only one or two variables in each. The equations use variables that are easily available to local planners from a variety of sources for both historical and future trends. Each of the variables is significant at the 95 percent confidence level, and all function in the proper direction; i.e., as the variables increase, travel increases. Equations 4 and 5 explain much more of the variance than Equation 6, but this is an expected result. The last type of rural road is much more abundant and serves many more purposes than the other, more specialized, types of roads. Therefore, it is expected that there would be much more variability in the data and much less explanatory power in a simple model. This variability is probably caused by local factors below the town level. Large traffic generators such as malls, drive-in fast food restaurants, or schools in the proximity of the counting station are examples of such a local effect.

There are several items of interest in Equations 4-6. First, in only one equation does a population variable enter, whereas a household variable is in every equation. This supports the contention that households, not population, are a better determinant of travel. This is especially significant because the previous procedure at NYSDOT relied exclusively on population as the determinant of future traffic volumes. Second, it is interesting to note that the energy variable did not enter any of the equations. In fact, its correlation with AADT was extremely small. This variable was a statewide value, whereas the rest of the data was of a finer detail. Unfortunately, more detailed information on fuel supply was not available. Perhaps with more detailed data energy factors may become significant in these equations.

By using the elasticities derived from the regression equations, it is now possible to complete the development of the forecasting model by substituting those elasticities into Equation 1. This model is presented in Equations 7-9. For Interstates,

$$AADT_f = AADT_p\ [1 + 0.228\ (\%\Delta \text{ county automobiles}) + 0.832\ (\%\Delta \text{ town households})] \qquad (7)$$

For principal arterials,

$$AADT_f = AADT_p\ [1 + 0.572\ (\%\Delta \text{ county households}) + 0.670\ (\%\Delta \text{ town population})] \qquad (8)$$

For minor arterials and major collectors,

$$AADT_f = AADT_p\ [1 + 0.314\ (\%\Delta \text{ town households}) \qquad (9)$$

To make the procedure even easier to use, nomographs were developed to provide faster estimates of the growth factor (called Z), that portion of the equation encompassing only the elasticities $(1 + e_1\ \Delta\%X_1 + \ldots)$. These nomographs are shown in Figures 2-4, along with example calculations demonstrating their use.

To use these nomographs, the user needs to compute the percentage change in the appropriate variables at the project site from the base year to the horizon year. By using Figure 2 (Interstates) as an example, the variables would be county automobile registrations and town households. The intersection of those lines in the graph yields the growth factor. In the example, a 35 percent change in county

Figure 2. Interstate nomograph.

INTERSTATES

Given: AADT (1980) = 25,600

County Autos, 1980 = 52,000

County Autos, 1990 = 69,160

Town HH, 1980 = 2,200

Town HH, 1990 = 2,620

$$\%\Delta \text{ County Autos} = \frac{69{,}160 - 52{,}000}{52{,}000} = 33\%$$

$$\%\Delta \text{ Town HH} = \frac{2{,}620 - 2{,}200}{2{,}200} = 19\%$$

From the nomograph = Z = 1.23

AADT (1990) = Z x AADT(1980)

= 1.23 x 25,600

= 31,488

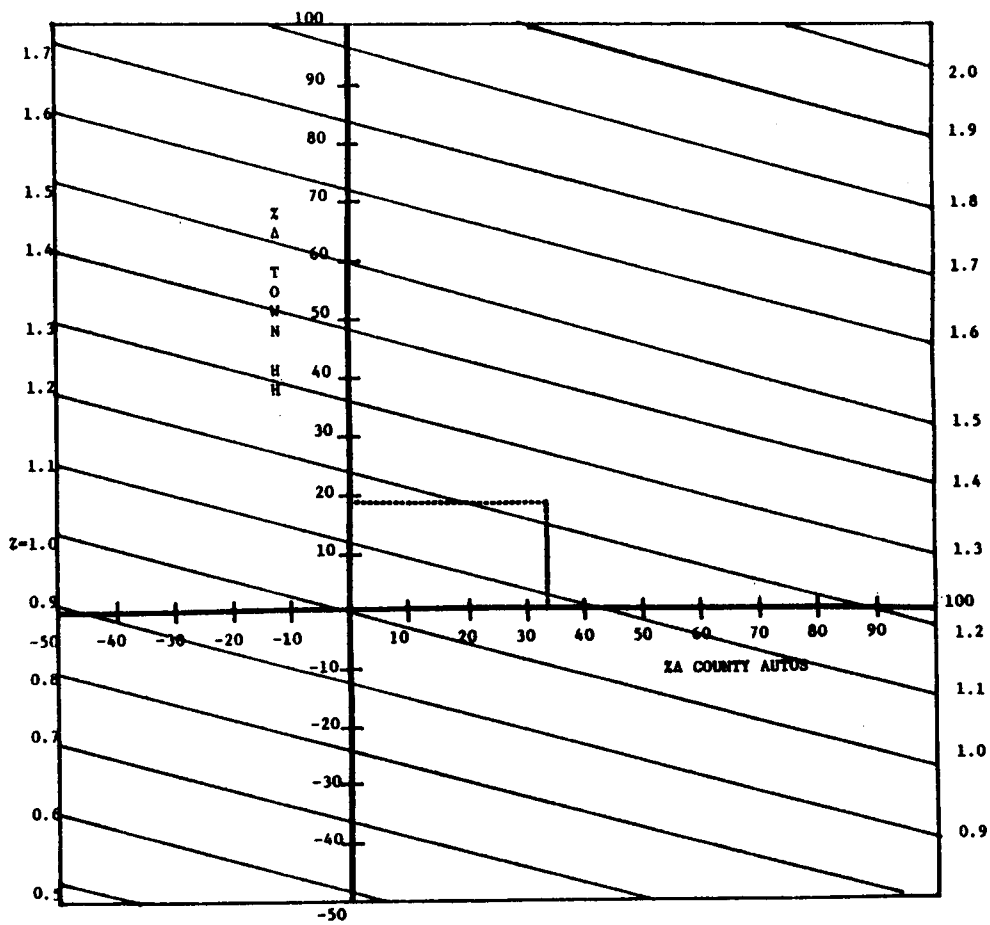


Figure 3. Principal arterial nomograph.

PRINCIPAL ARTERIALS

Given: AADT (1980) = 7,800

Town POP, 1980 = 5,700

Town POP, 1990 = 8,720

County HH, 1980 = 15,500

County HH, 1990 = 24,500

$$\%\Delta \text{ Town POP} = \frac{8{,}720 - 5{,}700}{5{,}700}$$

= 53%

$$\%\Delta \text{ County HH} = \frac{24{,}500 - 15{,}500}{15{,}500}$$

= 58%

From the nomograph, Z = 1.73

AADT (1990) = Z x AADT(1980)

= 1.73 x 7,800

= 13,494

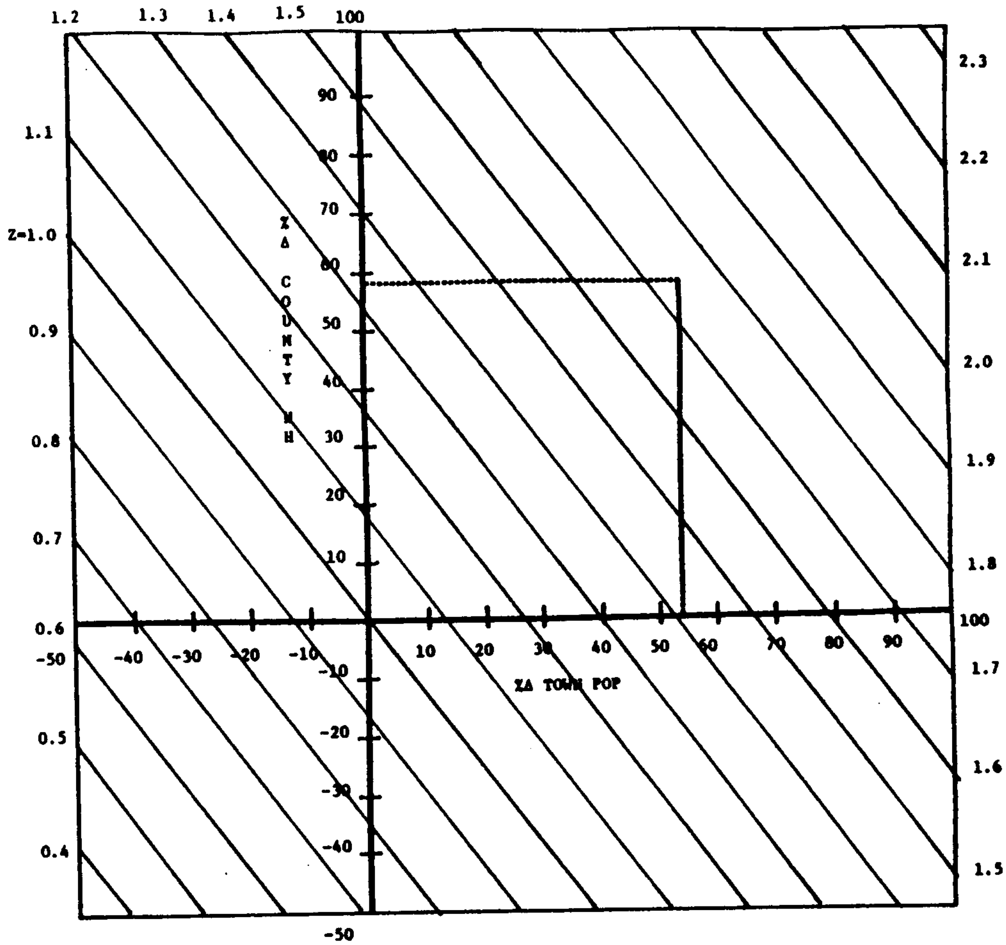


Figure 4. Minor arterial or major collector nomograph.

MINOR ARTERIALS & MAJOR COLLECTORS

Given: AADT (1980) = 1,500

Town HH, 1980 = 1,750

Town HH, 1990 = 2,485

$$\%\Delta \text{ Town HH} = \frac{2{,}485 - 1{,}750}{1{,}750}$$

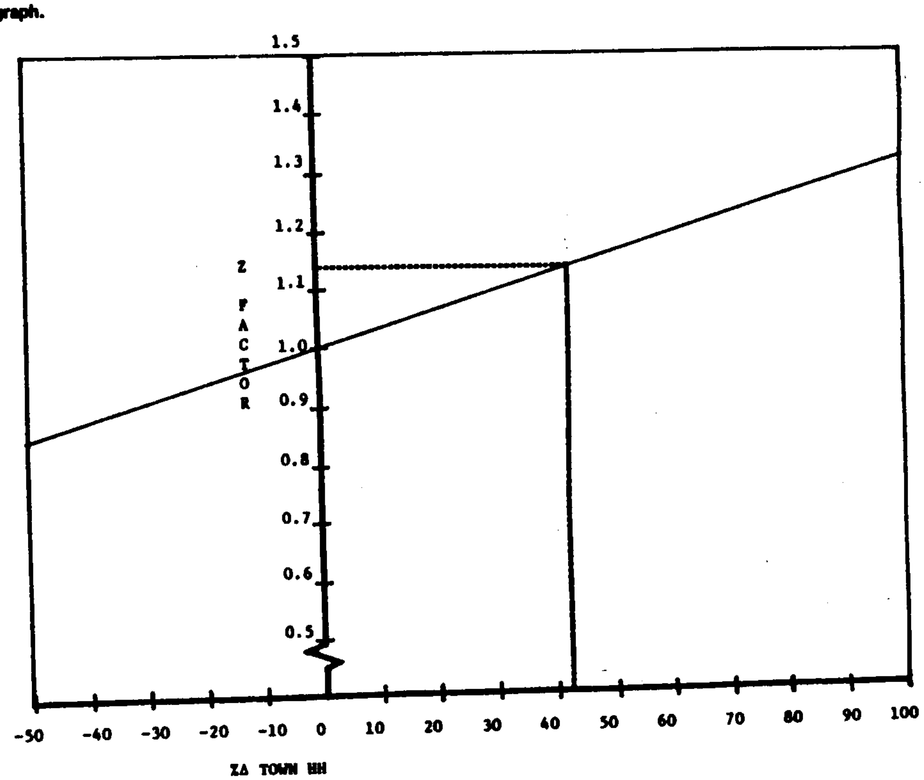
= 42%

From the nomograph, Z = 1.13

AADT (1990) = Z x AADT(1980)

= 1.13 x 1,500

= 1,695



automobile registrations and a 20 percent change in town households give a growth factor of approximately 1.23, which implies a 23 percent growth in traffic from the present to the future year.

These models satisfy all of the criteria specified earlier. The procedure is easily used by anyone with a hand-held calculator; no large computer system is necessary. With the nomographs, the forecasting procedure becomes even easier to use. The data needed to predict rural traffic volumes with these models are readily available at the local and regional levels. Historical trends for population and households are found in census publications, and automobile registration data are generally available from either the state transportation or motor vehicle departments. In addition, recent work at NYSDOT has been directed toward compiling a reference directory for gathering transportation and energy data at all levels of detail (3). This directory provides guidelines and suggestions for collecting this type of information at the local, regional, and state levels.

To use the forecasting procedure, a simple seven-step outline was developed:

1. Determine functional class of roadway,
2. Determine town and county of roadway,
3. Collect base year AADT,
4. Collect base and horizon year data for required variables,
5. Calculate percentage change for each variable,
6. Calculate (or use nomograph to estimate) Z factor, and
7. Calculate horizon year AADT.

The user's manual that describes this procedure was developed and distributed to the regional offices of NYSDOT (4). This manual included step-by-step instructions for using the procedure, the nomographs, an example calculation, and the necessary data to use the methodology.

TESTING THE METHODOLOGY

A sample of 100 sections from the state highway system were selected to test this procedure. These sections were selected because they were proportional to the total number of sections for each of the three service types, and each section had a traffic count performed in 1975 and 1980. By using the appropriate town and county values for the background variables, forecasts of AADT for 1980, based on 1975 AADTs, were computed and compared with the actual 1980 AADTs for each section.

The results indicated that the models performed satisfactorily.

Roadway	Avg Forecast Error (%)	Avg AADT
Interstates	-4.54	12,180
Principal arterials	14.49	5,415
Minor arterials and major collectors	6.93	3,865

The larger errors (for principal arterials, and minor arterials and major collectors) are associated with the smaller values of AADT. Errors of these sizes will not have a large impact on any design decisions.

The models overestimate future AADT on most of these sections, but it must be remembered that in the 1975-1980 time period an energy shortage caused a drop in travel of 5 percent or more. Therefore, the estimate of future AADT should be high. By adjusting the forecasts to account for the 1979 fuel shortage, the models would perform even better.

APPLICATIONS

There are many potential uses for the rural travel forecasting model. Several possible applications are presented in this section, and these deal primarily with the project development process, which is the main task for many state highway agencies.

The most obvious and direct use of this procedure is for the estimation of the benefits for specific highway system improvements. These projects can range from relatively simple road widenings to large-scale reconstruction of highway sections. The procedure estimates future traffic volumes reasonably quickly and accurately, and thus allows the analyst to examine many alternative projects with minimal expenditures of time and money.

A second, related application would be as an aid in the selection of the appropriate design for a project. Answers to questions such as the number of lanes and type of traffic control required are also determined by the volumes on that highway segment. The engineer can gain some insight into the future needs of the area in order to scale the project to meet those criteria.

The final application for the rural traffic forecasting model is the use of the procedure as a guide in the identification of potential problem segments of the state highway system (at least the rural portion). Because the model is based on town- and county-level variables, it is possible to identify the towns and counties where traffic growth will be the greatest and to focus on these areas for more detailed examination. This will be of great assistance in helping the planner estimate where the future problems will be. As a corollary to this use, it is possible to key the traffic counting program to this information by concentrating on the areas that show rapid growth (or decline) and by eliminating frequent counts in the areas that show a stable situation. As the available funds for all phases of highway work decline, this could be one of several ways to reduce the cost of the traffic count program without sacrificing much of its information.

PROBLEMS AND LIMITATIONS

Perhaps the most serious problem with the procedure is one that is common to all forecasting tools: the accuracy of the model is determined to a large degree by the accuracy of the inputs, especially for future values of the background variables. The state provides a set of forecasts for county population and households for 5-year intervals, but there is little information available for the other variables required in the procedure. Thus the question is how to estimate future values for county automobile registrations and town populations and households.

There are several ways to obtain future year estimates of the number of automobiles registered in the county. The first and most obvious way is to check with the state departments of transportation or motor vehicles to see if they have some forecasts of that sort. If that fails, or the local planner wishes to check those forecasts, there are other ways to forecast future automobile registrations. The easiest is to calculate the average annual growth rate from the historical data (in this case, 1973-1980 data), and assume an increasing, decreasing, or constant rate for the future. This method does not incorporate any concern about reaching a saturation point, but it may be reflected by altering the projected growth rate. Another way, which accounts for the saturation problem, is to examine the trend of historical automobiles per person in the county, and then carry that trend out to the future until this value reaches a predefined saturation point. Then, by multiplying this trend by the county population, estimates of future year county automobile registrations are developed.

The various ways to obtain future values for both town population and households are virtually identical, and will be considered together. These methods also parallel the ones used to estimate county automobile registrations in the future. The first and simplest way is to calculate an average annual growth rate for the town and carry it over into the future. Of course, the analyst can adjust this rate to more closely reflect the local situation. This method, however, does not guarantee that the sum of the town values will equal the county total (provided already) for a given year. This is not a real problem for localized projects, but it could prove to be a significant error in larger undertakings. Therefore, a somewhat more complex way may be considered. In this method, the town's proportion of the county total is calculated for two points in time. Depending on the difference between them, it may be assumed that the town's proportion increases, decreases, or remains constant out to the horizon year. Although these procedures are not elegant, they do provide several options for the local analyst to use to meet the data requirements of the rural travel forecasting models.

One other major problem encountered while using this new forecasting tool deals with the applicability of the model in various areas. How does the analyst decide that the project area is rural enough for the model? Obviously, the model should not be used to estimate future traffic volumes in the central city, but what about the rest of the areas? It is difficult to develop guidelines to assist in this decision. Perhaps the best advice to give here is to use this model in conjunction with any other travel forecasts (e.g., from the assignment network in the fringe of the urbanized area) that deal with the same area. If no other forecast exists, then the area may be assumed to be adequately represented by this model. As experience is gained in the use of this procedure, better guidelines may be developed.

Finally, the model formulation assumes that the elasticities are constant over time, but the regression derivations do not ensure this. Historically, travel has been growing at a fairly constant rate for many years. After the interruptions caused by the two fuel shortages, travel growth resumed that rate in a short time. Therefore, it was believed that assuming constant elasticities would not introduce any substantial errors.

In addition, a log-linear form to estimate the elasticities, which ensures constant elasticities, was tested. The form of the equation is

$$Y = a_0 + a_1 \ln X_1 + a_2 \ln X_2 + \ldots \qquad (10)$$

where a_1, a_2, . . . are the elasticities. The results were not significantly different from the original models (as shown in the following table), and this provides further evidence to support the assumption of constant elasticities from the linear regression formulation.

	Elasticities	
Roadway	Linear	Log-Linear
Interstates	-4.54	9.98
Principal arterials	14.49	13.19
Minor arterials and major collectors	6.94	6.50

Overall, few problems have been identified during the initial uses of these models. The problems previously identified were the only significant ones experienced to date. As local planners begin to use this procedure more often, some of the subtler shortcomings may surface, but they are not expected to be major concerns. It must be kept in mind that the end use for the forecasted volumes is the design of rural highway projects. These volumes are generally low enough so that large errors (on the order of 20 to 50 percent) will not cause a significant change in the design criteria.

Finally, it is important to note here that this model is not intended to be the perfect forecasting tool, if such a thing could ever exist. Rather, it is to be used by the analyst as one way, among many, to estimate future travel on the rural highway system. The user is expected to weigh the results in terms of the local situation, and adjust them according to his judgment of the specific area and application.

ACKNOWLEDGMENT

I wish to thank all those who provided assistance in this project, especially David T. Hartgen for his helpful guidance and comments during this study. Special thanks to Judy Khachadourian, Dianne DeVoe, Tedi Toca, and Charron Many for their excellent typing of the manuscript.

REFERENCES

1. T.F. Morf and F.V. Houska. Traffic Growth Patterns on Rural Highways. HRB, Bull. 194, 1958, pp. 33-41.
2. B. Tennant. Forecasting Rural Road Travel in Developing Countries from Land Use Studies. In Transport Planning in Developing Countries, Proc., Summer Annual Meeting, Planning and Transport Research and Computation Company, Ltd., Univ. of Warwick, Coventry, Warwickshire, England, July 1975.
3. N. Erlbaum. Data Sources and Reference Directory of Transportation and Energy Data. Planning Division, New York State Department of Transportation, Albany, Prelim. Res. Rept. 199, Oct. 1981.
4. A.J. Neveu. Quick Procedure to Forecast Rural Traffic: User's Manual. Planning Division, New York State Department of Transportation, Albany, Oct. 1981.

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

Notice: The author is responsible for any errors of fact or omission.

Respondent Trip Frequency Bias in On-Board Surveys

LAWRENCE B. DOXSEY

In this paper it is shown that on-board surveys are burdened with an inherent and serious sampling bias. The source and implications of this bias are presented, and a simple statistical correction procedure is developed. An example is used in which the bias leads to a 50 percent overestimate of average tripmaking by users and a 33 percent underestimate of the number of transit users in the population.

On-board surveys are the most commonly used mechanism for the collection of disaggregate data about public transit patrons. For many operators, on-board surveys are conducted on an annual basis and provide them with their only source of information on the users of their systems. The attractions of this survey technique are both strong and obvious. Compared with other survey procedures, on-board surveys are inexpensive to both develop and administer and they guarantee that all respondents will be transit users. Thus fairly modest resource expenditure can generate a substantial volume of user information.

Unfortunately, on-board surveys are also burdened with a number of disadvantages. Frequently acknowledged among these is the general low response rate, often 25 to 50 percent, with the possibility of severe nonresponse bias. Also widely recognized is the inherent need for brevity and hence the relative paucity of information on each respondent. Other difficulties relating both to the method of administration and to limitations on the information received have been identified and could be mentioned here. However, one fundamental and potentially serious drawback to on-board surveys has been quite generally overlooked. This is the problem of selection bias, which results from using transit passenger trips as the sampling frame for interviewing transit users. This particular form of selection bias is referred to in this paper as respondent trip frequency bias. It is the purpose of this paper to isolate the source and implications of respondent trip frequency bias in on-board surveys and to offer a simple statistical weighting procedure that can correct it.

SELECTION BIAS

Selection bias results when the probabilities with which sample units are actually drawn differ from the probabilities with which they are believed or perceived to have been drawn. The relationship of the sample to the population consequently differs from what it is thought to be and, in turn, estimates based on the sample are biased. Selection bias commonly occurs either if the pattern of nonresponse is such that the actual probability of an individual unit appearing in the sample is unknowingly correlated with variables under study, or if the actual sampling procedure differs from the sampling design. Selection bias constitutes a broad class of problems in survey sampling (1). The on-board survey respondent trip frequency bias addressed here belongs to the latter category. In other applications, identification of and correction for selection bias are routine steps in the analysis of survey data. However, with the exception of a few specific and sophisticated applications (2), the presence and implications of selection bias in on-board surveys has been commonly overlooked.

In an on-board survey the sampling frame is the set of passenger trips taken during the sample period. However, much of the subsequent analysis, and much of the motivation for conducting a survey at all, involves identifying the characteristics of the users of the system. It is significant that in conducting the analysis, the observations are treated as if the sampling frame had been system users rather than system trips, and each respondent is treated as if he had an equal probability of appearing in the sample. This is in error because the probability of an individual user appearing in the sample is directly proportional to the number of transit trips that user makes during the sample period.

Individuals who take many trips are far more likely to appear in the sample than are individuals who take few trips. Potentially severe selection bias occurs because the assumed design probabilities (i.e., those implicit in the analysis) differ markedly and systematically from the actual probabilities. From a sampling viewpoint, the relationship between trips and users can be regarded as an implicit stratification of users on the basis of their respective individual trip rates. This interpretation allows viewing respondent trip frequency bias in the endogenous variable stratification context of Hausman and Wise (3), although their work is couched in terms of explicit rather than implicit stratification.

WHAT RESPONDENT TRIP FREQUENCY BIAS MEANS FOR ON-BOARD SURVEY RESULTS

Because differences in individual travel frequencies give rise to the bias, it may be intuitively clear that its most critical impact is on estimates of patrons' mean transit use. Relative to the population, the sample has an overrepresentation of frequent users and an underrepresentation of infrequent users. A linear average of responses to the question of frequency of use will provide an estimate of the population mean that is biased sharply upward. An estimate of mean frequency of travel based on an on-board survey is often used with total boarding counts to estimate the total number of patrons and hence the market penetration of the transit system. An upward bias to the mean frequency estimate will imply a downward bias to the estimated total number of users and the degree of market penetration.

Although the consequences are greatest for estimates of mean travel frequency, bias also results for any characteristic that is correlated with travel frequency. For example, if an analyst wants to determine the income distribution of transit users, and if low-income people generally take fewer transit trips than do middle-income people, then the estimated income distribution of users will be biased from those with low incomes and toward those with middle incomes. Overall, the distortions can be large enough that the on-board survey provides a misleading picture of the user population.

CORRECT WEIGHTING PROCEDURE

Although the problem of respondent trip frequency bias in on-board surveys is serious, application of a relatively simple statistical correction can eliminate the bias. The correction involves the use of individual travel frequencies to develop weights for each observation, whereby observations are weighted by the ratio of their relative frequency in the population to their relative frequency in the sample. The individual weights are as follows:

$$w_i = n/\left(f_i \sum_{i=1}^{n} 1/f_i\right) \qquad (1)$$

where

w_i = weight assigned to the ith observation,
f_i = transit travel frequency of the ith respondent, and
n = total sample size.

The weights calculated from Equation 1 take larger values for observations on infrequent travelers and smaller values for observations on frequent travelers. In calculating the relative frequency of the occurrence of various characteristics in the user population, an observation contributes a share equal to the value of its weight rather than contributing a unit amount. Thus an unbiased estimate of the share of the user population in a given income bracket is equal to the sum of the weights for all respondents in that income bracket divided by the sum of the weights for all respondents. It may be worth noting that the weighting procedure is self-normalizing in that

$$\sum_{i=1}^{n} w_i = n.$$

That is, the sum of the weights taken over all respondents equals the number of respondents.

For variables that are not categorical, estimates of population values are made by using the weights multiplicatively with the variables. For example, if the on-board survey had continuous data on income rather than categorical variables, an unbiased estimate of the mean income level of users would be made by summing all observations of the product of the individual weight values and income level and then dividing this summation by the total number of users. This is also the procedure by which the mean transit trip frequency is calculated. Thus the following equation provides an unbiased estimate of average number of trips taken by users:

$$u = \left(\sum_{i=1}^{n} f_i w_i\right)/n \qquad (2)$$

where u is the mean frequency. This compares with an estimate of the mean calculated as

$$\sum_{i=1}^{n} f_i/n$$

in instances where the effect of respondent trip frequency bias is ignored. As a direct estimate of the mean trip frequency, Equation 2 can be simplified to

$$u = n/\left(\sum_{i=1}^{n} 1/f_i\right) \qquad (3)$$

It should be evident that with nearly any analysis, software package calculation and application of weights to correct for respondent trip frequency bias can be easily accomplished. In the next section an example that should underscore the importance of this correction is given.

ILLUSTRATIVE EXAMPLE

In this example data are used from an on-board survey conducted in Atlanta during May 1979 as part of a project sponsored under the Service and Methods Demonstration program of UMTA. The demonstration project was designed to study the impacts of fare integration of a monthly transit pass that had been introduced to the Metropolitan Atlanta Rapid Transit Authority (MARTA) system in March 1979. Interviews were conducted with 4,672 people during the on-board survey.

The survey results provide clear evidence of the importance of correcting for respondent trip frequency bias. When the observations are properly weighted, the average user is estimated to take eight trips per week on MARTA. If the weighting is ignored, the estimate is 12 trips per week. Thus in this example the influence of respondent trip frequency bias is to overstate the mean trip frequency by 50 percent. Bias of 30 to 60 percent could well be found in most on-board surveys.

Bias in estimating mean user trip frequency is reflected in estimates of the total number of users. For May 1979, MARTA counts indicated a total of 5.4 million boardings. If the unbiased estimate of mean trip frequency is used, a total of 161,000 persons are estimated to use the system. The uncorrected estimate of the mean implies an estimate of 107,000 system users. The indicated market penetration of the MARTA system differs substantially between the two estimates. The former suggests that 8.7 percent of the area's population are system users, whereas the latter indicates that only 5.8 percent are users (note that these data are based on U.S. Census estimates of 1.852 million people in the Atlanta standard metropolitan statistical area as of July 1, 1978).

The data in Table 1 give a further illustration of the effect of respondent trip frequency bias. In the table the household income distributions of users are presented based on unweighted, and hence incorrect, data and on the same data properly weighted. Also in the table are the respective within-group mean weekly transit trip frequencies. Although the share of riders in any one income group is not more than a few percentage points wrong, the income distribution calculated without correcting for respondent trip frequency bias is biased toward lower-income people. When corrected, people with household incomes of $15,000 and greater appear to compose 28 percent of the users as compared with the 22 percent they appear to compose with the unweighted data. This bias in the income distribution is the direct consequence of a lower average transit trip frequency at higher incomes.

It is also worth observing that the impact of respondent trip frequency bias is not constant across income groups. The overstatement effect of the bias on the within-group trip frequencies ranges from 39

Table 1. Income distribution of MARTA system users.

Household Income Range ($)	Without Weights to Correct for Respondent Trip Frequency Bias: Users (%)	Without Weights: Mean Weekly Trips	With Weights to Correct for Respondent Trip Frequency Bias: Users (%)	With Weights: Mean Weekly Trips
< 5,000	23	12.3	23	8.0
5,000-9,999	32	12.5	28	9.0
10,000-14,999	23	12.5	21	8.8
15,000-24,999	15	10.8	18	6.7
> 25,000	7	10.0	10	5.7

percent for the $5,000 to $9,999 group to 75 percent for the $25,000 and greater group. The difference results from differences among the groups in the underlying trip frequency distributions. In general, the greater the dispersion of trip frequencies across group members, the greater will be the relative bias.

SOME GUIDELINES FOR USE

Although no great difficulty is presented in calculating weights to correct for respondent trip frequency bias, the survey instrument must be written to provide information on individual trip frequency. A precise count of transit trips taken during the survey period is the ideal situation. Complete accuracy is, however, too much to expect, and an adequate alternative is the number of transit trips taken within the previous 7 days or the number typically taken in a week. It can be an aid to the thought process of the respondent to ask for total use through questions about its components. Thus a survey form could ask for the number of transit trips to work during the previous 7 days, the number of transit trips from work during the previous 7 days, and the number of transit trips to or from places other than work during the same time period. Note that although measurement error creeps in with any form of question, the need is not so much to distinguish the person who takes 8 trips from the one who takes 10 trips as it is to distinguish the person who takes 2 or 3 trips from the one who takes 10, 12, or more trips. Furthermore, even if questions are written precisely, the accuracy of responses to on-board surveys is sufficiently unsatisfactory, especially with the common practice of self-administration, so that it is unrealistic to expect the instrument to distinguish fine gradations. Thus substantial improvement can be made even when working with four or five categorical responses.

Lest the case appear to have been made too strongly, there are instances when unweighted data are appropriate. When the analyst's interest lies not with the users of the system but with the trips, then the unweighted data provide an unbiased picture. Nevertheless, care should be taken to distinguish between reporting that some 55 percent of all system trips are taken by people with household incomes less than $10,000 (which is the case in Table 1) and reporting that 55 percent of users have household incomes less than $10,000 (which exceeds the unbiased estimate by 4 percentage points). To the extent that the role of a transit system is the provision of service to a region's population, understanding the user population and measuring market penetration are crucial. Neither can be accomplished with unweighted on-board data.

FINAL COMMENTS

The focus of this paper has been exclusively on one fundamental and dramatic source of bias in on-board surveys. This is not to suggest that on-board surveys are otherwise above reproach. Among the avenues for improvements to on-board surveys are optimal use of the clustering implicit in drawing observations through bus runs, development of techniques to increase response rates, and application of procedures for efficient stratification so as to minimize the variance of estimates. Nevertheless, incorporation of the weighting procedure presented in this paper can do much to increase the validity of on-board surveys conducted, even without benefit of sophisticated sampling techniques.

ACKNOWLEDGMENT

Work on this paper was sponsored by the Office of Service and Management Demonstrations, UMTA.

REFERENCES

1. W.H. Williams. How Bad Can "Good" Data Be? The American Statistician, Vol. 32, No. 2, May 1978, pp. 61-65.
2. S. Lerman and S. Gonzalez. Poisson Regression Analysis Under Alternate Sampling Strategies. Transportation Science, Vol. 14, No. 4, 1980.
3. J. Hausman and D. Wise. Stratification on Endogenous Variables. In Structural Analysis of Discrete Data with Econometric Applications (C. Manski and D. McFadden, eds.), MIT Press, Cambridge, Mass., 1981.

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

Bus, Taxi, and Walk Frequency Models That Account for Sample Selectivity and Simultaneous Equation Bias

JESSE JACOBSON

A 2-year user-side subsidy experiment that provided the handicapped and the elderly with discounted coupons to be used on buses and taxis was conducted in a small northeastern metropolitan city. The effect of the user-side subsidy experiment on bus and taxi travel by the elderly population is described. As expected, the subsidy experiment increased the number of trips taken by bus and by taxi. Furthermore, able-bodied elderly persons who do not own automobiles and handicapped elderly persons who are either employed or students are more likely to purchase discounted bus coupons than the population of elderly persons as a whole. Also, the number of walk trips was not affected by the number of bus and taxi trips taken. Therefore, people who have participated in the subsidy program have enjoyed a net increase in mobility (in the form of additional bus and taxi trips) because bus and taxi trips have not simply replaced walk trips.

Starting in July 1978 and for 24 consecutive months thereafter, the U.S. Department of Transportation (DOT) conducted an experiment of user-side subsidies for public transportation in Lawrence, Massachusetts, a small metropolitan city north of Boston. A select group of individuals--the elderly (65 years and older) and the handicapped of all ages--was eligible to receive financial assistance in the form of a reduced bus fare (the regular bus fare for elderly and handicapped persons was $0.15, but only $0.01 if project coupons were used) and a 50 percent discount on taxi rides (the discount was limited to $1.25 per ride and $20 per month). To establish eligibility individuals were to register at a downtown office, which was also the only location where discount coupons for bus and taxi rides could be purchased.

In conjunction with the experiment, a sample of individuals who were eligible to receive the assistance was contacted and asked to report sociodemographic information and to record a diary of travel for May 1978 and May 1979 (before the experiment and during the tenth month of the experiment). Although the total sample included both elderly and transportation-handicapped persons, only the subsample of the elderly (handicapped and able-bodied persons) was selected for this study. From this group, 130 completed returns were available; 48 percent of these returns were from transportation-handicapped persons, and 40 percent of the returns were from individuals who chose to become project users.

The purposes of this paper are to measure the travel impact of the experiment on the elderly population and to understand the reasons that attracted some of the eligible population to purchase discounted coupons and to use bus and taxi for their travel.

There is a problem in measuring the impact of the project because the purchase of the discounted coupons is prompted by expected benefits and other exogenous factors that are not fully measurable. If the incidence of these factors was known, the variables that identify them could be used in the analysis. Unfortunately, these variables are often not known or measured; thus in this paper a method to represent their effect is presented.

In the following sections two models that measure bus and taxi trip frequency, and a model that measures the number of walk trips, are presented. The latter model is used to determine whether walk trips are being replaced by bus or taxi trips.

PROBLEM OF SELF-SELECTION TO TREATMENT

Although the goal of this research is to measure the effectiveness of the project in increasing travel mobility, it is recognized that the inevitable limitations of the data generate issues that the model has to deal with explicitly. This is so, in particular, because the choice of becoming a project user (i.e., registering in the project and purchasing the discount coupons) rests entirely on the individuals who participate in the survey. Therefore, a definition has to be found for the following dichotomous variable for individual t,

$$d_t = \begin{cases} 1 \text{ if individual purchases discounted coupons} \\ 0 \text{ otherwise} \end{cases} \qquad (1)$$

and for the following model of travel demand,

$$y_t = \beta' X_t + \delta d_t + \epsilon_t \qquad (2)$$

where

y_t = number of trips taken by individual t;
β = column vector of coefficients;
X_t = column vector of independent variables;
δ = a scalar, which is the coefficient of the dichotomous variable d_t; and
ϵ_t = stochastic component of the model.

At first glance it would appear that δ would represent the effect of the project. However, those who became project users did so because, as a general rule, they expected their travel demand to be higher than otherwise, and those who chose not to become users did so because they did not expect their travel to increase by becoming users. In other words, the benefits that users derive from purchasing the discounted coupons are larger than the benefits foregone by nonusers. This implies that d_t and ϵ_t are correlated; thus the model of trip generation that was proposed could not be estimated either by ordinary regression or by conventional cross-classification, a method that assumes, much like ordinary regression, independently distributed stochastic components.

As mentioned previously, if it was possible to measure all the variables that determine project participation, the variables could be incorporated in the analysis explicitly. However, because some of these variables are unmeasured, it is necessary to consider d_t as being an endogenous variable. Thus the estimation of a model that recognizes this endogenicity, which is also called selectivity bias, is presented. The theoretical justification for such a model is straightforward, and the reader is referred to the extensive literature on the subject (1,2) for more detail.

BUS FREQUENCY MODEL

Purchase of discounted coupons for bus travel is clearly a major factor in the frequency with which individuals take bus trips. However, as discussed earlier, the use of the variable that represents the observed purchase decision in the model could yield inconsistent estimates of the project effect because of the likely presence of sample selectivity. Accordingly, the bus frequency model is estimated by a two-stage procedure first proposed by Maddala and Lee (3). The procedure requires estimation of a probit model of the decision to purchase discounted bus coupons, and estimation of a model (which incorporates as an independent variable the expected value of the dependent variable of the probit) of bus trip frequency.

The probit model of purchase of discounted bus fares is estimated from data on the actual purchase of these fares in May 1979. The observed dependent variable of the model is equal to one if bus coupons were purchased (in May 1979) and zero otherwise. The probability of purchasing discounted bus coupons (i.e., the expected value of the dependent variable) is equal to $\Phi(\gamma' Z_t)$, where Z_t is a column vector of independent variables, γ is a column vector of coefficients, and $\Phi(.)$ is the cumulative of the standard normal distribution. The estimated coefficients (γ), together with some goodness-of-fit measures, are given in Table 1. Although the probit was formulated as a single-equation model, different coefficients were estimated for able-bodied and transportation-handicapped persons.

Table 1. Probit estimates of use of bus coupons.

User	Coefficient	Asymptotic t-Statistic
Able-bodied person		
Constant	−0.945	3.2
Zero automobiles in household	0.711	1.7
Bus trips in May 1978	0.0678	2.5
Transportation-handicapped person		
Constant	−0.856	3.1
Employed or student	2.09	2.9
Bus trips in May 1978	0.281	3.6

Note: Log-likelihood with estimated coefficients = −50.16, log-likelihood with constants alone = −74.86, log-likelihood ratio statistic (4 df) = 49.4, number of observations = 130, 85.4 percent of sample was correctly classified, 10.8 percent of sample was erroneously classified as nonuser, and 3.8 percent of sample was erroneously classified as user.

For able-bodied elderly persons, automobile ownership (a zero-one variable) was found to affect the purchase of bus coupons significantly, whereas for transportation-handicapped persons, the most important variable was that of employment and student status, again a zero-one variable. The log-likelihood ratio statistic is equal to 49.4, a value that allows rejection, with a large level of confidence, of the hypothesis of no effect of the independent variables.

The second-stage model--a limited dependent variable model of the number of bus trips--is estimated from bus trips reported in the May 1979 diary survey. As discussed earlier, instead of including a zero-one variable for actual coupon purchase (or nonpurchase), the probability of being a project user is included in this model, i.e., the expected value of the dependent variable from the probit model. This ensures that the coefficient for the bus coupon purchase variable is consistent because sample selectivity is accounted for. A single-equation specification is again used for the groups of able-bodied and transportation-handicapped persons. The estimated coefficients are given in Table 2.

To test the effectiveness of the program, further statistical tests are performed on the subsample of actual project users. Specifically, the expected number of bus trips of project users, had they been nonusers, is compared with the actual number of bus trips taken. Because the distribution of the number of trips is truncated normal, the probability that the expected number of bus trips (conditional on nonpurchase of the project coupons is lower than the actual number of bus trips) is written as $(X'\beta - \mu)/\sigma$, where β is a column vector of coefficients, X is a vector of independent variables, μ is the actual number of bus trips taken in May 1979, and σ is the standard deviation of the underlying non-truncated distribution of the stochastic component of the model. For the subsample of program users, this probability averages 80 percent, and the Pearson's P_λ is 252.90 with 70 df, a value that clearly permits rejection of the null hypothesis of no-project effect on bus travel. Note also that the mean number of bus trips for the individuals who purchased discounted bus coupons in May 1979 is 16.51, whereas the mean expected number of bus trips for the same individuals, had they been nonusers, is 4.70, a difference of approximately 12 monthly trips.

Table 2. Estimates of May 1979 bus trips (limited dependent variable model).

User	Coefficient	Asymptotic t-Statistic
Able-bodied person		
Constant	−6.07	2.5
Probability of being a user for individuals who are neither students nor employed	28.7	2.2
No. of bus trips in May 1978	0.791	1.9
Transportation-handicapped person		
Constant	−10.1	3.9
Probability of being a user	23.4	4.3
No. of bus trips in May 1978	0.630	4.2
σ	10.4	10.6

Note: $y^* = X'\beta + \epsilon$

$$y = \begin{cases} 0 & \text{if } y^* < 0.5 \\ y^* & \text{otherwise} \end{cases}$$

and log-likelihood with estimated coefficients = −270.36, log-likelihood with constants alone = −318.01, log-likelihood ratio statistic (4 df) = 95.3, and number of observations = 130.

TAXI FREQUENCY MODEL

The estimation of a probit model of taxi coupon purchases did not yield acceptable results. Specifically, standard statistical tests pointed to the low explanatory power of the model. Several different specifications of the probit model were tested, but those also met with little success. Although it would have been possible to investigate the failure of the probit formulation to yield a satisfactory model, doing so would have been beyond the scope of this research. As a consequence, the two-stage procedure adopted for the bus frequency model was replaced by a simpler model. This model, which measures the monthly taxi trips taken, includes as an independent variable the actual purchase (or nonpurchase) of taxi coupons in May 1979 (a zero-one variable) and not the expected value from a probit model.

It is recognized that the coefficient estimate of the coupon purchase variable will be biased because of its endogenicity. However, it should be mentioned that this endogenicity is expected to be much less severe in the taxi model than in the bus model, particularly because the subsidy is only 50 percent (versus 93 percent for bus trips) and it is more limited in availability (the maximum taxi subsidy is $1.25 per trip and $20.00 per month per person). Accordingly, although the model presented in the following paragraphs has some evident limitations, it was decided to include it in this paper for completeness.

The taxi frequency model, like the model for bus travel, is a limited dependent variable model. As for the previous model, the taxi trip rate cannot be negative, and 79 of the 130 persons in the sample (61 percent) did not take any taxi trips in May 1979. In addition to the zero-one variable for individuals who purchased taxi coupons, the number of household automobiles has, as expected, a significant effect on taxi trip frequency (see Table 3).

Table 3. Estimates of May 1979 taxi trips (limited dependent variable model).

Item	Coefficient	Asymptotic t-Statistic
Constant	-7.54	4.6
No. of household automobiles	3.48	2.1
No. of taxi trips in May 1978	0.812	8.5
Purchased taxi coupons (1 if yes, 0 otherwise)	9.10	5.3
σ	6.91	9.6

Note: $y^* = X'\beta + \epsilon$

$y = \begin{cases} 0 & \text{if } y^* < 0.5 \\ y^* & \text{otherwise} \end{cases}$

and log-likelihood with estimated coefficients = -199.56, log-likelihood with constant alone = -247.35, log-likelihood ratio statistic (3 df) = 95.6, and number of observations = 130.

To test the effectiveness of the program in increasing taxi travel, statistical tests identical to the ones used for the bus travel model are applied here. Specifically, Pearson's P_λ (which has a value of 167.92 for the subsample of the 30 individuals who are taxi coupon purchasers) allows rejection of the null hypothesis of no increase in taxi travel because of project participation. The analysis also indicates that the mean number of taxi trips taken in May 1979 by taxi coupon purchasers is 8.6, whereas the expected value conditional on nonpurchase is 3.45 taxi trips for the same group of individuals, a difference of approximately 5 trips per month.

WALK TRIPS FREQUENCY MODEL

Although vehicular trips in general, and bus and taxi trips in particular, increased as a result of the user-side subsidy, it was hypothesized that some of the new vehicular trips might have replaced what were formerly walk trips. To test this hypothesis a walk frequency model that includes bus and taxi trip frequency as explanatory variables is estimated.

Because bus and taxi trips are endogenous to the walk trips model (i.e., the models for each travel mode are part of a system of structural equations), it was decided to use the expected trip rates from the models presented in the previous two sections as instruments instead of using the observed trip rate for bus and taxi trips.

The specification chosen for the estimation is again a limited dependent variable model. As for the previous models, the walk trip rate cannot be negative, and 17 of the 130 persons in the sample (13 percent) did not take any walk trips in May 1979. The coefficient estimates for the model are given in Table 4. If bus and taxi trips were actually replacing potential walk trips, the coefficients of the frequency of bus and taxi trips would be negative (and statistically significant). The results, however, reveal these coefficients to be positive and not statistically different from zero, which indicates that the hypothesis of modal substitution is unlikely to be valid.

Table 4. Estimates of May 1979 walk trips (limited dependent variable model).

Item	Coefficient	Asymptotic t-Statistic
Constant	0.40	0.13
Expected no. of bus trips in May 1979	0.15	0.85
Expected no. of taxi trips in May 1979	0.43	1.3
No. of walk trips in May 1978	0.79	18.0
σ	20.7	15.0

Note: $y^* = X'\beta + \epsilon$

$y = \begin{cases} 0 & \text{if } y^* < 0.5 \\ y^* & \text{otherwise} \end{cases}$

and log-likelihood with estimated coefficients = -515.862, log-likelihood with constant alone = -602.458, log-likelihood ratio statistic (3 df) = 173.19, and number of observations = 130.

CONCLUSIONS

The models presented in this paper have confirmed quite strongly the a priori hypothesis regarding travel by bus, taxi, and walk. The large increases in bus and taxi travel observed in May 1979 by those individuals who purchased discounted coupons can be directly attributed to the project. Also, it was shown that the increase in bus and taxi trips was not achieved at the expense of walk trips. Rather, the additional bus and taxi trips were trips that would have not been taken in the absence of the subsidy project.

The data in Table 5 further confirm the findings of the models. Note in particular the increase (between 1978 and 1979) in bus trips for bus subsidy users (i.e., for those individuals who purchased bus coupons), and the increase in taxi trips for taxi subsidy users. These increases are much larger than the increases for the sample as a whole and for the subsample of nonusers of the program.

Table 5. Trip rates by mode and project participation status.

		Project Participation Status			
Mode	Month and Year	All Sample (n = 130)	Project Users, Taxi and Bus (n = 49)	Project Bus Users (n = 35)	Project Taxi Users (n = 30)
Bus	May 1978	3.52	7.18	9.97	8.27
	May 1979	6.22	12.82	16.51	11.77
Taxi	May 1978	2.43	3.69	3.83	5.33
	May 1979	3.22	5.57	4.37	8.6
Walk	May 1978	40.27	49.27	54.54	42.9
	May 1979	37.12	44.98	49.69	39.93
All modes	May 1978	109.05	100.00	103.63	100.67
	May 1979	106.69	99.61	104.14	99.37

Walk trips are mostly unaffected by program use, which confirms the findings of the model of walk trips. Note that only bus subsidy users take a larger number of walk trips than other groups.

ACKNOWLEDGMENT

This paper was prepared as part of an exploratory study of the travel behavior of the elderly and the handicapped; it was sponsored by the Service and Methods Demonstration program of UMTA. I wish to acknowledge various comments by Lawrence B. Doxsey of the Transportation Systems Center.

REFERENCES

1. B.S. Barnow, G.G. Cain, and A.S. Goldberg. Issues in the Analysis of Selectivity Bias. In Evaluation Studies Review Annual, Volume 5 (E.W. Stromsdorfer and G. Farkas, eds.), Sage Publications, Beverly Hills, Calif., 1980.
2. J. Heckman. Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables with or Without Structural Shift in the Equations. In Studies in Nonlinear Estimation (S.M. Goldfeld and R.E. Quandt, eds.), Ballinger Publishing Company, Cambridge, Mass., 1976.
3. G.S. Maddala and L.-F. Lee. Recursive Models with Qualitative Endogenous Variables. Annals of Economic and Social Measurement, Vol. 5, 1976, pp. 525-545.

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

Notice: This paper represents the views of the author and not necessarily those of U.S. Department of Transportation.

Effect of Sample Size on Disaggregate Choice Model Estimation and Prediction

FRANK S. KOPPELMAN AND CHAUSHIE CHU

Sampling error is one of several types of error in econometric modeling. The relationship between sampling error and sample size is well known for both estimation and prediction. The objective of this paper is to provide an empirical foundation for using these relationships to guide researchers and planners in the determination of sample size for model development. Analytic relationships are formulated for sample size, precision of parameter estimates, replication of parent population, and replication of an alternative (transfer) population. Application of these relationships to an empirical case indicates that the sample sizes required to obtain reasonably precise parameter estimates are substantially larger than the sample sizes generally considered to be needed for disaggregate model estimation. Nevertheless, these sample sizes appear to be adequate for obtaining reasonably accurate replication of observed choice behavior in the parent population. The corresponding results for prediction to a different population are complicated by the issue of intrapopulation transferability. Although the results reported in this paper should be validated in other contexts, it appears that accurate estimation requires the use of samples that are substantially larger than formerly believed. Samples on the order of 1,000 to 2,000 observations may be needed for estimation of relatively simple disaggregate choice models. Although some reduction in this requirement may be obtained by improved sample design, it is unlikely that the final sample requirements can be reduced to less than 1,000 observations.

Econometric model development is subject to errors in sampling, model specification, and measurement (1,2). In this paper the effect of sampling error is examined for model parameter estimates, prediction to the parent population, and transfer prediction to alternative populations. Sampling error can be avoided only by observation and analysis of the entire population. In practice, the resources needed to collect data for an entire population and to analyze such extensive data are not available. Thus there is concern with the magnitude of the errors that are introduced by use of samples of the population.

EXPECTED EFFECTS OF SAMPLE SIZE

The precision of parameter estimates for a given model structure depends on the estimation method used, the multidimensional distribution of the explanatory variables of the model, the range of observed behavior, the quality of model specification, and the sample size of the estimation data set. Maximum likelihood estimation obtains consistent estimators of the parameters of disaggregate choice models and provides estimates of the precision with which model parameters are estimated (3-5).

The relationship between parameter precision and sample size is well known. The variance-covariance matrix of estimated parameters in linear models is inversely proportional to sample size (3,6). The variance-covariance matrix of maximum likelihood estimated parameters for quantal choice models is asymptotically equal to the negative inverse of the Hessian of the log-likelihood function (3,7). The asymptotic expectation of this matrix is inversely proportional to sample size. Thus the error variance-covariance matrix for maximum likelihood estimations for quantal choice models is also inversely proportional to sample size.

Prediction accuracy describes how well the choice model replicates observed population behavior. Prediction performance of discrete choice models is a function of the validity of model theory, the validity of the derived model structure, the quality of model specification, the quality of variable measurement and prediction, and the accuracy of estimated parameters (8). As noted earlier, precision of model parameter estimates is proportional to sample size. It follows that the portion of prediction error attributable to errors in parameter estimation is inversely proportional to sample size. Specifically, the expected squared prediction error caused by errors in parameter estimates is inversely proportional to sample size (5, p. 189). Models estimated from large samples are more likely to accurately describe the behavioral process in the general population, and consequently such models will have satisfactory prediction performance. Thus it is expected that increased sample size in model estimation will yield improved prediction precision. When excessively small samples are used, both parameter estimates and parent population predictions will be highly variable.

Transferability of disaggregate discrete choice models is based on the argument that choice models describe the underlying behavioral response mechanisms or decision rules of decision makers in the selection among available alternatives (9,10). If the behavioral response or decision rules of decision makers is constant across contexts, models that describe this behavior will be transferable. Koppelman and Wilmot (11) define transferability of choice models as "the degree of success with which the predictions obtained by model transfer describe behavior in the prediction context." Transferability is a function of the quality of the model being transferred and similarity of behavior between the estimation and application contexts.

If choice behavior in the estimation and application contexts is based on the same behavioral process, the transfer predictive accuracy will be increased with increasing estimation sample size. In this case a model that is able to provide an accurate description of choice behavior in the estimation context will be able to provide an accurate description in the transfer or prediction context. However, if the behaviors are different between contexts, increasing sample size will not overcome these differences.

The objective of this paper is to examine the effect of sample size on parameter stability, parent population replication, and transferability of disaggregate discrete choice models of multinomial logit structure. In each case the expected relationship is formulated, an empirical analysis to scale the relationship is executed, the implications of the results obtained are identified, and the conclusions are stated. Also described in the paper are the data used and the structure of the empirical analysis undertaken.

DATA DESCRIPTION AND EXPERIMENTAL DESIGN

Data

The data used in this study are drawn from the Washington Council of Governments travel to work modal-choice data collected in Washington, D.C., in 1968. The data used describe the central business district (CBD) work trips of 2,236 persons. A total of 1,768 persons have drive-alone, shared-ride, and transit alternatives available, and 468 persons have only the shared-ride and transit alternatives because of a lack of driver's license or cars available in the household.

The data set is partitioned into three geographic sectors of the region according to worker residential location. Each sector includes approximately one-third of the sample observations. The partition allows for the examination of the first two relationships (parameter precision and parent population replication) within each sector and the investigation of the transferability prediction relationship for six possible transfers between sectors.

Experimental Design

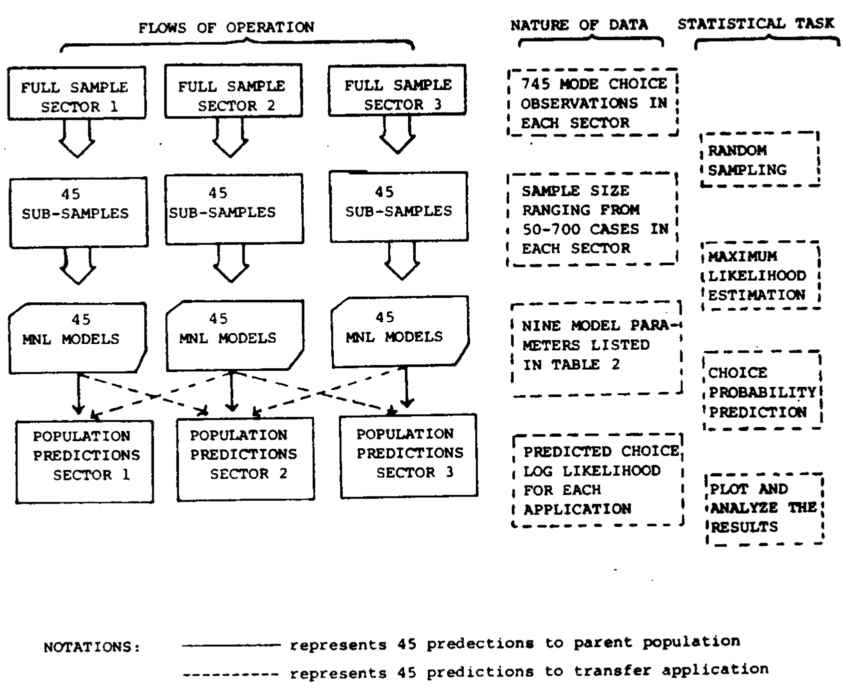
The experiment is constructed by defining the full sample in each sector as the population of interest, and then subsamples of varying size are selected. These subsamples are used to estimate multinomial logit model parameters, predict choice behavior for the population from which each sample is drawn, and predict choice behavior in each of the other populations (different sectors). The flowchart of this experimental design is shown in Figure 1, which describes the sampling and estimation process and also the data used in each step.

The first task of the experiment is to obtain subsamples of each data set with varying sample sizes. Forty-five sets of random subsamples are independently generated within each of the three sectors. Within each sector the number of individuals in samples varies from approximately 50 to approximately 700.

The second task is to estimate travel modal-choice models for each data sample. A nine-variable model previously used in a related study of model transferability (11) is used in this study. These variables are described in Table 1. By using a single-model specification, it is possible to examine the effect of sample size without any confounding effects caused by differences in model specification. The estimation results for these models that use the full set of cases (the population) in each sector, as well as additional data, are reported in Tables 2 and 3. These estimation results serve as a reference point for the models estimated with each data subsample. The subsample estimation results are discussed later in this paper.

The third step in this study is to use the 45 models estimated in each sector to predict travel choices for the full population in each of the three

Figure 1. Flowchart for experimental design.



sectors. Thus each estimated model is used for three predictions (one local and two transfer predictions). Population replication performance and transferability measures are developed for each of these predictions and used to interpret the model accuracy relationships.

EFFECT OF SAMPLE SIZE ON PARAMETER PRECISION

Parameter precision is the inverse of the variance of parameter estimates obtained in repeated samples. In this section the effect of sample size on parameter precision is evaluated by comparing estimated parameter values for each sample with the population parameters reported in Tables 2 and 3.

Relation Between Parameter Precision and Sample Size

The total available data sample is treated as the population of interest, and the difference between models estimated on subsamples and models obtained from the population (full sample) is examined. As all the data included in each subsample are also included in the full sample, the parameter estimates obtained from samples are not independent of parameter estimates obtained from the full data. The variance-covariance matrix of estimates of the difference between sets of parameter estimates is

$$\Sigma_z = \Sigma_s + \Sigma_p - 2\Sigma_{sp} \quad (1)$$

where

Σ_z = error variance-covariance matrix for difference between subsample and full sample parameters (i.e., $z = \beta_s - \beta_p$);
Σ_s, Σ_p = error variance for subsample and full sample parameter estimates, respectively; and
Σ_{sp} = covariance matrix of error between subsample and full sample parameter estimates.

When the subsample is a subset of the full sample $\Sigma_{sp} = \Sigma_s$ (see Appendix),

$$\Sigma_z = \Sigma_s - \Sigma_p \quad (2)$$

which is a positive semidefinite covariance matrix of the differences between parameter estimates obtained from the full and partial samples. The expected relationship between the full and partial sample error variances is

$$\Sigma_p = \Sigma_s (N_s/N_p) \quad (3)$$

Thus, from Equations 2 and 3,

$$\Sigma_z = [(N_p - N_s)/N_p]\ \Sigma_s \quad (4a)$$

and

$$\Sigma_z = [(N_p - N_s)/N_s]\ \Sigma_p \quad (4b)$$

A standardized variable of differences is formulated in parameter estimates (Q) by dividing observed differences (z) by the standard error in population estimates (s_p); i.e., square root of diagonal elements in Σ_p,

$$Q = z/s_p \quad (5)$$

where Q is the difference between sample parameter and population parameter values in units of standard error of estimate for population parameters. Then the variance and 95 percent confidence interval of Q are

$$V(Q) = (N_p - N_s)/N_s \quad (6a)$$

and

$$-1.96\ [(N_p - N_s)/N_s]^{½} < Q^* < 1.96\ [(N_p - N_s)/N_s]^{½} \quad (6b)$$

Table 1. Model specifications.

Variable Name	Variable Description
DAD, SRD	Dummy variable specific to drive-alone and shared-ride alternative; measures average bias between pairs of alternatives other than that represented by the included variables
CPDDA, CPDSR	Cars per driver included separately as alternative specific variables from the drive-alone and shared-ride modes; measures the change in bias among modes caused by changes in automobile availability within the household
OPTCINC	Round trip out-of-pocket travel cost divided by income (cents/$1,000 per year); measures the effect of travel cost on mode utility with cost effect modified by household income level
TVTT	Round trip total travel time in minutes; measures the linear effect of combined in- and out-of-vehicle travel time in mode utility
OVTTD	Round trip out-of-vehicle travel time divided by trip distance (minutes/mile); measures the additional effect of out-of-vehicle travel time in utility in addition to the effect represented in TVTT; this added effect is structured to decline with increasing trip distance
GWSR	Dummy variable that indicates if the breadwinner is a government worker specific to the shared-ride alternative; measures the effect on shared-ride utility of shared-ride incentives for government workers
NWORKSR	Number of workers in the household specific to the shared-ride alternative; measures the change in utility of shared ride when there is an opportunity to share ride with a household member

Table 2. Parameter estimates and standard errors.

	Sector 1		Sector 2		Sector 3		Region	
Variable	Estimated Parameter	Standard Error	Estimated Parameter	Standard Error	Estimated Parameter	Standard Error	Estimated Parameter	Standard Error
DAD	-3.30	0.425	-1.44	0.388	-2.73	0.402	-2.67	0.226
SRD	-2.62	0.321	-1.92	0.277	-2.52	0.345	-2.35	0.175
CPDDA	4.06	0.426	2.70	0.382	3.58	0.396	3.41	0.227
CPDSR	2.06	0.319	1.67	0.235	1.59	0.315	1.77	0.159
OPTCINC	-0.0138	0.0155	-0.0282	0.0139	-0.0280	0.0163	-0.0297	0.0084
TVTT	-0.0459	0.0070	-0.0110	0.0050	-0.0223	0.0049	-0.0233	0.0031
OVTTD	-0.0019	0.0668	-0.1068	0.0666	-0.0421	0.0781	-0.0588	0.0393
GWSR	0.775	0.179	0.481	0.166	0.680	0.163	0.648	0.096
NWORKSR	0.133	0.128	0.275	0.110	0.502	0.123	0.308	0.067

Table 3. Estimation statistics for sectors.

Item	Sector 1	Sector 2	Sector 3	Region
No. of cases	744	746	746	2,236
No. of observations	2,078	1,997	2,156	6,240
Log-likelihood at zero	-755	-722	-790	-2,266
Log-likelihood at convergence	-580	-636	-688	-1,928
Likelihood ratio statistic	350	171	203	678
Likelihood ratio index	0.232	0.118	0.129	0.150

Thus the variance of this standardized measure, given a particular model and population, is a function of sample size only.

Scattergrams of estimates for the standardized measure of the seven slope parameters defined in Table 1 have been plotted against estimation sample size for three different sectors, and they were found to be similar. The scattergram for one parameter in all three sectors is shown in Figure 2, along with the 95 percent confidence limits. As expected, the estimated parameters are distributed around the true parameters, with the range of the distribution decreasing as the number of cases in the estimation sample increases. It appears that the mean of Q is approximately zero (as expected), and its variance is described by Equation 6a. Further, approximately 95 percent of the reported deviations are within the expected range. Finally, as expected, the distribution appears to be independent of the estimation sector.

Parameter Precision and Required Sample Size

The deviations of sample parameter estimates from population parameters for each sector and variable are related to the standardized deviation (Q) by the population parameter standard deviation (s_p), as shown in Equation 5. Thus the variance of observed parameter deviations (z_s) is

$$V(z) = s_p^2 \times V(Q)$$
$$= s_p^2 \times [(N_p - N_s)/N_s] \qquad (7)$$

which is a function of the estimation precision of the parameter in the population and the sample size. Thus it is possible to determine a priori the sample size necessary to obtain a predetermined level of precision in parameter estimates if the population estimation precision and the population size are known.

The interpretability of this relationship can be improved by formulating an index of estimation precision that is independent of both population size and sample size. Thus,

$$s_*^2 = \underset{N\to\infty}{E} (s_N^2 \times N) \qquad (8)$$

This index, which can be estimated by

$$s_*^2 = s_p^2 \times N_p \qquad (9a)$$

or

$$s_*^2 = s_s^2 \times N_s \qquad (9b)$$

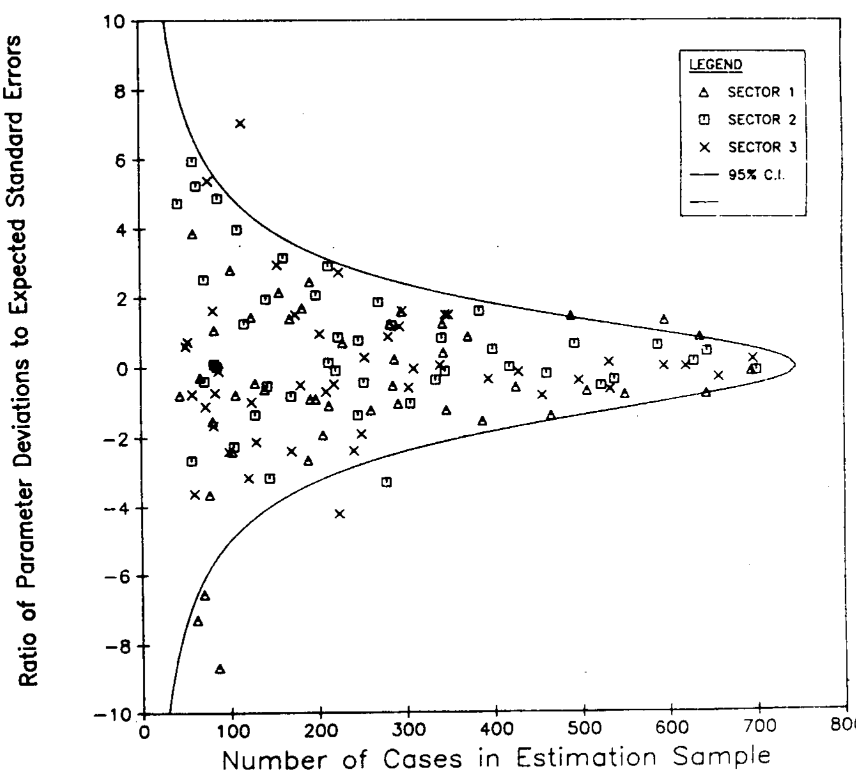
characterizes the underlying precision of a parameter independent of population or sample size. This index is used in Equation 7 to obtain

$$V(z) = s_*^2 \, [(N_p - N_s)/(N_p \times N_s)] \qquad (10)$$

By using this formulation, the sample size required to obtain a desired level of precision in parameter estimation can be obtained as a function of population size and as the precision index for the parameter of interest. Specifically, an N_s is sought that satisfies

$$t_{\alpha/2}\, s_* \, [(N_p - N_s)/(N_p \times N_s)]^{1/2} = z^* \qquad (11)$$

Figure 2. Scattergram of parameter precision with estimation sample size.



where $t_{\alpha/2}$ is the t value associated with the desired (α) confidence interval for z, and z* is the desired level of precision for parameter deviations. Thus the required sample size is

$$N_s^* = [N_p + N_p (z^*/s_* \; t_{\alpha/2})^2]/[1 + N_p (z^*/s_* \; t_{\alpha/2})^2] \quad (12)$$

which, when N_p is large, simplifies to

$$N_s^* \simeq (s_* \; t_{\alpha/2}/z^*)^2 \quad (13)$$

This relationship (Equation 12) is plotted in Figure 3 for the case where the parameter deviation (z) is to be within a prespecified fraction of the parameter precision index (s*) with 95 percent confidence.

Equation 12 (or Equation 13 for large populations) can be used to predetermine the sample size required to obtain a desired level of parameter estimation precision. This determination is based only on prior knowledge of population parameter precision (s*) and population size. Estimates of population parameter precision may be obtained by reviewing estimation results of similarly specified models in other contexts or by using a small data sample. The use of small data samples to obtain information to optimally design the sample collection procedure for a given sample size has been treated extensively by Daganzo (12).

The use of Equation 12 is demonstrated by calculating the sample size required to have 80 percent confidence so that the absolute value of z is less than 25 percent of the true parameter value. (More generally, this analysis can be undertaken by setting limits to the deviations of each parameter based on required or desired precisions in model sensitivity and the differences in the corresponding variable across plan alternatives. However, use of an arbitrary proportional range provides useful insight in an abstract context.) The calculation process and results are given in Table 4. These results illustrate again that as population increases, the number of sample observations needed to obtain parameter estimates in a prespecified range increases at a decreasing rate. When the population is large (i.e., more than 100,000), the required estimation sample size approximates that for an infinite population.

More important, the sample sizes required to obtain what would appear to be a modest level of parameter precision are substantially greater than those commonly used in the estimation of disaggre-

Figure 3. Required sample size with 95 percent confidence.

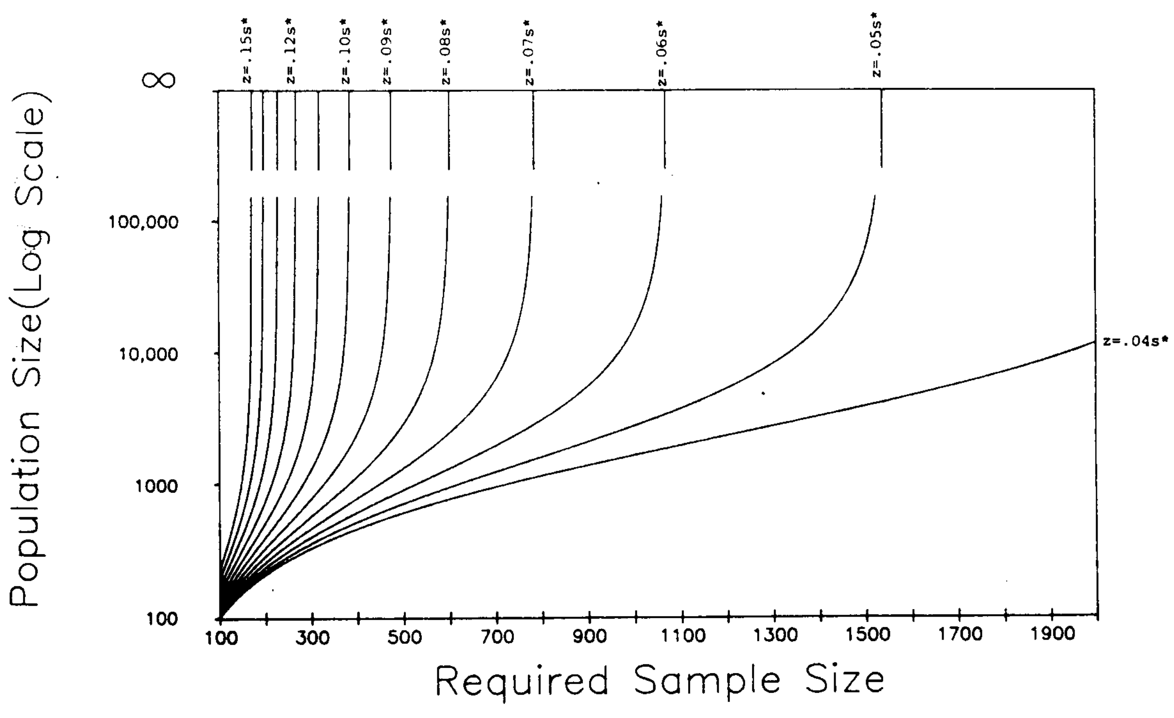


Table 4. Computation of required sample size to obtain parameter estimate with 80 percent confidence within 25 percent of true values.

Variable	β^* (see Tables 2 and 3)	Z ($\pm 0.25\ \beta^*$)	S_s (see Tables 2 and 3)	S_* (From Equation 8b)	Required Estimation of Sample Size for Different Population Sizes: N_p = 100,000	N_p = 1,000,000
DAD	-2.366	±0.5915	0.2261	10.69	533	536
SRD	-2.349	±0.5873	0.1747	8.26	324	325
CPDDA	3.047	±0.8518	0.2268	10.72	260	261
CPDSR	1.767	±0.4418	0.1593	7.53	475	477
GWSR	0.6477	±0.1619	0.0962	4.55	1,276	1,291
NWORKSR	0.3084	±0.0771	0.0674	3.19	2,721	2,789
OPTCINC	-0.0297	±0.0074	0.0084	0.395	4,422	4,605
TVTT	-0.0233	±0.0058	0.0031	0.147	1,032	1,042
OVTTD	-0.0588	±0.0147	0.0393	1,086.0	20,756	25,523

gate choice models. The sample size required is substantially greater than the 300 to 500 observations that are commonly believed to be adequate for estimation of disaggregate choice models (13,14) for more than half of the model parameters. Use of the smaller samples can be expected to produce parameter estimates that have a high probability of being different from the true parameters. This problem is most serious for level-of-service parameters in this data set.

Conclusions

Two important observations are drawn from these results. First, as expected from sampling theory, the variability of parameter estimates is inversely related to sample size in a nonlinear fashion. This relationship is described in Equation 6a and is shown in Figure 3. Second, the sample size needed to obtain a reasonable degree of precision for managerial policy analysis may be substantially larger than is commonly suggested for the estimation of disaggregate choice models. The commonly held belief that 300 to 500 observations are satisfactory seriously underestimates the sample size suggested in this analysis to be needed to obtain estimators with a reasonable level of precision, especially for service variables. The importance of these results, if verified in other studies, is heightened by noting that many studies use samples of 1,000 or less observations (15-20), whereas this study suggests a need for at least 1,000 observations to estimate the influence of travel time--a most important variable--within an error of 25 percent with 80 percent confidence.

EFFECT OF SAMPLE SIZE ON REPLICATION OF PARENT POPULATION BEHAVIOR

In this study an examination was made of the accuracy with which a model, based on a data sample, will replicate the choice behavior in the parent population.

Relation Between Replication Precision and Sample Size

A prediction test statistic was formulated to test the hypothesis that the subsample model β_s is equivalent to the population model β_p,

$$PTS_p(\beta_s) = -2[LL_p(\beta_s) - LL_p(\beta_p)] \tag{14a}$$

This statistic, which is approximately chi-squared, can be expressed as a quadratic function of the difference in parameter vectors (5):

$$PTS_p(\beta_s) \simeq (\beta_p - \beta_s)' \Sigma_p^{-1} (\beta_p - \beta_s) \tag{14b}$$

Entering the relationships of $z = \beta_s - \beta_p^*$ and $\Sigma_z = [(N_p - N_s)/N_s]\Sigma_p$ into Equation 14b:

$$PTS_p(\beta_s) \simeq [(N_p - N_s)/N_s]\, z' \Sigma_z^{-1} z \tag{15}$$

where the quadratic term has a chi-square distribution. Thus the mean, variance, and $1 - \alpha$ confidence limit of PTS are

$$E(PTS) = [(N_p - N_s)/N_s] \times DF \tag{16}$$

$$V(PTS) = 2\,[(N_p - N_s)/N_s]^2 \times DF \tag{17}$$

and

$$PTS_\alpha < [(N_p - N_s)/N_s]\, \chi^2_{DF,\alpha} \tag{18}$$

Thus both the average and the variance of PTS decrease at decreasing rates as estimation sample size increases and are asymptotic to zero as sample size approaches population size.

Empirical Population Replication Analysis

To empirically demonstrate the results derived in the previous subsection, the predicted population log-likelihood by subsample models was compared with the maximum population log-likelihood by the full sample model in each sector by using Equation 14a. To examine the distribution and the 95 percent confidence limit of the prediction test statistic, scattergrams were plotted of the prediction test statistic against the size of estimation subsamples in Figure 4, and different symbols were used to represent observations in three different sectors. The results were as follows. First, as expected, PTS is subject to large variance when estimation sample size is small. The variance decreases quickly as estimation sample size increases for observations in all three sectors. Second, the curve that represents the expected value of PTS appears to fit the data well in all three sectors. Third, it appears that approximately 95 percent of the observations are within the 95 percent confidence limit shown in the figure. Thus these observations are consistent with the analytic results in the previous subsection.

Next, a prediction index was formulated that defines the degree to which the model estimated from the sample describes the population choice behavior relative to a model based on the full population. First, the common sample-based rho-square measure was considered:

$$\begin{aligned}\rho_{ss}^2 &= [LL_s(\beta_s) - LL_s(NM)]/[LL_s^* - LL_s(NM)] \\ &= 1 - [LL_s(\beta_s)/LL_s(NM)]\end{aligned} \tag{19}$$

and then the corresponding population-based rho-square measure based on sample estimates was considered:

$$\begin{aligned}\rho_{ps}^2 &= [LL_p(\beta_s) - LL_p(NM)]/[LL_p^* - LL_p(NM)] \\ &= 1 - [LL_p(\beta_s)/LL_p(NM)]\end{aligned} \tag{20}$$

Based on population estimates,

$$\begin{aligned}\rho_{pp}^2 &= [LL_p(\beta_p) - LL_p(NM)]/[LL_p^* - LL_p(NM)] \\ &= 1 - [LL_p(\beta_p)/LL_p(NM)]\end{aligned} \tag{21}$$

Next, the prediction index as the ratio of Equations 20 and 21 were formulated to obtain

$$PI = [LL_p(\beta_s) - LL_p(NM)]/[LL_p(\beta_p) - LL_p(NM)] \tag{22}$$

The degree to which the sample-based model provides information about population behavior relative to that provided by the population-based model (when both referred to a common base or null model) is described by this ratio. To interpret this index, it was reformulated in terms of the population test statistic defined in Equation 14,

$$PI = 1 - \{PTS_p(\beta_s)/2[LL_p(\beta_p) - LL_p(NM)]\} \tag{23}$$

Note that the denominator in the second term is fixed for any population and model specification. Further, this term is the population model likelihood ratio statistic reported in Table 3 for each of the population models. These results can be used to obtain the expected value of the prediction index for fixed population size as

$$E(PI) = 1 - \{[(N_p - N_s)/N_s] \cdot DF/LRS\} \tag{24}$$

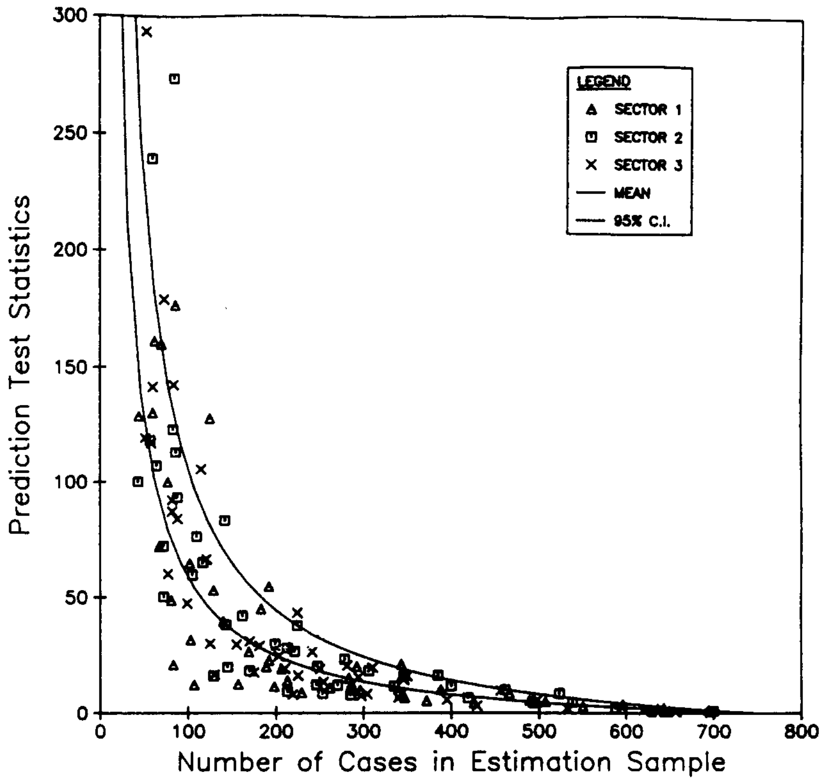


Figure 4. Scattergram of prediction test statistics with estimation sample size.

Finally, these results are modified for populations of varying size but otherwise identical characteristics by defining the likelihood ratio statistic per individual in the population (obviously, population data from which to compute the population model likelihood ratio statistic are not generally available; however, LRS* can be estimated by dividing sample likelihood ratio statistics by sample size):

$$LRS_* = LRS/N_p \tag{25}$$

to obtain

$$E(PI) = 1 - \{[(N_p - N_s)/N_s] \cdot DF/(N_p \times LRS_*)\} \tag{26}$$

which, when population size is much greater than sample size, is

$$E(PI) = 1 - (1/N_s)(DF/LRS_*) \tag{27}$$

The expected values of the prediction index for the three Washington sectors for different sample sizes are given in Table 5. The proportion of information provided by models estimated on samples of different sizes depends on the ability of the model to provide information about the behavior under study, as represented by the value of the likelihood ratio statistic per person. Sectors in which estimated models provide a higher level of information require smaller samples to achieve a specified level of relative accuracy. The results reported in Table 5 indicate that samples of 500 observations will provide 90 percent of the potential model information in each of the three Washington sectors.

Table 5. Expected value of prediction index (large population cases).

	Expected Value of Prediction Index		
Sample Size	Sector 1	Sector 2	Sector 3
50	0.62	0.21	0.34
100	0.81	0.61	0.67
200	0.90	0.80	0.83
300	0.94	0.87	0.89
500	0.96	0.92	0.93
1,000	0.98	0.96	0.97

Conclusions

The theoretical relationship between sample size and population description accuracy in the form of the prediction test statistic is developed in Equations 14-18. The empirical results reported in Table 4 are consistent with those relationships. The prediction index provides a somewhat more intuitive description of the relationship between sample size and descriptive accuracy. This relationship suggests that, in terms of descriptive accuracy alone, disaggregate samples of approximately 500 observations may be adequate. It is important to recognize the distinction between the ability to describe parent population choice behavior and prediction of behavior under different travel service conditions, which is most closely related to the precision of estimated parameters discussed previously.

EFFECT OF SAMPLE SIZE ON TRANSFERABILITY

Statistical Measure and General Expectation

Model transferability at the disaggregate level can be measured by indices formulated as a function of the difference in log-likelihood for the application sample of a transferred model [LL_i (β_j)] and the corresponding log-likelihood of a model estimated on that sample [(LL_i (β_i)]. The transfer test statistic formulated by Koppelman and Wilmot (11) is used to evaluate the transferability of disaggregate models. The transfer test statistic

$$TTS_i(\beta_j) = -2\,[LL_i(\beta_j) - LL_i(\beta_i)] \quad (28)$$

is chi-squared distributed with degrees of freedom equal to the number of model parameters under the assumption of fixed values of parameters for the transferred model. The smaller this statistic is, the more applicable is the transferred model to the application population.

This transfer test statistic is used to evaluate each of the sample-based models for transfer prediction of the population in each of the other sectors. Based on the results given previously, it is expected that the sample size of estimation subsamples will affect both the prediction accuracy and variability of a transferred model in the application context, according to a function that has a term of $(N_p - N_s)/N_s$ to reflect the sampling effect in the estimation context. It is also expected that there is a constant term in the transfer test statistic that reflects the real difference between the population of estimation and the population of prediction. These relationships are developed in the following subsection.

Relation Between Transfer Test Statistics and Estimation Sample Size

The transfer test statistic of a subsample-based model (j), predicted on an alternative population, is defined as

$$TTS_{ij} = -2\,[LL_i(\beta_j) - LL_i(\beta_i^*)] \quad (29a)$$

which is approximately (5),

$$TTS_{ij} = (\beta^* - \beta_j)'\,\Sigma_{p_i}^{-1}\,(\beta_i^* - \beta_j) \quad (29b)$$

Let TTS_{ij}^* represent the transfer test statistic of the population-based model,

$$TTS_{ij}^* = (\beta_i^* - \beta_j^*)'\,\Sigma_{p_i}^{-1}\,(\beta_i^* - \beta_j^*) \quad (30)$$

which is nonstochastic, and assume that $N_{p_i} \times \Sigma_{p_i} = N_{p_j} \times \Sigma_{p_j}$ (i.e., the underlying model parameter covariance matrices are equivalent) for the two populations; thus (21),

$$TTS_{ij} = TTS_{ij}^* + (N_{p_i}/N_{p_j}) \cdot [(N_{p_j} - N_{s_j})/N_{s_j}]\;[2\,(\beta_i^* - \beta_j^*)'\,\Sigma_{z_j}^{-1}\,z_j + z_j'\,\Sigma_{z_j}^{-1}\,z_j] \quad (31)$$

That is, the transfer test statistic for a model estimated on a sample from population j and used to predict population i is composed of a deterministic term that describes the difference between the two populations and a random variate composed of two terms. The first term, which is random because of the inclusion of z_j, is normally distributed with mean zero and variance-covariance matrix Σ_{z_j}. The second term, which is random because of the inclusion of $z_j' \Sigma_{z_j}^{-1} z_j$, is a chi-square variate with DF degrees of freedom. Thus TTS_{ij} is the sum of a fixed term, a normal variate and a chi-square variate. (Note that this breakdown of TTS_{ij} ignores the interaction between terms and the constraint required to ensure that TTS_{ij} is nonnegative.) The expected value and variance of TTS_{ij} are

$$E(TTS) = TTS_{ij}^* + (N_{p_i}/N_{p_j}) \times [(N_{p_j} - N_{s_j})/N_{s_j}] \cdot DF \quad (32a)$$

and

$$V(TTS) = 4\,(N_{p_i}/N_{p_j}) \times \{[(N_{p_j} - N_{s_j})/N_{s_j}]\;TTS_{ij}^*\} + 2\,(N_{p_i}/N_{p_j}) \times [(N_{p_j} - N_{s_j})/N_{s_j}] \times DF \quad (32b)$$

Thus both the mean and variance of the transfer test statistic increase with the difference between the two populations involved in the transfer process and decrease with the sample size of the estimation data set so that increased estimation sample size improves model transferability.

Empirical Analysis

The relationship between the transfer test statistic and sample size is examined empirically. The values of the population transfer test statistic (TTS*) are given in the following table:

Estimation Sector	Transfer Test Statistics by Prediction Sectors 1	2	3
1	--	67.2	72.2
2	48.6	--	29.0
3	52.6	27.2	--

A scattergram of the transfer test statistic is plotted with varying estimation sample size for transfers from sectors 1 and 3 to sector 2. This scattergram (Figure 5) can be used to examine the expected values and variances that were derived. In this figure, the expected value of the transfer test statistic, as defined by Equation 32a, is included. As expected, these lines fit the data in the respective transfer conditions satisfactorily. It was also observed that the variance of the transfer test statistic decreases as the estimation sample size increases, as suggested by Equation 32b. Further, it was noted that the sample values of TTS for transfers from sector 3 with the smaller value of TTS* have both lower mean and variance than the transfer from sector 1.

Conclusions

The expected relationship between sample size and transfer prediction accuracy is confirmed by the analytic decomposition of the transfer test statistic into a deterministic component that is independent of sample size and a stochastic component, the distribution of which is related to sample size for any given pair of populations. Empirical transferability tests are consistent with these analytically formulated relationships.

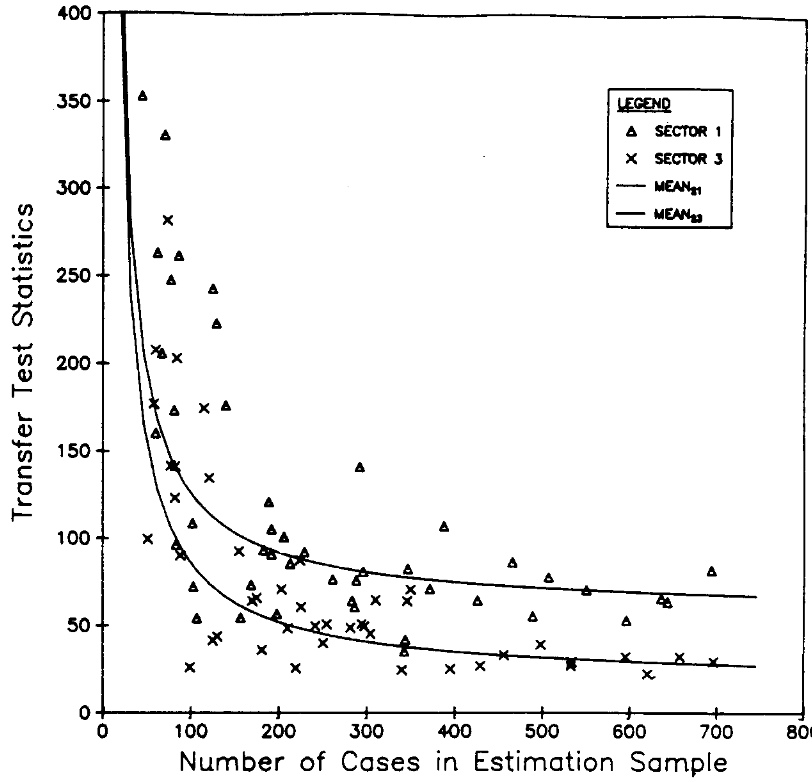
Increases in sample size cannot be used to offset real differences in the behavior of two populations reflected in TTS*. However, they can reduce the stochastic component. Additional analysis may be useful to clarify these relationships, but the empirical results suggest that samples in excess of 500 observations may be necessary to obtain transfer predictive accuracy that is close to that which might be obtained by a population-based model.

SUMMARY OF CONCLUSIONS

The conclusions reported in the preceding sections are summarized as follows.

1. Increased size of estimation samples leads to (a) parameter estimates that are likely to be closer to the true population parameters, (b) smaller standard errors of such parameter estimates, and (c) more accurate prediction of population choice behavior.
2. The sample size required to obtain choice model parameter estimates that are reasonably close

Figure 5. Scattergram of transfer test statistics predicted on sector 2.



to the true population parameters appears to be substantially larger than the sample sizes commonly prescribed for the estimation of disaggregate choice models.

3. The sample size required to obtain a model that accurately replicates parent population choice behavior appears to be somewhat smaller than that required to obtain accurate parameter estimates and accurate prediction under changed transportation service coordination.

4. Model transferability is a function of both the estimation sample size and the difference between the populations involved in the model transfer. Increasing estimation sample size has a positive effect on transferability at a decreasing rate. When the difference between two populations is large, it is expected that there will be large and highly variable transfer errors.

5. The required sample size needed to obtain a desired level of parameter estimation or prediction accuracy can be determined from pilot sample model estimation.

Overall, these results suggest the need to use data samples on the order of 1,000 to 2,000 observations rather than 500 observations as formerly believed. Although some reduction in sample size may be feasible when optimal sample stratifications are used (12, and paper by Sheffi and Tarem elsewhere in this Record), it is unlikely that samples as small as 500 observations can be adequate for model estimations.

Obviously, the importance of this issue suggests that additional research be undertaken to obtain further analysis of sample size requirements for models of different travel choices in different contexts. Further, transportation planners must formulate judgments about the desired precision of estimated model parameters and model prediction.

Appendix: Derivation of Sample Population Covariance Matrix

The population (full sample) estimation covariance matrix is the negative inverse of the Hessian (4) or

$$V_p = \left[\sum_{t \subset p} \sum_i (X_{it} - \bar{X}_t)' P_{it} (1 - P_{it}) (X_{it} - \bar{X}_t)\right]^{-1} \quad \text{(A1)}$$

where

- V = covariance matrix,
- Σ = summation,
- x_{it} = variable vector of alternative i for individual t,
- $\bar{X}_t$ = probability weighted average of x_{it},
- P = choice probability of alternative i for individual t,
- p = population,
- t = individual, and
- i = alternative.

Similarly, the sample estimation covariance matrix is

$$V_s = \left[\sum_{t \subset s} \sum_i (X_{it} - \bar{X}_t)' P_{it} (1 - P_{it}) (X_{it} - \bar{X}_t)\right]^{-1} \quad \text{(A2)}$$

where s is the sample indicator.

Finally, the covariance matrix between the population and sample estimates is given by

$$V_{sp} = \left[\sum_{t \subset s,p} \sum_i (X_{it} - \bar{X}_t)' P_{it} (1 - P_{it}) (X_{it} - \bar{X}_{it})\right]^{-1} \quad \text{(A3)}$$

where sp indicates the covariance matrix between population and sample estimations, and t C s,p implies summation over observations included in both the sample and the full population.

In this case, where the population includes all sample elements, the summation over s,p is equivalent to the summation over s and $V_{sp} = V_s$ or, by using the notation in the body of the paper, $\Sigma_{sp} = \Sigma_s$.

ACKNOWLEDGMENT

We acknowledge the support of the Office of University Research, U.S. Department of Transportation, for support of this work under a contract to investigate "Conditions for Effective Transferability of Disaggregate Choice Models", and James Ryan of UMTA, who as contract monitor has provided substantial stimulation to this research effort.

We also acknowledge earlier analysis undertaken by Joseph N. Prashker of Technion University in collaboration with Koppelman, and valuable discussions with Eric I. Pas of Duke University. Finally, we thank our colleagues at the Northwestern University Transportation Center for intellectual and analytic support in the research. These colleagues include Chester G. Wilmot, Peter Kuah, and Albert Lunde.

REFERENCES

1. C.F. Manski. The Analysis of Qualitative Choice. Massachusetts Institute of Technology, Cambridge, Ph.D. dissertation, 1973.
2. C. Chu. Structural Issues and Sources of Bias in Residential Location and Travel Mode Choice Models. Department of Civil Engineering, Northwestern Univ., Evanston, Ill., Ph.D. dissertation, 1982.
3. H. Theil. Principles of Econometrics. Wiley, New York, 1971.
4. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973.
5. C. Daganzo. Multinomial Probit. Academic Press, New York, 1979.
6. G. Maddala. Econometrics. McGraw-Hill, New York, 1977.
7. C. Rao. Linear Statistical Inference and Its Application. Wiley, New York, 1965.
8. F.S. Koppelman. Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models. TRB, Transportation Research Record 592, 1976, pp. 17-23.
9. T.J. Atherton and M.E. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. TRB, Transportation Research Record 610, 1976, pp. 12-18.
10. D. McFadden. Properties of Multinomial Logit Model. Univ. of California, Berkeley, Working Paper 7617, 1976.
11. F.S. Koppelman and C.G. Wilmot. Transferability Analysis of Disaggregate Choice Models. TRB, Transportation Research Record 895, 1982, pp. 18-24.
12. C. Daganzo. Optimal Sampling Strategies for Statistical Models with Discrete Dependent Variables. Transportation Science, Vol. 14, No. 4, 1980.
13. M.F. Richards. Application of Disaggregate Techniques to Calibrate a Trip Distribution and Modal-Split Model. TRB, Transportation Research Record 592, 1976, pp. 29-34.
14. M. Richards and M. Ben-Akiva. A Disaggregate Travel Demand Model. Saxon House and Lexington Books, Lexington, Mass., 1975.
15. A.P. Talvitie and D. Kirschner. Specification, Transferability, and the Effect of Data Outliers in Modeling the Choice of Mode in Urban Travel. Transportation, Vol. 7, No. 3, 1978.
16. K.L. Sobel. Travel Demand Forecasting by Using the Nested Multinomial Logit Model. TRB, Transportation Research Record 775, 1980, pp. 48-55.
17. D. McFadden. The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation. In Emergency Transportation Planning Methods (W.F. Brown et al., eds.), U.S. Department of Transportation, Rept. DOT-RSPA-DPB-50-78-2, 1978.
18. K. Train. A Validation Test of a Disaggregate Mode Choice Model. Transportation Research, Vol. 12, 1978, pp. 167-174.
19. C.F. Manski and L. Sherman. An Empirical Analysis of Household Choice Among Motor Vehicles. Transportation Research, Vol. 14A, No. 5-6, Oct.-Dec. 1980.
20. T.A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
21. F. Koppelman and C. Chu. Sample Size Effects on Parameter Stability, Prediction Precision, and Transferability of Multinomial Logit Models. In Identification of Conditions for the Effective Transferability of Disaggregate Choice Models, Transportation Center, Northwestern Univ., Evanston, Ill., 1982.

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

Mobility Enterprise: One Year Later

MICHAEL J. DOHERTY AND F.T. SPARROW

A mobility enterprise is a new transportation concept aimed at increasing the productivity of the automobile through use of mini or micro automobiles in conjunction with a shared fleet of intermediate and full-sized vehicles. The main objective of the enterprise is to provide a better matching of vehicle attributes to trip requirements and still maintain the personal freedom that appears to be so highly valued by the American driver. Although this concept was presented in detail in an earlier TRB Record (TRR 882), a view of the progress that has been made in taking the mobility enterprise from an innovative concept to an actual experiment is presented in this paper. The majority of the information deals with methods for observing consumer attitudes, designing the actual mobility enterprise, and measuring mini and micro automobile performance.

In January 1982 the Automotive Transportation Center at Purdue University unveiled an innovative transportation concept called the mobility enterprise (1). Briefly stated, the research examined the effects of mini and micro class automobiles and shared-vehicle fleets on the overall productivity of the personal automobile. This paper is designed to provide an update of the progress made during the last year and to discuss the experimental design and preliminary findings.

After years of promoting public transit and carpooling to conserve energy, it appears that the average consumer still prefers the convenience of the personal automobile. At the same time, although automobile efficiency (fuel economy) has undergone significant improvement, automobile productivity has remained disturbingly low (2,3). The concept presented here for improving productivity is based on a better matching of the trip requirements of an individual to the characteristics of the vehicle. Three interrelated features of a mobility enterprise--retained autonomy, easy access to an expanded fleet, and reduced expenditures--are the inferred keys to its success. An enterprise member's minimum attribute vehicle (a mini or micro automobile in these experiments) provides him, by definition, with the most economical means of accomplishing his most frequent trips. When a member's mini or micro automobile is inappropriate for a desired trip, he must seek access to an appropriate vehicle from the shared fleet. This process may involve delays, some advanced planning, paperwork, and out-of-pocket costs, depending on the procedures of the enterprise. A general description of the mobility enterprise that has been set up at Purdue University is as follows.

1. The following items are included in a set monthly fee: (a) an individually garaged mini or micro class vehicle that will satisfy most commuting and around-town driving, (b) access to a shared fleet of intermediate and full-sized vehicles for trips that the mini or micro vehicle would be unsuitable, (c) all insurance costs, (d) all maintenance costs, (e) all registration and licensing costs, and (f) taxes.
2. Gasoline costs are not covered in the monthly fee.
3. Cost per participating household for experiments is $165 per month.

The concept of a mobility enterprise requires careful examination of several behavioral parameters of the American as a driver. Judging from the underutilization of public transit systems and ridesharing programs, it appears that personal freedom and independence are highly valued attributes. If it is imperative that this independence be preserved, a key step in the design of proposed experiments must be an inventory of the current patterns of the U.S. driver and the use of his personal vehicle. The shape of the enterprise must come as close as possible to satisfying travel demands, with as little inconvenience as possible. However, because there may be some inconvenience (changes in travel behavior), it is important to gauge the value drivers place on the quality of travel provided by the shared fleet available through the enterprise. In other words, what would be the trade-offs between the current condition of automobile ownership and participation in a mobility enterprise?

Two key tools that have been used to acquire data pertaining to consumer acceptance and current travel behavior are the focus-group interview and a survey instrument (questionnaire). In addition to consumer and travel-behavior studies, a microprocessor-based data acquisition system, under development at Purdue University, will measure the stress on these small automotive engines when subjected to real-world missions. Such a system is necessary to determine the feasibility of using mini or micro automobiles for personal transportation in the United States.

FOCUS-GROUP INTERVIEWS

Focus-group interviews are predicated on the assumption that the mobility enterprise will be better understood and more efficiently designed when there are more data on how potential users, supporters, and detractors define its advantages and disadvantages and its significant and modifiable attributes (4). The content of each interview was analyzed for recurring themes. The attributes that account for decisions to join or not join the enterprise were schematized, and questions measuring the character and quality of these attributes were developed for the larger general survey instrument.

Focus-group interviews began in West Lafayette, Indiana, in March 1982. The length of the focus-group interviews varied from 1 to 1.5 hr. There were seven focus groups: one group of Purdue University faculty and staff, one group of Purdue University faculty and staff couples, one group of Purdue University faculty and staff as new car intenders (intention to buy a new car within 2 months), two groups of college students, and two groups of teenagers (one consisting of all male and one consisting of all female). A total of 62 individuals participated.

Data from the focus-group interviews were analyzed for issues raised, opinions expressed, and experiences reported and were then examined for recurrent significant themes. The focus-group interviews and subsequent analyses were based on the assumption that the study of consumer attitudes and interaction and the emphasis on analysis of themes should provide insight into the consumer decision-making process of automobile ownership, mini and micro vehicles, and the mobility enterprise (5,6). This in turn should improve the capability for planning and developing the mobility enterprise. The focus group interviews were divided into four content areas: (a) vehicle ownership and use, (b) the expense of owning and operating cars, (c) the mini or micro automobile, and (d) the mobility enterprise. The major findings in each of these content areas were as follows.

1. Vehicle ownership: Increasing costs are creating compromises concerning style; i.e., when purchasing a vehicle, people are settling for less car than they originally had planned to buy. Also, there was an overwhelming attitude that automobiles are synonomous with personal mobility and freedom.
2. Vehicle expenses associated with vehicle ownership: All groups knew that owning a car was expensive, but when probed they were relatively unaware of the actual cost. There was a strong belief that ownership costs would not get too high. Virtually all groups believed that some technological breakthrough would occur to keep automobiles affordable.
3. Mini and micro automobiles: Price (quoted as between $3,000 and $4,000) makes these cars attractive as a second car. Also, safety was dismissed as a realistic issue because the participants generally perceived drivers to be more important than automobiles with respect to safety.
4. Mobility enterprise: Generally, the shared-fleet concept was not well received, as most groups believed it was an infringement on their freedom of mobility; thus they tended to dwell on the negative aspects of sharing. But, continuous maintenance was almost universally viewed as the major point in favor of the mobility enterprise. Finally, the ability of membership for a trial period of time was seen as crucial.

Because this study uses a small population and is not truly representative, and because the findings are qualitative and subject to biases, the study should be viewed as exploratory in nature, thus making generalizations difficult. Nevertheless, it is anticipated that the validity of issues raised will be considerably strengthened as the hypotheses derived from the focus-group interviews are further explored by forthcoming surveys. Such has already been the case in two other papers (7,8).

SURVEY INSTRUMENT

The local survey was intended to help gather data pertaining to the acceptability of the mobility enterprise concept to a representative sample of households in the area where the first experiments were to be run. It also acted as a tool to compile an inventory of current vehicle use patterns in the sample area.

The Social Research Institute of Purdue University conducted the local survey. The sample size was 300 households. Tippecanoe County is a designated standard metropolitan statistical area (SMSA), and 80 percent of the sample was drawn from the urbanized area and 20 percent from the nonurbanized area. Within the urbanized area, four strata were selected based on socioeconomic status (SES): high, medium, low, plus a fourth category containing small blocks (four dwelling units or fewer). Three strata were selected from the nonurbanized area based on SES (high, medium, and low). The survey instrument was administered by personal interviews of 30 to 45 min each. Two additional subgroups of 30 households each were interviewed, which represented retirement communities and condominiums. General demographic information that characterize the sample population is given in Table 1. The attitudes of the respondents toward the mobility enterprise as a transportation mode are given in Table 2.

When the sample is broken down into two subgroups, one consisting of those interested in joining and the other consisting of those not interested (only two respondents were undecided), several intriguing differences with respect to age, automobile purchasing intentions, and the acceptability of small cars for everyday use are noted (see Table 3). In general, those interested in joining a mobility enterprise are younger, closer to making car purchase decisions, and find small cars more acceptable than those not interested in joining. Two other significant observations are that (a) no retirees were interested in joining, and (b) those who were interested in joining believed they would need to use a shared vehicle, on average, approximately 45 percent more often than those who were not interested (67 days per year versus 46 days per year).

The results presented here are merely preliminary findings. A more detailed report analyzing the local survey will be forthcoming. In addition, a national survey about the mobility enterprise con-

Table 1. General demographics of transportation survey.

Item	No. of Respondents
Total	360
Male	173
Female	187
Age (years)	
18-25	79
26-40	124
41-60	73
>61	83
Highest level of education	
Less than 12th grade	46
High school education	123
Some postsecondary	86
Four or more years postsecondary	102
Household income	
<$5,000	38
$5,000-$14,999	80
$15,000-$24,999	93
$25,000-$34,999	74
>$35,000	62

Table 2. Preliminary survey results from questionnaire.

Question	Positive Response (%)
Do you think the mobility enterprise is practical?	65.3
Do you think the mobility enterprise is complicated?	20.3
Would the mobility enterprise work for your household?	23.9
Would it be important to see others join the mobility enterprise before you would?	50.3
Would you be interested in joining the mobility enterprise?	14.3
Would you be willing to join the mobility enterprise for a trial period?	24.4
For your household, would owning your own car be better than being a member of the mobility enterprise?	88.3

Note: 360 respondents were asked these questions.

Table 3. Preliminary survey results.

	Willing to Join a Mobility Enterprise?	
Item	Yes (n = 51)	No (n = 309)
Mean age of respondent	31.6	44.6
Planning to purchase a vehicle within the next year (%)	37.2	11.0
Planning to purchase a used car within the next year (%)	62.8	30.7
A mini or micro automobile is acceptable as a vehicle for everyday use (%)	76.5	63.1
A subcompact is acceptable as a vehicle for everyday use (%)	96.1	72.5
It would be acceptable sharing a car with several other people (%)	88.2	63.6

cept will be conducted by J.D. Power & Associates of West Lake Village, California.

TRIP DIARIES

Although focus-group interviews and transportation surveys are helpful in identifying the inclination toward acceptance of a mobility enterprise concept and some of its critical attributes, another more direct measure of acceptance based on actual behavior was also needed. For this reason, the collection of trip diaries from potential experimental subjects began in August 1982. Thus nearly 6 months of actual travel behavior was collected before the initial experiments.

Because participation in the mobility enterprise involves changes in vehicle use, it is important to know whether the enterprise fits into the current travel behavior of the participants. Because the travel patterns of the participants both as a group and as a household are known up to this point, this data should prove to be extremely valuable. Significant changes in travel patterns caused by the accommodation of the operating system and restrictions of the mobility enterprise are detected with these data. A meaningful control group of trip diary participants who will not be enterprise members is being maintained for the duration of the experiments.

Trip diary results to date have revealed a remarkable degree of consistency for the test population from week to week. A summary of trip types and mileage for the first 12 weeks of the study is given in Table 4. The trip occupancy pattern for the population for the first 12 weeks is given in Table 5.

Table 4. Pretest trip diary results of trip type and mileage.

Trip Type	Trips per Week	Mean Mileage per Trip (one way)
Shopping (grocery and nongrocery)	2.66	4.97
Commuting (work or school)	5.19	7.64
Social-recreation	3.49	14.65
Personal business (errands, passenger ferry, and so on)	5.75	5.14
Return home	9.05	9.24

Note: 65.36 percent were multipurpose trips. Results cover a 12-week period.

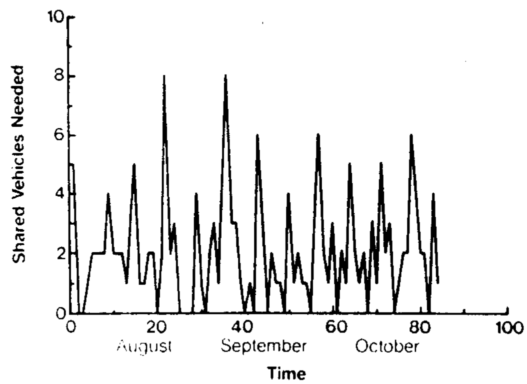
Table 5. Pretest trip diary results of trip occupancy.

	Occupancy per Trip (%) by No. of Occupants	
Trip Type	<2	>3
Shopping (grocery and nongrocery)	89.7	10.3
Commuting (work or school)	99.0	1.0
Social-recreation	79.1	20.9
Personal business (errands, passenger ferry, and so on)	91.2	8.8
All trips	91.1	8.9

Note: Results cover a 12-week period.

A final purpose for which the trip diary data may be useful is in the design of the shared fleet. One of the most critical design characteristics of a mobility enterprise is the size of the shared fleet for a given size of enterprise. How many cars would be too many? How many would be too few? For the purposes of the experiments currently being conducted, assume that a shared vehicle is required for a trip greater than 30 miles (one way) or transporting four or more occupants. By using these criteria, the expected use of shared vehicles for the first 12 weeks of the study is shown in Figure 1. Extrapolation of these data for a 20-member enterprise, run under the restrictions assumed here, appears to indicate that the enterprise is most efficient if it owns two vehicles in its shared fleet and uses an outside vendor for those times when additional vehicles would be needed. However, these questions must be more thoroughly examined during the actual experiment.

Figure 1. Hypothetical shared-fleet use.



Note: Data give expected need for shared vehicle for first 12 weeks of trip diary studies. These data are based on trip diary results from that time period for a hypothetical enterprise of 24 member households.

TECHNICAL DATA ACQUISITION SYSTEM

All mini and micro vehicles in the experiment are to be equipped with a data acquisition system (DAS) to collect information on the performance characteristics of these vehicles. The DAS has a standard configuration, with sensors mounted on the power plant that pass signals to the computer. The processor passes the data or processes it and sends the information to a digital recording device. Design specifications were developed to accommodate the harsh automotive environment. This work is not new; it is an extension of the basic work on internal combustion vehicles already performed for instrumentation of electric vehicles at Purdue University (9).

A mission use pattern will be developed through a series of plots, such as vehicle speed histograms (percentage of time spent in various velocity ranges), trip length histograms, number of trips per day versus day of the week, and so forth.

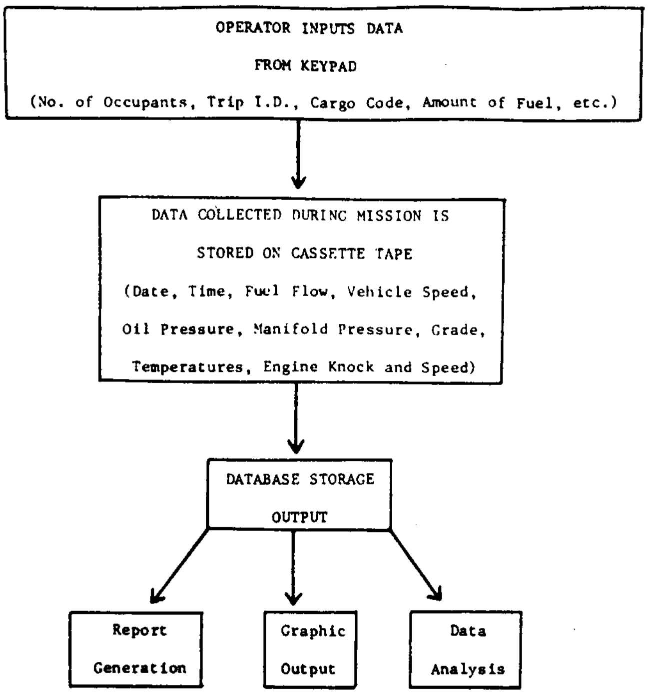
A mission severity index will be used to calculate the energy required for acceleration, constant speed, and idle periods. Data from engine fuel-consumption maps will also be used to characterize fuel consumption during a mission. A general schematic of this system is shown in Figure 2.

INITIAL EXPERIMENTS

The first mobility enterprise experiment became a reality on January 22, 1983. The enterprise initially consisted of seven participating households.

The basic service included an individually garaged mini or micro automobile and access to a shared fleet of one vehicle. Because of insurance restrictions resulting from the lack of safety data on the mini and micro automobiles, all such vehicles are prohibited from use on Interstate highways. All operating costs (excluding gasoline) are

Figure 2. Flowchart of DAS.



included in the monthly fee. In addition, each member receives approximately 10 coupons for use of the shared fleet. The coupons have a cash value of approximately $7.00. The basic rate for shared-fleet use varies according to peak or off-peak periods. The coupon exchange rate for shared vehicles is two coupons per weekday and three coupons per weekend day. Coupons may be accumulated for use at a later time, traded among members, or turned in at the end of the month for a credit toward their next month's bill. Maintenance of all vehicles and shared-fleet operations is administered through the Purdue University Transportation Services Department.

Trip diaries are being maintained for all vehicles in the mobility enterprise as well as in a control group of nonenterprise members. In April 1983 the mini and micro vehicles were equipped with the on-board DAS that measures various factors in engine performance. All test subjects are being closely monitored throughout the experiments.

SUMMARY

The purpose of this paper is to describe the progress that has been made in the past year in bringing the mobility enterprise from a hypothetical concept to a set of actual experiments designed to test its viability as a transportation mode. Many of the results presented here deal with research activities that must precede the actual experiments. The research emphasis to date has been in the area of consumer acceptance of the mobility enterprise concept, recruitment of experimental subjects, operational design of the Purdue University experiments, and methods for measuring mini and micro vehicle performance under U.S. driving conditions.

Thus far the data are encouraging because more than 20 percent of the random sample would be willing to try a mobility enterprise for a trial period and more than 10 percent said they would be willing to join such an organization. The data from the trip diaries appear to indicate that a mobility enterprise operation could satisfy a significant portion of the travel demands of the potential participants. This is particularly noteworthy because the data from the trip diary include August (a high vacation month) and September (Labor Day weekend). The focus-group interviews imply that there is no aversion to mini or micro automobiles (also indicated in the survey) and that continuous maintenance is a significant factor in favor of the mobility enterprise concept. The survey and focus groups have also indicated that the mobility enterprise, to be successful, must come close to the current state of automobile ownership. Other work currently under way deals with determining optimal shared-fleet size (10), which is crucial to the ultimate economic success of such a venture.

In addition to the data presented here, a great deal of the first year's effort has dealt with logistical considerations, such as obtaining waivers for importing the mini and micro automobiles, arranging insurance coverage and maintenance delivery systems, procuring vehicles for the shared fleet, and calculating costs to the participants. Although such efforts yield no experimental data, they are both time consuming and crucial to the performance of the actual experiments. Thus, because of the work described in this paper, the Purdue University mobility enterprise experiments were able to begin in January 1983.

ACKNOWLEDGMENT

This work has truly been of an interdisciplinary nature, and special thanks are due to Harry Potter (Department of Sociology), Richard Feinberg and Thelma Snuggs (Department of Consumer Sciences), Jon D. Fricker (Department of Civil Engineering), and Patrick McCarthy (Department of Economics) for all of their interest and active participation in the early and tedious data-gathering phases of the study. A special thanks is also due to Lori Berthold and Pat Sanders, without whose diligent work and skills this paper would not have been possible.

Thanks is also due to the three research teams that have assisted in the formulation of the focus-group interviews: the Purdue University team (headed by Richard Feinberg and Thelma Snuggs of the Department of Consumer Sciences), Avis Rent-A-Car of Garden City, New York (headed by Al Dold, executive vice president for marketing), and J.D. Power & Associates, consulting specialists in automobile marketing (headed by John Hemphill, executive vice president).

REFERENCES

1. F.T. Sparrow, J.D. Fricker, and R.K. Whitford. The Mobility Enterprise: Improving Automobile Productivity. TRB, Transportation Research Record 882, 1982, pp. 19-24.
2. The U.S. Automobile Industry, 1980. Report to the President, Secretary of the U.S. Department of Transportation, Jan. 1981.
3. A. Altschuler, J.P. Womack, and J.R. Pucher. The Urban Transportation System: Politics and Policy Innovation. MIT Press, Cambridge, Mass., 1979.
4. J. Higgenbottom and K. Cox. Focus Group Interviews. American Marketing Association, Chicago, 1979.
5. E. Fern. Why Do Focus Groups Work...A Review and Integration of Small Group Process Theories. In Advances in Consumer Research (S. Mitchell, ed.), Univ. of Michigan Press, Ann Arbor (in preparation).

6. R.A. Feinberg and T.L. Snuggs. The Mobility Enterprise: Focus Group Interviews. Automotive Transportation Center, Purdue Univ., West Lafayette, Ind., July 1982.
7. R.A. Feinberg and T.L. Snuggs. The Mobility Enterprise as Innovation. Automotive Transportation Center, Purdue Univ., West Lafayette, Ind., May 1982.
8. R.A. Feinberg and T.L. Snuggs. Psychological Pricing and the Mobility Enterprise. Automotive Transportation Center, Purdue Univ., West Lafayette, Ind., June 1982.
9. W.G. Blom et al. High Speed Digital Data Acquisition System for Electric Vehicles. Proc., Electric Vehicle Council Symposium VI, EVC Paper 8120, Oct. 1981.
10. J.D. Fricker and J.K. Cochran. A Queuing Demand Model to Optimize Economic Policymaking for a Shared Vehicle Enterprise. Presented at the Operation Research Society of America (ORSA) and The Institute of Management Science (TIMS) Conference, San Diego, Oct. 25, 1982.

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

Person-Category Trip-Generation Model

JANUSZ SUPERNAK, ANTTI TALVITIE, AND ANTHONY DeJOHN

A person-category model of trip generation is presented as an alternative to household-based trip-generation models. In this model a homogeneous group of persons is used as an analysis unit. The final description of the person categories is not arbitrary but results from the multistage, multivariate analysis of many potentially significant variables. The variables age, employment status, and automobile availability were found to be the most significant descriptors of a person's mobility. The final version of the model is based on eight person categories. Both theoretical discussion and empirical findings favor the proposed version of the person-category model over household-based models because it is more practical at the forecast stage, requires significantly less data, has better behavioral background, and is more compatible with the entire system of individually oriented travel-demand models.

The development and evaluation of a person-category trip-generation model as an alternative to household-based models are discussed in this paper. The individual-level approach was chosen for the following reasons. First, a person-level trip-generation model is compatible with other components of the four-step travel-demand model system that is based on tripmakers rather than on households. Second, it is extremely difficult to devise a household-based cross-classification scheme that uses all important variables and has a manageable number of classes [e.g., a British household cross-classification model (1) has 108 categories]. Predicting representations in so many classes is difficult.

Third, the sample size for the person-category model can be much smaller (10 to 40 times) than for the household-category model. Fourth, demographic changes can be more easily accounted for in the person- rather than household-category model, and some demographic variables (such as age) are virtually nondefinable for households. Finally, person categories are easier to forecast to the future than the household categories, which require forecasts about household formation and family size. With the person categories these tasks are altogether avoided. More importantly, because the bulk of the trips will be made by people older than 18 years of age, the task of predicting the tripmaking population 15 to 20 years ahead is much easier.

There are of course some limitations that a person-category model may have. Foremost among these is the difficulty of introducing household-interaction effects and household money costs and money budgets into the model. On the other hand, it is not clear how vital these considerations are and how they can effectively be introduced even in a household-category model. The methodology of the development and testing of the person-category model was based on previous work from Europe (2-6), where the person level of data aggregation was found to be successful for travel-demand analysis.

DATA AND DEFINITIONS

Data

The data used in preparing this paper were from the Baltimore home interview survey conducted in 1977 by the FHWA and from Minneapolis-St. Paul home interview data collected in 1970. Before the analyses, data were superfically cleaned. Workday records were separated from weekend-day records, and some persons were excluded from the original sample. For example, if in the original file a significant inconsistency was found (e.g., number of cars in the family = 7 and number of drivers = 0), the person was excluded. Outliers were also excluded. If the number of trips done by a person was greater than 10 and if total time spent on traveling during the day exceeded 150 min, then this person was suspected to be a professional driver (or similar category) and was excluded from the sample.

Definitions

The following definitions are used in the analyses:

- N_i = trip rate, that is, the daily number of one-way trips made by (average) person in category i; and N_{qi} = trip rate to purpose q in category i;
- T_i = daily travel time; that is, the time (in minutes) spent by (average) person in category i on traveling during the day;
- Y_j = total number of trips made anywhere by the inhabitants of zone j (all categories together);
- L_j = number of zone j inhabitants; and
- α_{ij} = percentage of inhabitants of zone j belonging to category i.

Thus the following basic relationship is given:

$$Y_j = L_j \sum_i \alpha_{ij} N_i \quad (1)$$

The method of calculating zonal productions (P_j) and attractions (A_j) is not presented in this paper. This method is briefly presented in Supernak (5).

In analyzing and calculating trip rates, trips are divided into

1. Home-based (HB) trips if origin (HBO) or destination (HBD) of the trip is the place of residence of the traveler, and
2. Non-home-based (NHB) trips if neither origin nor destination of the trip is at home.

Trips are further divided by trip purpose (q) as follows: work (W), education (E), shopping (S), personal business (Pb), and social-recreational and other purposes (Sr). This trip-purpose classification applies to both HB and NHB trips. Work and education trips are called obligatory trips, and all other trips are called discretionary trips. The traditional description of the trip links (instead of sojourns of trips) was chosen because it clearly relates the number of outside-the-home activities to the number of trips made (6,7).

An example of trip rates for category i is given in Table 1. Fifteen-element vectors of partial trip rates N_{qi} (i.e., separated by purpose, direction, and base) may be derived from the data, as shown in Table 1; they served as the trip characteristic of category i.

Table 1. Example of trip rate characteristic N for category i.

	Obligatory		Discretionary			
Trip	W	E	S	Pb	Sr	Total[a]
HBO	0.86	0.02	0.10	0.21	0.05	1.33
HBD	0.86	0.05	0.21	0.19	0.02	1.33
NHB	0.02	0.05	0.14	0.14	0.07	0.43
Total[a]	$N_i^{obl} = 1.86$		$N_i^{disc} = 1.24$			$N_i = 3.10$

[a]Note that some columns will not total because of rounding.

ANALYSIS PROCEDURE

The model development was done in four stages:

Stage 1--(a) arbitrary choice of many variables, which are expected to be important for explaining differences in a person's mobility, and definition of plausible person categories by using these variables; and (b) preliminary analysis of trip rates (N_i) and trip times (T_i) to find which variables have the least explanatory power and can be excluded from the model;

Stage 2--(a) detailed analysis of trip characteristics to find variables that define similar categories for stage 3; variables that do not give substantial explanation of the data variance or variables that duplicate an explanation of other better variables are excluded; (b) proposal for the final trip-generation categories, the number of which should not exceed a certain practical maximum (for example, 10); and (c) analysis of dependency of trip rates between trip purposes [not reported in this paper, see Supernak et al. (8)];

Stage 3--(a) final trip-generation characteristics of each category, as determined in stage 2, are analyzed in detail; and (b) transferability of the results within different sections of Baltimore and to other cities is examined; and

Stage 4--comparison with household-based trip-generation model, as presented in detail in DeJohn (9).

The statistical methods used in the analyses are simple and straightforward. At all times these statistical methods are supplemented by visual analysis of data that try to find patterns in the data that a blind application of statistical methods may not find.

In stage 1 of the model development only a pairwise comparison of total trips rates is performed. The Z-statistic for the trip rates of two categories i and j, which are differentiated by the analyzed variable only, is computed and compared with the critical Z-value at the 0.01 level of significance.

In the remaining stages three additional measures supported by histograms and analyses of variance are used. These three measures are the correlation coefficient, slope (m), and intercept (b) of the regression $N_{qi} = b_{ij} + m_{ij}N_{qj}$.

The categories i and j may be treated as similar if (a) the correlation coefficient between vectors of the partial trip rates (i.e., trip rates by purpose and base) N_{qi} and N_{qj}, and (b) the parameters of the regression coefficients (m_{ij} - slope, b_{ij} - intercept), satisfy the following conditions:

$$r_{ij} > 0.900 \quad (2)$$

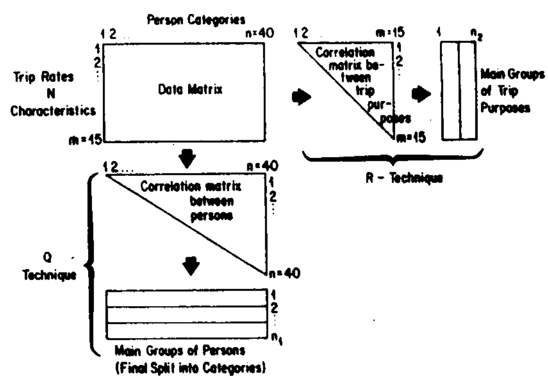
$$0.75 < m_{ij} < 1.25 \quad (3)$$

$$|b_{ij}| < 0.10 \quad (4)$$

These conditions are arbitrarily chosen and are quite demanding.

These three measures can be used to analyze the appropriate categories for both persons and trip purposes, as shown in Figure 1. The Q-type regression and correlation analysis is used for analyzing the best grouping of persons, and the R-type analysis is used for grouping trip purposes. These analyses are useful for both travel-demand analysis (3,4) as well as for nontransportation applications (10).

Figure 1. Q-type and R-type analysis of trip rates N_{qi}.



STAGE 1: CHOICE OF VARIABLES AND DEVELOPMENT OF CATEGORIES

For stage 1, the following variables (and strata) were used to form the categories.

1. Sex: The obvious choice of strata here is male and female.
2. Age: Age was used to describe the main activity at a given age (primary school pupils, high

school pupils, college students, employees, retired). Accordingly, the age groups used were 0 to 12, 12 to 18, 18 to 65, and older than 65. Age 40 is also used to divide the employable work force into two categories.

3. Car availability: In all known trip-generation models the variable car ownership was used and treated as a basic variable. Here a variable defined as car availability is used. The reason for this change comes directly from the general concept of the model. When using a traveler or a person as the analysis unit, car ownership of the family is not directly related to the car availability of different family members. Thus the following distinction was made (where N_c = number of cars in the household and N_d = number of drivers in the household). For a given person, car availability is (a) never available if $N_c = 0$ or $N_d = 0$ (person has no driving license) or (b) sometimes available if $N_c > 0$ and $(N_c/N_d) < 1$ $(N_d > 0)$ or (c) always available if $(N_c/N_d) > 1$.

4. Employment status: Status is divided by employed and not employed.

5. Income: Income is defined at the individual level rather than at the family level. Household income was converted to per capita income simply by dividing it with family size.

6. Race: The race variable (white versus nonwhite) was analyzed because of the significant percentage of nonwhite respondents in the Baltimore data set.

7. Employment types: Three strata are used--white collar, blue collar, and other.

8. Family type: Five family types were analyzed to understand how the family duties affected a person's tripmaking behavior. The strata of this variable were as follows: single person, childless couple, family with children younger than 5 years of age, family with children 5 to 12 years of age, and family with children older than 12 years of age.

These variables and strata resulted in the 100 categories shown in Figure 2. (Note that Figure 2 is read in the following way: each dot indicates which variable applies. For example, persons in category 24 are white, single, employed blue-collar males who have a car always or sometimes available and whose per capita income is between $1,500 and $4,000 per year; there are 11 such persons in the sample.) Note that in defining these categories many potentially important variables were included initially, and yet there was a desire to keep the number of categories reasonable (i.e., not to exceed 100). The eight variables could have produced 5,400 categories, whereas the sample size was only about 2,000. The categories were also defined in such a way so as to avoid impossible or improbable combinations of variables and to avoid extremely unequal representation in each category. Therefore, no computerized procedure to generate categories automatically, which would be otherwise useful, was applied. The initial arbitrary split into categories is presented in Figure 2.

The aim of the analysis at this stage was to discover which variables have the least effect on trip-generation rates and can be removed from consideration. A convenient method used was a series of pairwise comparisons performed for categories i and j, which differ with respect to one variable only. An example of such an analysis is given in Table 2.

The results of the stage 1 analyses are summarized in Table 3. Some variables always give a significant and regular explanation of patterns in tripmaking. These variables are car availability, employment status, age, and sex. Income might be significant if only two levels (higher, lower) were introduced and, therefore, deserved further investigation.

Other variables such as family status, race, and employment type gave unsatisfactory explanations and were excluded from the second stage of the model. The proposal for further analysis of the category definition is shown in Figure 3 and is analyzed in the next section.

It is worth dwelling on the significant result that household type does not appear to be an important descriptor of a person's tripmaking behavior. One of the major arguments made in favor of the household level of data aggregation is that family structure (e.g., number of children of different ages) affects travel behavior of adults in the household. It was claimed, therefore, that the family's needs (and consequently trips) should be analyzed together with special reference to interactions within the family.

The result here suggests that adults will fulfill their transportation needs (measured by trip rates) independently of their family situation; the sources of variation in data are outside the family-structure variable. This result supports the person level of data aggregation applied here. It is also worth noting that, with the exception of single-member households, the sample size is rather large (>250), and the result obtained should not be a statistical artifact.

STAGE 2: ANALYSIS OF TRIP RATES AND DEVELOPMENT OF FINAL PERSON CATEGORIES

Pairwise Analysis of Remaining Variables

The total trip rates (trips per person) and travel times (total daily travel time per person) by age groups, sex, automobile availability, employment status, and income, as well as the results of pairwise comparisons of trip rates for each strata, are given in Table 4. The accompanying figures (Figures 4-7) provide a graphic analysis of two or more factors that the pairwise comparison is unable to do. These graphs are useful in understanding basic relationships between variables.

The results given in Table 4 and shown in the accompanying figures suggest that the most important variables are age, employment status, and car availability. Sex and income appear to be weak variables. Their independent effect when analyzed together with car availability or employment status tend to disappear altogether (for example, see Figure 4, which is an analysis of employment and sex).

Traveling activity, measured by trip rates N and by daily travel times T, declines with age (Figure 5). Most dramatically this is true for the obligatory trip, which declines substantially after retirement.

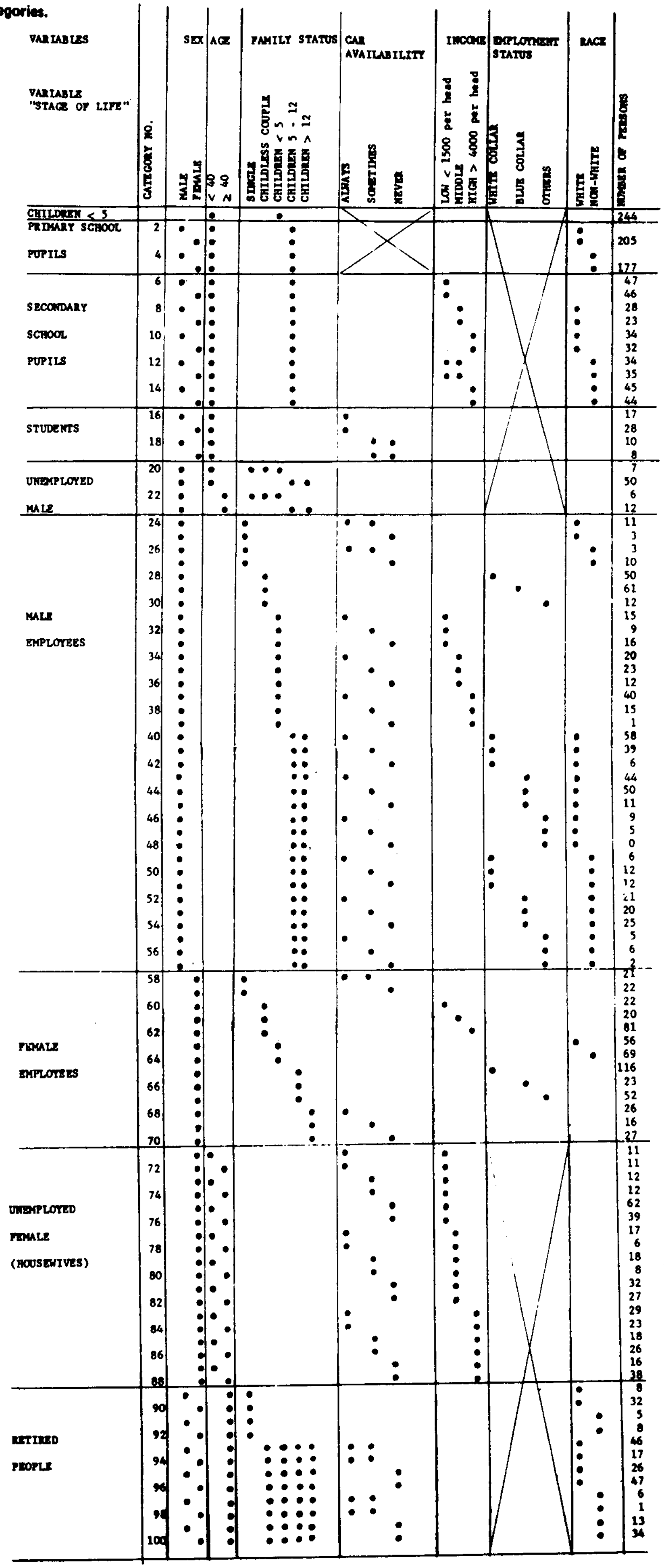
Employment (i.e., the existence of obligatory activity) is a basic factor for explaining the differences in trip rates and daily travel, as shown in Figure 6. Car availability is also of great significance; this is especially true for distinguishing the tripmaking patterns of those who do not have cars available from those who do have cars available (see Figure 7).

The obvious reasonableness of these conclusions supports the modeling approach by which they were derived. A more thorough analysis of data will be described next to define the final categories.

Q-Type Correlation Analysis of 40 Person Categories

Based on previous results, four versions of the final categories shown in Figure 8 might be considered. In these groupings age is divided into

Figure 2. Primary split into 100 categories.



three strata: younger than 18, 18 to 65, and older than 65. The pairwise analysis suggested that the age groups younger than 40, 40 to 65, and older than 65 may be most appropriate. However, plots in Figures 4-7, which consider more than one variable, as well as practical considerations, favor the first-mentioned age strata. The first stratum consists of (mostly) unemployable students, the second stratum includes the labor pool, and the third stratum includes retired people.

Four versions of category descriptions were analyzed (Figure 8). Version D is preferred because it is a parsimonious grouping of people into only eight categories; however, it must be based on a more

Table 2. Analysis of variable age: trip rates for younger versus older housewives.

Category No.		Trip Rates N_i	
Age < 40	Age > 40	Age < 40 (n = 215)	Age > 40 (n = 190)
71	72	3.36	2.81
73	74	3.00	2.33
75	76	1.42	0.85
77	78	2.18	1.50
79	80	1.89	1.13
81	82	1.12	0.70
83	84	3.86	3.65
85	86	2.72	2.27
87	88	1.50	1.52
		2.11[a,b]	1.73[a,c]

[a]Mean of total trip rate.
[b]$Z_{1,2} = 4.00$.
[c]$Z_{0.01} = 2.30$.

Figure 3. Stage 2 description of person categories.

CAT	AGE	SEX	AUTO	EMP	INC
0	1				
1	2	1			
2	2	2			
3	3	1	1	1	1
4	3	1	2	1	1
5	3	1	3	1	1
6	3	1	1	1	2
7	3	1	2	1	2
8	3	1	3	1	2
9	3	1		2	
10	3	2	1	1	1
11	3	2	2	1	1
12	3	2	3	1	1
13	3	2	1	1	2
14	3	2	2	1	2
15	3	2	3	1	2
16	3	2	1	2	
17	3	2	2	2	
18	3	2	3	2	
19	4	1	1	1	1
20	4	1	2	1	1
21	4	1	3	1	1
22	4	1	1	1	2
23	4	1	2	1	2
24	4	1	3	1	2
25	4	1		2	
26	4	2	1	1	1
27	4	2	2	1	1
28	4	2	3	1	1
29	4	2	1	1	2
30	4	2	2	1	2
31	4	2	3	1	2
32	4	2	1	2	
33	4	2	2	2	
34	4	2	3	2	
35	5	1	1		
36	5	1	2		
37	5	1	3		
38	5	2	1		
39	5	2	2		
40	5	2	3		

Variable Levels:

AGE	1. <12
	2. 12-18
	3. 19-40
	4. 41-65
	5. >65
SEX	1. Male
	2. Female
AUTO AVAILABILITY	1. Never
	2. Sometimes
	3. Always
EMPLOYMENT	1. Employed
	2. Non-employed
INCOME	1. < \$3000/cap
	2. ≥ \$3000/cap

Table 3. Pairwise comparison of trip rates by variable categories (stage 1).

		Total Trip Rate[a]								Z-Values[b] ($Z_{0.01}$ = 2.57, $Z_{0.05}$ = 1.96)			
		1		2		3		4					
Variable	Category	Mean	No.	Mean	No.	Mean	No.	Mean	No.	1,2	1,3	2,3	Comments
Sex	Male, female	2.65	811	2.20	1,093					7.89	–	–	Significant difference in trip occurred only for persons >65; this group alone may not warrant stratification by sex
Age	12-18, 18-65, >65	2.92	482	2.56	1,661	1.23	243			4.82	–	26.3	Younger persons travel more
Age, housewives only	<40, >40	2.11	215	1.73	190					4.00	–	–	Younger persons travel more
Car availability	Never, sometimes, always	1.38	309	2.78	289	3.23	341			15.4	–	3.8	Differences between car never, sometimes, and always available are significant; greater car availability means more trips
Employment status	Employed, not employed	2.85	1,183	1.85	478					17.0	–	–	Whether a person is employed or not is an extremely significant variable
Income	Low, middle, high	1.89	187	1.85	163	2.83	206			0.40	–	8.00	Trip rates between high and other income groups are different
Race	White, nonwhite	2.25	398	1.98	176					2.88	–	–	This is an extremely erratic variable; visual examination of data did not suggest stratification by race; difference caused by four categories (46, 59, 94, 98)
Employment type	White collar, blue collar, other	3.05	133	2.67	171	2.92	27			2.28	0.42	0.83	Not a significant variable
Household type	Single, couple, couple with children <5, couple with children >5	2.90	70	2.78	246	2.82	276	2.80	591	0.62	0.21	0.32	Family type is not significant ($Z_{3,4}$ = 0.10)

[a]The columns in this section are read as follows. The strata for each variable are defined under the Variable and Category columns; e.g., car availability—never, sometimes, always, and the trip rates in columns 1, 2, and 3 pertain to these strata in the codes shown (i.e., 1 for never, 2 for sometimes, and 3 for always).
[b]Z-values are calculated by comparing the mean trip rates for the columns shown.

Table 4. Pairwise comparison of trip attributes by category (stage 2).

Variable	Category	Attribute	Characteristics of Attributes: 1 Mean	1 SD	1 No.	2 Mean	2 SD	2 No.	3 Mean	3 SD	3 No.	4 Mean	4 SD	4 No.	Z-Value of Pairwise Comparison of Means of Attributes: $Z_{1,2}$	$Z_{1,3}$	$Z_{1,4}$	$Z_{2,3}$	Comments
Age	<18, 18-40, 41-65, >65	N	2.88	2.05	347	2.77	2.01	698	2.40	1.76	586	1.25	1.67	195	0.82	3.65	3.50	3.50	$Z_{2,4}$ = 10.72, $Z_{3,4}$ = 8.23
		T	51.8	37.2	347	52.8	38.0	698	47.9	35.2	586	22.0	31.2	195	0.40	1.17	9.94	2.40	
Sex	Male, female	N	2.69	2.05	816	2.37	1.98	1,010							3.37	–	–	–	Sex alone is a significant variable, but when plotted together with employment status its significance disappears
		T	53.8	38.9	816	42.7	36.5	1,010							6.57	–	–	–	
Automobile availability	Never, sometimes always	N	1.55	1.58	501	2.86	2.05	349	3.23	1.95	483				10.05	14.80	–	4.23	Important variable
		T	32.6	36.1	501	54.8	35.8	349	60.6	38.2	483				8.86	11.80	–	2.23	
Employment status	Employed, not employed	N	3.05	0.19	1,086	1.71	1.43	740							21.18	–	–	–	Important variable
		T	61.4	39.4	1,086	27.8	34.7	740							19.22	–	–	–	
Income	Low, high	N	2.78	2.05	217	3.27	2.23	522							2.88				Income is not a strong variable; for T it is not a significant stratifier even when considered alone
		T	62.8	39.1	217	67.2	42.3	522							1.42				

Note: This table is read in the same manner as Table 3. N = trip rate and T = total travel time.

Figure 4. Values of N and T as dependent on employment, sex, and age of persons.

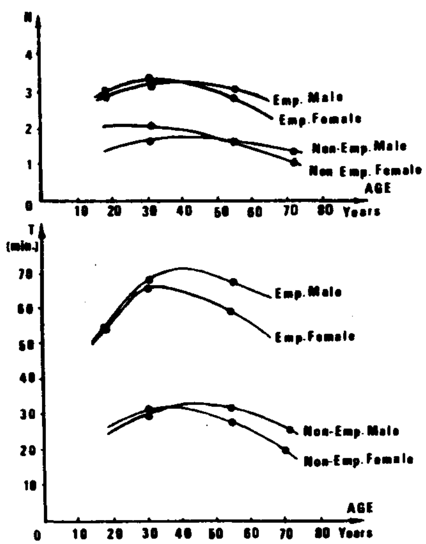


Figure 5. Values of N and T for obligatory and discretionary trip purposes as dependent on age of persons.

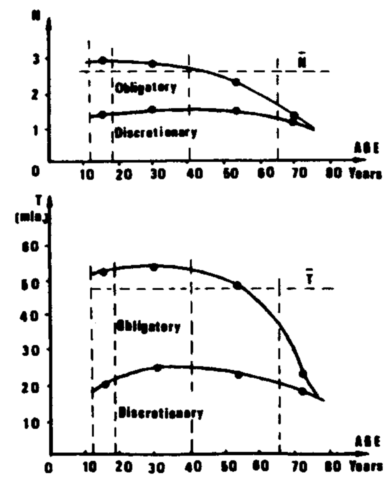


Figure 6. Values of N and T as dependent on age of persons and employment status.

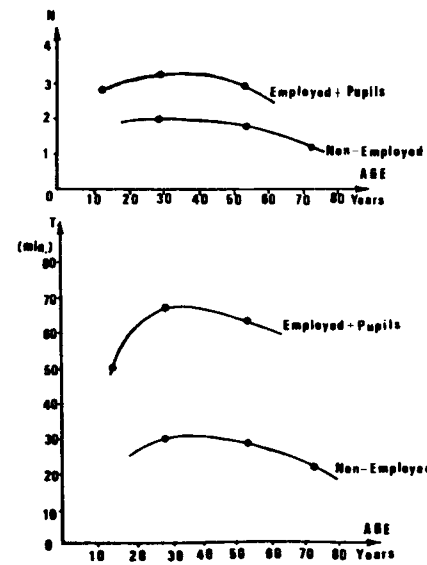


Figure 7. Values of N and T as dependent on car availability and age of persons.

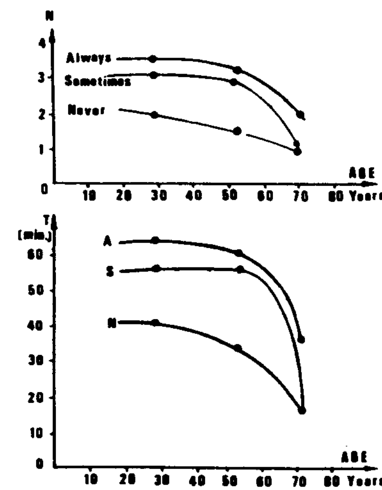
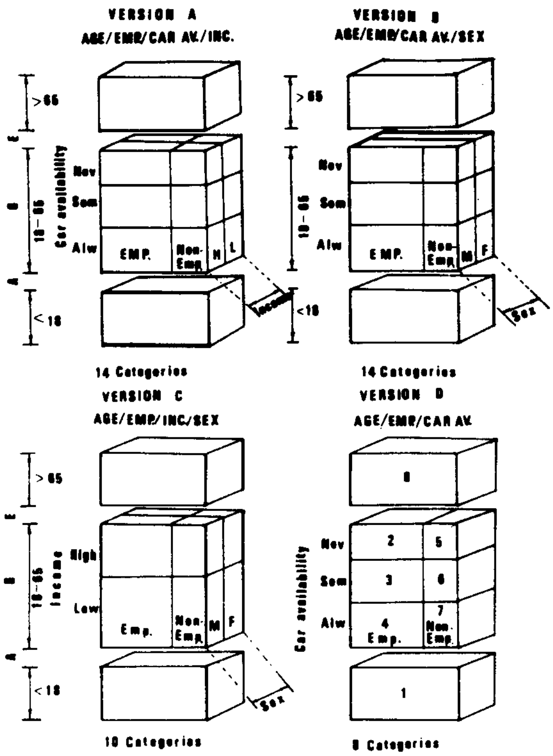


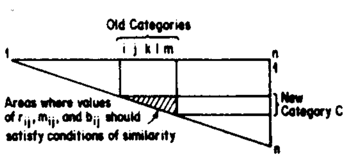
Figure 8. Four versions of person-category definition.



detailed examination of the data using the 15-element trip rate vector (N_{qi}) shown in Table 1 and calculated for each category.

For all four versions of the final category definition, respective triangle matrices of r_{ij}, m_{ij}, and b_{ij} were found (the Q-type analysis). From the analysis point of view, the interesting parts of these matrices are those near the hypothenuse, where the values of r_{ij}, m_{ij}, and b_{ij} are expected to satisfy conditions of similarity given earlier (Equations 2-4) for those old categories i, j,..., m, which will be combined in one new category C (Figure 9).

Figure 9. General idea of creating and evaluating new final person categories.



The shadowed triangles in Figure 9 that were near hypothenuses of the matrices r_{ij}, m_{ij}, and b_{ij} were examined carefully. As one of the possible measures of appropriateness for each four versions of the final category description, the average regression for pairs of categories in the shadowed areas was calculated.

The results of the regressions [see Supernak et al. (8) for details] indicated that a 14-category version is only slightly better than the 8-category version. This conclusion is also supported by visual inspection of the triangular matrices for r_{ij}, b_{ij}, and m_{ij} (8).

Further detailed examination of the matrices for r_{ij}, b_{ij}, and m_{ij} led to three specific comments. First, there are three main groups of travelers that have clearly different trip-generation characteristics: people under the age of 18 (mostly students), employed adults (age 18 to 65), and not employed adults and retired people. Second, the conditions taken as a measure of similarity ($r_{ij} > 0.900$, $0.75 < m_{ij} < 1.25$, $b_{ij} < |0.10|$) are satisfied for most pairs of old categories, which are consolidated into the final new categories. These criteria are better met by the student and employed adult categories than by the not employed and retired categories. It means that the existence of an obligatory activity (work, school) makes travelers' behavior more regular. Third, unsatisfactory values of r_{ij}, m_{ij}, and b_{ij} observed in some cases were regularly accompanied by small size in the categories.

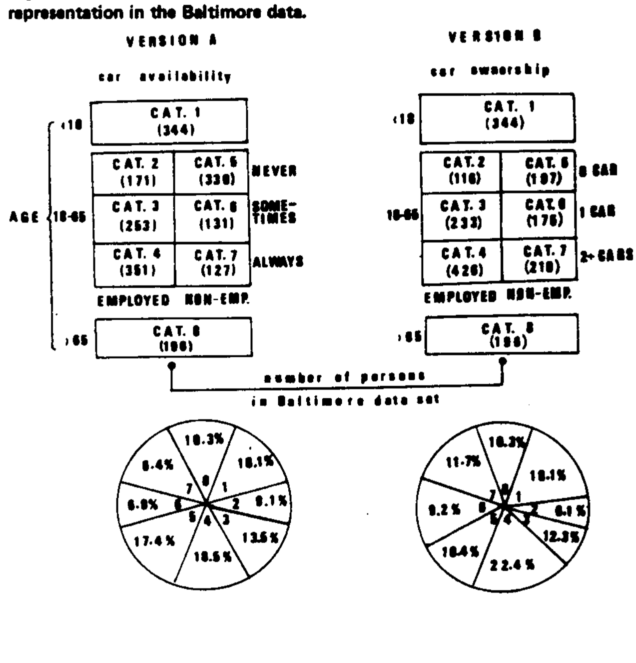
The correlation analyses and the pairwise comparisons strongly suggest that the final categories should be based on age (younger than 18, 18 to 65, older than 65) and employment status (employed, not employed). Of the remaining variables, either car availability or sex and income could be used. For practical reasons, to keep the numbers of categories low and variables compatible with other models, car availability was chosen to complete the list of variables for defining trip-generation categories. A two-dimensional analysis of variance was done to provide quantitative support for this choice; the results indicated that sex and income do not have much explanatory power when analyzed together with car availability.

STAGE 3: FURTHER ANALYSIS OF FINAL TRIP-GENERATION CATEGORIES

The final eight person categories were based on three variables: age, employment status, and car availability. These eight categories are analyzed in more detail.

Car availability data may be replaced in the model by car ownership, the latter in some cases being more readily available. The results of a version A (using car availability) and those of a version B (using car ownership) are compared in Figure 10. For practical model applications, both versions require estimation of category representa-

Figure 10. Two versions of final person-category description and their representation in the Baltimore data.



tions at the zonal level. This can be achieved by applying the person-category car-availability and ownership model, which is presented in detail in Supernak et al. (11). This model uses land use and level-of-service variables and thus takes into consideration the influence of these variables on both the category representations and final trip rates in the given area.

Figure 10 compares these two versions of the final trip-generation categories in the available sample. The weekday trip-generation rates for the two versions are given in Table 5 for all trips and in Table 6 for vehicular trips only. The data indicate that there is little difference whether car availability or car ownership is used. The biggest difference is in discretionary trips by car-owning persons. Generally, version A of the model formulation is recommended because it clearly refers to the person (a real or potential traveler) and his access to transportation models and his individual travel choices. The person-category car-availability model (11) is a direct input to the person-category trip-generation model. Both models require only routinely available data and are easy in practical application.

Table 5. Trip-generation rates (trips per person) for eight person categories, weekdays only (stage 2).

Category No.	Home Based: Obligatory A	Home Based: Obligatory B	Home Based: Discretionary A	Home Based: Discretionary B	Non-Home Based A	Non-Home Based B	Total A	Total B
1	1.47	1.47	1.13	1.13	0.38	0.38	2.98	2.98
2	1.40	1.27	0.59	0.70	0.51	0.57	2.50	2.54
3	1.77	1.69	0.85	0.85	0.55	0.59	3.17	3.23
4	1.67	1.72	1.05	0.90	0.76	0.68	3.48	3.30
5	0.13	0.15	0.89	0.93	0.31	0.35	1.33	1.43
6	0.34	0.23	1.74	1.39	0.47	0.43	2.55	2.05
7	0.30	0.27	2.10	1.66	0.59	0.43	2.99	2.36
8	0.12	0.12	0.93	0.93	0.43	0.43	1.48	1.48
Weighted avg of population	1.01		1.07		0.50		2.59	

Note: Categories in versions A and B are defined in Figure 10.

A comparison of the data in Tables 5 and 6 indicates the importance of walk and other nonvehicular trips (e.g., bike, horse, boat). For example, for persons not owning cars these trips account for 40 to 60 percent of all trips. For young people this percentage is greater. This is important because there clearly exist substitution possibilities between walk and bike and vehicular modes, and these should be accounted for in the models. It also appears that there is a distinct difference between employed and not employed persons' trip rates; the same is true for the car-ownership and car-availability groups.

For example, non-home-based vehicle trips (during weekdays) are more numerous for employed persons and increase with higher automobile availability level, which is an expected finding.

Also, modal choice is strongly related to the person category for both obligatory and discretionary trips. Employed persons are more likely to drive than not employed persons; public transit is rarely used by those with car always available, and the same applies to discretionary trips by persons with any access to a car; also the percentage of walk trips increases with decreasing car availability and is larger for discretionary trips. Again, the walk trips are of no small significance; they are more common than the transit trips (7, Figure 3).

Table 6. Trip-generation rates (vehicle trips per person) for eight person categories, weekdays only.

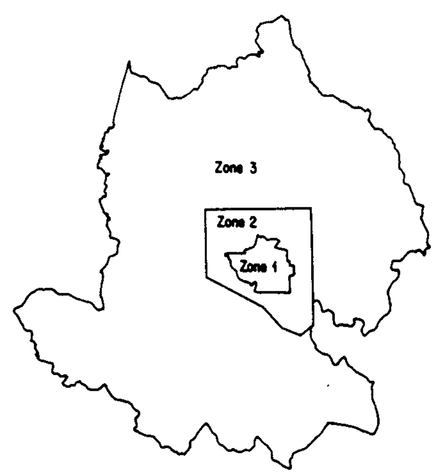
Category No.	Home Based: Obligatory A	Home Based: Obligatory B	Home Based: Discretionary A	Home Based: Discretionary B	Non-Home Based A	Non-Home Based B	Total A	Total B
1	0.63	0.63	0.48	0.48	0.15	0.15	1.26	1.26
2	1.15	0.98	0.28	0.28	0.28	0.27	1.71	1.53
3	1.64	1.57	0.76	0.83	0.49	0.52	2.89	2.91
4	1.61	1.61	0.96	0.82	0.71	0.63	3.28	3.09
5	0.06	0.06	0.40	0.31	0.14	0.11	0.60	0.48
6	0.28	0.16	1.39	1.04	0.38	0.32	2.05	1.52
7	0.24	0.21	2.03	1.50	0.57	0.39	2.84	2.10
8	0.12	0.12	0.60	0.60	0.28	0.28	1.00	1.00
Weighted avg of population	0.80		0.75		0.36		1.91	

Note: Categories in versions A and B are defined in Figure 10.

TRANSFERABILITY OF MODEL WITHIN THE BALTIMORE AREA

To examine the performance of the person-category trip-generation model, it was applied to three different areas of the Baltimore region. Area 1 is the central urban area (628 persons), area 2 is the remainder of the urban area (617 persons), and area 3 is the suburban area (622 persons) (see Figure 11).

Figure 11. Baltimore region divided into three areas.



A transferability error analysis of areawide trip rates, nonwork vehicle trip rates, and the automobile drive portion of modal split for subareas 1 and 3 is given in Table 7. The data indicate that the categorization of persons reduces the percentage error in the average trip rate, and thus in travel-demand prediction, often by more than 50 percent, which leaves the remaining error rather low. The remaining errors for total trip rates (N, N^{obl}, N^{disc}) are smaller for the recommended version A of the model formulation than for version B. The data also indicate that person categories provide a satisfactory explanation of automobile driver modal-split percentages (it can even be argued that these are better results than the results obtained with a sophisticated modal-split model).

Table 7. Comparison of transferability errors for subareas 1-3 in Baltimore with and without category division.

			Zone 1: Central Urban		Zone 3: Suburban	
No.	Value	Category Split Version	Percentage Error without Category Split[a]	Percentage Error with Category Split[b]	Percentage Error without Category Split[a]	Percentage Error with Category Split[b]
1	N	A	+23.3	+10.5	-14.2	-6.6
		B		+12.8		-8.6
2	N^{obl}	A	+29.1	+9.3	-11.2	-3.2
		B		+8.1		-3.2
3	N^{disc}	A	+19.4	+11.9	-16.4	-1.7
		B		+19.4		-15.8
4	$N^{nonwalk}$	A	+57.0	+26.5	-26.4	-14.3
		B		+18.7		-15.4
5	Percentage of discretionary nonwalk trips	A	+51.9	+31.2	-18.2	-13.3
		B		+29.9		-16.0
6	Percentage of drive-alone trips	A	+86.8	+7.4	-37.6	-15.3
		B		+16.9		-11.4

[a]Calculated as $(\bar{N}_{ave} - \bar{N}_j)/\bar{N}_j$, where j = area.
[b]Calculated as $\left(\sum_i \alpha_{ij}N_{ij} - N_j\right)/N_j$, where α_{ij} = percentage of sample in category i who reside in area j.

The numbers in Table 7 also call for caution in treating walk trips. The data indicate that there is an overprediction of nonwalk trips in the urban area by about 30 percent, and an underprediction of nonwalk trips in the suburban area by about 15 percent, even when person categories are used; thus walking is an important mode.

Overall, this analysis demonstrates the usefulness of categorization of the population into eight segments. The conclusion from the data in Table 7, however, should not be that trip-generation forecasts based on person categories provide a substantial improvement over trip-generation forecasts based on average (one category) trip rates. This would be a trivial finding. Rather, the conclusion is that the remaining transferability errors are low, keeping in mind that sample size in Baltimore subareas is only about 600.

Another transferability test was performed between Baltimore and the Twin Cities of St. Paul-Minneapolis (12). Unfortunately, this comparison could be made for travelers only and their vehicular trips because the data records in the Twin Cities were not complete. The trip rates of eight categories appeared to be similar for those two cities, and the transferability errors were low. However, because the analysis unit traveler is not recommended for trip-generation analyses, this part of the research is not presented in this paper. More details about transferability of the person-category trip-generation model are given in Supernak (13).

COMPARISON WITH HOUSEHOLD CATEGORY MODEL AND CONCLUDING REMARKS

For comparison purposes, a household-category model was developed in the same way as the person-category model (9). Because there were only 609 households (but 1,825 individuals) in the Baltimore data (weekdays), the analyses lacked the richness of the person-category model.

Based on previous research (1,14,15), three variables were chosen for the analyses: household size (one, two, three, four, five or more), car ownership (zero, one, two or more), and number of employed household members (one, two, three or more). Unfortunately, other variables such as income and race could not be included because the chosen variables already yielded 51 categories, and the sample size was only 609.

Some results of the pairwise comparison of trip rates are given in Table 8. One unexpected result is noticed. The household-size variable is the only one that gives expected, consistent results. Household size appears to overshadow all other differences; this of course is a trivial finding (i.e., more people, therefore more trips). This result is substantial because it indicates the inefficiency and simplicity of the household-category model. The person-category model totally avoids these types of trivialities and the difficulty of predicting household size [for substantial errors in predicting household size, see Talvitie et al. (16)].

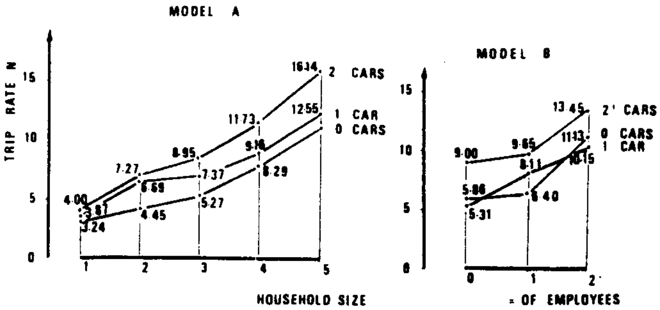
Table 8. Results of pairwise comparison of trip rates for different variable strata.

Variable Examined	Stratum i	Stratum j	Z_{ij}
Car ownership	0	1	2.24
	1	2+	4.24
Household workers	0	1	1.48
	1	2	1.76
	2	3+	4.17
Household size	1	2	4.80
	2	3	2.70
	3	4	3.39
	4	5+	3.89

Note: $Z_{0.01}$ = 2.57.

The two models discussed next are two-dimensional combinations of the three variables. The first model, model A, has 15 categories of household size (one, two, three, four, five or more) and car ownership (zero, one, two or more). Model B has nine categories of workers (zero, one, two or more) and car ownership (zero, one, two or more). Trip rates for these models are shown in Figure 12. Model A shows consistency; that is, trip rates increase with car ownership and family size. Model B does not show consistency; that is, the trip rate for one-car families is less than the zero-car households when there are zero or two or more workers in the household. This outcome is difficult to explain and suggests that model A is the better model because introduction of one more variable (e.g., household size) would increase the number of categories to make the model impractical. It may be recalled that

Figure 12. Household model trip rates.



employment status was the key variable in the person-category model.

Examination of the performance of model A was difficult. Because of reasons of data incompatibility, a transferability check with Minneapolis-St. Paul data was impossible. The scarcity of data required that the Baltimore region be divided only into two areas, instead of the three used with the person-category model, to examine the transferability properties of the model. The remaining transferability error between the two zones was approximately 15 percent, or slightly more than for the person-category model (6 to 12 percent for the recommended version). Nevertheless, the findings are not comparable because the Baltimore subareas were defined differently.

Principally, then, the person-category model is favored for the following reasons. First, it classifies people in a manner that is logical and eliminates the necessity of predicting household formation and, especially, household size with their attendant difficulties. The research also indicated that household type was an unimportant variable in explaining person trip generation. Second, data are used much more efficiently in the person-category model than in household-category model, or, alternatively, less data are needed for developing the person-category model. Third, fewer categories may be used in the person-category model. Because household size is the key variable in household-category model, it precludes the introduction of real behavioral variables (such as age, employment status, and others) if the number of categories is to be kept within practical limits. This renders the household-category model trivial.

Finally, the person-category model has a better behavioral background because the analysis unit is identical with the traveling unit. This makes the person-category trip-generation model compatible with other models in the entire travel-demand model system. The person-category car-availability model, which is fully compatible with the person trip-generation model, makes references to the land use and level-of-service variables that where found to be significant in previous aggregate models, but were not present in most household-category trip-generation models. Therefore, the person-category trip-generation model reported in this paper is considered to be useful and practical and superior to a household-category model.

ACKNOWLEDGMENT

The research described in this paper was supported in part by a contract from the FHWA, U.S. Department of Transportation, to the State University of New York at Buffalo. We wish to thank Preston Luitweiler and Marilyn Macklin for their help in preparing this paper.

REFERENCES

1. H.Y. Wooton and G.W. Pick. A Model for Trips Generated by Households. Journal of Transport Economics and Policy, No. 1, 1967.
2. U. Böhme. Grundlagen zur Berechnung des städtischen Personenverkehrs. Technische Universität Dresden, Dresden, German Democratic Republic, Ph.D dissertation, 1970.
3. E. Kutter. Demographische Determinanten städtischen Personenverkehrs. Technische Universität Braunschweig, Braunschweig, Federal Republic of Germany, 1972.
4. J. Supernak. Travel Demand Models for Polish Cities. Technical University of Warsaw, Warsaw, Poland, Ph.D dissertation, 1976 (in Polish).
5. J. Supernak. A Behavioral Approach to Trip Generation Modeling. Presented at the 7th Planning and Transport Research and Computation, Ltd. Summer Annual Meeting, Univ. of Warwick, Coventry, Warwickshire, England, 1979.
6. J. Supernak. Travel Demand Models. WKIL, Warsaw, Poland, 1980 (in Polish).
7. J. Supernak. Travel-Time Budget: A Critique. TRB, Transportation Research Record 879, 1982, pp. 15-28.
8. J. Supernak, A. Talvitie, and A. DeJohn. Analysis of the Person Category Trip Generation--Split into Forty Categories. State Univ. of New York, Buffalo, Working Paper 782-11, 1980.
9. A. DeJohn. The Comparison Between the Household and Person Category Trip Generation Models. State Univ. of New York, Buffalo, M.S. thesis, 1981.
10. K. Überla. Faktorenanalyse. Springer-Verlag, Berlin, New York, 1971.
11. J. Supernak, A. Talvitie, and B. Bogan. A Person Category Car Availability/Ownership Model. State Univ. of New York, Buffalo, Working Paper 782-17, 1981.
12. J. Supernak, A. Talvitie, and A. DeJohn. Transferability of the Person Category Trip Generation Model Between Baltimore and Twin Cities. State Univ. of New York, Buffalo, Working Paper 782-13, 1980.
13. J. Supernak. Transferability of the Person Category Trip Generation Model. Presented at the 9th Planning and Transport Research and Computation, Ltd. Summer Annual Meeting, Univ. of Warwick, Coventry, Warwickshire, England, 1981.
14. C.R. Fleet and A.B. Sosslair. Trip Generation Procedures: An Improved Design for Today's Needs. Traffic Engineering, Arlington, Va., Nov. 1976, pp. 17-25.
15. D.Y. Ashley. The Regional Highway Traffic Model: The Home-Based Trip End Model. Presented at the Planning and Transport Research and Computation, Ltd. Summer Annual Meeting, Univ. of Warwick, Coventry, Warwickshire, England, 1978.
16. A. Talvitie, M. Morris, and M. Anderson. An Assessment of Land Use and Socioeconomic Forecasts in the Baltimore Region. State Univ. of New York, Buffalo, Working Paper 782-04, 1980.

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

Notice: The views expressed in this paper are those of the authors and not necessarily those of the FHWA.

Trip Generation by Cross-Classification: An Alternative Methodology

PETER R. STOPHER AND KATHIE G. McDONALD

An alternative methodology for calibrating cross-classification models, namely multiple classification analysis (MCA), is described. This technique, which has been available in the social sciences for some time, does not appear to have been used in transportation planning before, although it appears to be able to overcome most of the disadvantages normally associated with standard cross-classification calibration techniques. The MCA procedure is described briefly, and its merits—in terms of statistical assessment, ability to permit comparisons among alternative models, and lack of susceptibility to small samples in individual cells—are discussed in detail. In addition, the method is based on analysis of variance (ANOVA), which provides a structured procedure for choosing among alternative independent variables and alternative groupings of the values of each independent variable. These procedures are contrasted with standard procedures for cross-classification that estimate cell values by obtaining the average value of the dependent variable (e.g., a trip rate) for those samples that fall in the cell and are unable to use any information from any other cell. The process of selecting independent variables and selecting groupings of the chosen variables by ANOVA is illustrated with a case study. In this study the way in which this process works, and the degree to which there is statistical information provided to guide the analyst's judgment, is shown. In the case study the confirmation of intuitive selections of variables is noted, and also a more surprising result is produced that shows that the best household grouping is one that combines two- and three-person households. A second case study illustrates the use of MCA to calculate trip rates. A comparison of the conventional procedure of cell-by-cell averaging, a MCA design that does not account for interactions among the independent variables, and a MCA design that corrects for interactions is given. It is shown that the MCA allows trip rates to be computed for some cells that are empty of data, and that MCA removes some possibly spurious rates that arise in the conventional method from small sample problems in some cells. It is concluded that MCA provides a strong methodology for cross-classification modeling and that the procedure is effective in surmounting most of the drawbacks of conventional estimation of such models.

In the 1950s and 1960s most of the transportation planning studies developed trip-generation equations that used linear regression, particularly for person trip-production models. Linear regression was so strongly favored that it was the central method in the FHWA guide to trip-generation analysis (1). Initially, most of the trip-production models were formulated to provide an estimate of zonal trips as a function of zonal variables that describe households. These models were increasingly the subject of criticism, particularly because of the loss of variance from the extremely aggregate nature of these models (2,3). As a result, household models of trip production were developed, in which the dependent variable became average daily trips per household, possibly by purpose, as a function of attributes of the household. These models remained, however, predominantly linear-regression models.

In a few instances an alternative method of modeling trip generation appeared. This method was known in the United States as cross-classification and in the United Kingdom as category analysis (1,4). This method went through the same development as the linear-regression models, with the earliest procedures being zonal trip estimators and subsequent models being based on household rates. For the most part, however, the household-based cross-classification models were still aggregate in that the classes were defined by average zonal values for household characteristics, and the trip rates were applied simply to the total number of households in the zone. Thus a cross-classification model based on household size and car ownership might have the first variable classified into ranges, such as less than 1.5 persons per household, 1.5 to 2.5 persons per household, 2.5 to 3.5 persons per household, and more than 3.5 persons per household; car ownership was defined similarly in ranges. Then the average zonal values of each variable would be determined and a look-up table would be used to select one cell rate for the zone based on these average values.

Although the cross-classification method was widely used in Europe, it was used in relatively few instances in North America. However, with the growing interest in and use of disaggregate modal-choice models, there has been a resurgence of interest in the cross-classification model, formulated now in a substantially more disaggregate form. Currently, the model uses categorized variables, such as household size, vehicle ownership, and so on, as integer values to describe individual households. The rates in the cells of the table are then average rates for households of that type. The correct application of the model is to estimate the number of households in each category within a zone and to multiply the trip rates by those numbers of households. In general, this procedure leads to greater disaggregation than any other method of modeling trip generation, and has the potential to provide more policy responsiveness than alternative methods.

It is important to note that the standard method for computing cell rates is to group households in the calibration data to the individual cell groupings and total, cell by cell, the observed trips by purpose groups. The rate is then the total trips in a cell by purpose divided by the number of households in the cell. In mathematical form it is as follows:

$$t^p_{mn} = T^p_{mn}/H_{mn} \tag{1}$$

where

t^p_{mn} = trip rate for the pth purpose for households of type mn,
T^p_{mn} = observed trips made by households of type mn for purpose p, and
H_{mn} = observed number of households of type mn.

The advantages that can be claimed for the disaggregate cross-classification methods are as follows:

1. Cross-classification methods are independent of the zone system of a region,
2. They do not require prior assumption about the shape of the relationships (which do not even need to be monotonic, let alone linear),
3. Relationships can differ in form from class to class of any one variable (e.g., the effect of household size changes for zero car-owning households can be different from that of one car-owning households), and
4. The cross-classification model does not permit extrapolation beyond its calibration classes, although the highest or lowest class of a variable may be open-ended.

The models also have several disadvantages, which are common to all traditional cross-classification methods:

1. There is no statistical goodness-of-fit measure for the model, so that closeness to the calibration data cannot be ascertained;
2. Cell values vary in reliability because of different numbers of households being available in each cell for calibration;
3. For the same reason as the preceding problem, the least-reliable cells are likely to be those at the extremes of the matrix, which may also be the most critical cells for forecasting;
4. There is no effective way to choose among variables for classification or to choose best groupings of a given variable, except to use an extensive trial-and-error procedure not usually considered feasible in practical studies; and
5. The procedure suppresses information on variances within a cell (5).

An alternative computational method is put forward and illustrated in the balance of this paper. This method--multiple classification analysis (MCA)--is well known to quantitative social scientists, but appears not to have been used by transportation analysts. As will be shown, MCA overcomes most of the disadvantages of cross-classification models without compromising their advantages.

MULTIPLE CLASSIFICATION ANALYSIS

MCA is based on a simple extension of analysis of variance (ANOVA), and ANOVA (6) also provides a statistically powerful procedure for selecting the variables and their categories for the cross-classification models. MCA is a rather simple development out of ANOVA, with application primarily for two-way and greater ANOVA problems.

Although a number of alternative methods have been suggested for analyzing cross-classification models and for determining cell values (7), there remains little change in the practice of estimating cross-classification cell values. Generalized linear models and regressions with dummy variables have been suggested as alternative methods, but they have not found wide acceptance in practice. The method suggested here is more readily accessible than most others because it is contained in some statistical packages that are available to transportation planners. Nevertheless, like many of the other methods that have been suggested recently, there is no treatment of this method in the statistical texts most frequently used by engineers and by courses taken by transportation planners. Indeed, no reference to the method could be found in any of the statistical texts most likely to be found on the bookshelf of a transportation planner or an engineer. Therefore, a brief description of the method is provided here.

Consider a two-way ANOVA design in which the dependent variable is a continuous variable, such as a trip rate, and the two independent variables are two integer variables that describe households, such as household size and vehicle ownership. First, a grand mean can be estimated for the dependent variable, where this grand mean is estimated over the entire sample of households. Second, group means can be estimated for each group of each independent variable, without regard for the other; in other words, means are computed from the row and column sums of the cross-classification matrix. Each of the group means can be expressed as a deviation from the grand mean. Observing the signs of the deviations, a cell value can now be estimated by adding the row and column deviations of the cell to the grand mean.

An example may help to clarify this. Suppose the dependent variable is home-based work trips, and the independent variables are cars owned and household size. The grand mean is 1.49 trips per household. Deviations for cars owned are -0.97 for zero cars, -0.26 for one car, and +0.88 for two or more cars. Deviations for household size are -1.06 for one person, -0.33 for two persons, +0.49 for three persons, +0.55 for four persons, and +0.70 for five or more persons. For a household with one car and three people, the trip rate would be estimated as 1.72 (= 1.49 - 0.26 + 0.49). That is, it is the grand mean plus the deviation for one car plus the deviation for three persons. Note that, in contrast to standard transportation cross-classification models, the deviations are computed not only for households in the cell three persons with one car, but rather the car deviations are computed over all household sizes, and the household deviations are computed over all car ownerships.

If interactions are present, then these deviations need to be adjusted to account for the interactive effects. This is done by taking a weighted mean for each of the group means of one independent variable over the groupings of the other independent variables, rather than a simple mean, which assumes that variation is random over the data in a group. These weighted means will decrease the sizes of the adjustments to the grand mean when interactions are present. The cell means of a multiway classification are still based on means estimated from all the available data, rather than being based on only those data points that fall in the multiway cell. Furthermore, there is no over-compensation resulting from a false assumption of total lack of correlation between the independent variables.

Because it is based on ANOVA, MCA also has statistical goodness-of-fit measures associated with it. Primarily, these consist of an F statistic to assess the entire cross-classification scheme, an eta-square statistic (8) for assessing the contribution of each classification variable, and an R-square for the entire cross-classification model. These measures provide a means to compare among alternative cross-classification schemes and to assess the fit to the calibration data.

Without pursuing some further advantages offered by the statistical context within which MCA is applied, it is apparent that MCA overcomes effectively several of the disadvantages cited for other types of cross-classification models. First, there are statistical goodness-of-fit measures available for the MCA models that permit selection from among alternative classification schemes and that permit overall assessment of fit to the calibration data. Second, the cell values are no longer based only on the size of the data sample within a given cell; rather the cell values are based on a grand mean derived from the entire data set, and two or more class means are derived from all data in each class of the classification variables, where the intersection of those classes defines the cell of interest. This also tends to reduce the uncertainty of forecasting outlying households. For example, if a critical cell is the five or more person household with two or more cars available, for which the original data might have provided less than 2 percent of the sample, MCA will provide a cell rate that is based on the grand mean (from all the data) adjusted by deviations for all five or more person households and all two or more car households, where the first of these might comprise 10 percent or more of the data and the second more than 20 percent. Clearly, there is far greater reliability in this cell rate than would be obtained from traditional methods.

SELECTING CLASSIFICATION VARIABLES AND CLASSES

In current computer software packages that compute an MCA (9), the MCA is usually provided after performing ANOVA. In turn the use of ANOVA provides the appropriate method for selecting variables and classes within variables. After developing a series of hypotheses about possible variables and classes of variables that might be used for the cross-classification scheme, a series of ANOVAs can be performed, from which several pieces of information are obtained that indicate better or worse classification schemes.

Several pieces of information are provided by a standard ANOVA that enable this evaluation to be made. First, there is an F statistic available for each main effect and for the interaction effects. A highly significant F statistic for the main effects indicates that the variable is strongly associated with the trip-rate variations in the data. A highly significant F statistic for the interaction effects suggests that the independent variables may be too highly intercorrelated to be useful, and it is likely to be necessary to choose among alternative independent variables and reduce as much as possible the interaction effects. There is also an overall F statistic for the entire cross-classification scheme that indicates the extent of covariation between the trip rates and the set of classified independent variables.

By trial-and-error procedures, or nested hypotheses, it is also possible to compare alternative independent variables and to compare alternative classifications. Of course, as the number of classes is changed, there is a consequent change in the number of degrees of freedom of the ANOVA problem and a consequent change in the expected F statistic. Obviously, this must be taken into account in assessing alternative schemes, but it then becomes possible to determine the amount of information loss occurring by aggregating classes, or the amount of added information obtained by disaggregating classes.

Thus ANOVA provides a structured and statistically sound procedure for selecting both the independent variables and the best groupings of those variables from those available. There is no claim of optimality in this, and clearly there are countervailing tendencies from aggregating and disaggregating variables, which demand the application of judgment to the results rather than blind acceptance of the statistical indicators. Also, the method is only as good as the initial and subsequent hypotheses of model structure. This may be interpreted as an advantage to the method over linear regression. The latter method permits too readily the abrogation of judgment to stepwise or similar regression procedures that may build models that appear to perform well, based on statistical measures and the R-square values, but which make no conceptual sense, whereas the application of ANOVA is far more demanding of the structuring of conceptually sound hypotheses, particularly because of its rather low efficiency in selecting good structures from blind application.

Finally, with each ANOVA it is possible to obtain the MCA results. These can also be revealing because they provide the additional statistics of an R-square and the eta-square for each variable, and they indicate the size of the deviations from the grand mean provided by each class of each independent variable. These data items may illuminate, clarify, or support the results from the ANOVA and should generally lead to a more rapid closure on a good structure for the model.

In summary, the use of the ANOVA that accompanies the MCA procedure resolves the remaining disadvantage of traditional cross-classification methods, namely the lack of a sound method for choosing among alternative variables and alternative classes within a variable.

There is, however, one disadvantage incurred as a result of the use of MCA. MCA averages the effect of the relationships of one variable over classes of the other variables. Because the deviations are based on row and column means, there is no longer the capability for the shape of the relationship to differ from class to class of each variable as exists in traditional cross-classification methods. There does remain, however, no limitation on the average shape of the relationship for each independent variable, which still is not required even to be monotonic, let alone linear. This appears to be a relatively small price to pay for the advantages obtained, particularly when taking into account that many of the variations in functional form between classes in traditional models may derive from spurious small-sample effects.

USE OF ANOVA TO SELECT VARIABLES AND CLASSES

A case study application of this method used data on 2,446 households from a metropolitan area in the Midwest. For initial variable selection, several candidates were identified and classifications were proposed for each of these variables. As a precursor to the multiway analyses, one-way ANOVAs were performed between trip rates and each candidate variable.

There are two bases for selecting variables in travel-forecasting models that hold true for any model. This first is conceptual or behavioral justification that the variable has a causal effect on the phenomenon being modeled, and the second is statistical justification that the variable shows a significant and measurable empirical association with the phenomenon being modeled.

Given 30 years of travel forecasting at the regional level, considerable experience and information exists now on variables that affect trip production, so that extensive concept formulation is not necessary. Based on past experience, the following variables were considered:

1. Household size (persons per household),
2. Automobile ownership or availability,
3. Housing type,
4. Household life cycle or structure,
5. Number of workers,
6. Number of licensed drivers,
7. Income, and
8. Area type.

Each of these variables is described briefly, together with its expected effects on trip production.

Household size is defined as the number of persons in the household without regard to age. Household size is expected to cause increases in tripmaking for all trip purposes, although not in a uniform manner. Trips per person is expected and has been shown to be relatively stable; hence the more people in the household, the more trips are likely to be made by the household.

Automobile ownership or availability is measured as the number of automobiles, vans, or lightweight trucks usable for personal travel by household members, either owned by the household or available to members of the household. A well-documented phenomenon is that acquisition of a vehicle increases substantially the number of trips and motorized trips made by a household. This arises both from substitution of vehicular trips for walk trips and from satisfaction of previously unsatisfied demand for travel. The tripmaking rate of increase is nonlinear, with a decreasing rate of increase with increasing automobiles. Vehicle availability is likely to be the more appropriate measure than ownership because it is a more accurate measure of the potential to satisfy demand for vehicular trips.

Housing type is usually defined as single-family or multifamily dwellings, and hotel and motel units when tourists and nonresidents are to be included. It has a weak conceptual link, deriving principally from density considerations and some aspects of vehicle availability associated with vehicle storage space.

Recent research (10) suggests that a household-structure variable correlates more strongly with trip rates than almost any other variable. The categories of this variable are described elsewhere (see paper by McDonald and Stopher elsewhere in this Record), as are the arguments for its conceptual effect on tripmaking (10), and they are not described in this paper.

Number of workers may be defined as all workers, or as full-time workers only, where worker is restricted to work outside the home. Clearly, the number of workers will be in direct proportion to and is causative of the number of household work trips. Also, as more members of a household of a given size work, the number of trips for all other purposes is likely to be fewer, except for non-home-based trips, because more activities are likely to be undertaken on the way to or from work.

To the extent that a household has more licensed drivers than vehicles, more licensed drivers than workers, and more vehicles than workers, the number of licensed drivers would be expected to have a positive relationship to all nonwork trip purposes.

Income is usually defined as income groups of fairly broad income ranges. As income increases (all other things being equal), it is expected that tripmaking would increase because purchasing trips requires available monetary budgets and, as these increase, so does the potential to satisfy previously unsatisfied demand.

Area type has been defined in a variety of ways and is designed to differentiate between areas with markedly different intensities of development and activity. Therefore, either explicitly or implicitly, it is related to employment and residential densities. Where densities are higher, motorized trips are likely to be fewer because opportunities for satisfying activities are closer and both congestion and parking price may be significantly higher, whereas parking availability is lower. In addition, various services and home deliveries may be more available, thus reducing the need for some trips. The effect of area type is likely to be greatest on discretionary travel (home-based social-recreational, home-based other) and least on mandatory travel (home-based work or school).

The purpose of the one-way ANOVAs was both to determine which variables appeared to have the strongest relationships to tripmaking by purpose and to determine the best grouping of data to use. The results of these procedures were as follows.

1. Number of cars available was consistently one of the most significant variables for all trip purposes. It always performed better than number of cars owned.
2. Household size was also consistently a significant variable for all trip purposes.
3. Area type, which was defined as two groups--high density of either residences or employment, and low density of both residences and employment--was ranked third in significance across most trip purposes.
4. Housing type, denoted as single family and multifamily, ranked about fourth in significance across most trip purposes.
5. Household structure, which was defined in terms of the relationships among household members, presence or absence of children, and some aspects of both household size and ages of members, was found to be inferior to household size alone and to number of cars available.
6. Other variables examined included number of workers, number of licensed drivers, and income. Each of these variables was significant for at least one purpose in the most disaggregated form of the variables, but they did not perform satisfactorily across a majority of the purposes.

In experiments on groupings, the results were as follows.

1. Vehicle ownership or availability could be specified as zero, one, and two or more without significant loss of power of the variable.
2. The optimal grouping of household size appeared to be one, two and three, four, and five or more. Examination of some other recent models (11) revealed a small difference in tripmaking rates for most purposes between two- and three-person households, which tended to confirm this grouping.
3. Income is best grouped into low (less than $15,000), medium ($15,000 to $34,999), and high (more than $35,000) categories.
4. Household structure should be grouped into five categories: single-person households, one-parent households, adult households with children and more than one adult, adult households without children and more than one adult, and households of unrelated individuals.
5. Number of workers can be grouped so as to aggregate households of four or more workers into one class, yielding categories of zero, one, two, three, and four or more.
6. Number of licensed drivers can also be aggregated to a set comprising zero, one, two, three, and four or more.

These results should not be considered indicative of general rules of classification. They are for the case study data and are provided here to illustrate the way in which ANOVA can be used for this type of analysis. Details of the runs are not provided here, because the results were derived from use of six trip purposes and involved running a rather large number of ANOVAs. Furthermore, it is not the purpose of this paper to produce specific recommendations on the structure of trip-generation models or to develop conclusions about the inclusion of one or another variable in the model. This is left to other papers that may use the approach described here to make more detailed studies of the performance of alternative variables. Despite the number required to be run, neither setup time to run them nor central processing unit (cpu) time on the computer to complete them were large.

The results of some of the multiway ANOVAs used to select the cross-classification scheme are given in Tables 1-4. The data in Table 1 give five purposes by using car ownership, housing type, and household size, whereas the data in Table 2 are the same except for the use of car availability in place of ownership. For all purposes except shopping, the F statistics are higher, although not significantly so, in most cases. The R-squares for the MCA tables and the eta-squares for the vehicle variable follow the same pattern. There are also two fewer significant interaction terms for car availability than for car ownership. This led to the selection of car availability in preference to car ownership, thus confirming the results from the one-way ANOVAs.

The data in Table 3 give the replacement of the partly insignificant housing type by total employment. Only the home-based work model is clearly better in this specification, the models for all other purposes being virtually indistinguishable from the model with housing type. The data in Table 4 give the use of income in place of housing type. Confirming the NCHRP results (10), income is apparently able to add little once vehicle availability is included. In all purposes, none of the statistical measures for the ANOVAs is as good for this specification as for the one that uses housing type.

An additional interesting result is given in Table 5. In the ANOVAs presented in Tables 1-4, household size was left disaggregated for two- and three-person households. In Table 5 the best speci-

Table 1. ANOVA results for model structure 1.

	Purpose				
Statistic	HBWORK	HBSHOP	HBSOCR	HBOTHR	NHB
F	28.0	6.0	5.7	33.8	10.5
df					
Within group	2,240	2,240	2,240	2,240	2,240
Between groups	29	29	29	29	29
Significant	–[a]	–[a]	–[a]	–[a]	–[a]
R^2	0.255	0.065	0.059	0.291	0.103
Eta-square					
Vehicles owned	0.34[b]	0.14[b]	0.09[b]	0.10[b]	0.16[b]
Housing type	0.06[b]	0.05[b]	0.01	0.02	0.05[b]
Household size	0.25[b]	0.16[b]	0.20[b]	0.50[b]	0.22[b]
Significant interactions	Vehicles owned and household size	None	None	Vehicles owned and household size; housing type and household size	Vehicles owned and household size

Note: Independent variables are vehicles owned, housing type, and household size. F = F-score, df = degrees of freedom, HBWORK = home-based work, HBSHOP = home-based shopping, HBSOCR = home-based social-recreation, HBOTHR = home-based other, and NHB = non-home-based trips.

[a]Significant at 99 percent or beyond. [b]Significant at 95 percent or beyond.

Table 2. ANOVA results for car availability.

	Purpose				
Statistic	HBWORK	HBSHOP	HBSOCR	HBOTHR	NHB
F	29.5	5.9	6.0	35.1	11.4
df					
Within group	2,292	2,292	2,292	2,292	2,292
Between groups	29	29	29	29	29
Significant	–[a]	–[a]	–[a]	–[a]	–[a]
R^2	0.261	0.062	0.060	0.295	0.113
Eta-square					
Vehicles available	0.36[b]	0.12[b]	0.10[b]	0.11[b]	0.20[b]
Housing type	0.05[b]	0.05[b]	0.00	0.01	0.04
Household size	0.24[b]	0.16[b]	0.19[b]	0.50[b]	0.21[b]
Significant interactions	None	None	Vehicles available and household size	Housing type and household size	None

Note: Independent variables are vehicles available, housing type, and household size. Statistics and purposes are defined in Table 1.

[a]Significant at 99 percent or beyond. [b]Significant at 95 percent or beyond.

Table 3. ANOVA results with employment.

	Purpose				
Statistic	HBWORK	HBSHOP	HBSOCR	HBOTHR	NHB
F	37.0	4.3	5.2	25.9	9.5
df					
Within group	2,402	2,402	2,402	2,402	2,402
Between groups	42	42	42	42	42
Significant	–[a]	–[a]	–[a]	–[a]	–[a]
R^2	0.376	0.058	0.061	0.295	0.126
Eta-square					
Vehicles available	0.22[b]	0.15[b]	0.11[b]	0.10[b]	0.16[b]
Workers	0.40[b]	0.04	0.02	0.05	0.14[b]
Household size	0.16[b]	0.17[b]	0.20[b]	0.49[b]	0.19[b]
Significant interactions	Workers and vehicles available; workers and household size	None	Household size and workers; household size and vehicles available	Workers and household size	Workers and household size

Note: Independent variables are vehicles available, workers, and household size. Statistics and purposes are defined in Table 1.

[a]Significant at 99 percent or beyond. [b]Significant at 95 percent or beyond.

Table 4. ANOVA results with income.

Statistic	Purpose HBWORK	HBSHOP	HBSOCR	HBOTHR	NHB
F	23.7	4.1	3.2	22.8	10.5
df					
Within group	2,153	2,153	2,153	2,153	2,153
Between groups	41	41	41	41	41
Significant	–[a]	–[a]	–[a]	–[a]	–[a]
R^2	0.298	0.053	0.046	0.284	0.119
Eta-square					
Vehicles available	0.21[b]	0.13[b]	0.08[b]	0.08[b]	0.13[b]
Income	0.31[b]	0.00	0.02	0.07[b]	0.18[b]
Household size	0.19[b]	0.15[b]	0.08[b]	0.49[b]	0.17[b]
Significant interactions	None	None	None	Income and household size	Income and household size; vehicles available and household size

Note: Independent variables are vehicles available, income, and household size. Statistics and purposes are defined in Table 1.
[a]Significant at 99 percent or beyond. [b]Significant at 95 percent or beyond.

Table 5. ANOVA results with aggregated household size.

Statistic	Purpose HBWORK	HBSHOP	HBSOCR	HBOTHR	NHB
F	34.2	7.2	7.3	41.2	13.9
df					
Within group	2,298	2,298	2,298	2,298	2,298
Between groups	23	23	23	23	23
Significant	–[a]	–[a]	–[a]	–[a]	–[a]
R^2	0.244	0.061	0.058	0.284	0.112
Eta-square					
Vehicles available	0.37[b]	0.12[b]	0.11[b]	0.12[b]	0.20[b]
Housing type	0.05[b]	0.05[b]	0.00	0.01	0.04
Household size	0.19[b]	0.15[b]	0.19[b]	0.49[b]	0.21[b]
Significant interactions	Vehicles available and household size	None	Vehicles available and household size	None	None

Notes: Independent variables are vehicles available, housing type, and household size. Statistics and purposes are defined in Table 1.
[a]Significant at 99 percent or beyond. [b]Significant at 95 percent or beyond.

fication from the previous structures is used, but with the two- and three-person households aggregated into a single group. Because there is a decrease in the number of degrees of freedom, it is expected that the F score will increase. However, the increase is larger than would be expected just from this effect. Housing type still appears to be an ineffective variable, but the use of the more aggregated household size appears to be indicated quite clearly.

DERIVATION OF CROSS-CLASSIFICATION TRIP-GENERATION MODELS

A useful example of the MCA procedure is provided by the use of some data from a trip-generation modeling process used in San Juan, Puerto Rico (12). Figure 1 provides a set of trip rates computed in the standard procedure by using individual cell means. Note that cells 9 and 21 do not have trip rates because the available data lacked observations in these two cells. Figure 2 shows the numbers of households in each cell, and it can be seen that these range from a low of 4 to a high of 133. This range indicates clearly a significant range of reliability in the estimates of rates. If conventional wisdom is adopted, in that a mean and variance can be estimated with some element of reliability from a minimum of 50 observations, 14 of the 24 possible cells are estimated with too few data points.

As the next step in the procedure, a manual estimation of a noninteractive MCA was undertaken. This was done at the time because of the lack of availability of the computer software to undertake a full MCA, but it is useful because it traces out the procedure for MCA. First, a grand mean was computed for the entire set of home-based work trips; it was found to be 1.49. Then deviations were computed for each of the three variables. For the four household-size groups, the group means were found to be 0.33, 1.26, 1.85, and 1.84; for the two area types, they were 1.41 and 1.60; and for the three vehicle-ownership groups, they were 0.65, 1.51, and 2.36. The deviations are computed in each case by expressing the group means as values that deviate from the grand mean. To compute the cell value for area type 1, vehicle ownership of 1, and household size of four persons, the value is 1.98 (= 1.49 + 0.11 + 0.02 + 0.36). The complete set of cell values is shown in Figure 3. Note that there are values now in both cell 9 and cell 21.

Several points are worth noting from a comparison of Figures 1 and 3. First is the one already mentioned of the existence of rates for the empty cells of Figure 1 that appear in Figure 3. Second, some counterintuitive progressions in Figure 1 are removed or decreased substantially in Figure 3. These progressions appear to have been caused by problems from the small sample size. From examining the data in Figure 2, it can be seen that the grand mean is estimated from 1,178 observations, and that the least-reliable deviation (for one-person households) is based on 81 observations. All other deviations are based on more than 120 observations. Although there are still some large variations in the sample size used to compute the deviations, the range of 81

Figure 1. Conventional trip rates: home-based work.

cross class — Area Type	Vehicles/DU	Persons/DU: 1	2,3	4	5+
1 Rural Low Density	0	(1) 0.00	(2) 0.48	(3) 1.35	(4) 1.39
	1	(5) 1.50	(6) 1.46	(7) 1.88	(8) 1.65
	2+	(9) -	(10) 2.10	(11) 2.23	(12) 2.36
2 Urban High Density	0	(13) 0.10	(14) 0.62	(15) 1.00	(16) 0.70
	1	(17) 0.80	(18) 1.29	(19) 1.58	(20) 1.69
	2+	(21) -	(22) 2.19	(23) 2.70	(24) 2.59

Figure 2. Number of households by cell of cross-classification.

cross class — Area Type	Vehicles/DU	Persons/DU: 1	2,3	4	5+
1 Rural Low Density	0	(1) 17	(2) 60	(3) 17	(4) 28
	1	(5) 4	(6) 38	(7) 48	(8) 69
	2+	(9) 0	(10) 43	(11) 40	(12) 70
2 Urban High Density	0	(13) 40	(14) 110	(15) 34	(16) 40
	1	(17) 20	(18) 133	(19) 55	(20) 93
	2+	(21) 0	(22) 58	(23) 43	(24) 63

Figure 3. Noninteractive MCA trip rates: home-based work.

cross class — Area Type	Vehicles/DU	Persons/DU: 1	2,3	4	5+
1 Rural Low Density	0	(1) 0.00	(2) 0.52	(3) 1.12	(4) 1.10
	1	(5) 0.45	(6) 1.38	(7) 1.98	(8) 1.96
	2+	(9) 1.30	(10) 2.23	(11) 2.83	(12) 2.81
2 Urban High Density	0	(13) 0.00	(14) 0.33	(15) 0.93	(16) 0.91
	1	(17) 0.26	(18) 1.19	(19) 1.79	(20) 1.77
	2+	(21) 1.11	(22) 2.04	(23) 2.64	(24) 2.62

Figure 4. Full MCA trip rates: home-based work.

cross class — Area Type	Vehicles/DU	Persons/DU: 1	2,3	4	5+
1 Rural Low Density	0	(1) 0.12	(2) 0.62	(3) 1.01	(4) 0.96
	1	(5) 0.86	(6) 1.36	(7) 1.75	(8) 1.70
	2+	(9) 1.63	(10) 2.13	(11) 2.52	(12) 2.47
2 Urban High Density	0	(13) 0.10	(14) 0.60	(15) 0.99	(16) 0.94
	1	(17) 0.84	(18) 1.34	(19) 1.73	(20) 1.67
	2+	(21) 1.61	(22) 2.11	(23) 2.50	(24) 2.45

to 689 observations represents a much less-significant variation in reliability than in the data used for Figure 1.

Figure 4 presents the results from a full-interaction MCA for the same data. There are clearly some major interactions in this specification of the model, as shown by the differences in the rates between Figures 3 and 4. The anomalous decrease in rate between four and five or more person households remains and is of a similar order of magnitude, which suggests that this result is structured in the data. For the remaining differences, some rates are higher than before, whereas others are lower. As is expected from the theory, the range of trip rates is lower in Figure 4 than in Figure 3 because accounting for interactions decreases the net effect of each variable. Thus the highest trip rate in Figure 3 is 2.83, whereas the highest rate in Figure 4 is 2.52. Similarly, the lowest value has increased from 0.00 in Figure 3 to 0.10 in Figure 4. Perhaps the most marked difference in the two figures is between the one and two or more vehicle households. The large differences at all household-size values between these two have decreased markedly in Figure 4, and the values of the one-vehicle households are substantially higher in the one-person households, and lower in the largest households for Figure 4 compared with Figure 3.

Some statistical comparisons among the results serve to illustrate the differences better than can be seen from a visual inspection. First, root mean square (RMS) errors were calculated between Figures 1 and 3, Figures 1 and 4, and Figures 3 and 4. For Figures 1 and 3, it is 0.47; between Figures 1 and 4 it increases to 0.51; but it is only 0.24 between Figures 3 and 4. This is about as expected. The largest difference is between the conventional rates and the MCA rates with full interactions. The difference between MCA with full interactions and without is by far the least of the differences. Given an average trip rate of around 1.45, the differences between the conventional method and the MCA methods are on the order of one-third of the average trip rate.

Chi-square contingency tests between values close to 1.0 are notoriously misleading because the value of chi-square is necessarily small in such a case. This case is no exception, with the three comparisons producing chi-squares of 1.88, 4.22, and 1.30, each with 21 degrees of freedom. These values would not be considered significant. However, if the rates are multiplied by the number of households in the sample (Figure 2), the chi-square test would be for differences in the numbers of trips produced for work. In this case the chi-squares are 55.5, 19.0, and 41.4, respectively. The degrees of freedom are the same as before, and all values except the second one are significant beyond 95 percent. The low chi-square between Figures 1 and 4 appears to arise purely by chance, where two of the larger groups of households are associated with a small difference in trip rates, fortuitously. It is not clear whether this result should lead to a conclusion of no significant difference in trip rates between the two cases. Thus these results indicate some real differences in trip rates that are likely to lead to significant differences in forecasts.

CONCLUSIONS

The two case studies presented in this paper serve to illustrate the potentials provided by the MCA method and ANOVA from which it stems. This procedure overcomes a number of the criticisms that have been made before about cross-classification models. Specifically, the method permits a statistically based selection of variables for the cross-classification model, and also allows comparisons to be made between alternative groupings of any given variable. From this it is possible to provide a model structure that has both conceptual and statistical merit, rather than relying only on a conceptual selection.

Second, the method provides a statistically sound procedure for estimating cell means, which reduces the inherent variability of rates computed from different size samples of households and is capable of providing estimates for some cells where data may be lacking in the base data set (although the use of this capability does reduce some of the available statistical information). Third, there are goodness-of-fit statistics from all of these steps in the process that permit more specific comparisons to be made, good hypothesis-testing procedures to be followed, and results to be assessed in terms of the amount of the variability of the dependent variable that is captured in the model. Finally, and most important, the method takes into account the interactions among the alternative independent variables, which have never been taken into account in standard cross-classification models.

It should be noted that similar models have been developed for predicting vehicle availability, as well as for trip productions by a variety of purposes. There is no reason why such cross-classification models should not be built for any other phenomenon that is appropriately modeled by this procedure. Principally, any phenomenon that has a nonlinear, and possibly discontinuous, functional form, and that is most readily related to variables that are categorical in nature, would be a prime candidate for the method.

REFERENCES

1. Guidelines for Trip Generation Analysis. FHWA, U.S. Department of Transportation, 1967.
2. G.M. McCarthy. Multiple-Regression Analysis of Household Trip Generation--A Critique. HRB, Highway Research Record 297, 1969, pp. 31-43.
3. H. Kassoff and H.D. Deutschman. Trip Generation: A Critical Appraisal. HRB, Highway Research Record 297, 1969, pp. 15-30.
4. R. Lane, T.J. Powell, and P. Prestwood-Smith. Analytical Transport Planning. Halsted Press, Wiley, New York, 1973.
5. P.R. Stopher and A.H. Meyburg. Urban Transportation Modeling and Planning. D.C. Heath and Co., Lexington Books, Lexington, Mass., 1975.
6. N.L. Johnson and F.C. Leone. Statistics and Experimental Design: Volume II. Wiley, New York, 1964.
7. D. Segal. Discrete Multivariate Model of Work-Trip Mode Choice. TRB, Transportation Research Record 728, 1979, pp. 30-35.
8. P.R. Stopher. Goodness-of-Fit Measures for Probabilistic Travel Demand Models. Transportation, Vol. 4, 1975, pp. 67-83.
9. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and H.D. Bent. Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.
10. P.M. Allaman, T.J. Tardiff, and F.C. Dunbar. New Approaches to Understanding Travel Behavior. NCHRP, Rept. 250, Sept. 1982, 147 pp.
11. Schimpeler-Corradino Associates. Review and Refinement of Standard Trip Generation Model. Florida Department of Transportation, Tallahassee, Final Report Task B, June 1980.
12. Schimpeler-Corradino Associates. Urban Transportation Planning Models. Department of Transportation and Public Works, Commonwealth of Puerto Rico, San Juan, Final Rept., July 1982.

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

Some Contrary Indications for the Use of Household Structure in Trip-Generation Analysis

KATHIE G. McDONALD AND PETER R. STOPHER

The variables used to predict household trip-generation rates have long been an area of concern for transportation planners; these variables included household size, number of vehicles owned, and income. However, a recent NCHRP study that used linear regression analysis has proposed that a household-structure variable would correlate more strongly with trip rates than almost any other variable, except vehicle ownership. In particular, this should improve the model significantly where it is combined with vehicle ownership and used as a substitute for household size. The results of a trip-generation analysis performed on data from the Midwest by using multiple classification analysis (MCA) in contrast to linear regression are described. The household-structure variable was tested by using both analysis of variance and MCA to determine how well the variable performs in various model structures when compared with other variables. The other variables tested were number of cars or vehicles available to the household, household size, housing type, total number of employed persons, household income, and total number of licensed drivers. It was concluded that the household-structure variable did not perform significantly better than the other variables tested.

With the increasing acceptance into practice of behavioral models for travel forecasting, recent research by the NCHRP has focused on enriching travel-forecasting models with theories and procedures from the behavioral sciences. [Note that these research results are from work done at Boston College for NCHRP Project 8-14 (New Approaches to Understanding Travel Behavior); the report is available on request from NCHRP.] One of the first potential directions examined for translation into practice is the incorporation of behavioral concepts in trip-generation modeling at the household level. As part of this research, Charles River Associates (CRA) proposed that a household-structure variable would significantly improve the performance of such a model (1).

This proposal was based on the premise that households with differing structures, in terms of adults, children, and personal roles, would have differing activity requirements, mobility constraints, and opportunities for trade-offs with other household members or for trip chaining. Thus proposed changes in household structure, such as an increasing percentage of single and single-parent households as well as adult households with no children, as is expected within the next decade, would have a significant effect on trip-generation rates within a population. It is argued that such a variable should add behavioral content that is lacking from traditional trip-generation models, which generally have included such variables as household size, number of vehicles owned, and income to predict household trip rates. Furthermore, a household-structure variable would be more significant in capturing changes in the future than many of the more traditional variables used.

The household-structure categories proposed were based on the age, gender, marital status, and last names of each household member. These variables determined the presence or absence of dependents within the household, the number and type of adults present, and the relationships among and of household members.

The results of an application of this household-structure variable in trip-generation analysis in a Midwest study area are described. The value of this variable is compared with other variables that were tested at this time by using multiple classification analysis (MCA) (see paper by Stopher and McDonald elsewhere in this Record). MCA is an extension of analysis of variance (ANOVA) that, for a set of classified data, expresses group means as deviations from the grand mean.

HOUSEHOLD-STRUCTURE CONCEPT

The household-structure variable defined by CRA comprises eight household categories: male and female single-person households, single-parent households, couples, nuclear families, adult families with children, adult families without children, and unrelated individuals. Age 20 was used as the cutoff to distinguish between children and adults. These categories were determined by using the method shown in Figure 1.

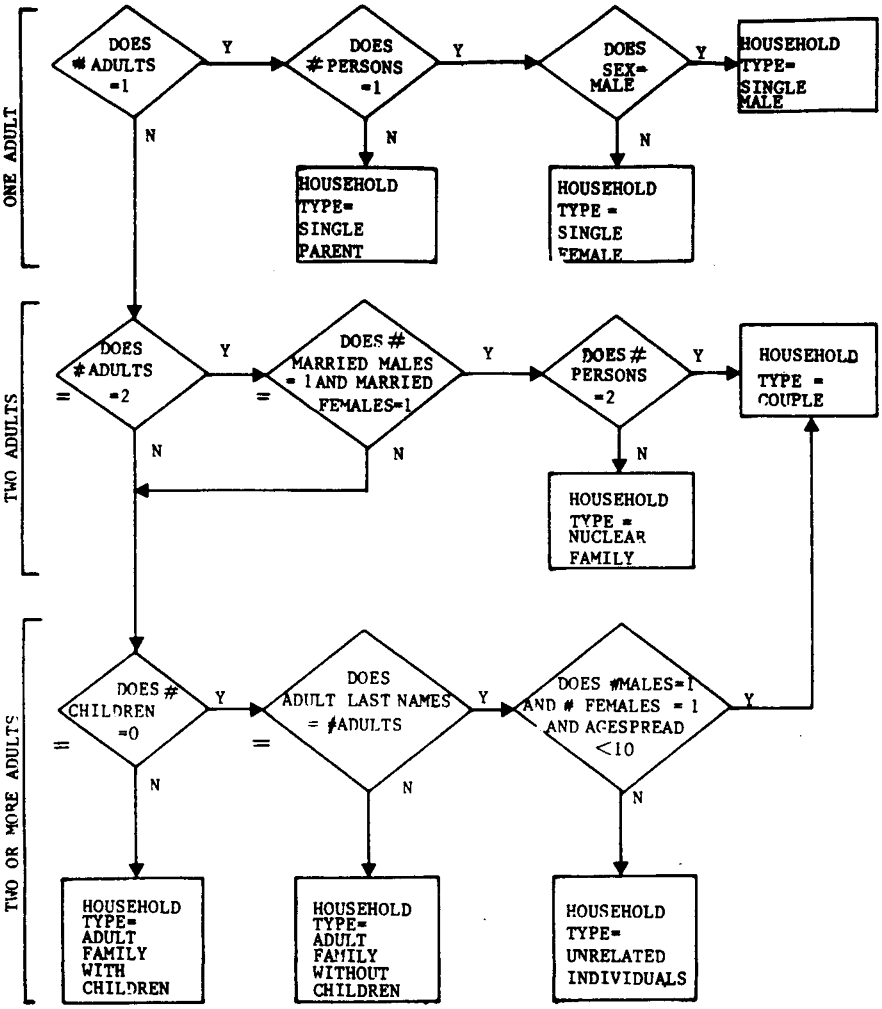
It was expected that these categories would have varying effects on trip rates. Adults living alone would be less mobility constrained than those adults living with children; but they would have none of the opportunities for trip coordination produced by living with other adult members. Single-parent families would have both increased mobility constraints as well as no opportunities for trip coordination, whereas couples would have the advantages of the opposite of both of these. An adult family would have further increased opportunities for trip coordination, but would perhaps differ from an adult household of unrelated individuals where individual activities would possibly be less influenced by other household members.

More specifically, when trip-generation rates are analyzed by purpose groups, differences between the trip-generation rates of these household categories would be expected. Those households with children would be expected to have a greater proportion of school trips and trips serving passengers than those households without children, whereas the latter would probably have a greater proportion of social-recreation trips.

CRA examined this household-structure concept by using Baltimore survey data with regression analysis, where the dependent variables were trip-generation rates by purpose mode, and the independent variables tested included, in addition to household structure, vehicles owned, income, number of persons older than 12, age structure of household, housing type, number of preschoolers present, number of gradeschoolers present, employment status, race, population per residential acre, a city limit classification, and length of residence at that address. The trip-purpose groups defined as the dependent variables were as follows: total home-based trips, home-based work trips, home-based shopping trips, home-based personal business trips, home-based entertainment and community trips, home-based visit and social trips, and home-based service and accompany-traveler trips.

CRA concluded that the household-structure variable was significant in predicting trip frequency. It should be noted, however, that the regressions were constrained to use all independent variables to permit comparability, even though varying numbers of independent variables were highly insignificant. Potentially, intercorrelations among variables could have masked some of the true underlying relation-

Figure 1. CRA flowchart of household typology.



SOURCE: Charles River Associates, 1980.

ships. CRA concluded that, of two commonly used trip-generation variables--number of vehicles owned and income--only number of vehicles owned out-performed the household-structure variable.

CASE STUDY

The analysis of travel data collected in the Midwest examined the household-structure concept. The data were collected from a stratified random sample of the population in seven counties (2). The principal purposes of the survey were to provide

1. The means to update trip-generation rates and modal-split models,
2. Attitudes of the population toward transportation and energy,
3. Attitudes toward possible changes in the transit system, and
4. Preferred methods of obtaining information on carpooling.

The data were collected by using an in-home interview and a 24-hr travel diary and included the variables age, gender, possession of a driver's license, employment status, and income of household members, all of which were available for use in trip-generation analysis.

The final data set consisted of 2,446 households. Of these households, the average household size was 2.9 persons per household, where less than 50 percent (1,656) of all households had two or less persons; 60 percent (1,483) had no children; and 53 percent (1,300) were two adult person households. In addition, almost 80 percent (1,952) of all households had at least one car available for use, and 30 percent (734) had more than one; 80 percent (1,875) occupied single-family dwellings; and 87 percent (2,124) of all households had at least one licensed driver. Seventy percent (1,724) of all households had at least one person employed, 63 percent (1,537) had at least one person employed full-time, and 60 percent (1,468) of all households had 1980 incomes greater than $15,000, with 14 percent (341) greater than $35,000.

The household-structure variable defined by CRA was derived from the data by the method shown in Figure 2. This differs slightly from the CRA flowchart because of the definition of the variables within the Southeastern Michigan Transportation Authority (SEMTA) data set. These differences include the following: (a) the cut-off age between children and adults is 18 years instead of 20, and (b) relationship codes were used to distinguish between adult families without children and households of unrelated adults; the last name of each person was not ascertained in the survey.

Figure 2. Flowchart of household typology used in analyzing SEMTA data.

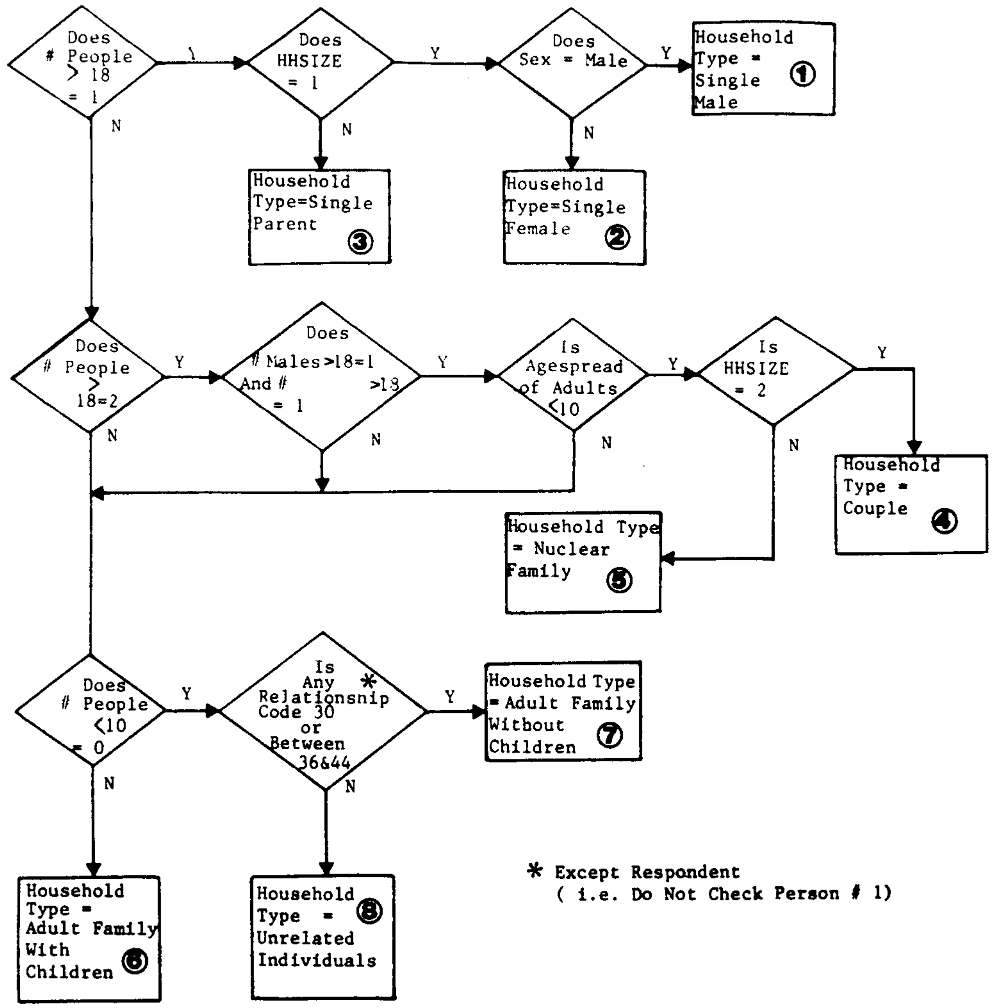


Table 1. Household-structure characteristics of SEMTA data.

Household-Structure Category	No. of Households	Percentage of Households
Single male	200	8.3
Single female	254	10.5
Single parent	150	6.2
Couple	502	20.8
Nuclear family	483	20.0
Adults with children	347	14.4
Adults with no children	420	17.4
Unrelated individuals	56	2.3
Missing	34	–
Total	2,446	

The final breakdown of the data into these household categories is given in Table 1. Almost 19 percent are single-person households, with slightly more single females than single males (2 percent). Single-parent households comprise only 6 percent, whereas couples and nuclear families comprise 21 and 20 percent, respectively. Adults with children make up slightly fewer households than those without children (14 percent compared with 17 percent), but households of unrelated individuals form the smallest category--2 percent of all households. Thirty-four households could not be classified. These included 17 single-person households where the person was younger than 18 years old.

To analyze the role of the household-structure variable in trip-generation analysis, this variable and seven other variables that were also thought to play a significant role in trip-generation rates were selected from the data set. The other variables selected were car ownership, household size, housing type, licensed drivers, household income, and total number of employed persons in the household (see Table 2). These eight variables were first analyzed by using one-way ANOVAs to determine how well they performed against the household-structure variable. Subsequently, the variables were analyzed by using one-way ANOVAs to determine the effects of varying grouping strategies on the categories within each variable.

The household-structure variable was grouped in three ways. The least-aggregate grouping combined the single-male and single-female categories because it was believed that there would be no significant difference between the overall tripmaking characteristics by gender, although there might be small differences for specific trip purposes. The least-aggregate grouping also combined nuclear families with adult families with children, based on the theory that additional adult members in the household would not significantly change the pattern of tripmaking. The second grouping strategy further combines all adult households, except single persons and couples. This assumes that adult households that consist of related persons will have little difference in tripmaking characteristics than those

Table 2. Variable grouping strategies used in SEMTA trip-generation analysis.

Variable Name	Grouping	No.	Categories Used in Grouping
LIFE I (household structure)	1-6	1	Single persons
		2	Single parents
		3	Couples
		4	Families with children
		5	Adult families without children
		6	Unrelated individuals
LIFE II (household structure)	1-5	1	Single persons
		2	Single parents
		3	Couples
		4	Families with children
		5	Other adult households with no children
LIFE III (household structure)	1-3	1	Single persons
		2	Families with children
		3	Households with no children
NUMCAR (number of cars available to household)	0-2	0	No cars available
		1	One car available
		2	Two or more cars available
HHSIZ I (household size)	1-5	1	One-person household
		2	Two-person household
		3	Three-person household
		4	Four-person household
		5	Five or more person household
HHSIZ II (household size)	1-4	1	One-person household
		2	Two- and three-person household
		3	Four-person household
		4	Five or more person household
HOUSTYP (housing type)	0-1	0	Multifamily
		1	Single family
TOTEMP (total number of employed persons)	0-2	0	No employed persons
		1	One employed person
		2	Two or more employed persons
INC80 (1980 household income)	1-3	1	\$0-\$14,999
		2	\$15,000-\$34,999
		3	>\$35,000
TOTLIC (total number of licensed drivers)	0-2	0	No licensed drivers
		1	One licensed driver
		2	Two or more licensed drivers

households that consist of unrelated individuals. Thus the theory of a coordination of tripmaking decisions between related household members was examined. The most severe grouping strategy separates households with children from households without children, identifying this characteristic as the most important in trip decision making. Only single-person households are further distinguished to reflect unique trip-generation characteristics.

Other variable groupings are also given in Table 2. The model II household size grouping, which combines two- and three-person households, was examined after initial analysis indicated little difference in trip rates of these households. Income was grouped into high-, medium-, and low-income categories.

Finally, MCA (3, and paper by Stopher and McDonald elsewhere in this Record) was used to compare different combinations of these grouped variables in trip-generation analysis. MCA derives trip rates within a standard trip-generation matrix by using deviations from the grand mean of the data set. Thus it improves on the traditional method of computing individual cell means because it permits estimation of trip rates for cells that contain no data. In addition, MCA, by using version 6, 7, or 8 of the Statistical Package for the Social Sciences (SPSS) (3), is able to take into account the interactive effects between independent variables where these variables have nonzero correlations with each other. This corrects for the overestimation of adjustments from the grand mean when these correlations are ignored. This use of MCA and the cross-classification structure is different from the CRA approach, which was to use least-squares regression analysis to predict the trip-generation measures. The effects of household structure were analyzed both in terms of the additional level of variance explained by the household-structure variable as well as the level of variance explained when substituting household structure for another variable.

The models examined in trip-generation analysis are given in Table 3. It can be seen that the number of vehicles (NUMVEH) available to the household was substituted for number of cars in some models because this variable performed significantly better across all purpose groups.

Table 3. MCA models used in SEMTA trip-generation analysis.

Trip Purpose	No.	MCA Models
Home-based work, home-based shopping, home-based social-recreation, home-based other, and non-home-based trips	1	NUMCAR, HHSIZ I, HOUSTYP
	2	NUMVEH, HHSIZ I, HOUSTYP
	3	NUMCAR, LIFE II, HOUSTYP
Home-based work, home-based shopping, home-based other, and non-home-based trips	1	NUMVEH, HHSIZ II
	2	NUMVEH, HHSIZ II, LIFE II
	3	NUMVEH, HHSIZ II, HOUSTYP
	4	NUMVEH, HHSIZ II, TOTEMP
	5	NUMVEH, HHSIZ II, INC80

In all three types of analysis previously discussed, trip-generation models were examined for motorized trips by specific trip purpose. Initial analysis distinguished social-recreation trips, but the final trip-purpose categories examined were home-based work, home-based shopping, home-based school, home-based other, and non-home-based trips. These final trip-purpose categories differ from the categories used by CRA that (a) do not examine non-home-based trips, and (b) break down the other category into more specific purpose groups.

DESCRIPTION OF RESULTS

The results of the ANOVA for ungrouped variables are given in Table 4; the results indicate that across all purpose groups the number of cars available to the household explains more variation than any other variable. This result is consistent with results obtained by CRA. Household size and housing type are the next most significant variables across all purpose groups; and whereas the number of employees in the household explains the most variation for home-based work trips, it does not perform well for all other purpose groups. Household structure and income appear to be of equal strength, although they perform better on different purpose groups. Income is most effective in explaining the total number of non-home-based trips, whereas household structure is most effective in explaining the number of home-based school trips. The licensed-driver variable ranks no better than third in explained variation for any purpose group.

The ANOVA results of the grouping strategies performed on the household-structure variable are given in Table 5. The most effective grouping is the model II grouping: single-person households, single-parent households, couples, other families with children, and other adult households. There appears to be little difference between the travel considerations of adult families that consist of related individuals and those that consist of unrelated individuals, because there is a large increase in the F-ratio across all purpose groups when these are combined, whereas the change in the within-group

Table 4. ANOVA results for ungrouped variables.

Variable	Original Category Values	Statistic	Purpose HBWORK	HBSCHL	HBOTHR	NHB
LIF8	1,8	F	65.6	112.7	50.9	23.6
		SS	5,711.1	6,062.2	19,116.3	17,464.1
		df				
		Between group	7			
		Within group	2,402			
NUMCAR	0,4	F	179.0	24.1	68.5	54.0
		SS	5,296.5	7,809.3	19,902.0	17,265.3
		df				
		Between group	4			
		Within group	2,440			
HHSIZ	1,8	F	60.3	181.8	62.0	29.5
		SS	5,842.1	5,334.5	18,794.9	17,327.0
		df				
		Between group	7			
		Within group	2,438			
HOUSTYP	0,1	F	125.4	22.6	73.9	52.9
		SS	6,195.7	7,842.4	20,792.9	18,008.8
		df				
		Between group	1			
		Within group	2,321			
TOTEMP	0,8	F	205.6	12.7	24.6	30.8
		SS	4,551.3	7,873.5	20,878.2	17,472.1
		df				
		Between group	6			
		Within group	2,439			
TOTLIC	0,8	F	112.3	30.3	57.6	39.3
		SS	5,006.4	7,383.5	18,624.0	16,649.3
		df				
		Between group	8			
		Within group	2,437			
INC80	1,7	F	148.4	15.8	32.7	41.0
		SS	5,020.3	7,814.4	20,493.5	17,075.2
		df				
		Between group	6			
		Within group	2,439			

Note: F = F-score, SS = sum of squares, and df = degrees of freedom. HBWORK = home-based work, HBSCHL = home-based school, HBOTHR = home-based other, and NHB = non-home-based trips.

Table 5. ANOVA results for grouped variables.

Variable	Statistic	Purpose HBWORK	HBSCHL	HBOTHR	NHB
LIFE I	F	85.5	148.8	69.1	30.3
	SS	5,745.0	6,150.0	19,195.0	17,557.7
	df				
	Between group	5			
	Within group	2,406			
LIFE II	F	110.2	186.0	86.2	37.6
	SS	5,748.9	6,150.0	19,200.0	17,564.8
	df				
	Between group	4			
	Within group	2,407			
LIFE III	F	128.4	354.4	135.6	61.8
	SS	6,146.4	6,221.0	19,730.9	17,752.0
	df				
	Between group	2			
	Within group	2,409			
NUMCAR	F	315.7	36.4	112.6	91.1
	SS	5,445.6	7,883.5	20,273.5	17,491.1
	df				
	Between group	2			
	Within group	2,443			
HHSIZ I	F	104.2	278.9	98.2	50.5
	SS	5,853.4	5,572.0	19,219.6	17,405.2
	df				
	Between group	7			
	Within group	2,438			
INC80	F	358.5	37.1	78.1	108.7
	SS	5,298.1	7,879.5	20,812.4	17,259.0
	df				
	Between group	2			
	Within group	2,443			

Note: Statistics and purposes are defined in Table 4.

variance is small. This contrasts with the model III grouping (single persons, families with children, and other families without children) where, although there is a large increase in the F-ratio across all purpose groups, and most particularly with home-based school trips, this is accompanied by a significant increase in the within-group variance.

The ANOVA results of the other grouped variables are also given in Table 5. It is clear that the number of cars available to the household remains the most significant variable in household trip-generation analysis. Once again, the F-scores are substantially greater across all purpose groups, even taking into account the difference in the degrees of freedom. Model II household size, which combines two- and three-person households, improves on model I household size by increasing substantially the F-ratio without increasing substantially the within-group variance. Household income (1980) is also effective in explaining trip-generation rates for all purpose groups except home-based school and home-based other, and thus may be useful when applied to specific trip-purpose models. The total number of licensed drivers, a variable that performed so poorly in earlier analyses, was not tested as a grouped variable.

The MCA results for the two sets of trip-purpose groups are given in Tables 6 and 7. From the first set of purpose groups (Table 6), the basic model consists of number of cars or vehicles available to the household and model I household size. Of the variables used as additions to this basic model, housing type clearly performs the best across all purpose groups. In addition, this model performs better than the model that uses number of cars, household structure, and housing type, where household structure is used as a substitute for household size, an alternative suggested by CRA (1). Further improvements are made by using number of vehicles available to the household instead of number of cars available.

The results of the models analyzed for the second set of trip purpose groups are given in Table 7. An initial examination of these MCA results gives the impression that the model that uses household structure, household size, and number of vehicles is the best model, particularly from an examination of the F-ratios. This is, however, a misleading impression. The F-ratio for an entire model is usually based on all main effects and interactions. If data are missing in some cells of the matrix that define the ANOVA problem, SPSS (3) is unable to calculate the interactions and computes an F-ratio on the main effects only. This F-ratio has substantially fewer degrees of freedom than one on the main effects and interactions, and therefore it must be a larger numeric value for the same significance level.

The household-structure model generated empty cells for some combinations of household structure, household size, and vehicle availability (e.g., the household structure of a couple can occur only for two-person households) and resulted in suppression of interactions in the ANOVA. The model that uses household structure is the only model in Table 7 for which this happened, and leads to an inflated F-ratio compared with all other models. When F-ratios are calculated on main effects only for the other models (as indicated by a footnote in Table 7), the F-ratios are almost all larger than those for the household-structure model. Thus the addition of household structure to the basic model of number of vehicles available to the household and household size does not improve its performance for any trip-purpose group.

Of the other variables examined as additions to the model, the total number of workers in the household improves the model for home-based work trips. Household income (1980) and the model II household-size variable are both improvements over the household-structure variable. Income is better in explaining home-based work trips and non-home-based trips, and housing type is better in explaining the other trips. Thus, unless a separate model is developed for home-based work trips by using the employment variable, the model of number of vehicles per household, household size, and housing type still remains the best approach. These conclusions support those found with the previous set of purpose groups, with the exception that the model II household size performs better than, and thus replaces, the model I household size.

CONCLUSIONS

In the trip-generation analysis of the case study data, the household-structure variable did not per-

Table 6. MCA results of set I models used in analyzing SEMTA data.

Model	Statistic	Purpose				
		HBWORK	HBSHOP	HBSOC	HBOTHR	NHB
NUMCAR, HHSIZ I, HOUSTYP	F	29.5	5.9	6.0	35.1	11.4
	df					
	Between group	29				
	Within group	2,292				
	SIG	0.000	0.000	0.000	0.000	0.000
	R^2	0.261	0.062	0.060	0.295	0.113
NUMVEH, HHSIZ I, HOUSTYP	F	29.1	5.6	5.0	35.5	11.5
	df					
	Between group	29				
	Within group	2,244				
	SIG	0.000	0.000	0.000	0.000	0.000
	R^2	0.261	0.060	0.053	0.298	0.116
NUMCAR, LIFE II, HOUSTYP	F	28.3	5.2	5.1	26.4	9.5
	df					
	Between group	29				
	Within group	2,259				
	SIG	0.000	0.000	0.000	0.000	0.000
	R^2	0.254	0.056	0.054	0.238	0.096

Note: SIG = significance, HBSHOP = home-based shopping, HBSOC = home-based social-recreation, and the rest are defined in Table 4.

Table 7. MCA results of set II models used in analyzing SEMTA data.

Model	Statistic	Purpose: HBWORK	HBSHOP	HBOTHR	NHB
NUMVEH, HHSIZ II	F	74.1	14.0	15.9	29.1
	df				
	Between group	11			
	Within group	2,434			
	SIG	0.000	0.000	0.000	0.000
	R^2	0.246	0.057	0.060	0.109
NUMVEH[a], HHSIZ II, LIFE II	F	94.1	16.5	21.0	34.0
		94.1[b]	16.5[b]	21.0[b]	34.0[b]
	df				
	Between group	9			
	Within group	2,402			
	SIG	0.000	0.000	0.000	0.000
	R^2	0.261	0.058	0.073	0.113
NUMVEH, HHSIZ II, HOUSTYP	F	34.8	7.0	7.3	13.9
		127.2[b]	24.9[b]	24.2[b]	47.3[b]
	df				
	Between group	23			
	Within group	2,299			
	SIG	0.000	0.000	0.000	0.000
	R^2	0.246	0.061	0.059	0.108
NUMVEH, HHSIZ II, TOTEMP	F	38.7	5.0	5.5	9.7
		176.2[b]	19.2[b]	14.2[b]	38.6
	df				
	Between group	33			
	Within group	2,268			
	SIG	0.000	0.000	0.000	0.000
	R^2	0.348	0.057	0.055	0.105
NUMVEH, HHSIZ II, INC80	F	32.4	5.2	5.6	10.4
		148.1[b]	21.1[b]	22.3[b]	44.3[b]
	df				
	Between group	33			
	Within group	2,412			
	SIG	0.000	0.000	0.000	0.000
	R^2	0.298	0.057	0.060	0.112

Note: Statistics and purposes defined in Tables 4 and 6.
[a]Interactions suppressed. [b]F-ratios calculated on main effects only.

form as well as was expected from the CRA analysis of Baltimore data. This may, however, be a result of the different methodologies that were used in the two analyses. The analysis reported in this paper applied traditional cross-classification models that used MCA to predict cell-by-cell trip rates. The final model consisted of number of vehicles, household size, and housing type. However, subsequent analysis not discussed in this paper has revealed that the use of an area-type variable instead of housing type may improve the models even further.

Figures 3 and 4 show the results of the automatic interaction detection (AID) analysis performed on 1973 Niagara Frontier Transportation Committee (Buffalo) and a 1974 Genesee transportation travel survey (Rochester) data for all trips and for home-based nonwork trips (4). The number of vehicles represents the first cluster. This supports both the conclusions drawn by CRA and by the authors. This is followed by number of children (usually a function of household size) and age of the oldest child. The final clusters are based on household size, vehicles per licensed driver (a function of both vehicles per household and household size), household employment status, and number of vehicles available to the household. Although the various age classifications may be a function of household structure, they may also be a function of other variables (for example, household size).

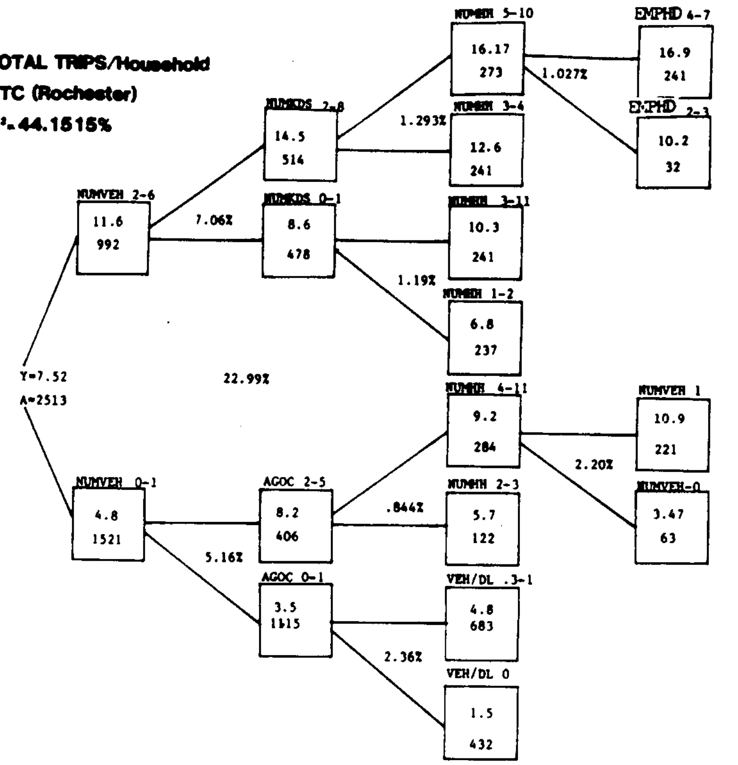
It is also pertinent to note that even had the household-structure variable performed satisfactorily in this trip-generation analysis, there would be problems implementing it in trip-generation models. When presented with a possible trip-generation design that used the household-structure variable, a metropolitan planning organization (MPO) was reluctant to implement it. Although CRA stated that the household-structure variable could be easily obtained from census data, the MPO expressed doubts that it could be. Forecasting at a zonal level, particularly to obtain distribution of households by household-structure category, appears fraught with problems. Possibly, forecasts could be made at the regional level of the constituent elements of household structure, but current analysis-zone forecasts in most metropolitan areas do not include these components and would possibly be difficult to add to current forecasts. In addition, household structure cannot be used as a policy variable, whereas other variables, particularly housing type, could be used. This also helped in the decision to exclude the household-structure variable from the SEMTA trip-generation models.

REFERENCES

1. P.M. Allaman, T.J. Tardiff, and F.C. Dunbar. New Approaches to Understanding Travel Behavior. NCHRP, Rept. 250, Sept. 1982, 147 pp.
2. R. Parvataneni, P. Stopher, and C. Brown. Origin-Destination Travel Survey for Southeast

Figure 3. AID analysis of Rochester survey data.

TOTAL TRIPS/Household
GTC (Rochester)
R^2=44.1515%



WHERE
NUMVEH = NUMBER OF VEHICLES AVAILABLE TO THE HOUSEHOLD
NUMKDS = NUMBER OF CHILDRED PER HOUSEHOLD
NUMHH = HOUSEHOLD SIZE
AGOC = AGE OF OLDEST CHILD
VEH/DL = VEHICLES PER LICENSED DRIVER
OCCUP = OCCUPATION
EMPHD = HOUSEHOLD EMPLOYMENT STATUS
LOC = LOCATION

Home Based - Non Work Trips
GTC
R^2=41.9916%

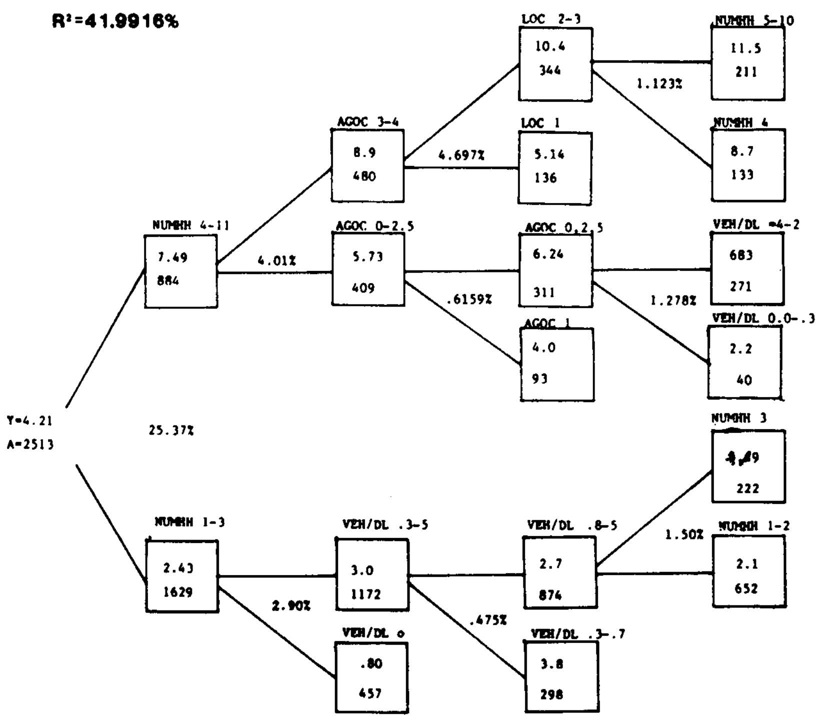
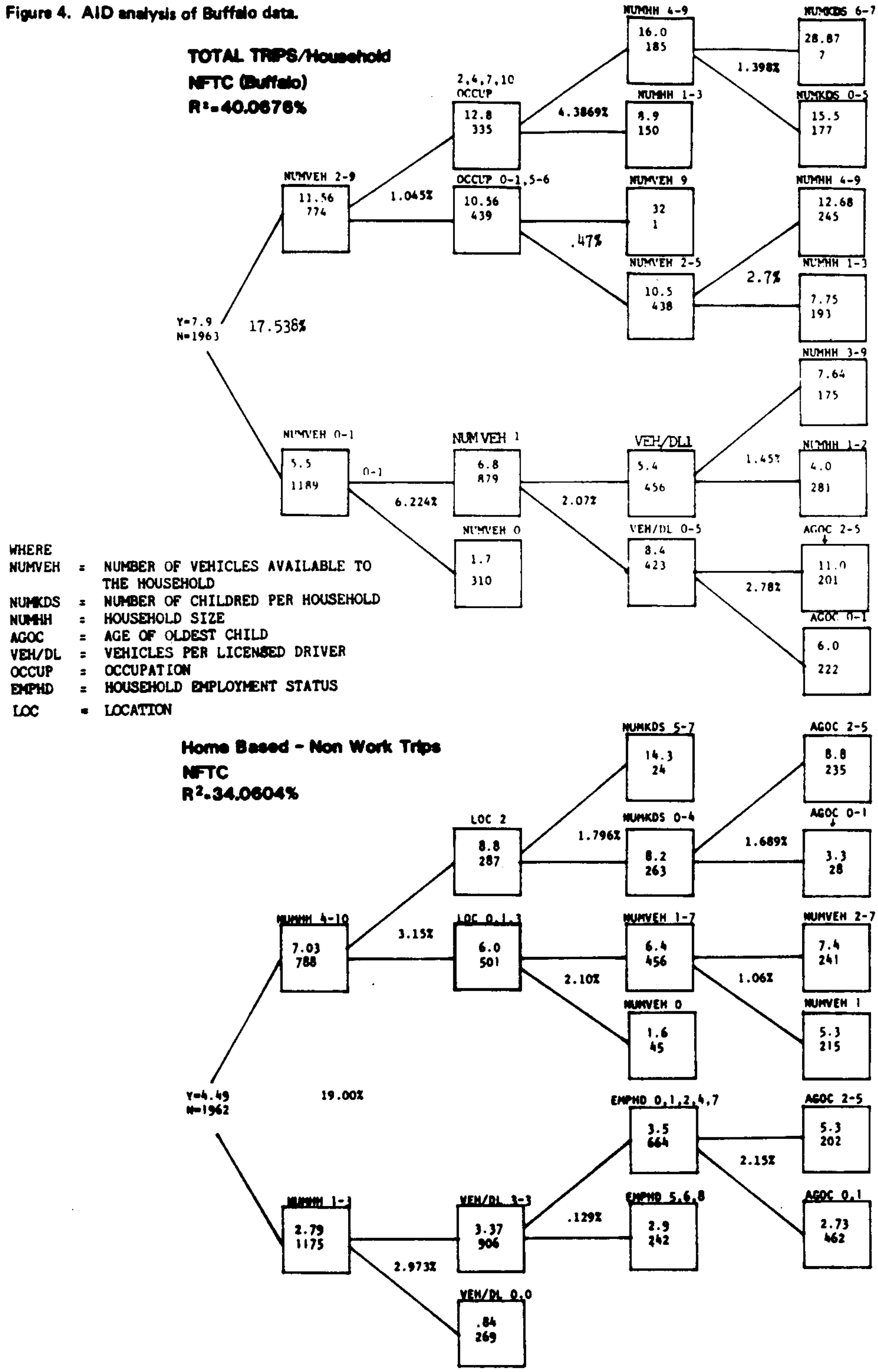


Figure 4. AID analysis of Buffalo data.



Michigan. TRB, Transportation Research Record 886, 1982, pp. 1-8.

3. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent. Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.
4. J. Chicoine. The Life Cycle Concept: A Practical Application to Transportation Planning. New York State Department of Transportation, Albany, Draft Paper, Aug. 1982.

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

Maximum-Likelihood and Bayesian Methods for the Estimation of Origin-Destination Flows

ITZHAK GEVA, EZRA HAUER, AND UZI LANDAU

The design of traffic management schemes usually requires knowledge of the pattern of trips on the system under scrutiny. This pattern is ordinarily described by an origin-destination (O-D) flow matrix. One common task of this type of matrix is the estimation of flows between the intersection approaches on a stretch of road. Estimation is based on intersection flow counts that are supplemented by a license-plate survey. In this paper a procedure is developed to obtain the most likely O-D flow estimates by using both intersection counts and results of the license-plate survey. The procedure is described in detail on the basis of a numerical example. An earlier paper reported a method of estimation that relies on intersection counts only and does not require the conduct of a sample license-plate survey. An empirical examination is conducted to test how estimation accuracy increases when the added information from the license-plate survey is used. This examination reveals that when the supplementary license-plate survey is small, the maximum-likelihood method yields unsatisfactory estimates. This deficiency is rectified by the use of a Bayesian method. The resulting solution procedure is simple, and satisfactory estimates are produced.

A variety of transportation planning and management tasks require the knowledge of the pattern of trip flows between origins and destinations. This pattern is usually described by an origin-destination (O-D) flow matrix. One common task of this type of matrix is the estimation of flows between the intersection approaches on a stretch of road. The estimation is based on a license-plate survey that is factored up to match counts of intersection flows.

In recent years attention has been given to the problem of estimating an O-D matrix by using traffic counts as the main source of information (1-4). A recent paper (5) describes a method that departs from previous work, in that travel behavior is brought into estimation by information contained in small O-D samples obtained by a survey. It is therefore not necessary to rely on speculative microstates (as in entropy models) or to assume that actual route choice is correctly captured by available models. Rather, the purpose is to find that matrix of O-D flows that is consistent with the observed traffic counts and that is most probable in view of the O-D samples observed.

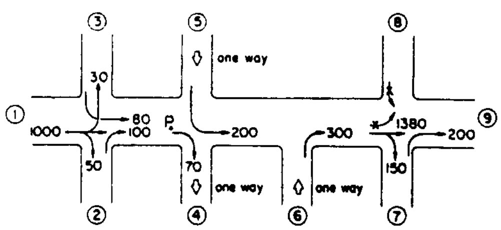
This approach is used in the present paper, in which a procedure to estimate flows between the intersection approaches on a stretch of road is developed based on intersection flow counts and a license-plate survey. The effect of sample size on estimation accuracy is explored in a real-life example.

In the first section of the paper two alternative likelihood models, which capture the manner in which data are obtained in the field, are presented. The normal equations that identify the maximum-likelihood estimate are obtained, and an algorithm for their numerical solution is described. A numerical example is presented in the second section. The example is intended to illustrate the how-to of the method and to assist the practitioner in its application. As noted earlier (6), estimates of O-D flows can be obtained from traffic counts alone, without having to resort to tedious license-plate surveys. The increase in estimation accuracy obtained as a function of sample size is examined in the third section. The results of this examination lead to the development of a new procedure based on Bayesian statistics. This procedure is presented and examined in the fourth section.

PROBLEM FORMULATION AND SOLUTION

Consider a street section as shown in Figure 1. The intersection approaches are thought of as origins and destinations. Estimates of O-D flows are desired.

Figure 1. Example of street section with eastbound flows.



The method most commonly used for this purpose in traffic engineering practice is to count traffic volumes at every intersection and to conduct a license-plate survey of a sample of vehicles entering and exiting the street of interest. Usually several digits of the license-plate number are recorded and later matched so as to obtain a sample O-D pattern. The sample is later factored up in an attempt to make the appropriate sums of O-D flow estimates match the corresponding volume counts. The purpose of this paper is to suggest an estimation procedure to replace the arbitrary and often ambiguous factoring. The merit of the procedure is that it identifies the O-D flows that are most likely in view of the results of the license-plate survey and the intersection volume counts.

In formulating the problem, the following basic notations are used:

- O_i = number of vehicles entering the street at entry approach i ($i = 1,2,\ldots,m$) during a specified period of time,
- D_j = number of vehicles leaving the street at exit approach j ($j = 1,2,\ldots,n$) during the same period of time,
- t_{ij} = number of license plates matched between records obtained at entry approach i and exit approach j, and
- T_{ij} = number of vehicles that enter the street by approach i and exit it by approach j.

The objective of the exercise is to obtain estimates of T_{ij} by using the data O_i, D_j, and t_{ij}. The estimation logic is of the customary maximum-likelihood kind. Thus the O-D sample matrix (t_{ij}) obtained from license-plate matching is thought to be a random sample drawn from the matrix of O-D flows (T_{ij}). The probability of observing this sample can be captured by an appropriate mathematical model. A search is made for the estimates of T_{ij} that maximize this probability and at the same time fit all the intersection volume counts. These are the most likely O-D flows to have prevailed at the time of the license-plate survey and intersection volume counts.

Two points deserve mention. First, for traffic planning and management purposes, O-D flow estimates are needed that represent average conditions rather than estimates of flows that have prevailed at the time of the survey. To do so, t_{ij}, and also O_i and D_j, would have to be regarded as random variables (7). Because the focus in this paper is the effect of the sampling ratio for the license-plate survey on O-D estimation accuracy, the estimation of O-D flows that prevailed at the time of the survey are sought. This is what practitioners have been doing anyway. The second point has to do with a discrepancy between the model and the practicalities of traffic surveys. In the model the analyst pretends that the intersection volume counts, as well as the license-plate survey, are conducted during the same time period. But because of personnel limitations, this is seldom true. With these qualifications, the random nature of the license-plate sample is described by using an appropriate probability model.

The probability model chosen must fit the manner in which the random sample is selected from the population. Thus the essential details of the license-plate survey procedure used have to be stated. To reduce survey personnel requirements and to keep errors of recording in check, it is usually best to specify beforehand some part of the license-plate number to serve as the sampling criterion. Thus if all even-numbered plates are recorded, the sampling ratio is 50 percent; if all plates ending with the digit 0 are recorded, the sampling ratio is 10 percent; and so forth. Provided that the digits selected to serve as a sampling criterion are uniformly distributed in the population of license plates, the sampling ratio is established when the sampling criterion is specified.

Two alternative probability models are suggested to capture the stochastic nature of this survey procedure. First, each license-plate match can be viewed as a success of a Bernoulli trial in which the probability of success is dictated by the sampling ratio and the rate of errors of recording and coding. The unknown flows T_{ij} correspond here to the number of Bernoulli trials. Thus the likelihood function is a product of binomial probability mass functions. Second, the license plates recorded at a certain survey point can be viewed as a random sample drawn (with replacement) from the constituent O-D flows passing that point. This leads to the multinomial probability model. Both models are considered and their merits are discussed.

Starting with the binomial model, let U_{ij} denote the number of license plates within T_{ij} that satisfy the sampling criterion. The probability distribution of U_{ij} can be described by the binomial model. Thus

$$p(U_{ij}) = \binom{T_{ij}}{U_{ij}} r^{U_{ij}} (1-r)^{T_{ij}-U_{ij}} \tag{1}$$

where r is the sampling ratio.

Equation 1 would be a reasonable description of the state of affairs if observers in the field were able to record all license plates that should be recorded and do so without error. In reality, errors occur. Thus instead of obtaining U_{ij} matching license plates for a stream of vehicles, only $t_{ij}(t_{ij} \leq U_{ij})$ is obtained.

Now the conditional probability mass function (PMF) of t_{ij} is given by

$$p(t_{ij}|U_{ij}) = \binom{U_{ij}}{t_{ij}} q^{t_{ij}} (1-q)^{U_{ij}-t_{ij}} \tag{2}$$

where q is the probability that nothing goes wrong and the license plate is obtained and processed correctly at both entry and exit.

This case is known in the literature as partial ascertainment (8). In such a case the original distribution will be distorted. If the model underlying the partial destruction of original observations (or the survival distribution) is known, the distribution of the observed values can be derived. It was shown that where the original distributions are Poisson, binomial, or negative binomial, the modified distribution is of the same form.

Therefore, the PMF of t_{ij} is also binomial and given by

$$p(t_{ij}) = \binom{T_{ij}}{t_{ij}} (rq)^{t_{ij}} (1-rq)^{T_{ij}-t_{ij}} \tag{3}$$

An expression analogous to Equation 3 can be written for every possible flow. It can be shown (9) that if $X_1,\ldots,X_t$ are binomial variates with sample size $N_1,\ldots,N_t$, respectively, and a common probability of success in each trial, then the distribution of $X = (X_1,\ldots,X_t)$ conditional on $\sum_{i=1}^{t} X_i = n$ is multivariate hypergeometric with parameters n, N, and $(N_1 \ldots N_t)$. Therefore, the probability of obtaining a matrix of (t_{ij}) if the matrix of flows is (T_{ij}) is given by

$$p(t_{ij}) = \left[\prod_{i=1}^{m} \prod_{j=1}^{n} \binom{T_{ij}}{t_{ij}}\right] \Big/ \binom{\sum_i \sum_j T_{ij}}{\sum_i \sum_j t_{ij}} \tag{4}$$

The identification of the array T^*_{ij} for which this probability (or the logarithm of this probability) is maximum is needed. However, the solution must satisfy the traffic count constraints

$$\sum_{j=1}^{n} T_{ij} = O_i \quad \text{for } i = 1, 2, \ldots, m \tag{5}$$

and

$$\sum_{i=1}^{m} T_{ij} = D_j \quad \text{for } j = 1, 2, \ldots, n \tag{6}$$

By forming the Lagrangean, using Stirling's formula, taking derivatives, and equating to zero (6), the following equation is formulated:

$$T^*_{ij} = t_{ij}/(1 - A_i B_j) \quad \begin{array}{l} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{array} \tag{7}$$

To obtain numerical values for the estimates T^*_{ij}, the unknown values $A_1, A_2, \ldots, A_m$ and $B_1, B_2, \ldots, B_n$ first must be found. This can be accomplished by a simple algorithm described in the next section.

The alternative manner of describing the survey by a probability model is to consider the random sample $t_{i1}, t_{i2}, \ldots, t_{in}$ obtained at station i as drawn from the flows $T_{i1}, T_{i2}, \ldots, T_{in}$, which are unknown. Only their sum (O_i) is given. The probability of observing this sample is given by the multinomial model:

$$\left[\left(\sum_j t_{ij}\right)! \Big/ \prod_j (t_{ij}!)\right] \prod_j (T_{ij}/O_i)^{t_{ij}} \tag{8}$$

(The multinomial model is only approximate because it assumes sampling with replacement. As long as the sample is a small fraction of the population, the assumption appears proper.)

Accordingly, the probability of observing all (t_{ij}) when the matrix of flows is (T_{ij}) is given by

$$\prod_{i=1}^{m} \left\{ \left[(\sum_j t_{ij})! / \prod_j (t_{ij}!) \right] \prod_j (T_{ij}/O_i)^{t_{ij}} \right\} \qquad (9)$$

The solution must satisfy the same constraints (Equations 5 and 6). By forming the Lagrangean and taking derivatives (5), the following equation is given:

$$T^*_{ij} = t_{ij}/(A_i + B_j) \qquad \begin{array}{l} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{array} \qquad (10)$$

The next task is to solve a system of $(m + n)$ simultaneous nonlinear equations with $(m + n)$ unknowns: $A_1, \ldots, A_m$; $B_1, \ldots, B_n$. The simplest solution algorithm consists of repeated balancing of the vectors A_i and B_j and is named after Kruithof (10). The algorithm is described and illustrated by a numerical example in the following section.

NUMERICAL EXAMPLE

To illustrate the procedure, consider the road section described in Figure 1, on which the eastbound flows are obtained from ordinary intersection counts. A license-plate survey is conducted with a sampling ratio of 50 percent $(r = 0.5)$. To achieve this sampling ratio, only vehicles with even license numbers were recorded. The number of vehicles that were matched in the survey (t_{ij}) are shown in the upper left corner of each of the 16 cells in Figure 2.

Figure 2. O-D matrix corresponding to street section in Figure 1.

From \ To	2	3	4	7	9	Σ	A_i
1	16 / 50	10 / 30	14 / 56	28 / 79	246 / 785	1000	1.0000
2			2 / 10	5 / 15	21 / 75	100	1.0494
3			1 / 4	2 / 6	21 / 70	80	1.0189
5				8 / 22	55 / 178	200	1.0047
6				10 / 28	86 / 272	300	0.9961
7					41 / 200	200	
Σ	50	30	70	150	1580	1880	
B_j			0.7513	0.6430	0.6867		

The flows T_{12}, T_{13}, and T_{79} are 50, 30, and 200, respectively, because these values can be obtained directly from the counts. Therefore, the estimation problem consists of the 13 empty cells that have to be filled with estimates so as to satisfy the 8 row and column sums. These sums are listed under the heading Σ and obtained from the intersection counts.

The solution algorithm begins by obtaining initial estimates of A_i. A starting guess may be $A_i = 1.0$. By using these tentative values for A_i, the first estimates of each B_j can be obtained. For example, for $j = 4$, the sum $T_{14} + T_{24} + T_{34}$ must be 70. Thus by using Equation 7,

$$[14/(1 - B_4)] + [2/(1 - B_4)] + [1/(1 - B_4)] = 70.$$

In this case $B_4 = 0.7571$. The values of B_7 and B_9 are obtained similarly by using Equation 7 to fit the given sums of columns 7 and 9. Then new estimates for A_i are calculated from the given sums of the appropriate rows and the current estimates of B_j. The new estimates of A_i are compared with the previous ones. Unless the desired closure is attained, a new round of computations is carried out. In this example, after a few iterations, the solution in Figure 2 is reached. The values of A_i and B_j are shown in the rightmost column and the lowest row, respectively. The final estimates of T_{ij} are shown in the lower right corner of each of the 13 cells. (A listing of a FORTRAN program for this procedure is available.)

The solution for the multinomial model (Equation 10) is obtained by the same algorithm. Both models produced slightly different results, which vanish after rounding to integers. Therefore, it is immaterial which model is used for the estimation.

ESTIMATION ACCURACY AND EFFECT OF SAMPLING RATIO

One of the purposes of this work has been to explore the accuracy of estimates obtainable by the method as a function of sample size. This is done empirically by comparing estimates obtained when different sampling ratios are used with the 100 percent sample. The information was provided by a detailed license-plate survey conducted on a section of a four-lane collector road with five intersections in Toronto. In the survey four digits of the license-plate code were recorded for 2 hr. The matched license-plate records were converted into O-D flows. For this investigation, these results were considered as the true matrix. It had to be pretended first that the survey was conducted with different sampling ratios by considering only license plates ending with certain digits. Flow estimates obtained by the suggested method are then compared with the true matrix.

Estimates were obtained for different sampling ratios and also for the case of zero sample [i.e., from the traffic counts only by the method described by Hauer and Shin (11)]. The error measure chosen was the average absolute error (AAE), which is defined as follows:

$$AAE = (1/N) \sum_{ij} |T^*_{ij} - T_{ij}| \qquad (11)$$

where

T^*_{ij} = estimated flow from i to j,
T_{ij} = true flow from i to j, and
N = number of nonzero cells.

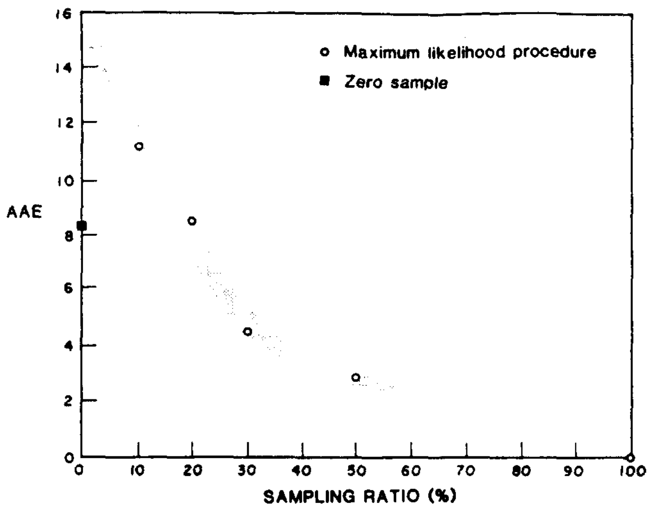
The results are shown in Figure 3 (similar results were found when other error measures were used). Some observations follow.

First, as expected, estimation accuracy increases with sample size. Initially, the improvement in accuracy is considerable. As higher sampling ratios are reached, the law of diminishing returns exerts strong influence.

Second, even without an O-D sample, reasonable flow estimates can be obtained. In this case none of the models described here can be used. The analyst has to rely on the assumption of equally likely microstates and use the method described by Hauer and Shin (6,11). The accuracy of estimation in this case (sampling ratio = 0) is shown by a square and is comparable to what can be obtained by using Equations 7 and 10 with a 20 percent O-D sample.

The reason for the unsatisfactory performance when the sample is small is inherent in Equations 7 and 10. When the flow between an O-D pair (ij) is not captured by the sample (i.e., $t_{ij} = 0$), then, of necessity, the estimate $T^*_{ij} = 0$. When the sample of license plates recorded is sufficiently small for this to occur often, estimation accuracy is likely to suffer. Thus it is not so much the sampling

Figure 3. Effect of sampling ratio on AAE of estimated matrix using the maximum-likelihood procedure.



ratio as the absolute sample size that governs estimation accuracy. When the sample size is small, the analyst can do better by ignoring it altogether because it uncovers a deficiency in the maximum-likelihood method of estimation described in the section Problem Formulation and Solution; it forces the analyst to assign zero values to flows, even though it is known that this is highly unlikely to be a satisfactory estimate. It is unwise to disregard this prior knowledge. A method that makes use of both the prior knowledge and the information contained in the O-D sample should be sought. The next section is aimed at developing such a procedure that bridges the existing discontinuity and improves estimation accuracy when relatively small samples are used.

BAYESIAN APPROACH TO ESTIMATION

The essence of Bayesian methods (12) is to apply the information contained in the outcome of an experiment to the knowledge about the probability distribution of some parameters that are available before the experiment in order to generate a new, posterior probability distribution function about these parameters.

In the present case the experiment is the license-plate survey that yields the sample realizations (t_{ij}). The prior probability distribution, denoted by $p^o(T_{ij})$, describes the probability of obtaining the matrix of flows T_{ij}. With this, and using Bayes' theorem, the posterior probability is given by

$$p(T_{ij}) \propto p(t_{ij}|T_{ij})p^o(T_{ij}) \quad (12)$$

The conditional probability component of Equation 12 has already been stated by Equation 4 (for the binomial model) or Equation 9 (in the case of the multinomial model). Thus the prior probability distribution component $p^o(T_{ij})$ must be specified.

In the absence of other information, it may be assumed that the probability of observing a certain matrix (T_{ij}) is proportional to the number of elementary events (microstates) from which it can arise (6). If all elementary events are equally likely, it can be shown that

$$p^o(T_{ij}) \propto \left[1/\left(\prod_{i=1}^{m}\prod_{j=1}^{n} T_{ij}!\right)\right] \quad (13)$$

Therefore, the posterior probability distribution function can be written as

$$p(T_{ij}) \propto \prod_{i=1}^{m}\prod_{j=1}^{n} [1/(T_{ij} - t_{ij})!] \quad \text{(binomial model)} \quad (14)$$

or

$$p(T_{ij}) \propto \prod_{i=1}^{m}\prod_{j=1}^{n} (T_{ij}^{t_{ij}}/T_{ij}!) \quad \text{(multinomial model)} \quad (15)$$

From here on, the procedure follows the logic explained in the section Problem Formulation and Solution. A search is made for that matrix T_{ij}^* that makes the posterior probability in Equations 14 and 15 as large as possible. Again, by using the method of Lagrange multipliers and Stirling's approximation,

$$T_{ij}^* = t_{ij} + A_iB_j \quad \text{(binomial model)} \quad (16)$$

and

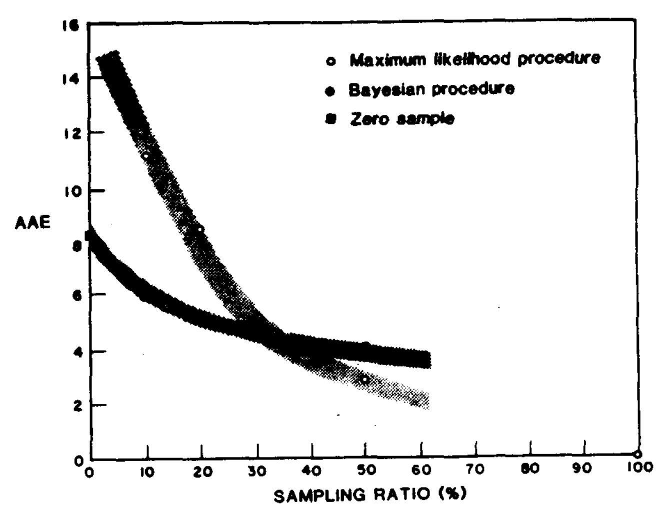
$$T_{ij}^* = \exp(t_{ij}/T_{ij}^*)A_iB_j \quad \text{(multinomial model)} \quad (17)$$

Note that when $t_{ij} = 0$, both equations produce the same result ($T_{ij}^* = A_iB_j$), which is also the general solution for zero sample (6,11). In this manner the discontinuity problem near the origin (Figure 3) is eliminated.

Examination of Equations 16 and 17 reveals that the first is easily solved. Equation 17 requires a complex iterative algorithm. Both equations were used to obtain O-D estimates for the case of the street section described in the previous section. For sampling rates of up to 50 percent, both models produced almost identical estimates. For higher sampling rates, however, there is a difference between them. This can be illustrated by considering the extreme case of a 100 percent sample. At this point, Equation 16 gives the natural result $T_{ij}^* = t_{ij}$ (which is the same as Equations 7 and 10). However, Equation 17 leads to different estimates.

The effect of sampling rate on the level of accuracy, by using the maximum-likelihood procedure (Equations 7 or 10) and the Bayesian procedure (Equations 16 or 17), is presented in Figure 4. It can be seen that for sampling rates of up to 30 percent, the Bayesian method improves estimation accuracy. The maximum-likelihood procedure is appropriate for the higher sampling rates.

Figure 4. Effect of sampling ratio on AAE of estimated matrix using the maximum-likelihood and the Bayesian procedure.



SUMMARY

Two coherent methods for the estimation of O-D flows from traffic count and license-plate survey information are presented. The first estimation method identifies the most likely set of flows that agrees with the observed intersection approach flow counts on a stretch of road and the results of a sample license-plate survey.

The effect of the sample size on the accuracy of O-D flows obtained by this procedure is examined by using data from a comprehensive license-plate survey conducted on a stretch of road in Toronto. As was expected, accuracy increases with sample size. However, for small samples, better accuracy can be obtained by estimating from traffic counts only. Therefore, a second procedure based on the Bayesian approach has been developed. This procedure significantly improves the accuracy of O-D flow estimates obtained from traffic count and small sample license-plate survey information. The procedure is capable of producing relatively satisfactory estimates from small samples and thus is an aid in the performance of a common task in practice.

It appears that this procedure is preferable because of its consistency and capability, whereas the maximum-likelihood procedure should be used when high sampling rates are available at all survey stations. The Bayesian procedure described here was applied only to simple systems, such as street sections, freeway sections, and subway or bus lines. Further research is required for the application of the procedure to cases in which there are multiple paths between an O-D pair.

REFERENCES

1. L.G. Willumsen. Estimation of an O-D Matrix from Traffic Counts--A Review. Institute for Transport Studies, Univ. of Leeds, Leeds, England, Working Paper 99, 1978.
2. V.K. Chan, R.G. Dowling, and A.E. Willis. Deriving O-D Information from Routinely Collected Traffic Counts: Phase 1, Review of Current Research and Application. Institute of Transportation Studies, Univ. of California, Berkeley, 1980.
3. Y. Gur et al. Determining an O-D Trip Table Based on Observed Volumes. Presented at the Operation Research Society of America (ORSA) and The Institute of Management Science (TIMS) Annual Meeting, New York, 1978.
4. H.J. Van Zuylen and L.G. Willumsen. The Most Likely Trip Matrix Estimated from Traffic Counts. Transportation Research, Vol. 14B, 1980, pp. 281-293.
5. U. Landau, E. Hauer, and I. Geva. Estimation of Cross-Cordon Origin-Destination Flows from Cordon Studies. TRB, Transportation Research Record 891, 1982, pp. 5-10.
6. E. Hauer and B.T. Shin. Estimation of O-D Matrices from Observed Flows: Simple Systems. Univ. of Toronto/York Univ. Joint Program in Transportation, Toronto, Ontario, Canada, Res. Rept. 65, 1980.
7. H.R. Kirby and J.D. Murchland. Gravity Model Fitting with Both Origin-Destination Data and Modeled Trip Engineering Estimates. Proc., 8th International Symposium on Transportation and Traffic Theory, Toronto, Ontario, Canada, June 1981.
8. C.R. Rao. On Discrete Distribution Arising Out of Methods of Ascertainment. In Classical and Contagious Distributions (G.P. Patil, ed.), Statistical Publishing Society, Calcutta, 1965.
9. Y.M.M. Bishop, S.E. Feinberg, and P.W. Holland. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Mass., 1975.
10. J. Kruithof. Calculation of Telephone Traffic. De Ingenieur, Vol. 52, No. 8, 1937.
11. E. Hauer and B.T. Shin. Origin-Destination Matrices from Traffic Counts: Application and Validation on Simple Systems. Traffic Engineering and Control, Vol. 22, No. 3, 1981.
12. L.D. Phillips. Bayesian Statistics for Social Scientists. Thomas Nelson and Sons, London, England, 1972.

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

Trip Table Synthesis for CBD Networks: Evaluation of the LINKOD Model

ANTHONY F. HAN and EDWARD C. SULLIVAN

Origin-destination (O-D) synthesis methods deal with the problem of deriving trip O-D patterns from traffic counts. A reliable O-D synthesis model for small area [e.g., central business district (CBD)] applications has great potential to help evaluate alternative transportation system management measures. Among various models reviewed, LINKOD was selected for in-depth evaluation because of its apparent suitability for CBD applications. A 1975 San Jose, California, CBD O-D data set with traffic counts estimated by the microassignment model was used to test the performance of LINKOD. Significant differences were found between the synthesized trip table and the base trip table; nevertheless, when assigned to the network by the microassignment model, both trip tables predicted similar flow patterns. Based on these tests, LINKOD was judged to be an acceptable tool for pragmatic applications in CBDs. An extensive sensitivity analysis of the performance of LINKOD was also made to investigate the effects of different initial target trip tables and incomplete link volume counts. Although LINKOD performed best with data on 100 percent of the turning movements, it was found that with 25 percent coverage (plus all cordon-station volume counts) there existed only a 10 to 20 percent loss in synthetic O-D table accuracy. It was also determined that the geographic pattern of the traffic count data affected the outcome considerably. Because a better CBD data set is indispensable for conducting a more complete validation of O-D synthesis models as well as other traffic models, a comprehensive CBD travel data-collection effort appears warranted.

In the past the standard technique to obtain origin-destination (O-D) information was to conduct an O-D field survey. These O-D surveys were expensive and sometimes disruptive. Such difficulties caused many different investigators to seek techniques for deriving O-D information from routinely collected field data such as traffic counts. These substitute approaches, which do not require an O-D field survey, are generally called O-D synthesis techniques.

Potential applications of O-D synthesis techniques can be divided into three categories (1): single-path, corridor, and multipath applications. This categorization is based on the relative complexity of the route-choice problem. For a single-path network, such as a section of urban freeway, there is only one path between each O-D pair; thus these O-D synthesis techniques do not have to consider route choice. A multipath network, such as the street system of a central business district (CBD), contains a large number of paths for each O-D pair and thus requires an O-D synthesis technique with a carefully selected route-choice assumption. Corridor applications are between these two extremes, and solution techniques are often hybrids of the single-path and multipath approaches.

Among multipath applications, the CBD is among the most complex of operating environments for applying an O-D synthesis technique. It is a small, heterogeneous study area with a potential for significant congestion, numerous route and modal-choice options, and a high percentage of external trips among the total trips observed within the study area.

The recent emphasis of planners and traffic engineers on improving the performance of the CBD transportation system has caused a great demand for improved analytical tools. High-impact transportation system management (TSM) measures, such as bus malls and automobile-free zones, must be evaluated with respect to their impacts on local circulation and ultimately in terms of the economic health of the CBD. However, available tools for analyzing the performance of the CBD street system [such as microassignment (3-5)] require a great deal of detailed O-D information. This requirement has inhibited the wide use of such analytical tools. Thus an O-D synthesis technique appropriate for CBD applications has great potential to help improve decision making for TSM measures.

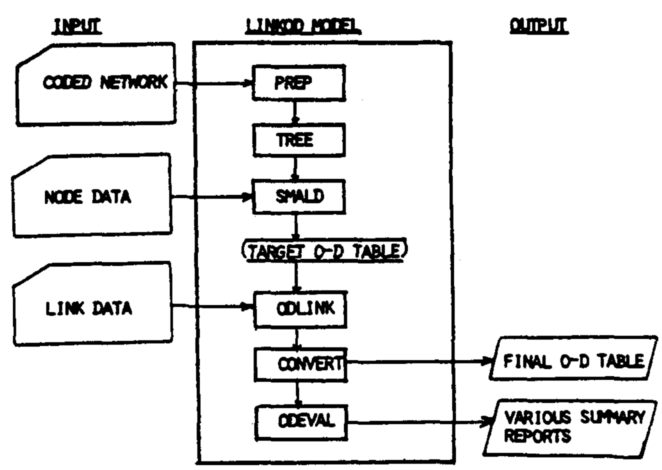
Among the many existing O-D synthesis models that were reviewed (9-13), a model called LINKOD, which was developed for the FHWA, was selected for in-depth evaluation, principally because it was designed specifically for small and congested area analysis (7). The objective of this study was to evaluate the performance of LINKOD for CBD applications. The study considered only the ability of LINKOD to synthesize an O-D table of vehicle trips from available trip-generation estimates and to link vehicular traffic counts.

The remainder of this paper is organized as follows. First, the structure of the LINKOD model is briefly reviewed. Second, the validation of LINKOD by using a 1975 San Jose, California, CBD O-D data set together with traffic counts estimated by the microassignment model is described. Third, a sensitivity analysis of the performance of LINKOD is presented. Finally, conclusions and suggestions for further research are given.

LINKOD MODEL STRUCTURE

The overall structure of the LINKOD computer programs is shown in Figure 1. As input data, the user

Figure 1. LINKOD model structure.



must provide a coded network, contained in a load-node file and a link file. The load-node file gives the node types (e.g., boundary or internal) and associated trip productions and attractions. The link file includes the length, type, number of lanes, and observed traffic volume for each link. Although the theory underlying LINKOD calls for traffic volume information on 100 percent of the links, the model has the ability to insert artificial counts for those links where actual counts are unavailable. (The impact of using incomplete link count data will be discussed in a later section.)

Three principal steps are involved in using LINKOD to develop a trip table from available data.

1. Prepare a network representation of the study area transportation system. Optionally, this involves coding turning movements as network links (program PREP).
2. Create a target trip table that subsequently will be adjusted to conform to the observed traffic counts. This step is performed by using a specialized small area gravity model that incorporates numerous adjustment factors to deal with high proportions of external and through travel (programs TREE and SMALD).
3. Through an iterative procedure, use available link counts to adjust the target trip table such that observed link counts are reproduced when the adjusted trip table is assigned to the transportation network by using an equilibrium traffic assignment procedure (program ODLINK).

LINKOD also contains two utility programs (CONVERT and ODEVAL), which are used for managing data files and generating printed reports, respectively. FHWA documents (6,7) should be consulted for details of the LINKOD algorithms and their theoretical bases.

INITIAL VALIDATION

A 1975 San Jose, California, CBD data base was used to evaluate the performance of LINKOD. Because of a lack of full coverage of actual traffic counts and turning movements, a well-validated set of traffic counts estimated by the microassignment model (3,4) was used to create the input link file. A 1975 trip table that contained data updated from a 1964 O-D survey was used as input to the microassignment model. Because this was considered to be the best available estimate of the true trip table, it was used as the base table against which the synthetic trip table was compared.

The San Jose CBD network is shown in Figure 2. It includes about 69 city blocks and covers about 1 mile². To convert the data from the microassignment format to the LINKOD format consistently and correctly, turning movements were defined as separate links. These are indicated by the dotted links in Figure 2. The coded network includes 857 one-way links, 233 network nodes, and 156 load nodes. Among them, 113 are internal load nodes, each of which represents a block face. The network includes both arterials and local streets, many of which are one-way streets. For the initial validation, 100 percent of link counts and turning movements were provided as input. Trip productions and attractions used to estimate the target trip table were obtained by summing the rows and columns of the base trip table, respectively.

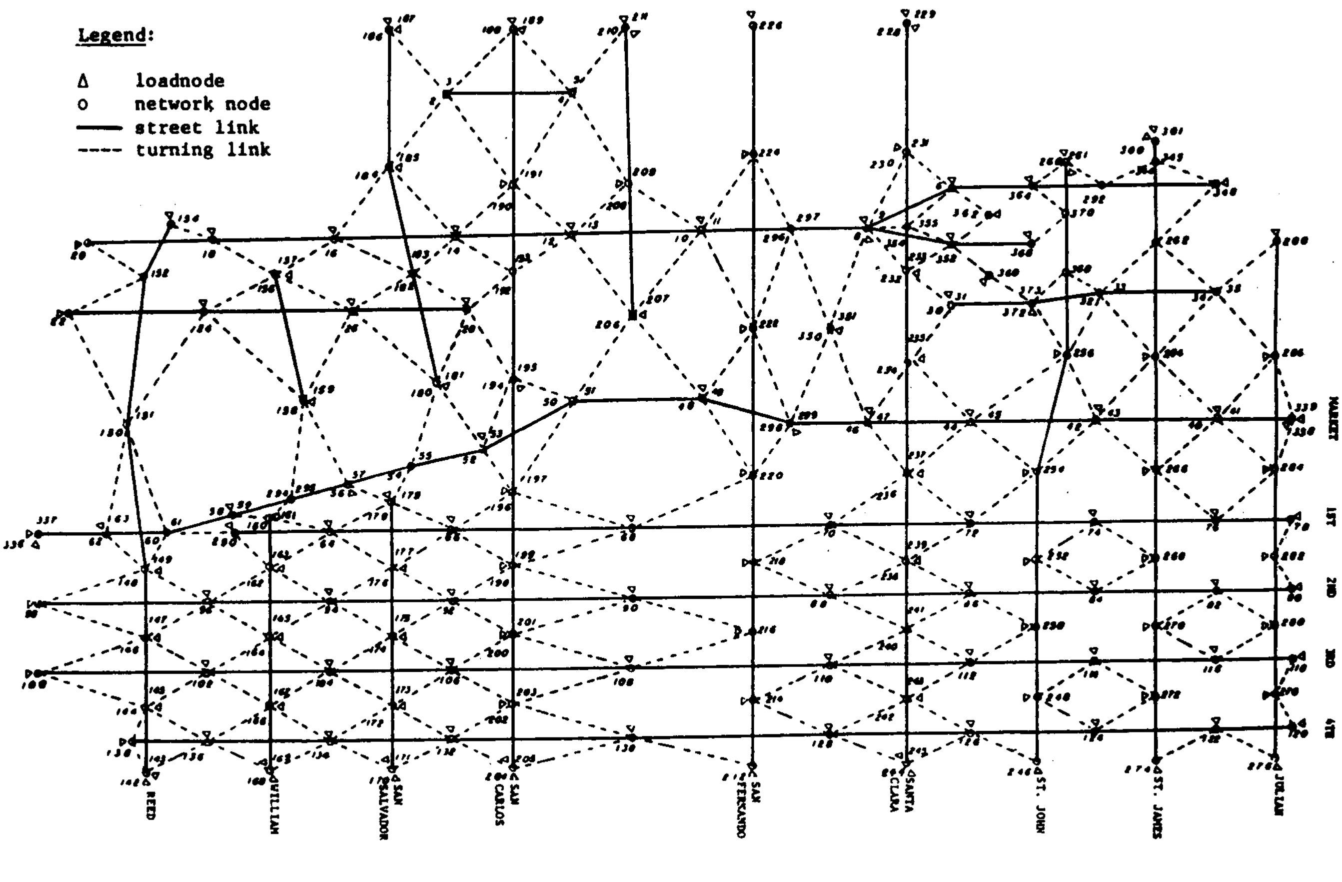
Several goodness-of-fit statistics were used to measure the cell-to-cell differences between the synthetic LINKOD table and the base table. Specifically, four cell-by-cell comparison statistics used are defined as follows. For the mean absolute error per cell (MABSE/cell),

$$\text{MABSE/cell} = \sum_i \sum_j \left(|T_{ij} - T^*_{ij}|/N\right) \quad (1)$$

For the mean absolute error per trip (MABSE/trip),

$$\text{MABSE/trip} = \sum_i \sum_j \left(|T_{ij} - T^*_{ij}|/T\right) \quad (2)$$

Figure 2. San Jose CBD network.



For the root mean square error (RMSE),

$$RMSE = \left\{ \left[\sum_i \sum_j (T_{ij} - T^*_{ij})^2 \right] / N \right\}^{1/2} \quad (3)$$

And for the RMSE as a ratio of average cell value (RMSE/AVGT),

$$RMSE/AVGT = (RMSE/T) \times N \quad (4)$$

where

T_{ij} = value (number of trips in the ijth cell in the base table),
T^*_{ij} = corresponding cell value in the produced table,
N = total number of cells in both tables, and
AVGT = average number per cell for the base table.

In addition, a chi-square statistic was used to measure the difference between two trip length distributions. This statistic is defined as follows:

$$\chi^2 = \sum_{i=1}^{11} [(O_i - T P_i)^2 / T P_i] \quad (5)$$

where

O_i = number of trips in the comparison trip table with length in the ith group (the full range of trip length is divided into 11 groups),
T = total number of trips in the comparison table, and
P_i = percentage of trips in length group i for the base trip table.

To detect any systematic distortions in the model, comparisons were made separately for different O-D groupings based on whether one or both load nodes were internal or on the study area boundary. The results, which are given in Table 1, were not satisfactory. Significant differences existed between the two trip tables. However, without knowing the true O-D table, a definite conclusion cannot be reached.

As a second basis for comparison, both the synthesized O-D table and the base O-D table were input to the microassignment model and the differences in the assigned traffic flows were measured. The data given in Table 2 indicate that the assigned traffic flows from the two O-D tables were close to each other. From this viewpoint, the LINKOD software package is considered to be an acceptable tool for pragmatic applications. Detailed descriptions of the data and results of this case study can be found in Han et al. (8).

SENSITIVITY ANALYSIS

Sensitivity analysis was performed to determine which features of the model are most critical to successful application and to assess how the model reacts to variations in input. The San Jose data set was used to investigate the sensitivity of LINKOD to different model parameter values, to changes in the initial target trip table, and to the extent and coverage pattern of available traffic counts.

Sensitivity to Control Parameters and Adjustment Factors

Four runs were made to find appropriate values for the control parameters that determine the number of program iterations. Little improvement, in terms of synthetic trip table accuracy compared with the base O-D table, resulted from allowing the program to exceed 3 equilibrium assignment iterations and 10 link flow correction iterations. These parameter settings can save appreciable computer time relative to the values proposed in the published documentation (7).

Six runs were made to test model sensitivity to adjustment factors of the small area gravity model (SMALD) used to create the target trip table. Varying fixed penalties and directional change factors were found to have little impact on the final trip table. However, the default values for these adjustment factors appeared to be slightly better than other values tested.

Sensitivity to Different Target Trip Tables

The accuracy impact of different target trip tables was also investigated. The modular structure of LINKOD (Figure 1) makes it an easy matter to run the program with other than the built-in gravity model. For convenience, T5 and T7 are used to denote the target trip table and the final trip table, respectively [these notations are adopted from the LINKOD user's manual (7)]. Besides the internally generated T5, denoted by SMALD T5, two alternative target trip tables, GRAV T5 and AVG T5, were used and evaluated. GRAV T5 denotes an O-D table generated by a simple origin-constrained gravity model that uses the input trip production and attraction data as internally calculated node-to-node travel times. AVG T5 denotes a trip table in which all cell values are equal to the average number of trips per cell.

Let SMALD T7, GRAV T7, and AVG T7 denote the final trip tables created from the target trip tables SMALD T5, GRAV T5, and AVG T5, respectively. The performance of these final trip tables, in terms of the four error measures defined earlier, is shown in Figure 3. As seen in the figure, SMALD T7, the final trip table that results from the internally

Table 1. Goodness-of-fit statistics: synthetic versus base trip table.

Trip	No. of Trips		Ratio[a] (%)	Chi-square (df = 10)	Mean Absolute Error		Root Mean Square Error	
	Base Table	ODL510			Per Cell	Per Trip	RMSE	RMSE/AVGT
Internal-to-boundary	3,977	3,544	89.1	122.2	1.6441	1.0096	4.0019	2.4634
Internal-to-internal	141	1,039	736.9	196.9	0.0982	8.0426	0.4475	36.6398
Boundary-to-boundary	10,309	10,234	99.3	516.8	15.2835	0.6849	34.8227	1.5606
Boundary-to-internal	1,753	1,599	91.2	309.8	1.0302	1.2835	4.9478	6.1643
Total trips	16,180	16,416	101.5	480.4	0.8694	0.8937	6.2770	6.4524

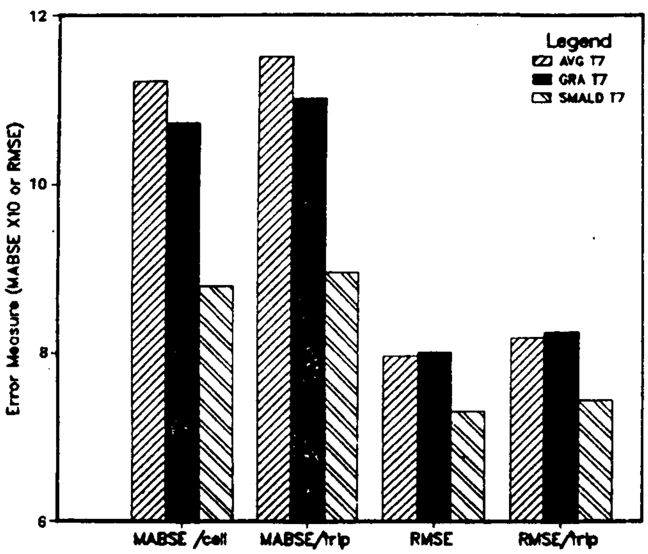
[a]Ratio = (ODL510/base table) x 100 percent.

Table 2. Synthetic link flows versus gound counts: the Second Street screen line.

Street	Direction[a]	Base O-D Table	LINKOD O-D Table[b]	Difference No.	Difference Percent
Reed	EB	275	282	+7	2
	WB	261	285	+24	9
William	EB	24	14	-10	42
	WB	20	19	-1	5
San Salvador	EB	92	68	-24	26
	WB	200	213	-13	6
San Carlos	EB	633	521	-112	18
	WB	358	353	-5	1
San Fernando	WB	410	368	-42	10
Santa Clara	EB	785	816	+31	4
	WB	492	494	+2	0
St. Johns	WB	304	285	-19	6
St. James	EB	1,019	1,076	+57	6
Julian	WB	522	473	-49	9
Subtotal	EB	2,828	2,777	-51	2
	WB	2,567	2,490	-77	3
Total		5,395	5,267	-128	2

[a]Direction is divided into eastbound (EB) and westbound (WB).
[b]LINKOD O-D table with 100 percent turning movements.

Figure 3. Accuracy impact of target trip tables on final trip tables.



generated target trip table, is consistently and significantly closer to the base O-D table than the others. The trip length distributions of the final trip tables were also compared with that of the base trip table. Results again showed that SMALD T7 yielded the closest comparison.

Therefore, it was concluded that the internal SMALD performs better than alternatives such as the simpler gravity model and the maximum entropy model used in this study. Until a more cost-effective alternative is found, the user is advised to use the internal model.

Sensitivity to Incomplete Link Data

The accuracy impact of incomplete link data has been analyzed for the single-path network case (11) and for small multipath networks (2). However, practical guidelines for real-world application have not been developed. Thus a systematic investigation of the sensitivity of LINKOD to incomplete link volume data was made in this study to provide guidelines to help users collect and prepare efficient data sets that, although incomplete, can generate satisfactory solutions.

Note that in a micronized network, such as that shown in Figure 1, each link represents a single through or turning movement. For simplicity, the terms turning movements (which also include through movements) and links are used interchangeably in the following discussion.

To define a strategy for collecting link-count data from a network, two factors must be included: location (where the surveyed links are) and coverage (the percentage of links counted). Relative to location, three sampling strategies were considered: random sampling (R), major link selection (M), and geographic pattern schemes (GP). R means that the counted links are randomly selected from all links of the network. When the M scheme is used, only links that carry the highest traffic flows are selected. Links selected by a GP scheme form a particular geographic pattern, e.g., a cordon or screen line(s), or a combination of these two.

These three link selection schemes, together with different coverage levels, were used to define the 16 experiments summarized in Table 3. Here the number associated with each experiment specifies the percentage of links for which link volumes were input. For example, R30 is the experiment that used a link file that contained turning movement volumes for a randomly selected 30 percent of the links of the network. Brief descriptions of the geographic patterns associated with the six GP experiments are given in Table 4.

Accuracy measured by RMSE and by MABSE/trip for the final trip tables produced in the 16 experiments are plotted in Figures 4 and 5. As shown in these figures, the GP scheme is the preferred sampling strategy because it yields the closest match to the base O-D table for almost all link coverage levels.

Table 3. Experiments with different sampling strategies.

Percentage of Observed Turning Movements[a]	Random Sampling	Major Link Selection	Geographic Pattern Scheme[b]
0-15	—[c]	—[c]	GP13
16-29	—[c]	M25	GP25
30-39	R30	M30	GP37
40-59	R50	M50	GP50
60-69	R60	M60	GP60
70-79	R75	M75	GP75
80-89	—[c]	M85	—[c]

[a]Turning movements include through movements.
[b]See Table 4 for a more detailed description of each GP experiment.
[c]Not tested.

Table 4. Descriptions of the six GP test runs.

Experiment	Description of Geographic Pattern
GP13	A broken cordon including some turning links
GP25[a]	A complete cordon with about half of the turning links connecting to the cordon
GP37	GP25 plus the through movements on three screenlines (Market, San Carlos, and Santa Clara)
GP50	GP37 plus one more screenline (San Fernando), all the turning movements between the four screenlines, and the other half of turning links at the cordon
GP60	GP60 plus three more screenlines (Vine, Almaden, and Notre Dame) and all turning movements between these three streets and all other screenlines
GP75	GP60 plus three more screenlines (St. James, William, and Third) with their turning links

[a]See Figure 8 for locations of the links selected.

Figure 4. RMSE of experiments with different sampling strategies.

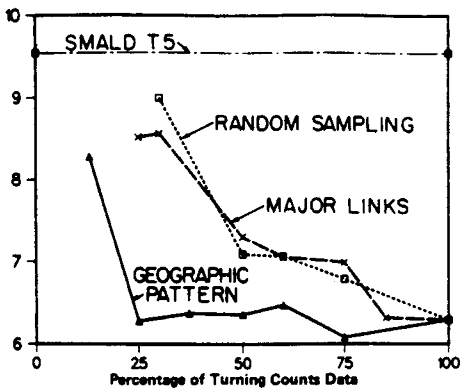
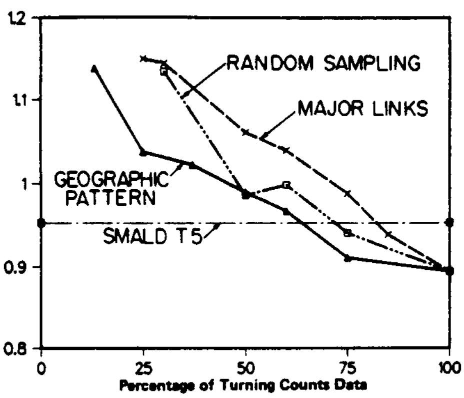


Figure 5. MABSE of experiments with different sampling strategies.



It was also found that when the link file contains turning counts for more than 60 percent of the links, the final table T7 produced by the GP experiment is closer to the base table in both RMSE and MABSE/trip than the target trip table SMALD T5. However, when the link file contains less than 60 percent of the total turning counts, the results are ambiguous. In such cases the final table has a better RMSE and yet a worse MABSE/trip than the SMALD T5. This is shown in Figures 4 and 5. It implies that a target trip table generated from a complete load-node file may be even better than a final trip table adjusted to correspond to a scanty link file.

When comparing the final trip table (T7) against the target trip table (SMALD T5) within the range of 40 to 60 percent available turning counts, the gain in the RMSE measure is much larger than the loss in the MABSE/trip measure. It implies that, in this range, the correction procedure tends to correct the bad cells in the trip table while sacrificing some overall goodness of fit.

Because both error measures are meaningful, the alternative data-collection schemes based on a single measure cannot be evaluated. Although the trade-off between these two error measures is still unclear, a combined measure of effectiveness (MOE) was defined based on the following assumptions:

1. The users are more concerned with the relative (percentage) improvement rather than the absolute improvement in the error measures, and
2. Both error measures are of equal importance.

On this basis, the MOE is defined as follows:

$$\begin{aligned} MOE_i &= \text{(percentage improvement in RMSE)} + \text{(percentage improvement in MABSE/trip)} \\ &= \left\{ [(RMSE_0 - RMSE_i)/RMSE_0] + [(MABSE_0 - MABSE_i)/MABSE_0] \right\} \times 100\% \end{aligned} \quad (6)$$

where

MOE_i = combined MOE for experiment i,
$RMSE_i$ = RMSE of the final trip table produced by experiment i,
$RMSE_0$ = 9.5438 = RMSE of SMALD T5,
$MABSE_i$ = MABSE/trip of the final trip table produced by experiment i, and
$MABSE_0$ = 0.9527 = MABSE/trip of SMALD T5.

The combined MOEs for the various experiments are plotted in Figure 6. It is clear that the geographic pattern scheme is the most effective data-collection scheme among the three tested. It can also be observed that, when using this link selection scheme, a minimum of 15 percent turning count data is required to produce a better final trip table than the initial SMALD T5.

Suppose that the data-collection cost is proportional to the number of links for which counts are available. The horizontal axis of Figure 6 then serves as a proxy for cost. The effectiveness/cost (E/C) ratio can thus be illustrated for each experiment, as shown in Figure 7. Again the six GP experiments are more cost effective than the others. Among these experiments, GP13 has a negative E/C ratio, presumably because less information is contained in the link file than is provided in the complete load-

Figure 6. Overall effectiveness of different alternatives.

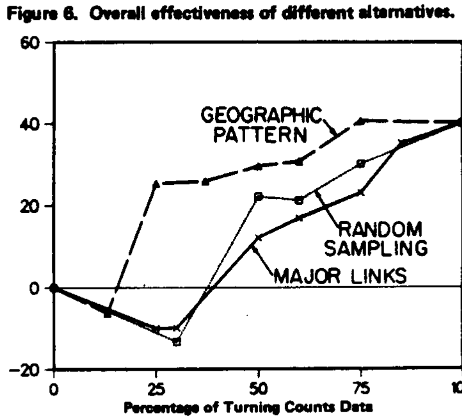
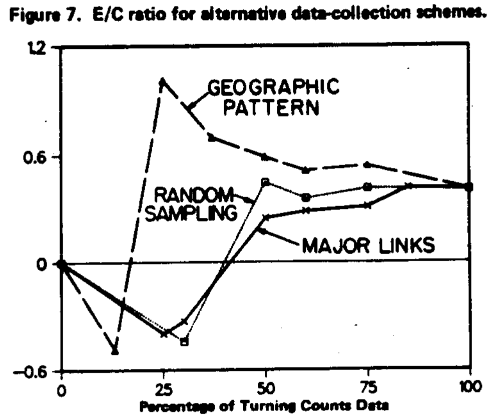


Figure 7. E/C ratio for alternative data-collection schemes.



node file. Finally, the alternative GP25, in which the selected links located along the cordon of the study area are emphasized by black lines in Figure 8, is found to be the most cost-effective data-collection scheme.

The complete results of each experiment and a detailed description of this sensitivity analysis can be found in Han et al. (8, Chapter 7).

SUMMARY AND CONCLUSIONS

O-D synthesis techniques deal with the problem of deriving trip O-D patterns from traffic counts. Among the many applications considered for O-D synthesis, the CBD is among the most complex because of the potential for congestion and its varied choice alternatives (including routes, modes, and vehicle occupancy). In this study the performance of a leading O-D synthesis technique was examined when it was applied to the estimation of vehicle trips in a 1 mile^2 portion of a major California CBD.

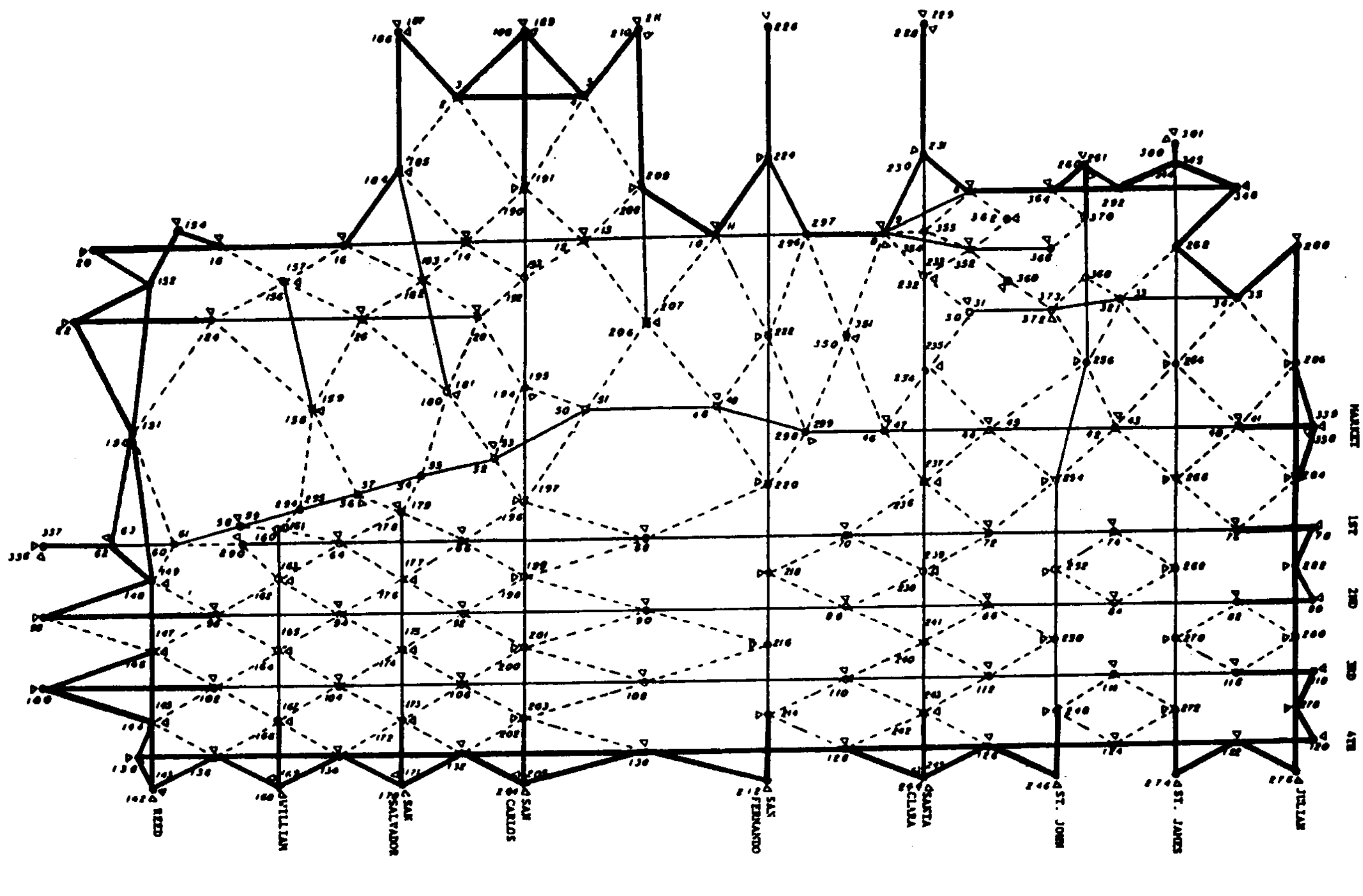
Among various models reviewed, the LINKOD model that was designed primarily for small and congested area analysis was selected for in-depth evaluation. A 1975 San Jose CBD O-D data set with traffic counts estimated by the microassignment model was used in the evaluation. Significant differences were found between the synthetic LINKOD trip table and the base trip table. However, when assigned to the network that used the microassignment model, both O-D tables produced similar flow patterns. LINKOD is thus considered as an acceptable tool for pragmatic applications in CBDs.

An extensive sensitivity analysis was also made. Among three alternative target trip tables, the internal SMALD performed much better than a simpler gravity model and a naive maximum entropy distribution. It was concluded that the trip-distribution model internal to LINKOD should be used whenever possible. The sensitivity of the model to incomplete link volume sampling strategies was tested to find the most effective way to collect this type of data. The alternative that used traffic counts solely along the study area cordon was found to be the most cost-effective data-collection scheme of those tested.

In the realm of future research, there are several topics that merit further investigation.

1. A comprehensive data-collection effort should be launched for a CBD study area. The data collection should include a field survey of vehicular O-D patterns and simultaneous collection of travel time and traffic volume information. Such a data set would be vastly superior to the San Jose data set used in this study, which was model derived. With this improved data set, the following topics can be investigated in greater detail.
2. Further research should be undertaken regarding O-D patterns of external vehicle trips traveling through a CBD. Techniques for estimating the external trip O-D based on minimal external network data should be investigated, as well as techniques for forecasting the changes in external trip O-D patterns that result from TSM measures in the internal network.
3. A cost-effective combination of manual turning movement counts and machine volume counts needs to be determined. The current research focused on turning counts only. Initial attempts to investigate the trade-off between turning counts and machine counts were inconclusive and require further investigation.

Figure 8. Links selected in test run GP25.



4. The effect of inaccurate traffic counts on the accuracy of the synthetic O-D table needs to be examined. The current research dealt with consistent, accurate count information only.

5. Finally, research is needed to expand the equilibrium framework to permit estimation of multimodal trip tables and the analysis of shifts in vehicle occupancy.

ACKNOWLEDGMENT

The work described in this paper was performed under contract with the Division of Transportation Planning, California Department of Transportation (Caltrans). The data in this paper are adapted from the project final report produced by a research team that included Rick Dowling, Alan Willis, and Adolf D. May, the project's principal investigator. We would like to express our appreciation to numerous individuals in Caltrans, the U.S. Department of Transportation, and elsewhere who offered their advice and assistance during the course of this study.

REFERENCES

1. V. Chan, R. Dowling, and A. Willis. Deriving Origin-Destination Information from Routinely Collected Traffic Counts--Phase I: Review of Current Research and Applications. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 80-1, Jan. 1980.
2. Y.-P. Chan. Inferring an Origin-Destination Matrix Directly from Network Flow Sampling. Pennsylvania Transportation Institute, Pennsylvania State Univ., University Park, Working Paper, 1981.
3. D.C. Chenu. San Jose Central Business District Project Validation of Microassignment and Evaluation of Two Loading Procedures. Division of Transportation Planning, California Department of Transportation, Sacramento, June 1977.
4. S.M. Easa, T.-L.A. Yeh, and A.D. May. Traffic Management of Dense Networks--Volume I: Analysis of Traffic Operations in Residential and Downtown Areas. Institute of Transportation Studies, Univ. of California, Berkeley, Res. Rept. 80-5, June 1980.
5. S. Easa, T.-L.A. Yeh, and A.D. May. Traffic Management of Dense Networks--Volume II: User's Guide of the Refined Microassignment Model. Institute of Transportation Studies, Univ. of California, Berkeley, Res. Rept. 80-6, June 1980.
6. Y.J. Gur, M. Turnquist, M. Schneider, L. Leblanc, and D. Kurth. Estimation of an Origin-Destination Trip Table Based on Observed Link Volumes and Turning Movements--Volume I: Technical Report. FHWA, Rept. FHWA/RD-80/034, May 1980.
7. Y.J. Gur, P. Hutsebat, D. Kurth, M. Clark, and J. Castilia. Estimation of an Origin-Destination Trip Table Based on Observed Link Volumes and Turning Movements--Volume II: User's Manual. FHWA, Rept. FHWA/RD-80/035, May 1980.
8. A. Han, R. Dowling, E. Sullivan, and A. May. Deriving Origin-Destination Information from Routinely Collected Traffic Counts--Volume II: Trip Table Synthesis for Multipath Networks. Institute of Transportation Studies, Univ. of California, Berkeley, Res. Rept. 81-9, Aug. 1981.
9. P. Hogberg. Estimation of Parameters in Model for Traffic Prediction: A Nonlinear Regression Approach. Transportation Research, Vol. 10, 1976, pp. 263-265.
10. J. Holm, T. Jensen, S. Neilsen, A. Christensen, B. Johnson, and G. Ronby. Calibrating Traffic Models on Traffic Census Results Only. Traffic Engineering and Control, Vol. 17, April 1976.
11. A.E. Willis and V.K. Chan. Deriving Origin-Destination Information from Routinely Collected Traffic Counts--Phase II: Comparison of Techniques for Single-Path and Multipath Network. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 80-10, Sept. 1980.
12. L.G. Willumsen. Estimation of an O-D Matrix from Traffic Counts: A Review. Institute for Transport Studies, Univ. of Leeds, Leeds, Great Britain, Paper 99, Aug. 1978.
13. H.J. Van Zuylen. The Information Minimizing Method: Validity and Applicability to Transport Planning. In New Developments in Modeling Travel Demand and Urban Systems (G.R. Jansen et al., eds.), Gower Publishing, Ltd., Brookfield, Vt., 1979.

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

Estimating Trip Tables from Traffic Counts: Comparative Evaluation of Available Techniques

YEHUDA J. GUR

Methods for estimating trip tables from traffic counts are potentially useful because of their relative efficiency in data requirements. Two techniques for estimating existing trip tables in urban highway networks—the information theory (IT) technique and the LINKOD model—are analyzed in this paper. The separate description of the two techniques is followed by a formulation of an algorithm that is designed for application of the two techniques as well as other variations. By using the algorithm, extensive experimentation with the various techniques is made by using artificial data. Both the convergence speeds and the ability of the techniques to stay close to the target trip table are evaluated. The main contribution of the paper is its presentation of the two major techniques within an easily understood, unified format. It opens a way for extending the IT techniques for equilibrium assignment problems.

Much work has been done in recent years in developing procedures for estimating trip tables from traffic counts. These methods are potentially useful because of their efficiency in terms of data requirements compared with the available alternatives. Chan et al. (1) and Willis and Chan (2) recently compiled a comprehensive survey of the various estimating methods and the types of problems that they solve.

One type of problem is dealt with in this paper--estimating an existing trip table for a typical urban highway network, based primarily on traffic counts on many links. Two different approaches to the problem have been reported. The first is the information theory (IT) approach, developed independently by Van Zuylen (3) and by Willumsen (4), and later described by Van Zuylen and Willumsen (5). The second is the network equilibrium approach proposed by Nguyen (6-8) and extended by Gur et al. (9) into the LINKOD system.

The two methods have been developed independently from each other. Both have been developed primarily (but by no means exclusively) for estimating trip tables for "windows" in city centers. Recently, van Vliet and Willumsen (10) have reported the testing of the IT model on data from the center of Reading, England. Test application of LINKOD in downtown Washington, D.C., is reported by Gur et al. (9). Recently, a large-scale validation of LINKOD on data from downtown San Jose, California, has been reported by Han et al. (11).

The purpose of this paper is to present the two methods by using a common basis, and to evaluate them comparatively. As a result of the evaluation, a third method, which uses some elements of each, is developed and tested.

DESCRIPTION OF PROBLEM

Consider a road network that consists of nodes connected by links; some of the nodes are load nodes, where trips originate or terminate or both. It is assumed that trips between the load nodes are the only cause for traffic on the links. Given volume counts on some of the links, the problem is to find the true trip table $\hat{T} = (\hat{t}_i)$ that is served by the network. (Note that for simplicity of notation, t_i denotes the ith cell in the table, giving the number of trips between two load nodes, e.g., k and (ℓ).

There are three important attributes inherent to the problem. First, the solution requires assumptions regarding the assignment rule, which describes how travelers select their paths. Two different types of assignment assumptions are possible. The first is the proportional assignment where link volumes are directly proportional to the interchanges served by them. This happens where path selection does not depend on link volume, as in an all-or-nothing assignment. Alternatively, with nonproportional assignment rules, path selection is a function of link volumes as in equilibrium assignment. Proportional assignment assumptions make the solution process simpler, but this assumption might be unrealistic in congested networks. The main body of this paper deals with all-or-nothing assignments.

A second important attribute of the problem is that in most cases there is no accurate solution; i.e., there is no trip table that, when assigned (according to the assumed assignment rules), satisfies exactly the given set of counts. This can happen both because of data imperfections (e.g., the counts are taken in different time periods) and modeling imperfections (e.g., the assumed assignment rule only approximates the actual route selection).

Third, in most cases the problem is underspecified; i.e., if there exists one table that satisfies a given set of flows, then there exist many other tables that, when assigned, produce those same flows. A complete solution method must address all these issues. It must be based on a realistic assignment assumption; it must be robust enough to withstand data inaccuracies and to estimate a table that approximates (rather than duplicates) the counts. It should also identify the best table among those that satisfy the counts.

Both the IT and the LINKOD models satisfy these requirements; although LINKOD can operate for both proportional and equilibrium assignment assumptions, the current version of the IT model operates only for proportional assignment. The problem of multiplicity of solutions is addressed in the two models in a similar way, i.e., the input to the model includes a target trip table--a trip table that describes the best estimate of the true table without traffic count information. The LINKOD model corrects this table as little as possible to approximate the observed flows. The IT model looks for the most likely, closest table to the target trip table that approximates the observed flows.

INFORMATION THEORY MODEL

Willumsen (4) developed a solution method based on entropy maximization considerations. The model (as well as a variation of it) is described by Van Zuylen and Willumsen (5). The problem is to find the maximum entropy trip table among those that satisfy the observed flows. Entropy of a table is defined as the number of micro states associated with it, weighted by probabilities that reflect the target trip table.

Van Zuylen and Willumsen (5) indicate that for the all-or-nothing assignment, the solution to the problem is of the form

$$t_i^n = f_i * \pi_{a\epsilon r_i} x_a \quad (1)$$

where

t_i^n = ith element of the final trip table,
f_i = ith element of the target trip table,

r_i = set of links that are included in the path of the ith interchange, and
x_a = link-specific parameters.

Van Zuylen and Willumsen suggest that Equation 1 can be solved [i.e., the values of $X = (x_a)$ can be found] by using iterative proportional fitting and an algorithm that will be described later. They note that even though the convergence of the method has not been proven, numerous experiments with artificial data failed to show a case of nonconvergence. As will be shown later, a case of nonconvergence was found, which was rectified by a minor change to the algorithm.

LINKOD APPROACH

The LINKOD model is described by Gur et al. (9) and by Turnquist and Gur (12). The theory was developed by Nguyen (6,7). Nguyen specifies a nonlinear optimization problem; it is shown that any solution to that problem is a trip table that, when assigned by using equilibrium assignment, replicates the observed flows. The optimization problem is similar to the problem connected to equilibrium assignment with elastic demand.

As in any other equilibrium assignment problem, the LINKOD model uses volume-delay functions. However, here both the link volume and the impedance at load are known. It can be shown that the correct solution to the problem is arrived at regardless of what function is used as the volume-delay function, as long as it is a strictly increasing function and it gives the correct impedance at the observed load. For convenience, LINKOD uses linear, or bilinear, functions, e.g.,

$$c_a(v_a) = \hat{c}_a + b_a(\hat{v}_a - v_a) \tag{2}$$

where
$c_a(v)$ = impedence of link a at volume v,
$\hat{v}_a$ = observed volume,
$\hat{c}_a = c_a(\hat{v}_a)$ = impedence at the observed volume, and
b_a = a parameter.

Those functions operate like error functions, where the error measure $\hat{c}_a - c_a(v)$ is directly related to the difference between the observed and assigned volumes.

Another important attribute of the model is that the theory does not provide for a unique solution to the problems; i.e., all the trip tables that satisfy the observed flows have exactly the same value as the objective function. To overcome this problem the solution algorithm was designed to keep the final trip table as close as possible to an input target trip table. Thus the LINKOD model actually corrects the target trip table so that it approximates, as close as possible, the observed flows.

ALGORITHM FOR SOLVING THE ALL-OR-NOTHING PROBLEM

In spite of the different theoretical backgrounds of the IT and LINKOD models, their solution algorithms are similar. The following algorithm describes the solution process by the two models and various possible combinations of them. This version of the LINKOD model is a special case, where it can be assumed that only one path is used for each origin-destination (O-D) pair (for example, travel on an expressway).

1. Given the target trip table (F), the observed volumes ($\hat{V}$), the link impedence at load ($\hat{C}$), and the link error (volume-delay) functions $[c_a(\cdot)]$, determine the minimum impedence path for each O-D pair. Denote by r_i the set of all links that serve the ith O-D pair and determine the skim trees at load $\hat{U}$:

$$\hat{U}_i = \Sigma_{a\epsilon r_i} \hat{c}_a \tag{3}$$

Assign the target trip table to the network and obtain V^n.

2. Set m = 0, $V^o = 0$, $T^n = F$, and $T^o = 0$.
3. Evaluate the solution (V^n, T^n). If it is satisfactory, go to step 10.
4. Set m = m + 1, $V^o = V^n$, $T^o = T^n$, and $C^n = C(V^n)$.
5. Calculate for each link the link error measure:

$$y_a = y\,[\hat{v}_a, v_a^n, c_a(\cdot)] \tag{4}$$

(The definition of y is given later.)

6. Calculate a correction factor for each interchange:

$$s_i = s(y_a; a\epsilon r_i) \tag{5}$$

That is, the interchange correction factor (s) is a function of the volume errors of the links along the path that serves the interchange. Calculate a corrected or a correction trip table:

$$t_i^c = t(s_i) \tag{6}$$

7. Assign T^c and get V^c.
8. Find λ such that

$$T^n = (1-\lambda)\,T^o + \lambda T^c \tag{7}$$

$$V^n = (1-\lambda)\,V^o + \lambda V^c \tag{8}$$

where $0 < \lambda < 1$ and λ minimizes the value of the objective function.

9. Go to step 3.
10. The solution to the problem is the trip table T^n. Stop.

In the LINKOD model, steps 5 and 6 use linear relations:

$$y_a^L = \hat{c}_a - c_a(v_a^n) \tag{9}$$

In cases where $c(\cdot)$ is linear (Equation 2):

$$y_a^L = (\hat{v}_a - v_a^n) * b_a \tag{9a}$$

$$s_i = \Sigma_{a\epsilon r_i} y_a \tag{10}$$

and

$$t_i^c = t_i^o \left\{1 + 2 * [s_i/(\hat{u}_i - u_i^o - s_i)]\right\} \tag{11}$$

where u_i^o is the skim trees that use the impedances

$$c_a^o = c_a(o) \tag{12}$$

In Willumsen's IT model, multiplicative relationships are used, i.e.,

$$y_a^W = \hat{v}_a / v_a^n \tag{13}$$

$$s_i^W = \pi_{a\epsilon r_i} y_a^W \tag{14}$$

and

$$t_i^c = t_i^o * s_i^W \tag{15}$$

Table 2. Bias coefficients for modal-choice disutility equations.

Income Group[a] (I)	Automobile Access Penalty Coefficient	Income Coefficients			
		Drive Alone (1, I)		Group Automobile (G, I)	
		Coefficient	t-Ratio	Coefficient	t-Ratio
Home-based work trips					
1	1.4165	1.4014	13.74	1.6733	21.14
2	1.0683	0.7979	8.24	1.2677	18.89
3	0.4943	−0.0750	−0.32	0.8939	10.01
4	−0.2245	−0.6783	−6.59	0.6140	7.83
Home-based other trips					
1	2.9661	0.0934	9.06	−1.5281	−24.40
2	2.3095	−1.1802	−21.00	−2.2168	−35.62
3	1.9305	−2.1397	−30.61	−2.7419	−44.41
4	1.4125	−2.9294	−38.33	−3.1109	−50.70
Non-home-based trips					
1		−1.3447	−11.73	−1.3496	−11.52
2		−1.9311	−17.53	−2.1027	−17.19
3		−2.6904	−24.48	−2.5040	−21.67
4		−3.0689	−27.57	−2.7298	−23.35

Note: See Table 1 for equations used for bias coefficients.
[a]Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Table 3. Variables used in modal-choice calibration.

Acronym	Description of Variable	Units of Measure
Transit variables		
TRN RUN	In-vehicle time from the transit network, not including automobile access time	Minutes
AUTO ACC	Automobile access time from the transit network	Minutes
WALK	Walk access time from the transit network	Minutes
WAIT1	Transit boarding time for the first transit vehicle from the transit network	Minutes
WAIT2	Time spent transferring from the transit network	Minutes
TXFER	Number of transfers from the transit network	Number
FARE	Transit fare	Cents
AUTO CONN	Dummy variable signifying if an automobile was required to access the transit system (0 is no, 1 is yes)	–
TRN DACC 25	Percentage of regional employment within 25 min of total transit time from destination zone	Percent
Highway variables		
HWY RUN1	Highway in-vehicle time from highway network for one person per car (drive alone) trips	Minutes
HWY RUNG	Highway in-vehicle time for group automobile trips (same as HWY RUN1 plus an additional time for each passenger)	Minutes
HWY CST1	Highway operating cost for one person per car trips	Cents
HWY CSTG	Highway operating cost for group trips	Cents
PRK CST1	Avg parking cost for one person per car trips	Cents
PRK CSTG	Avg parking cost for group trips	Cents
HWY EXC	Time spent parking and unparking an automobile; the sum of highway terminal time at the origin zone and the destination zone (also called highway excess or terminal time)	Minutes

$$\text{C.I.} = K / \sum_{i=1}^{3} [A(i) + C] \tag{1}$$

where

C.I. = value of composite impedance,
A(i) = modal choice disutility function for mode i (i = 1,2,3),
C = constant chosen such that all A(i)'s are positive, and
K = constant chosen such that all C.I.'s are between 1 and 127, inclusive.

This formula is simply the reciprocal of the sum of the modal impedances, scaled to represent suitable values. The second formulation sums the exponential of the disutility function for all modes, takes the reciprocal of the sum, and takes the natural logarithm of this reciprocal. This is called the log sum method, and is described as follows:

$$\text{C.I.} = K * \ln \left\{ C / \sum_{i=1}^{3} \exp[-A(i)] \right\} \tag{2}$$

Both of these functions meet the criteria previously described, but little was known about the ability of either to perform as a measure of spatial separation. Therefore, both measures were tested by calibrating the home-based work trip-distribution model twice, each time by using a different measure. The choice would then be made on the basis of whichever formulation provided the closer match to observed conditions, based on average trip length and other such measures.

CALIBRATION TECHNIQUE

The New Orleans distribution model uses the standard gravity model form (15). This model postulates that the number of trips for a given zone interchange is proportional to the number of trip productions at the origin zone and the number of trip attractions at the destination zone, and inversely proportional to the travel impedance between the two zones. The relationship with impedance is generally described by a nonlinear function that relates impedance to a nondimensional F factor (also called friction factor).

The usual calibration process involves determining the relationship between the impedance values and the F factors such that the distribution of estimated trips by impedance matches that of the observed trips. Additional adjustment factors (K factors) are used to help match observed and estimated trips by geographic stratification (such as districts). For this project, separate models were developed for each trip purpose and for each of four income levels. Observed person trips came from the home interview survey.

Initially, it was assumed that the UTPS program AGM, operating in the so-called SAC mode, would be able to automatically calculate the proper F factors. However, this function of program AGM was not operating correctly at that time and an ad-hoc method of calibrating the F factors was developed. This method used essentially the same technique as described in the AGM program documentation. F factors are calculated by using a gamma function, i.e.,

$$F(I) = A * I^{B} * \text{EXP}(G * I) \tag{3}$$

where

F(I) = F factor for impedance value I,
A,B, and G = calibrated coefficients, and
EXP = exponential function.

This function was judged to be adequate because there is considerable documentation that it simulates the relationship between F factors and impedance adequately. Calibration of a distribution model consists mainly of fitting this curve. This was done as follows:

1. Apply program AGM in the apply-and-calibrate (AC) mode, which reports the observed and estimated trips stratified by each unit of impedance.
2. The observed and estimated trips and the F

Tables 1 and 2. As the equations in Table 1 indicate, travel disutility is a linear function of the time and cost of the transit, drive alone, and group automobile modes, and other service characteristics such as number of transfers and transit accessibility. Also, the income level of the traveler is a prime influence on perceived disutility. The differential effect of walk access to transit versus automobile access to transit on modal choice is defined through the use of an automobile access penalty dummy coefficient in the transit disutility equation. The variables are described in more detail in Table 3. For the work trip purpose, peak-hour impedance values were used; for home-based other and non-home-based purposes, off-peak values were used.

The mode and variable definitions for these equations are similar to other modal-choice models recently developed for Minneapolis-St. Paul (9), Seattle (10), Houston (11), St. Louis (12), and Buenos Aires (13). The group mode consists of persons in automobiles with two or more occupants. A separate logit submodel is used to estimate the proportion of two-person, three-person, and four or more person trips in order to determine the average group occupancy for each interchange. The transit and highway variables are created from standard Urban Transportation Planning System (UTPS) network analysis programs (14) and special submodels are used to estimate accessibility, terminal time, and parking cost. The calibration data consisted of a comprehensive, home interview origin-destination survey conducted in the New Orleans region in 1960.

The coefficients and the final list of variables were developed by using ULOGIT on a sample of the survey file, followed by disaggregate validation and adjustment by using the full survey file. The coefficients are comparable to coefficients from other cities, exhibit internal consistency, and have acceptable t-ratios (see Tables 1 and 2). The following observations support the reasonableness of these equations:

1. The out-of-vehicle time coefficients exceed those for in-vehicle time;
2. The model is much more sensitive to automobile access time to transit than to time spent on the transit vehicle;
3. The ratio of the time coefficient to the cost coefficient, which is the implied value of travel time, is approximately one-third to one-half the average 1960 regional income in cents per minute; and
4. The income bias coefficients indicate that as income level increases, there is a lower propensity to use transit, and within the automobile mode, a higher propensity to be a driver rather than a passenger.

COMPOSITE IMPEDANCE CALCULATION

The previous section describes how impedance is defined for each mode. The remaining challenge is to combine the three impedances into one value. For this task, the following conditions must be met.

1. The combined value must decrease as any of the modes becomes better, i.e., declines in time or cost.
2. The combined value must increase if a mode is not available [i.e., an interchange with even unsatisfactory transit service must have a better (lower) impedance than one with no service at all].
3. The value must lie between 1 and 127, inclusive. The UTPS program AGM assumed that the input impedance values are stored as 1-byte matrix elements. The highest value that can be represented in this format is 127.
4. The distribution of values within this range should be reasonable; i.e., they should not be concentrated at the top or bottom of the range.

It was ascertained that at least two mathematical formulations meet these criteria. One formulation is a variation of the harmonic mean function:

Table 1. Modal-choice disutility equations.

Mode	Equation
Home-based work trips	
Transit disutility	0.0332 • WALK + 0.0769 • WAIT1 + 0.0319 • WAIT2 + 0.0078 • FARE + 0.0145 • TRN RUN + 0.1005 • AUTO ACC (4.07) (20.21) (8.85) (10.45) (6.72) (2.59) + 0.0588 • TXFER + Auto Access Penalty (I) • AUTO CONN (3.59)
Drive-alone disutility	0.0693 • HWY EXC + 0.0145 • HWY RUN1 + 0.0078 • HWY CST1 + 0.02145 • PRK CST1 + Income Coefficient (1, I) (4.94) (6.72) (10.45) (10.45)
Group automobile disutility	0.0174 • HWY EXC + 0.0145 • HWY RUNG + 0.0078 • HWY CSTG + 0.02145 • PRK CSTG + Income Coefficient (G, I) (1.74) (6.72) (10.45) (10.45)
Home-based other trips	
Transit disutility	0.0165 • (WALK + WAIT1 + WAIT2) + 0.0116 • FARE + 0.0066 • (TRN RUN + AUTO ACC) – 0.0183 • TRN DACC25 (7.45) (9.55) (–22.91) + Auto Access Penalty (I) • AUTO CONN
Drive-alone disutility	0.3403 • HWY EXC + 0.0066 • HWY RUN1 + 0.0116 • HWY CST1 + 0.0319 • PRK CST1 + Income Coefficient (1, I) (25.98) (7.45) (9.55) (9.55)
Group automobile disutility	0.2828 • HWY EXC + 0.0066 • HWY RUNG + 0.0116 • HWY CSTG + 0.0319 • PRK CSTG + Income Coefficient (G, I) (28.50) (7.45) (9.55) (9.55)
Non-home-based trips	
Transit disutility	0.0328 • (WALK + WAIT1 + WAIT2) + 0.0047 • FARE + 0.0131 • (TRN RUN + AUTO ACC) + 0.0750 • TXFER (9.41) (2.75) (9.41) + 2.7472 • AUTO CONN (4.91)
Drive-alone disutility	0.2423 • HWY EXC + 0.0131 • HWY RUN1 + 0.0047 • HWY CST1 + 0.0291 • PRK CST1 + Income Coefficient (1, I) (20.14) (9.41) (2.75) (2.75)
Group automobile disutility	0.3048 • HWY EXC + 0.0131 • HWY RUNG + 0.0047 • HWY CSTG + 0.0291 • PRK CSTG + Income Coefficient (G, I) (25.58) (9.41) (2.75) (2.75)

Note: Disutilities must be multiplied by –1 before taking the exponential in the logit equation. Numbers in parentheses represent t-ratios. T-ratios were not calculated for the work and other automobile access penalty coefficients, or the non-home-based coefficient on TXFER. See Table 2 for explanation of bias coefficients used for the equations.

Trip Distribution Using Composite Impedance

WILLIAM G. ALLEN, JR.

In this paper the theory and results of a trip-distribution model that uses a multimodal composite definition of impedance as its measure of separation, instead of highway time, are presented. The distribution model is part of a complete travel-demand model chain developed for the New Orleans region. This model chain is briefly described, and its special features of income stratification and connectivity among programs are emphasized. The disutility functions of a three-mode logit modal-choice model are used to develop modal impedance values. The structure and coefficients of these equations are discussed. Two alternative methods for combining these modal impedances are presented: harmonic mean and log sum. A special technique for calibrating the F factor curves was developed to circumvent shortcomings in the urban transportation planning system (UTPS) software. The results of the calibration are presented. These results indicated that the log sum formula produced better results than the harmonic mean formula, based on various observed and estimated comparisons. In addition, the log sum composite impedance-based model proved suitable only for home-based work trips. Unsatisfactory results for the other trip purposes led to the use of off-peak highway time for those purposes. Results for home-based other and non-home-based models are also presented. The conclusions of this analysis are that a distribution model can be successfully calibrated by using composite impedance; that, at least in this case, the log sum formula worked better than harmonic mean; and that a successful alternative to the standard AGM gravity model calibration process can be developed.

The theory and results of a trip-distribution model that uses a composite definition of impedance as its measure of separation, instead of highway time, are presented in this paper. The premise that such a model is inherently logically superior to a gravity model based on highway time is accepted as a given. This superiority involves a composite impedance-based model that is sensitive to the characteristics of all modes and provides for improved connectivity between the distribution and modal-choice models. This should, in theory, produce more reasonable estimates of trip distribution. The distribution model is part of a complete travel-demand model chain developed for the New Orleans region. Previous work is reviewed here; the accompanying logit modal-choice models are described; and alternative methods of combining impedances, a different technique for calibrating gravity models, and the final results are presented.

PRIOR RESEARCH

The use of composite impedance in distribution models is not new. For example, an early reference to a generalized resistance formulation for the gravity model is a 1973 paper by Manheim (1) based on his earlier work (2). Wilson (3) also describes a composite generalized cost function. Much of the recent work in this field has focused on the joint choice type of model. By combining destination choice and modal choice (and often trip frequency) into a single model (generally by using a logit structure), this type of model effectively incorporates the impedances of all modes and the socioeconomic status of the traveler into the trip-distribution process. There are numerous references to and examples of this model type in the literature (4,5), with perhaps the best known of these being the Metropolitan Transportation Commission (MTC) model set (6).

However, the New Orleans model chain uses the traditional sequential application of models, and there appears to be but one previous attempt at using composite impedance in this context. In 1975 a similar set of models was developed for the Regional Transportation District in Denver (7). That study used modal-choice logit coefficients to define impedance. Alternative methods of combining impedances were reviewed, and a parallel resistance (harmonic mean) formulation was selected.

Basically, the New Orleans distribution models are a direct extension of the Denver work. The major changes are that separate models are developed for each income level and the log sum method of combining impedances was used. The log sum method, which is simply the natural logarithm of the denominator of the modal-choice logit equation, was also used in the San Francisco MTC models (6).

MODEL CHAIN

The distribution model can best be described by placing it in the setting of the entire travel model chain (see paper by Schultz elsewhere in this Record). The New Orleans model chain consists of the traditional generation, distribution, and modal-choice models. What distinguishes these models is that they are entirely income stratified and highly connected with each other. The generation models use an elaborate cross-classification structure, including the capability of estimating trip productions and attractions for each of four income levels (quartiles). The modal-choice models consist of a three-mode logit structure, which contains bias variables based on income level.

One of the criticisms of the traditional type of travel-demand models is that the models are applied sequentially, independent of each other. It is generally recognized that actual travel decisions are seldom made in this fashion. Rather, decisions on frequency, destination, mode, and route tend to be interrelated. The use of composite impedance is an attempt to address this concern. The modal-choice and distribution models are tied together because the coefficients of the logit models are used to define the composite impedance value. Therefore, the distribution of trips is sensitive to both highway and transit service levels, travel cost as well as time, and the income level of the traveler. The high level of transit service in New Orleans makes this multimodal definition of impedance especially meaningful. This multimodal sensitivity is also essential to one of the goals of this model chain: to be able to respond more accurately to the existence of transit guideways, high-occupancy vehicle (HOV) facilities, and a wide range of transportation policy variables.

The results of the model calibration indicated that the composite impedance formulation was suitable only for the work trip purpose. For the home-based other and non-home-based trip models, composite impedance could not successfully be used, and thus highway time was used. For the work model, the log sum method of combining impedance gave better results than the harmonic mean formulation. Finally, all three models were calibrated to a high degree of accuracy, with K factors used sparingly and only for trips crossing major geographic barriers.

MODAL-CHOICE DISUTILITY FUNCTIONS

As previously mentioned, a three-mode logit modal-choice model was calibrated for each trip purpose (8). These models are defined in terms of their disutility equations for each mode, as given in residual of 70 after 6 iterations, and a residual of 15 after 15 iterations.

4. The three multiplicative versions of the model always resulted in similar trip tables, which tended to be slightly different than that of the LINKOD algorithm.

5. Adding the λ weighting (Equation 8, model 3) increased the convergence speed only slightly.

CLOSENESS OF FINAL TRIP TABLE TO TARGET TRIP TABLE

The data in Table 3 give the value of ϕ for the final and target trip tables for the different algorithms and target trip tables. Figure 3 shows the ratios of ϕ for LINKOD (model 1) and the square root version of the IT algorithm (model 4).

1. All the algorithms succeed in producing final trip tables that are close to the target tables. Different target trip tables result in completely different final trip tables that, nevertheless, are similar in their ability to reproduce the observed flows.

2. The different multiplicative algorithms result in final trip tables whose distances from the target trip tables are similar. This is particularly significant relative to the algorithm with the λ weighting (model 3); its divergence from the basic form of the IT model (Equation 1) does not appear to harm its performance.

3. In most cases the multiplicative algorithms result in final trip tables that are slightly closer to the target trip table compared to LINKOD. This can be seen clearly in Figure 3.

4. In cases where the algorithms show convergence difficulties (Table 2f), the final trip table is not the feasible solution closest to the target. To confirm this point, a systematic search for the closest solution was made by using linear combinations of the eight basic solutions. The best trip table had $\phi = 0.274$ compared with $\phi = 0.553$ for the final trip table of the algorithm. In all cases without convergence difficulties, only slight differences between the two ϕ's were found.

CONCLUSIONS

In this paper the two major models for estimating trip tables based on traffic counts that have been verified in full-scale applications are compared. The analysis concentrates on all-or-nothing assignment problems. It is shown that the two models are similar, both in the structure of their algorithms and in their performance. LINKOD uses additive terms for the table correction steps, whereas the IT models use multiplicative terms. The different versions of the IT model produce similar results. They tend to produce final trip tables that are slightly closer to the target tables when compared with LINKOD.

The target trip table is shown to have major effects on all aspects of the solution. It dictates the structure of the final trip table and the speed of convergence. In any application of the model, the selection of a target trip table should be made with care.

The standard IT algorithm (model 2) failed to converge in one case. It should be used with care. All the other algorithms performed satisfactorily in all cases.

A significant result is the successful performance of model 3--multiplicative corrections with λ weighting (Equation 8). The λ weighting step is an essential element in any equilibrium assignment algorithm; the success of the model that includes this step gives a strong indication that it can perform successfully under equilibrium assignment assumptions. Development work in this direction is under way.

REFERENCES

1. V. Chan, R. Dowling, and A. Willis. Deriving Origin-Destination Information from Routinely Collected Traffic Counts--Phase I: Review of Current Research and Applications. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 80-1, 1980.
2. A.E. Willis and V.K. Chan. Deriving Origin-Destination Information from Routinely Collected Traffic Counts--Phase II: Comparisons of Techniques for Single-Path and Multipath Networks. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 80-10, 1980.
3. H.J. Van Zuylen. A Method to Estimate Trip Matrix from Traffic Volume Counts. Presented at Planning and Transport Research and Computation Company, Ltd. Summer Annual Meeting, Univ. of Warwick, Coventry, Warwickshire, England, 1978.
4. L. Willumsen. Estimation of an O-D Matrix from Traffic Counts: A Review. Institute for Transport Studies, Univ. of Leeds, Leeds, England, Working Paper 99, 1978.
5. H.J. Van Zuylen and L.G. Willumsen. The Most Likely Trip Matrix Estimated from Traffic Counts. Transportation Research, Vol. 14B, 1980, pp. 281-293.
6. S. Nguyen. Estimating an O-D Matrix from Network Data: A Network Equilibrium Approach. Centre de Recherche sur les Transport, Universite de Montreal, Montreal, Quebec, Canada, Publ. 60, 1977.
7. S. Nguyen. On the Estimation of an O-D Trip Matrix by Equilibrium Methods Using Pseudo Delay Functions. Centre de Recherche sur les Transport, Universite de Montreal, Montreal, Quebec, Canada, Publ. 81, 1977.
8. S. Nguyen. On Estimating O-D Matrices from Traffic Data. Presented at the Canadian Transport Research Forum Meeting, Winnipeg, Manitoba, Canada, June 12-14, 1978.
9. Y. Gur, M. Turnquist, M. Schneider, L. Leblanc, and D. Kurth. Estimation of an O-D Trip Table Based on Observed Link Volumes and Turning Movement, Volumes 1-3. FHWA, Rept. FHWA/RD-80/035, 1980.
10. D. van Vliet and L.G. Willumsen. Validation of the ME2 Model for Estimating Trip Matrices from Traffic Counts. Presented at the 8th International Symposium on Transportation and Traffic Theory, Univ. of Toronto, Toronto, Ontario, Canada, 1981.
11. A.F. Han, R.G. Dowling, E.C. Sullivan, and A.D. May. Deriving O-D Information from Routinely Collected Traffic Counts--Volume II: Multipath Networks. Institute of Transportation Studies, Univ. of California, Berkeley, 1981.
12. M. Turnquist and Y. Gur. Estimation of Trip Tables from Observed Link Volumes. TRB, Transportation Research Record 730, 1979, pp. 1-6.
13. A.E. Willis and A.D. May. Deriving O-D Information from Routinely Collected Traffic Counts--Volume I: Trip Table Synthesis for Single-Path Networks. Institute of Transportation Studies, Univ. of California, Berkeley, 1981.

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

small enough to permit complete analytical solutions. Out of the 15 cells in the trip table, 4 are always zero because of the structure of the network. The 10 volume counts provide 8 independent equations. Those equations, when combined with nonnegativity constraints on the cells of the trip table, can be solved with eight different basic trip tables, each with eight positive cells, which satisfy the observed flows. The data in Table 2b mark the cells that can be zero. The data in Table 2c are an example of a basic solution. Every scaled linear combination of the eight basic tables also satisfies the observed flows.

There exist a number of measures for the distance between two matrices. These measures are described by Willis and May (13). For the present project, the following distance measure was selected:

$$\Phi = (1/\Sigma_j t_j^n)\ \Sigma_i [t_i^n \cdot |\log (t_i^n/f_i)|] \tag{18}$$

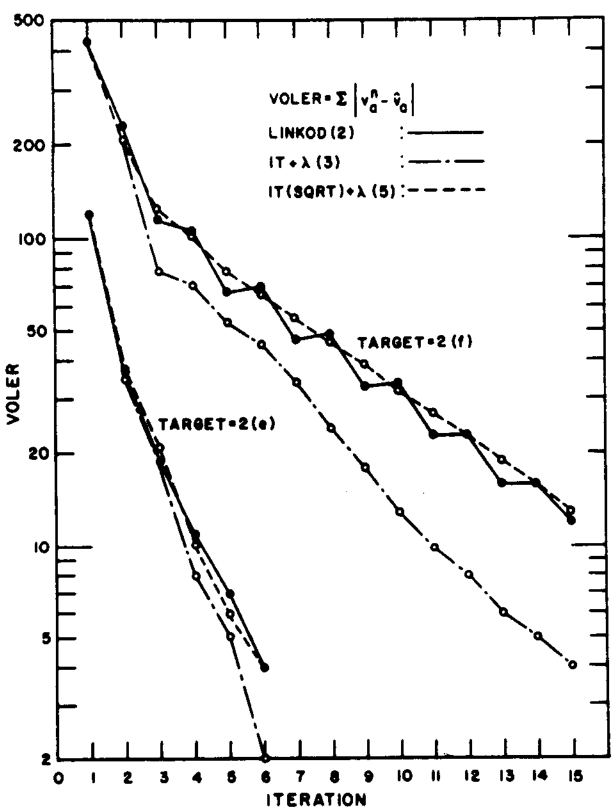
where F is the target trip table and T^n is the final trip table. This measure is a normalized equivalent to the distance measure used in developing the IT model.

The extent to which the final trip table approximates the observed flows was measured by two variables: the LINKOD objective function and the sum of absolute volume errors, i.e.,

$$\text{VOLER} = \Sigma_a |\hat{v}_a - v_a^n| \tag{19}$$

For the main body of the experiments, a number of different target trip tables were specified, and a set number of iterations (5 or 15) were run by using the different models. The statistics of the different runs were used for model evaluation. The major results are shown in Figures 2 and 3 and are given in Table 3.

Figure 2. Convergence speed for the different algorithms.



CONVERGENCE CHARACTERISTICS

The data in Table 3 give the values of the various error measures that use different target trip tables. Residual errors after each iteration for two sample tables are shown in Figure 2. The main conclusions are as follows.

1. At least in one case (target table as specified in the data in Table 2d), the simple IT algorithm (number 2) failed to converge. The other three algorithms always converged.
2. The LINKOD algorithm tends to improve the solution more than the IT algorithms during the first one or two iterations. However, the multiplicative algorithms tend to be more efficient when the errors are small. In general, after five or more iterations, all the algorithms show similar residual errors.
3. The speed of convergence depends strongly on the target trip table. It is interesting to note that all of the algorithms display convergence difficulties exactly for the same target trip tables. The data is Tables 2e and f give the two target tables whose convergence patterns are shown in Figure 2; these patterns display that behavior. Although the data in Table 2e give a residual of about 4 after 6 iterations, the data in Table 2f give a

Figure 3. Distance ratios for LINKOD and IT models.

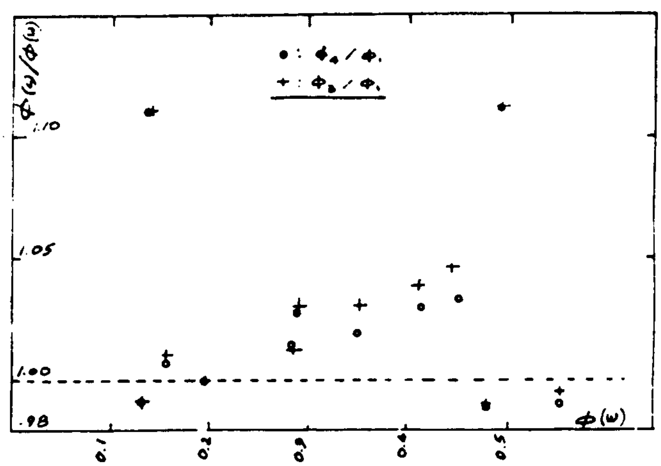


Table 3. Summary of performance measures by trip table and algorithm.

Target Table	Residual Error (VOLER[a]) LINKOD	IT with	IT[c] (square root)	Distance from Target (Φ^b) LINKOD	IT with	IT[c] (square root)
1	25	22	28	0.159	0.157	0.158
2	29	28	31	0.154	0.138	0.138
3	20	15	15	0.289	0.284	0.284
4	60	56	56	0.426	0.410	0.415
5	13	4	5	0.300	0.292	0.292
6	26	17	26	0.363	0.371	0.356
7	16	8	10	0.476	0.488	0.481
8	4	4	4	0.131	0.131	0.131
9	12	4	13	0.553	0.555	0.557
10	70	66	65	0.467	0.446	0.452
11	9	8	7	0.195	0.195	0.195

Note: The table presents values of the performance measures after five iterations.
[a] VOLER is the sum of absolute link volume errors, VOLER = $\Sigma|v_a - v_a^n|$.
[b] Φ is defined as described in Equation 18.
[c] By using analytical techniques, solutions with Φ of 0.130 and 0.268 were found for target trip tables 8 and 9, respectively.

Another difference between the two algorithms is the need for the λ weighting (step 8). In the LINKOD model the table T^c is a correcting trip table that points to the direction of the needed correction in any iteration. Therefore, the λ weighting is an essential part of the algorithm. In the IT algorithm T^c is a corrected trip table; thus the λ weighting is not a necessary part of the algorithm; it might even be harmful.

Note that in the IT algorithm, without λ weighting (x from Equation 1) is

Figure 1. Test network.

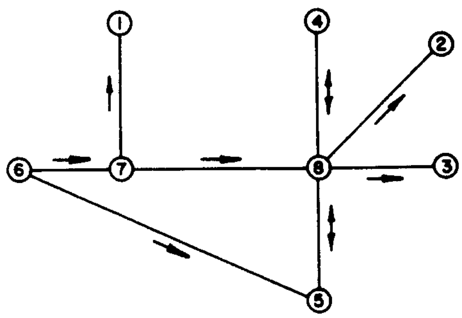


Table 1. Test network attributes.

Node A	Node B	c_a	b_a	v_a
4	8	6.9	0.03	130
5	8	9.2	0.04	130
6	5	31.0	0.05	20
6	7	19.6	0.04	290
7	1	8.0	0.05	60
7	8	13.9	0.03	230
8	2	13.5	0.05	170
8	3	12.0	0.03	200
8	4	11.4	0.06	90
8	5	6.5	0.05	30

$$x_a = \pi_m\, y_a^{w,m} \tag{16}$$

where $y_a^{w,m}$ is the value of y from Equation 13 at the mth iteration. However, if λ weighting is used, then the solution trip table (T^n) cannot be expressed in terms of Equation 1. Thus it is doubtful whether the solution with the λ weighting is the best solution, as specified by the IT criteria. In the pure IT model, λ in Equation 8 is always 1.

EXPERIMENTATION

The algorithm as described in the previous section was programmed, and a set of experiments with the two models and some other variations were performed. The experiments were designed to answer the following questions.

1. How do the various algorithms perform in terms of speed of convergence to a trip table that approximates the observed flows? Are there cases where the algorithms fail to converge?
2. How do the various algorithms perform in terms of finding a solution that is close to the target trip table?

The experimentation started with three models: (a) LINKOD, (b) IT with λ = 1, and (c) IT with optimal λ (step 8 of the algorithm). After a few experiments, it was found that in certain circumstances the standard IT model overcorrects the trip table and fails to converge. A fourth version of the model was added, where Equation 13 was replaced by Equation 4:

$$y_a^{w^1} = \mathrm{SQRT}\,(\bar{v}_a/v_a^n) \tag{17}$$

and λ = 1.

The experiments use the network shown in Figure 1, with the link attributes given in Table 1. The data in Table 2a give an example of a trip table that satisfies the observed flows. The problem is

Table 2. Test trip tables.

a. A solution (an example)

	1	2	3	4	5
3	0	60	40	0	30
4	0	30	60	40	0
5	60	80	100	50	20

b. Structure of a solution

1	2	3	4	5	
–	0	0	–	x	130
–	0	0	0	–	130
x	0	0	0	x	310
60	170	200	90	50	570

Note: – means it must be zero, 0 means it can be zero, and x means it must be positive.

c. A basic solution (an example)

	1	2	3	4	5
3	0	0	100	0	30
4	0	0	40	90	0
5	60	170	60	0	20

d. A target that cannot converge with algorithm 2

	1	2	3	4	5
3	0	52	52	0	52
4	0	52	52	52	0
5	52	52	52	52	52

e. A quickly converging target

	1	2	3	4	5
3	0	60	50	0	40
4	0	40	50	40	0
5	50	70	90	60	50

f. A slowly converging target

	1	2	3	4	5
3	0	0	50	0	40
4	0	0	90	80	0
5	80	60	70	0	30

factors used in that run were then keypunched into a file that could be used by Statistical Package for the Social Sciences (SPSS) programs (16).

3. The SPSS subprogram REGRESSION was then used to obtain a least squares fit for the coefficients A, B, and G (after suitable transformation of the variables).

4. New F factors were calculated by using the new coefficients, and program AGM was reapplied. The observed and estimated trip lengths were then compared, and if the results were inadequate, steps 2-4 were performed again.

The results were judged by a visual inspection of the impedance distribution and by comparing the average values for composite impedance and trip length. By using this technique, a satisfactory set of F factors could be obtained in between six and nine iterations.

RESULTS

The results of the impedance calculations are given in Table 4. The composite impedance values differ markedly by income level and are biased in the proper direction. That is, the lower-income levels are associated with higher impedance. This reflects the fact that lower-income persons tend to have lower mobility (for example, they are less likely to own automobiles). The composite values also indicate a larger spread than the time values, which may suggest more specific relationships between composite impedance and F factors. These statistics indicate that the composite impedance formulation behaves mathematically. This increases the confidence with which it can be used in gravity model development.

Table 4. Summary of observed impedance values.

Purpose and Income Level[a]	Average Value[b]	Lowest Value[c]	Highest Value[c]	Standard Deviation
Home-based work trips				
1	62.189	44	111	9.908
2	56.338	38	98	8.695
3	47.171	29	84	8.596
4	38.063	21	77	8.670
Home-based other trips				
1	7.579	1	43	4.942
2	7.965	1	39	5.334
3	8.214	1	44	6.118
4	7.692	1	44	5.679
Non-home-based trips				
1	7.720	1	36	5.263
2	7.671	1	43	5.451
2	7.710	1	39	5.392
4	7.520	1	39	5.130

[a] Income level is divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

[b] The values for home-based work represent log sum composite impedance. All other values represent highway time. All highway times used were off-peak highway times without terminal time.

[c] Highest and lowest values that contain observed trips.

The results of calibrating the home-based work model with both sets of composite impedance functions are given in Tables 5-7. These comparisons indicate that the log sum results are superior to those obtained with the harmonic mean formula. The basic philosophy of these comparisons was that, if the model could be calibrated by using one type of impedance and could be shown to properly replicate the means of a different (but related) type of impedance, the calibration would be considered successful. The log sum formulation estimates average highway travel time and distance considerably better than did the harmonic mean formulation, except for the lowest income quartile. When intrazonal trips are compared, the log sum approach is superior for total trip estimation, but slightly inferior for the low and high-middle income quartiles. In comparing major trip patterns, such as trips across the Mississippi River, the harmonic mean model overestimated the observed data by 69 percent, whereas the log sum model overestimated by only 44 percent (before K factors were applied, in both cases). In addition, a comparison was made of the number of district interchanges (there are 20 districts) for which the difference between observed and estimated trips was greater than 100 trips and the percentage difference was greater than 15 percent. The harmonic mean model had 68 such district interchanges, whereas the log sum model had 56.

Table 5. Comparison of two composite impedance formulas: highway running time (min).

	Harmonic Mean			Log Sum		
Income Group[a]	Observed	Estimated	Percent Error	Observed	Estimated	Percent Error
1	10.17	10.67	4.92	10.17	10.82	6.39
2	10.18	11.02	8.25	10.18	10.59	4.03
3	10.87	11.59	6.62	10.87	10.97	0.92
4	11.16	11.87	6.36	11.16	11.26	0.90
All income groups	10.68	11.39	6.65	10.68	10.94	2.43

Note: These values represent the home-based work purpose gravity model runs, without K factors.

[a] Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Table 6. Comparison of two composite impedance formulas: highway distance (mile).

	Harmonic Mean			Log Sum		
Income Group[a]	Observed	Estimated	Percent Error	Observed	Estimated	Percent Error
1	4.29	4.53	5.59	4.29	4.64	8.16
2	4.29	4.70	9.56	4.29	4.52	5.36
3	4.72	5.15	9.11	4.72	4.80	1.69
4	4.91	5.29	7.74	4.91	4.96	1.02
All income groups	4.61	4.99	8.24	4.61	4.75	3.04

Note: These values represent the home-based work purpose gravity model runs, without K factors.

[a] Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Table 7. Comparison of two composite impedance formulas: number of intrazonal trips.

	Harmonic Mean			Log Sum		
Income Group[a]	Observed	Estimated	Percent Error	Observed	Estimated	Percent Error
1	936	882	−5.77	936	857	−8.44
2	2,202	1,854	−15.80	2,202	2,396	8.81
3	2,895	2,471	−14.65	2,895	3,437	18.72
4	3,113	2,374	−23.74	3,113	2,866	−7.93
All income groups	9,146	7,581	−17.11	9,146	9,556	4.48

Note: These values represent the home-based work purpose gravity model runs, without K factors.

[a] Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Based on this analysis, the log sum formulation was chosen to complete the calibration of the distribution model. Because the log sum formula worked well for home-based work trips, this method was used for home-based other and non-home-based trips as well. An F factor equation for home-based other trips was calibrated, and the model was applied for a validation check. However, in this case, in comparing observed and estimated trips with respect to highway time and distance, the estimated trips showed a much higher trip length. The estimated trips were considerably less than the observed trips in the 1-, 2-, and 3-min time range and considerably higher in the 4-, 5-, and 6-min time range. At times greater than approximately 7 min, the two distributions were similar. Considerable thought was given to correct this imbalance in distribution, but no methodology appeared to offer any reasonable chance of successful calibration. Because of these results, the home-based other distribution model was calibrated by using off-peak highway travel time rather than composite impedance.

Similar results were obtained for the non-home-based model calibration, leading to the same solution: use of off-peak highway time instead of composite impedance. Off-peak highway time was also used for the remaining models (taxi, internal-external, and truck).

There is speculation that the lack of success in using composite impedance in the nonwork models is related to the nature of nonwork trips compared with work trips. Work trips are methodical and repetitive, and the commuter may actually have more knowledge than the nonwork traveler about his modal options and their associated impedances. Nonwork trips are less structured, and perhaps less thought is given to alternative modes for such trips. That is, cost considerations and the availability of transit service may not strongly affect nonwork destination choice.

The calibrated F factor equations for all trip purposes are given in Table 8, with the F factors being defined by the three coefficients of the gamma distribution. All coefficients are statistically significant, and the correlation coefficient (R^2), which compares the required F with the calculated F, was greater than 0.90. The regression program equations have been adjusted, where necessary, to ensure that the highest F value is not more than 999,999, in order for the data to be acceptable to AGM.

Table 8. F factor equations.

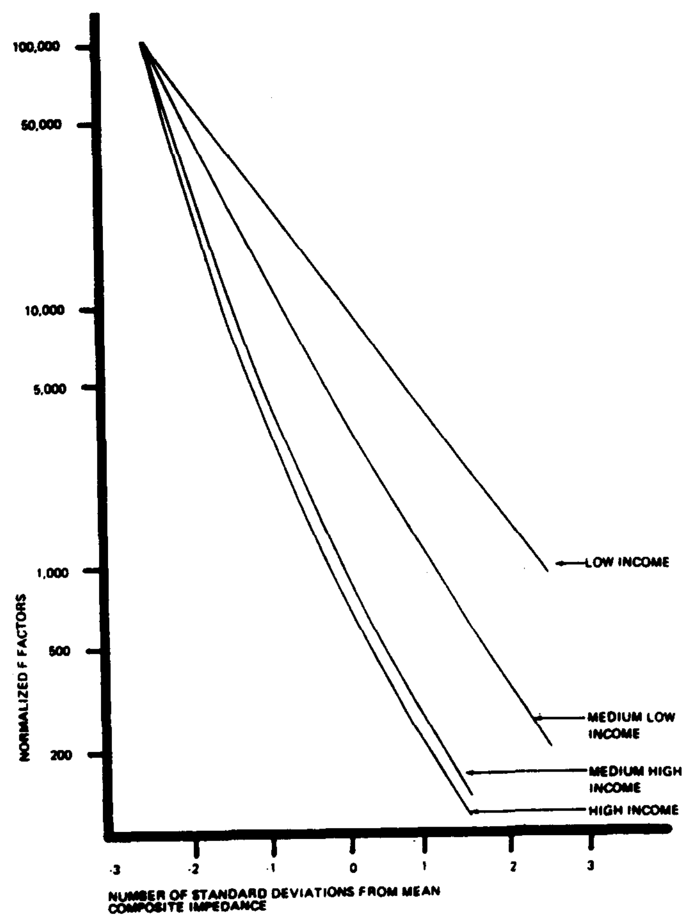
Purpose and Income Group[a]	Equation Coefficient Values[b] A	B	G
Home-based work trips			
1	4,296,752	0	-0.09300397
2	EXP (26.82271)	-3.153498	-0.0836755
3	EXP (34.10976)	-6.800698	-0.024841
4	EXP (28.39026)	-4.819197	-0.041024
Home-based other trips			
1	1,064,302	-1.055559	-0.1054066
2	1,070,772	-1.292004	-0.09307232
3	647,077	-1.838836	-0.03701391
4	1,033,560	-1.838298	-0.05231526
Non-home-based trips			
1	663,504	-0.6655663	-0.1231575
2	869,114	-0.9009789	-0.1125171
3	267,378	-0.9540237	-0.1127642
4	371,881	-0.7850539	-0.138105

[a] Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.
[b] F factors are calculated by using the equation: $F(I) = A \cdot I^B \cdot EXP(G \cdot I)$; where I is the composite impedance for work trips and highway time for the other trip purposes.

The primary reason for calibrating the distribution models by income quartile was the hypothesis that tripmakers in different income levels would react differently to the impedance measure. Although the gamma function coefficients are different for the four income levels, it is hard to ascertain the true difference because the F factors are relative, and the mean composite impedance values are different by income level. To test the hypothesis that the F factors are truly different by income level, a set of normalized F factors were calculated by using the mean composite impedance values and the standard deviations from the mean. Normalized F factors were developed by adjusting the constant term (the A coefficient in the gamma equation) so that the F factor would (arbitrarily) equal 100,000 at a composite impedance value, which was 2.5 standard deviations less than the mean value. This comparison is shown in Figure 1. In essence, the comparison shows that F factors for the lower incomes are less sensitive to the impedance values. It would not appear reasonable, though, to use this comparison to draw the conclusion that poorer people like to travel more than richer people. Perhaps a better explanation is that the lower-income traveler has fewer destinations to choose from, thereby reducing the impact of travel impedance on travel behavior, at least on distribution.

Most calibration reports on distribution models give the observed and estimated trips stratified by highway travel time. For distribution models calibrated by using highway time, these comparisons nor-

Figure 1. Home-based work normalized factors plotted against standard deviation units of composite impedance.



mally indicate a great deal of agreement between the observed and estimated trips, which is only reasonable because the F factors are directly related to highway time. Because the spatial measure used in this model was composite impedance, of which highway travel time was only one component, a comparison of the observed and estimated trips measured against highway travel time would be a useful validation test, as mentioned previously. These comparisons are shown graphically in Figures 2-6. As can be seen from the data in these figures, the estimated trips agree with the observed trips extremely well. The comparisons by income level are similar to normal gravity model trip-distribution comparisons. When the trips for all incomes are combined (Figure 6), the observed trip pattern is much smoother and the estimated trips compare extremely well with the observed trips.

After calibrating the F factors, the next step in the calibration procedure was to ascertain the trip movements that were inadequately simulated and that had specific attributes that would be identifiable

Figure 2. Comparison of trip distributions for low income home-based work trips.

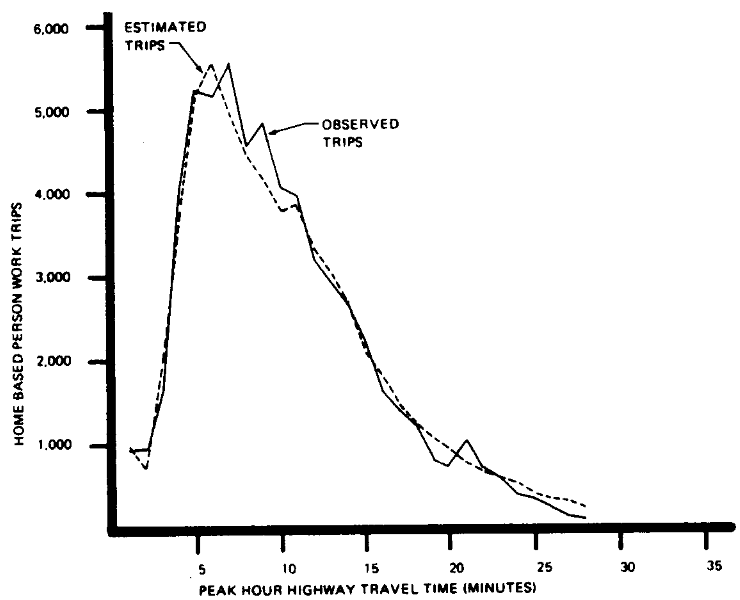


Figure 3. Comparison of trip distributions for low-medium income home-based work trips.

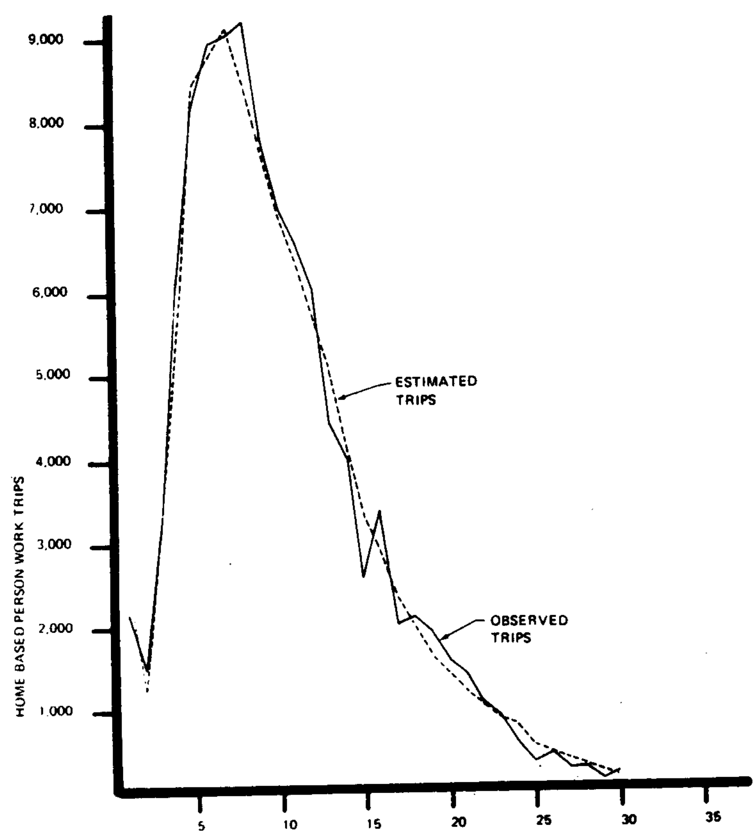


Figure 4. Comparison of trip distributions for high-medium income home-based work trips.

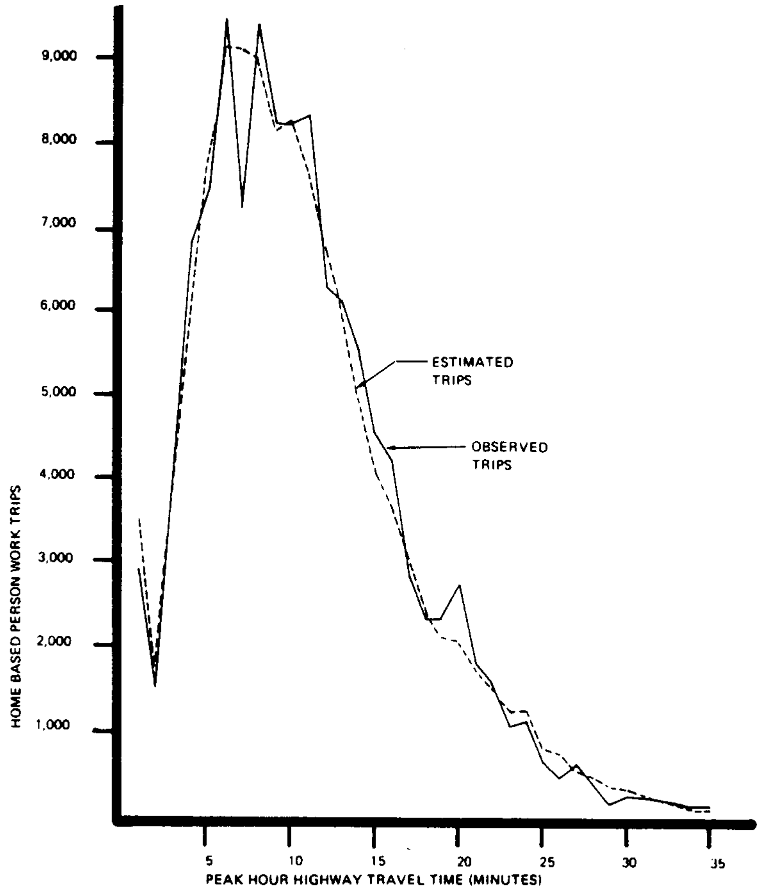


Figure 5. Comparison of trip distributions for high income home-based work trips.

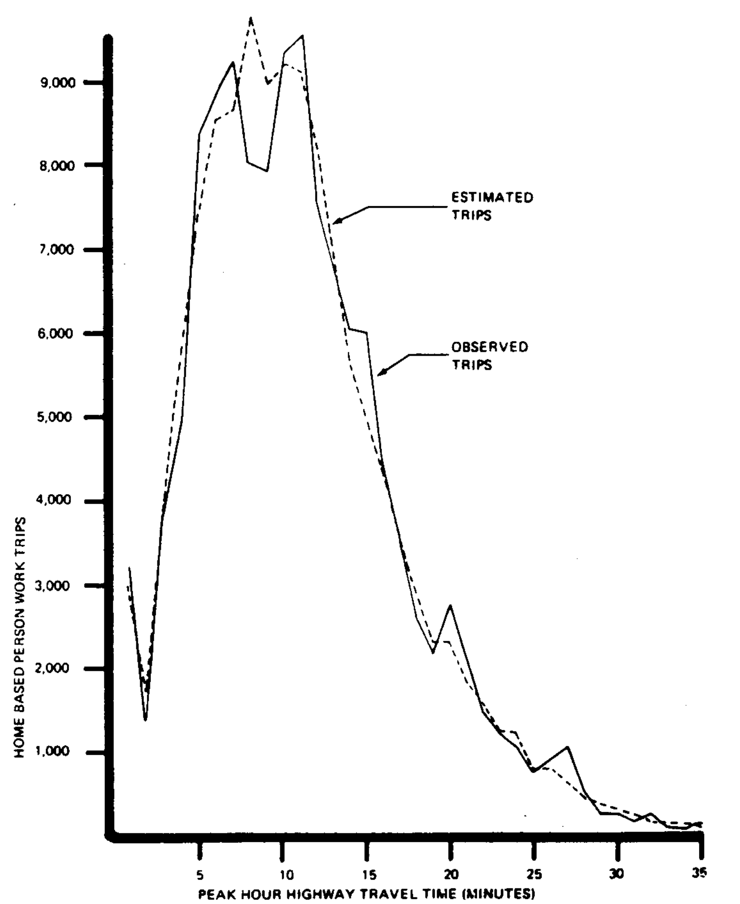
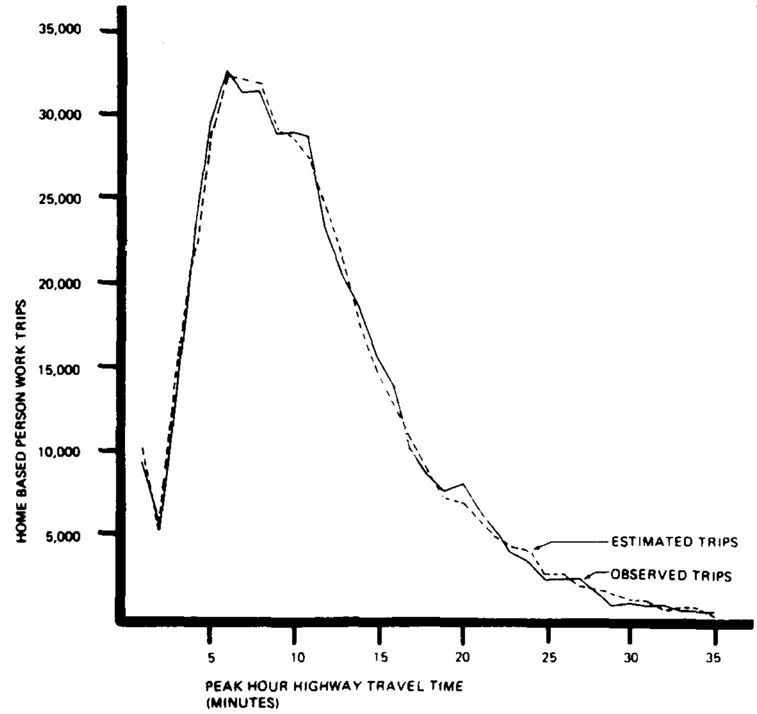


Figure 6. Comparison of trip distributions for all home-based work trips.



in the future. The two most important movements meeting these criteria were the water crossings, specifically the trip movements across the Mississippi River and the Navigational Canal. As can be seen from the data in Tables 5-7, even the log sum approach overestimated these trip movements for work trips. River crossings are traditionally difficult to estimate because there is a psychological factor associated with crossing this type of barrier. The calibration method to estimate the K values was to summarize the observed and estimated trips crossing the barrier and calculate the K value as a ratio of these two values. Because the K value appears both in the numerator and the denominator of the distribution formula, this formulation does not estimate a correct K factor in one iteration. Several iterations were required to develop K factors that produced adequate results. The final K factors are given in Table 9.

Table 9. Final K factors.

Purpose and Income Group[a]	K Factors: Across Mississippi River	K Factors: Across Navigational Canal
Home-based work		
1	0.496	0.798
2	0.463	0.962
3	0.703	0.896
4	0.660	0.895
Home-based other		
1	0.197	0.972
2	0.184	0.897
3	0.241	0.899
4	0.241	1.000
Non-home-based		
1	0.365	0.702
2	0.316	0.860
3	0.368	0.818
4	0.351	0.805

Note: For trips that cross both waterways, the Mississippi K's are used.

[a] Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Once the F factors and K values were calibrated, the full distribution model was applied by using AGM. The resulting trip table was then compared with the observed trip table by using several tests. The results of these tests for the work trip purpose are given in Tables 10-12. A primary check on the distribution model was to ascertain if the estimated trips had the same distribution as the observed trips when the impedance measure was highway time or highway distance or both. For the work trip model, this comparison was excellent. Total estimated work trips had an average highway travel time and highway distance that differed from observed trips by less than 0.2 percent. When the average travel time was compared by income level, the results were slightly less accurate but well within normal limits of acceptability. The number of intrazonal trips was also compared, and the results were favorable. Three screen-line checks were made: trips across the Mississippi River, trips across the Navigational Canal, and trips between Orleans Parish and Jefferson Parish. The model overestimated the latter by 1.44 percent, and most of this error was in the lowest income quartile.

The home-based other and non-home-based results are given in Table 13. The average travel time and distance for observed and estimated trips were similar. The model tended to underestimate intrazonal trips, but estimated travel across both major waterways (the Mississippi River and the Navigational Canal) extremely well. However, the movements between Orleans Parish and Jefferson Parish were overestimated.

The income-related sensitivity of the home-based other models to travel time is similar to that of the work models in that low-income travelers are less sensitive than high-income travelers, as shown in Figure 7. However, this sensitivity is less pronounced than for work trips.

Similar models were calibrated for internal-external person trips, taxi vehicle trips, internal-external truck trips, and internal-internal truck trips. They are discussed in the more detailed report on distribution models for New Orleans (15).

Table 10. Final calibration results of home-based work gravity model: average impedance.

Income Group[a]	Highway Running Time Observed	Highway Running Time Estimated	Highway Running Time Percentage Error	Highway Distance Observed	Highway Distance Estimated	Highway Distance Percentage Error	Composite Impedance Observed	Composite Impedance Estimated	Composite Impedance Percentage Error
1	10.17	10.56	3.83	4.29	4.49	4.66	62.19	62.18	-0.02
2	10.18	10.31	1.28	4.29	4.35	1.40	56.34	56.14	-0.35
3	10.87	10.77	-0.92	4.72	4.68	-0.85	47.17	46.85	-0.68
4	11.16	11.04	-1.08	4.91	4.84	-1.43	38.06	37.85	-0.55
All income groups	10.68	10.70	0.19	4.61	4.62	0.22	48.94	48.73	-0.43

[a]Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Table 11. Final calibration results of home-based work gravity model: number of intrazonal trips.

Income Group[a]	Observed	Estimated	Percentage Error	Total Trips	Intrazonal Trips as a Percentage of Total Trips: Observed	Intrazonal Trips as a Percentage of Total Trips: Estimated
1	936	884	-5.56	61,994	1.51	1.43
2	2,202	2,438	10.72	105,327	2.09	2.31
3	2,895	3,477	20.10	120,191	2.41	2.89
4	3,113	2,906	-6.65	127,533	2.44	2.28
All income groups	9,146	9,705	6.11	415,045	2.20	2.34

[a]Income groups are divided as follows: 1 = low, 2 = low-middle, 3 = high-middle, and 4 = high.

Table 12. Final calibration results of home-based work gravity model: major movement comparison.

Movement	Observed Trips	Estimated Trips	Percentage Error
Across Mississippi River	25,269	26,639	5.42
Across Navigational Canal	38,770	39,985	3.13
Between Orleans and Jefferson Parishes	71,143	72,164	1.44

SUMMARY

A complete set of distribution models was calibrated for the New Orleans region. The original intent of this calibration was to prepare a set of distribution models stratified by income level and using a combined impedance measure that would adequately reflect the travel time and cost of all models. This design proved to be feasible for home-based work trips, thus producing an excellent trip-distribution model. The log sum method of combining impedances yielded better results than the harmonic mean method.

For home-based other and non-home-based trips, the use of a combined impedance measure produced a model that overestimated long trips. For this reason, these models were calibrated by using off-peak highway times. For the home-based work and home-based other distribution models, the income stratification produced F factors that were substantially different by income level, but were logical, in that the higher income strata were more sensitive to travel impedance. The F factors for the non-home-based model showed only minor differences among income strata. In all cases the F factors were calibrated by using the function described in the documentation of program AGM and by using standard statistical regression techniques. A set of K values was required for all models for trip movements across the Mississippi River and the Navigational Canal. The final screen-line checks were quite accurate, with the exception of the home-based other trip movements between Orleans and Jefferson parishes.

With respect to model validation, the results of applying the models to 1980 conditions proved quite satisfying. As reported by Schultz (see paper elsewhere in this Record), the changes between 1960 and 1980 in New Orleans have been substantial. Nonetheless, the 1980 estimates of vehicle miles of travel were within 5 percent of the observed data, which indicate that the distribution (and modal-choice) models performed adequately.

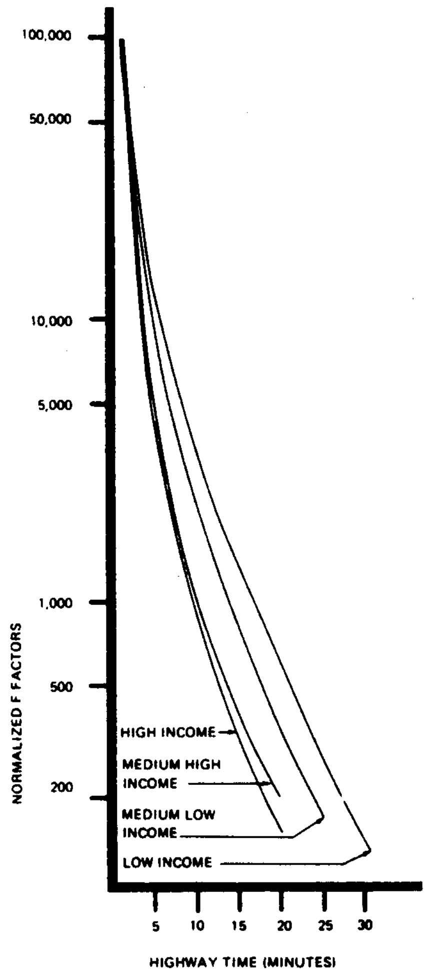
ACKNOWLEDGMENT

The results given in this paper are from a study performed for the Regional Planning Commission (RPC) of Jefferson, Orleans, St. Bernard, and St. Tammany parishes, Louisiana, which was funded in part by

Table 13. Final calibration results of home-based other and non-home-based gravity models.

Measure	Home-Based Other Observed	Home-Based Other Estimated	Home-Based Other Percentage Error	Non-Home Based Observed	Non-Home Based Estimated	Non-Home Based Percentage Error
Highway running time (min)	7.891	7.864	-0.34	7.630	7.642	0.16
Highway distance (mile)	3.186	3.225	1.22	3.089	3.118	0.94
No. of intrazonal trips	90,632	72,593	-19.90	17,827	17,817	-0.06
Major movements						
Across Mississippi River	14,941	15,022	0.54	5,520	5,575	1.00
Across Navigational Canal	39,288	39,515	0.58	7,746	7,825	1.02
Between Orleans and Jefferson Parishes	56,631	68,044	20.15	21,512	23,678	10.07

Figure 7. Home-based other normalized F factors plotted against highway travel time.



UMTA, U.S. Department of Transportation. I would like to thank Dennis D. Striz of the RPC and Foster de la Houssaye of the Office of Transit Administration, City of New Orleans, for their assistance, comments, and support in the completion of the work reported herein.

REFERENCES

1. M.L. Manheim. Practical Implications of Some Fundamental Properties of Travel-Demand Models. HRB, Highway Research Record 422, 1973, pp. 21-38.
2. M.L. Manheim. Fundamental Properties of Systems of Demand Models. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, 1970.
3. A.G. Wilson. Travel-Demand Forecasting: Achievements and Problems. In Urban Travel Demand, HRB, Special Rept. 143, 1973, pp. 283-306.
4. D.A. Hensher and P.R. Stopher, eds. Behavioral Travel Modelling. Croom Helm, London, England, 1979.
5. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
6. Cambridge Systematics, Inc. Travel Model Development Project. Metropolitan Transportation Commission, San Francisco, 1977.
7. System Management Contractor. SMC Multi-Modal Patronage Model. Regional Transportation District, Denver, 1975.
8. Barton-Aschman Associates, Inc. Development and Calibration of the New Orleans Mode Choice Models. Regional Planning Commission, New Orleans, 1981.
9. R.H. Pratt and Associates, and DTM, Inc. Development and Calibration of Mode Choice Models for the Twin Cities Area. Twin Cities Metropolitan Council, Minneapolis-St. Paul, 1976.
10. Barton-Aschman Associates, Inc. (under subcontract with Parsons, Brinckerhoff Quade, and Douglas in Seattle). Development and Calibration of the Mode Choice Models. Municipality of Metropolitan Seattle, Seattle, 1979.
11. Barton-Aschman Associates, Inc., and Texas Transportation Institute. Development and Calibration of Travel Demand Models for the Houston-Galveston Area. Metropolitan Transit Authority, Houston, 1979.
12. Barton-Aschman Associates, Inc. Calibration of the St. Louis Mode Choice Models. East-West Gateway Coordinating Council and the Bi-State Development Agency, St. Louis, 1982.
13. Barton-Aschman Associates, Inc. (under subcontract to ATEC, S.A., in Buenos Aires). Analisis de Alternativas para la Ampliacion de la Red de Subterraneos de la Ciudad. Municipality of Buenos Aires, Buenos Aires, Argentina, 1981.
14. Urban Transportation Planning System. UMTA and FHWA, 1972.
15. Barton-Aschman Associates, Inc. Development and Calibration of the New Orleans Distribution Models. Regional Planning Commission, New Orleans, 1981.
16. N.H. Nie et al. SPSS: Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.

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

Development of a Travel-Demand Model Set for the New Orleans Region

GORDON W. SCHULTZ

A complete set of travel-demand models was calibrated for the New Orleans region by using the 1960 origin-destination survey. The general form of the model set was sequential, with care being taken to include transportation system characteristics in all submodels of the modeling set. Other unique features of the model set were that (a) all submodels were stratified by income quartiles; (b) the distribution model used a composite impedance that combined travel time and costs for all modes, (c) the generation model used accessibility and locational measures, and (d) the exogenous input data, required in forecasting, were limited to six data items. The calibrated models were applied to 1980 conditions, and the resulting travel estimates were compared with ground counts. This comparison indicated that the model set could produce reasonably accurate 20-year forecasts.

In 1980 the New Orleans Regional Planning Commission (RPC) decided to update its travel-demand modeling procedures to support ongoing transportation planning in the New Orleans region. A previous set of models was developed in 1972. A review of these models indicated a number of deficiencies that made them inappropriate for the current planning environment, especially with respect to the modeling of substantial new transit service and high-occupancy vehicle (HOV) incentives.

Because of limited resources available for this model update, it was necessary to use an existing home interview survey, which was taken in 1960, rather than to conduct a limited new origin-destination survey. A conservative estimate of the cost of a limited survey indicated that more than a third of the available resources would be required to conduct this survey. It was also observed that a set of models based on the 1960 survey would allow the study team to immediately make a 20-year forecast, i.e., to 1980, which could be validated by using existing ground counts.

It was judged that the available resources were sufficient to develop a set of sophisticated models that could be applied by using the standard transportation planning computer programs. An initial decision was made that the model set would be implemented by using the Urban Transportation Planning System (UTPS) developed by UMTA and FHWA. Another initial decision was that the general model structure would be the sequential model form (generation, distribution, mode choice). It was believed that this model structure gave the best assurance of successfully calibrating the model set within the resources available, and that by proper specification most of the shortcomings of a sequential model structure could be overcome or minimized.

In this paper the general philosophy and structure of the New Orleans travel-demand model set are described, and the results of applying this model set to the 1980 conditions are presented.

MODEL STRUCTURE AND PHILOSOPHY

The goals of the New Orleans travel-demand model update were to develop a model set that would include transportation system characteristics for all major travel functions, would be reasonably easy to apply in the forecast mode, and would require a minimal amount of exogenous data in the forecast mode. The goal of incorporating transportation system characteristics into all major travel function submodels (i.e., generation, distribution, and mode choice) is a fairly standard objective for a travel-demand set, but in many cases the goal is not realized. The ease-of-application goal is reasonable and obvious, but there are many urban area model sets that require extremely large amounts of computer resources and person hours to implement.

In many ways the goal of minimizing exogenous input data is the key to producing logical forecasts with a reasonable amount of resources. Model sets that require extremely detailed exogenous data simply shift the possibility for errors to other modeling efforts, impose a large expenditure of resources on other planning groups, and contribute to the phenomenon of adjusting the data so that the answer is correct. It was the objective of this study to constrain the exogenous input data to elements that are normally forecasted and can be evaluated for reasonableness by using other forecasts or by using standard reasonableness checks.

The stated goals for the New Orleans travel-demand model update led to the establishment of the following objectives:

1. The trip-generation element of the model set should include not only socioeconomic and land use data, but it should also include locational measures that describe the transportation system and the urban form of the area;
2. The distribution element of the model set should incorporate all relevant transportation system characteristics for all modes of travel;
3. The modal-choice element of the model set should be properly sensitive to transportation system characteristics, socioeconomic measures, and land use form, and the model should be applicable to planned HOV incentives;
4. All elements of the model set should be stratified by a socioeconomic characteristic that measures the wealth of the traveler;
5. The model set should require a minimal amount of exogenous data in the forecast mode, and this data should lend itself to reasonableness checks; and
6. The procedures for forecasting with the model set should use straightforward computer programs, either UTPS programs or programs compatible with the UTPS system, and these programs should be relatively easy and inexpensive to apply.

There are two general types of model forms that meet the first three objectives and that have been developed in other urban areas: direct-demand models and sequential models. The direct-demand structure is theoretically the better structure for including transportation system characteristics in all elements of the model set. For this study, though, it was believed that the resources required to calibrate a direct-demand model set would probably exceed the project's budget, and that a sequential structure could be developed to meet all the objectives. In addition, the sequential structure allowed the project to have a fallback position in the event that the initial model specifications were impossible to implement within the budget constraints (the fallback position being the standard sequential model specification).

The objective to stratify all the model elements--generation, distribution, and mode choice--by a socioeconomic characteristic that measures wealth is not a unique proposal. Most trip-generation production models use this type of stratification, and many modal-choice models also have a stratification based on wealth. The deficiency with most of these model sets is that the distribution model is not stratified by the wealth measure, and therefore there is no connectivity among the submodels with respect to the wealth measure. By performing a complete stratification by the wealth measure, the model set would have complete connectivity with respect to this measure. That is, low-wealth trip ends would be distributed by using a low-wealth impedance measure, and these person trips could then be allocated to each mode by using a low-wealth modal-choice formulation. The development of a distribution model stratified by a measure of wealth presented no theoretical or practical problems. The major impediment in the development of a fully stratified set of travel-demand models was the development of a stratified trip-generation attraction model. It was hypothesized at the beginning of the project, though, that a wealth-stratified attraction model could be developed if proper attention was given to locational variables.

The last two objectives--minimal data input and ease in application--were essential if the model set was to be frequently used in the forecasting mode. Model sets that require extremely large resources, both in person hours and computer costs, have little usefulness, regardless of their level of accuracy, because most planning organizations have constrained resources and tend to implement these expensive model sets only once every 2 or 3 years. It should be the intent of all organizations developing travel-demand models that these models can be reasonably used at least three or four times a year.

In summary, the philosophy for developing the New Orleans travel-demand model set was to (a) develop a sequential set of models completely stratified by a measure of wealth, (b) have transportation system characteristics present in each submodel, and (c) require a minimal amount of exogenous input data. Locational measures were anticipated to be significant variables in the trip-generation model, and measures representing time and cost for all modes were to be explored as independent variables for the distribution model. Care was to be taken in the development of the models to ensure a resource-efficient application methodology.

MODEL DEVELOPMENT

The final New Orleans travel-demand model set consisted of three major models--generation, distribution, and mode choice--and six auxiliary models. The study team was able to develop a model set by using only six socioeconomic and land use data items along with the normal set of transportation system data items. The following list gives a summary of the exogenous data input items:

1. Socioeconomic and land use data (at the zone level)--population, households, retail employment, nonretail employment, area of zone, and mean zonal household income;
2. Highway system data (link specific)--distance, facility type, number of lanes, and toll; and
3. Transit system data--distance (link specific), facility type (link specific), travel time for nonlocal route links, headway (route specific), and fares (interchange specific).

The use of population and households to estimate travel demand is normal. The study team would have preferred to use a more detailed classification for employment than retail and nonretail, but the base year data did not allow any finer stratification. Traffic analysis zones were used to calculate gross density measures, such as employment per acre. The mean household income of a zone was chosen as the only exogenous socioeconomic variable and was primarily used to estimate the number of households in each income quartile by zone. The project team considered whether to use income or automobile ownership as the primary socioeconomic variable. Although automobile ownership appears to have a greater effect on tripmaking and mode choice than income, automobile ownership was not chosen for the following reasons.

1. There are many variables that influence automobile ownership. Some of the obvious variables are household income, the availability and magnitude of the transit system, the structure of the city in terms of density, and general economic conditions. The use of automobile ownership as a variable would require a fairly detailed forecasting model (including the use of an income measure), which was considered to be a difficult model to calibrate.
2. There are a considerable number of independent forecasts of national and regional income levels that can be used to evaluate the income estimates used in the forecasts.
3. A recent study (1) has indicated that household trip rates are declining over time for a given level of automobile ownership. In some cases the decline is more than 30 percent in a 10-year period. This lack of temporal stability suggests that automobile ownership and trip generation may not be as firmly related as previous studies indicated.

Because the model set requires only six socioeconomic and land use data items, the effort required to develop forecasts should be minimized, thereby allowing for a more rigorous assessment of the input data.

The specification of minimal exogenous data means that this model set had to include a set of auxiliary models that would estimate values of variables that in other model sets are simply specified as required data inputs. A summary of these auxiliary models is given in Table 1. The data developed from these models include parking cost, highway terminal time, an area-type classification, the stratification of households by income quartile and family size, and network speeds. Perhaps the most important auxiliary model was the procedure to stratify zonal households by family size and income quartile. This model was calibrated by using data from the 1960 origin-destination study and the 1960 census; the model consisted of a set of stratified curves and a procedure to ensure that the regional household and population totals were balanced. The area-type model classified zones into five urban area types: central business district (CBD), CBD fringe, urban residential, suburban residential, and exurban. The technique used to assign area types to zones was developed with the aid of discriminant analysis (2) and a standard statistical computer software package (3). These area types were used in developing highway and transit link speeds. The auxiliary models also contained procedures to estimate both highway and transit link speeds. The highway network used the UTPS program UROAD speed-capacity tables, which allowed the user to specify highway speeds by area type and highway facility type. Transit speeds were developed similarly, in that local transit speeds were a function of area type and the highway facil-

Table 1. Summary of auxiliary models.

Model	Measures Estimated	Independent Variables	Estimated Measures Are Used in
Parking cost model	Daily and hourly parking cost	Employment density	Modal-choice model
Highway terminal time model	Production and attraction terminal times	Employment and population density	Modal-choice model
Area-type model	Stratification of zones into five types of areas	Employment and population density	Highway and transit speed models
Income and family size stratification model	Stratification for each zone of households by income quartile and family size	Households, population, and mean household income	Trip-generation model
Transit speed model	Peak and off-peak transit speeds for local transit routes	Area type and highway facility type	Preparation of transit networks and travel times
Highway speed model	Off-peak highway speeds	Area type and highway facility type	Preparation of highway networks and travel times

ity type. A special program was required to implement this model.

The trip-generation models were calibrated by using a combination of cross-classification analysis and regression analysis. The normal socioeconomic and land use data were used in the model, but accessibility and locational variables were also found to be significant. The accessibility measures were defined as the number of jobs or households within a given highway or transit travel time. The locational variable used was the number of jobs or households within 0.75 mile. This was interpreted as a measure of the potential of a traveler to use a nonmotorized mode, i.e., walk. Obviously, as the potential for using a nonmotorized mode is increased, the probability for using a motorized mode should decrease. It was found that for almost all of the trip-generation submodes, this locational variable had to be included in the model to obtain logical coefficients on the accessibility measures. It was also found that the accessibility and locational measures were essential in estimating attractions by income level.

A detailed description of the trip-generation model would be too long for this paper, but a short description of the final home-based work trip equations will illustrate the use of the locational and accessibility measures. The home-based work production equations are given in Table 2. There are five linear equations, one for each household size group; each contains a constant, three income quartile dummy variables, and three locational variables. The constant and dummy variables are analogous to a cross-classification model with family size and income quartiles being the independent variables. The locational variables are (a) the number of jobs (employees) within walking distance of the household, with the walking distance being defined as 0.75 mile; (b) the percentage of all jobs within 30 min of highway driving time; and (c) the percentage of all jobs within 25 min of transit travel time. The walk potential measure (i.e., employees within walking distance) will reduce the number of motorized trips as the number of employees increase, whereas the two accessibility measures will show an increase in the trip rate as the accessibility increases.

Point elasticities were calculated for each of the three locational variables for each strata of household size and income. Although these elasticities varied for each strata, in general the walk potential variable and the transit accessibility measure had the same elasticity (with, of course, opposite signs), whereas the highway accessibility elasticity was approximately 3 times as large as the other two elasticities.

To estimate home-based work attractions by income quartile, it was first necessary to estimate the employment by income quartile. The equations for estimating this employment are as follows (note that in application, estimated employees by income are normalized to total employment):

$$\mathrm{ESTIEMP}(1) = \mathrm{TOTEMP} \times 0.09562 + 0.025532[\mathrm{DURAT}(1)] + 0.046435[\mathrm{ACRAT3}(1)] \quad (1)$$

$$\mathrm{ESTIEMP}(2) = \mathrm{TOTEMP} \times 0.19560 + 0.021294[\mathrm{DURAT}(2)] + 0.056881[\mathrm{ACRAT1}(2)] \quad (2)$$

$$\mathrm{ESTIEMP}(3) = \mathrm{TOTEMP} \times 0.25138 + 0.073811[\mathrm{DURAT}(3)] - 0.028823(\mathrm{DURAT}) + 0.052197[\mathrm{ACRAT1}(3)] \quad (3)$$

$$\mathrm{ESTIEMP}(4) = \mathrm{TOTEMP} \times 0.21657 - 0.004334(\mathrm{DURAT}) + 0.042297[\mathrm{ACRAT4}(4)] \quad (4)$$

where

ESTIEMP(i) = estimate of income i employees;
TOTEMP = total zonal employment (mean = 881.15);
DURAT(i) = ratio of income i dwelling units within 0.75 mile to employment

Table 2. Home-based work production equations.

Family Size	Constant	Income Dummy Variables[a] 1	2	3	EMPWK2[b]	PHWYACC3[c]	PTRNACC1[d]
1	0.1215	-0.20750	0.01960	-0.07919	-0.000001949	0.0040951	0.0038508
2	1.2614	-0.73882	-0.23995	-0.03878	-0.000012201	0.0040951	0.0038508
3	1.8393	-1.07462	-0.45938	-0.24010	-0.000019252	0.0040951	0.0038508
4	1.7926	-0.94112	-0.25897	-0.16738	-0.000012401	0.0040951	0.0038508
≥5	1.9193	-0.87939	-0.50307	-0.24072	-0.000014707	0.0040951	0.0038508

[a] Income dummy variables are defined as follows: 1 = lowest income quartile, 2 = medium-low income quartile, and 3 = medium-high income quartile.
[b] EMPWK2 = employees within 0.75 mile (mean = 5962.2).
[c] PHWYACC3 = percentage of regional employment within 30-min peak highway time (mean = 92.77).
[d] PTRNACC1 = percentage of regional employment within 25-min peak transit time (mean = 22.38).

within 0.75 mile (weighted means: income 1 = 0.2172, income 2 = 0.2077, income 3 = 0.1932);

DURAT = ratio of dwelling units within 0.75 mile to employment within 0.75 mile (weighted mean = 0.8082);

ACRAT1(i) = ratio of percentage of income i dwelling units within 25-min peak-hour transit time to percentage of all dwelling units within 25-min peak-hour transit time (weighted means: income 2 = 1.0449, income 3 = 0.9113);

ACRAT3(i) = same as ACRAT1(i), except for 35-min peak-hour transit time (weighted mean for income 1 = 1.1253); and

ACRAT4(i) = same as ACRAT1(i), except for 40-min peak hour transit time (weighted mean for income 4 = 0.9273).

These equations use two types of locational variables: (a) the ratio of dwelling units within walking distance (0.75 mile) to the number of employees within walking distance, and (b) the ratio of one income strata of household to all households within a given transit travel time range. These independent variables are relative variables in that they describe the mix of land use rather than the absolute value of the land use. The walk potential variable--the ratio of dwelling units to employees within walking distance--describes the mix of residential units and employment within a given area. For the lower income categories, the employment for these categories increases as the number of households in these categories increases, whereas for the highest income quartile the employment will decrease for this category when the number of total households increases. In other words, the model is showing that there is a relationship between low-income employment and low-income households, but the high-income employment tends to be in areas with little or no residential units. The accessibility variable--the ratio of one income strata of households to all households for a given transit travel time range--is always positive; that is, as the number of households for a given income group increases, the number of employees for the same income group increases.

When the number of employees for each income quartile is known, estimating home-based work attractions by income quartile is fairly simple. The equations for this model are as follows:

$$\mathrm{ESTATR}(1) = \mathrm{EMP}(1) \times \left\{1.3279 - 2.6367 \times 10^{-6}[\mathrm{DUWLK}(1)]\right\} \tag{5}$$

$$\mathrm{ESTATR}(2) = \mathrm{EMP}(2) \times \left\{1.3463 - 1.4483 \times 10^{-5}[\mathrm{DUWLK}(2)]\right\} \tag{6}$$

$$\mathrm{ESTATR}(3) = \mathrm{EMP}(3) \times \left\{1.3419 - 5.8307 \times 10^{-6}[\mathrm{DUWLK}(3)]\right\} \tag{7}$$

$$\mathrm{ESTATR}(4) = \mathrm{EMP}(4) \times \left\{1.3573 - 1.7085 \times 10^{-5}[\mathrm{DUWLK}(4)]\right\} \tag{8}$$

where

ESTATR(i) = estimated work attractions by income i employees,

EMP(i) = number of income i employees (weighted means: income 1 = 132.59, income 2 = 225.14, income 3 = 254.20, income 4 = 269.22), and

DUWLK(i) = number of income i dwelling units within 0.75 mile (weighted means: income 1 = 2604.3, income 2 = 1140.7, income 3 = 1075.6, income 4 = 1122.5).

This model is a set of linear equations that contains a constant and a locational variable--percentage of dwelling units within walking distance. The constant can be considered the average number of attractions per employee if no households are within walking distance. The locational variable has the correct sign, in that, as the number of households increases, the number of motorized work attractions decreases, but it does not contribute significantly to the trip rate; at the mean, the change in the trip rate is less than 2 percent.

The distribution model was specified as a normal gravity model. Attempts were made to use the modal-choice model equations to calculate a composite impedance by combining travel times and costs for all modes. This attempt was extremely successful for the home-based work trips, but it was not completely successful for other trip purposes. Highway travel time was thus used as the impedance measure for these other purposes. All the distribution models were stratified by income quartiles, and it was found that the low-income travelers were less sensitive to the impedance measure than were the high-income travelers. The modal-choice model was a multinomial logit model that used three modes: transit, drive alone, and group automobile. A submodel was used to split the group mode into integer automobile occupancies (two persons per automobile, three persons per automobile, and so forth). The initial modal-choice model was calibrated on a disaggregate level by using the UTPS program ULOGIT and then validated at the aggregate level. The use of integer automobile occupancies allowed the application methodology to be configured in a manner that would allow HOV incentives to be explicitly considered.

Because of the model specification, the normal forecasting procedure sequence (i.e., generation, distribution, and mode choice) was not applicable. For the New Orleans model set, the modal-choice model must be applied before distribution in order to generate the composite impedances; the general flow of the model application is shown in Figure 1. The modal probabilities, generated by the modal-choice model, can be saved and used to split the person trip distribution or, if computer time is less costly than storage, the modal-choice model can be applied again after the distribution model. Although the entire model set is fairly intricate, it does not use excessive computer resources. The central processing unit (CPU) time for the entire chain (468 traffic analysis zones) is approximately 1.5 hr on an IBM system 370 model 158.

In summary, the New Orleans travel-demand models were developed within the framework of the goals and objectives specified for the model set. The developed models are unique in that all models are stratified by income quartiles, the generation model includes accessibility and locational measures, and the home-based work distribution model uses a composite impedance measure. The goal of using transportation system characteristics in all major submodels was essentially met, although the inability to use the composite impedance measures for the nonwork trip-distribution models was somewhat disappointing. The development of six auxiliary models minimized the number of exogenous data items required for the model set, thereby reducing the effort required to apply the models and maximizing the objectivity of the forecasts.

MODEL APPLICATION RESULTS

A practical advantage of calibrating a travel-demand model set by using an old origin-destination survey was that the first forecast could use data for the present year and this forecast could be validated by using ground counts and other data sources. The New Orleans model set, which was calibrated by using the 1960 origin-destination survey, was applied for the year 1980. The resulting estimates compared quite favorably with actual ground counts and preliminary census data.

The comparison of the 1980 estimated data with observed data is given in Table 3. The number of households and the population for 1980 had been estimated before the publication of the preliminary 1980 census data, and these estimates appear to be slightly low (approximately 5 percent for households and 1 percent for population).

The Louisiana Department of Transportation and Development developed a 1978 estimate of daily vehicle miles of travel (VMT) primarily from ground counts, and this estimate is approximately 5 percent higher than the model estimates. The model overestimated transit trips by approximately 3 percent. These rather gross comparisons indicate that the model set was able to forecast trips for a 20-year time period with a reasonable degree of accuracy,

Figure 1. General model application flow diagram.

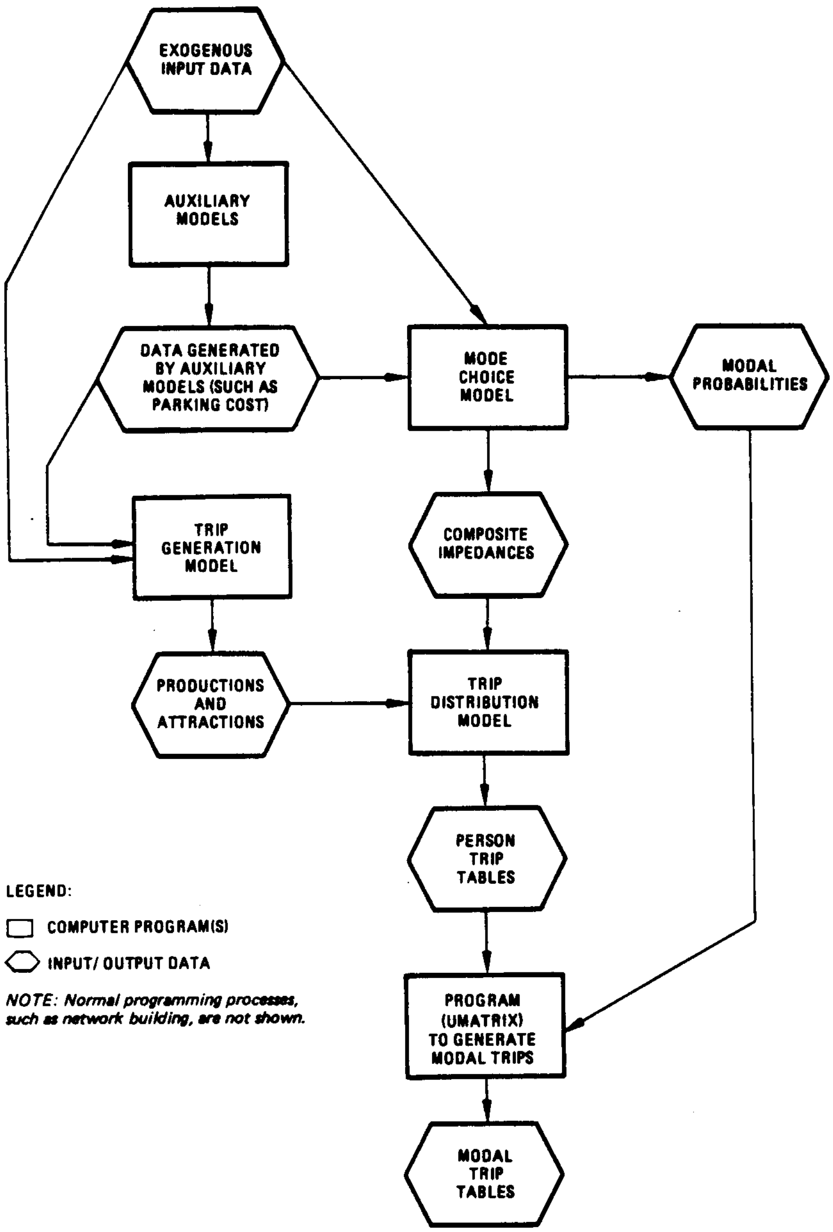


Table 3. Comparison of 1980 estimated data with data from other sources.

Item	Estimate	Data from Other Sources	Percentage Difference	Sources
Households	365,182	385,351	-5.2	Preliminary 1980 census
Population	1,064,876	1,076,171	-1.1	Preliminary 1980 census
Daily vehicle miles of travel (VMT)	7,922,045	8,325,000	-5.1	1978 estimate by Louisiana Department of Transportation and Development
Transit trips (not including school trips)	197,577	191,542	+3.1	Office of Transit Administration, city of New Orleans

although there is some indication that VMT may be slightly underestimated.

In the past 20 years there has been a significant change in travel patterns in most urban areas. The data in Table 4 present some travel measures indicative of these changes. The estimated increase in travel per person is approximately 16 percent, whereas the average trip length has increased by more than 20 percent. More significantly, the average VMT per person is estimated to have increased by approximately 100 percent during the past 20 years.

Table 4. Comparison of 1960 data with 1980 estimates.

Item	1960	1980	Percentage Change
Person trips per person	1.41	1.63	+15.6
Person trips per household	4.62	4.75	+2.6
Avg trip length (minutes of highway time)	8.44	10.36	+22.7
Daily VMT per person	3.71	7.44	+100.5
Percentage transit (total)	25.43	12.53	-50.7
Percentage transit (CBD)	54.14	36.71	-32.2
Vehicle occupancy (total)	1.477	1.480	+0.2
Vehicle occupancy (CBD)	1.487	1.365	-8.0

This growth, which represents approximately a 3.5 percent per year increase, was so substantial that growth rates from other urban areas were obtained to ascertain the reasonableness of this increase. The annual growth rate of VMT per person for the Virginia suburbs of Washington, D.C. (the counties of Arlington, Fairfax, and Prince William) was determined to be approximately 2.3 percent per year between 1968 and 1978 (4,5). This increase is not quite as large as the forecasted New Orleans increase, but it is in the same range. Transit ridership as a proportion of the total travel market decreased significantly between 1960 and 1980. The percentage of transit for the region decreased by 50 percent, whereas the percentage of transit to the CBD decreased by more than 30 percent. The model estimated only minor changes in vehicle occupancy, which was unexpected. Higher gasoline and parking costs probably account for the stable vehicle occupancies, in spite of rising incomes.

Vehicle assignments were compared with ground counts for five screen lines. In all cases the assignment volumes were lower than the ground counts. This occurred, in part, because highway assignments cannot always replicate double screen-line crossings and short (intrazonal) trips; the 1960 survey data revealed a 10 percent difference in assignment versus ground counts for one of these screen lines. The available ground counts were also simple tube counts, with no correction factor for multiaxle vehicles. The study team identified a range of errors that could be associated with the ground counts and the computer assignments, and two sets of error corrections were prepared. The ratio of assignments to ground counts for five screen lines, with the two error ranges, is given in Table 5. Perhaps the significant element of the screen-line comparisons is that the ratio of assigned volumes to ground counts are similar, which indicates that the model set estimated the distribution of travel correctly.

In summary, the New Orleans model set, calibrated on 1960 data, was used to estimate 1980 travel. This is equivalent to a 20-year forecast. The resulting travel patterns were similar to observed data, thereby providing regional planners with greater assurance that this model set could be used to forecast future travel.

Table 5. Screen-line comparisons.

Screen-Line Description	Assigned Volume to Ground Count Ratio: Forecast/ Ground Count	With Least Error Correction	With Highest Error Correction
Mississippi River crossings	0.870	1.086	1.206
Navigational Canal	0.746	0.930	1.034
Jefferson Parish—Orleans Parish Boundary on East Bank	0.717	0.894	0.993
Harvey Canal	0.603	0.753	0.836
Donner Canal	0.714	0.890	0.989

CONCLUSIONS

A complete set of travel-demand models was calibrated for the New Orleans region by using 1960 travel data. These models were successfully applied to 1980 conditions within a reasonable degree of accuracy, although the observed data were only available at an aggregate level. Although the physical changes in the transportation system between 1960 and 1980 were not radical (consisting primarily of a few freeway additions), the changes in aggregate travel patterns were substantial. The average VMT per person increased by approximately 100 percent, whereas the transit market share decreased by 50 percent. There was also a substantial change in economic conditions between 1960 and 1980. The consumer price index increased by more than 170 percent, whereas per capita income increased by more than 60 percent, in constant dollars. Most assuredly, changes of this magnitude would be considered significant changes for any forecast. The successful application of the model to 1980 conditions, coupled with the substantial changes in the travel patterns and economic conditions between 1960 and 1980, would imply that an appropriately specified travel-demand model set may indeed be temporarily stable (within reason), and that the use of old survey data is not appropriate in investigating travel behavior and in calibrating travel-demand models.

The calibrated travel-demand model set is fairly unique in that all submodels were stratified by income quartiles. Other noteworthy aspects of the model were the use of a composite impedance measure in the distribution model, the use of accessibility and locational factors in the trip-generation model, and the use of minimal exogenous input data.

ACKNOWLEDGMENT

The data in this paper are based on the results of a study performed by the RPC of Jefferson, Orleans, St. Bernard, and St. Tammany parishes in Louisiana. This research was funded in part by UMTA, U.S. Department of Transportation. I would like to thank Dennis D. Striz of the RPC and Foster de la Houssaye of the Mayor's Office of Transit Administration, city of New Orleans, for their assistance, comments, and support in the completion of the work reported herein.

REFERENCES

1. G.S. Cohen and M.A. Kocis. Components of Change in Urban Travel. New York State Department of Transportation, Albany, July 1979.
2. M.M. Tatsouka. Multivariate Analysis. Wiley, New York, 1971.
3. N.H. Nie et al. SPSS: Statistical Package for the

Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.

4. Average Daily Traffic Volumes on Interstate, Arterial, and Primary Routes. Traffic and Safety Division, Virginia Department of Highways and Transportation, Richmond, 1968.
5. Average Daily Traffic Volumes on Interstate, Arterial, and Primary Routes. Traffic and Safety Division, Virginia Department of Highways and Transportation, Richmond, 1978.

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

Estimation and Use of Dynamic Transaction Models of Automobile Ownership

IRIT HOCHERMAN, JOSEPH N. PRASHKER, AND MOSHE BEN-AKIVA

Models of automobile ownership level and type choice are described by using a dynamic transactions model structure. The functional form of the model is two-stage nested logit: the higher level in the hierarchical decision process is a decision on the type of transaction in the car market. The lower-level decision is on type of car, which is conditional on the decision to buy a car. Automobile type alternatives are defined by make, model, vintage, and body type. The model was estimated with data from the Haifa urban area in Israel. The sample consisted of a choice-based (stratified) sample of 500 households that did not buy a car in 1978 and 800 households that bought a car during the same year. Each stratum was drawn at random from the respective population of the Haifa urbanized area. The models estimated in this paper are sensitive to attributes of the type of car, household characteristics, and accessibility by public transit and private car. The models take explicit account of the transaction costs that are incurred when operating in the car market.

The purpose of this paper is to develop and test a dynamic demand model for automobiles. Understanding the demand for cars has always been an important consideration in transportation studies. In recent years the composition of the car market has become a key factor in the evaluation of energy-consumption policies. The relative roles of purchase price and usage costs in determining car choice are of interest to policy decision makers. This is especially true in a country such as Israel, where cars and fuel are taxed at high levels. Thus changes in the structure of these taxes can be used to achieve policy goals, such as increasing the share of small cars. In Israel, car purchase and use also affect the balance of payments, because almost all the cars sold and all the oil consumed are imported.

The market for private cars in Israel is characterized by two major aspects. First, the level of ownership is relatively low compared with North America and Western Europe, where a third of the households (40 percent in the major urban areas) own cars, and of these only about 6 percent (2 to 3 percent of the total population) own more than one car. Growth of the private car fleet still occurs mainly by purchase of a first car.

The second important characteristic of the Israeli car market lies in the composition of the car stock. Most of the cars in Israel are small European cars, with only a small percentage of U.S. made cars, one popular Japanese brand (Subaru), and two domestic models that are assembled in Israel. The Israeli car fleet is heterogeneous and includes scores of different makes. The typical car in Israel is older than in the United States. About 60 percent of all cars are more than 5 years old, with 20 percent more than 10 years old.

These characteristics imply that the usual categorization of cars into subcompact, compact, and so forth, used in some models of car type choice (1,2) is not valid for the Israeli market, as almost all cars fall in the subcompact category. Also, the relevant ownership levels are zero and one. Ownership of two or more cars may become of interest in the future, but any attempt to model this phenomenon now will require special data-collection efforts.

In summary, a practical model of the Israeli car market may confine itself to zero- and 1-car households; should deal with holding or purchase of all cars, new or old; and should be able to describe the determinants of growth in the market.

MODELING APPROACH

The model developed in this study is a disaggregate, dynamic transactions model for level of ownership and type of car owned. As its name implies, the decision process involved in buying or replacing a car at the household level is the model studied. The model is dynamic in the sense that level of ownership and type of car owned during the previous time period are assumed to influence decisions about transactions made during the current (modeled) time period.

The key aspects of the model developed here are as follows.

1. The model is dynamic. It uses data on previous car holdings and includes a detailed treatment of transaction costs.
2. It is a transaction model that concentrates on changes in automobile holdings.
3. It is a nested logit model of the decision to transact and then the choice of car type given a transaction.
4. It describes the Israeli market, which may be more representative of conditions in some European or developing countries than in North America in terms of type, composition, and levels of ownership.

THEORETICAL FRAMEWORK

Previous Disaggregate Automobile Ownership Models

The development of the discrete choice econometric techniques facilitated a disaggregate approach to the modeling of car ownership. The first studies dealt with level of ownership, usually as a joint decision with mode to work (3-7).

Lave and Train (1) studied the choice of new vehicles by size class. Manski and Sherman (8) developed a car type choice model where the car alternatives were defined by make, model, and vintage. Hensher and Manefield (9) suggested a nested logit model of automobile acquisition and type choice and presented some preliminary results. The car types were grouped into three classes according to fuel consumption. A different approach to model automobile market shares was applied by Cardell and Dunbar (10) and by Boyd and Mellman (11). They estimated a logit choice model with random coefficients by using aggregate market share data (12).

Almost all the studies mentioned used static holding models. Manski and Sherman (8) used an aggregate estimate of the proportion of cars purchased during the previous year as an external estimate of a constant transaction cost.

Rationale for a Dynamic Model

The importance of a dynamic model structure stems from the following observations.

1. Transaction costs: With time, the car owned by a household gets older and some of its attributes change, and the car may no longer match the requirements of the household. Also, the characteristics of the household may alter, thereby causing further changes in the relative attractiveness of the various car alternatives. Nevertheless, cars are usually kept for a number of years (in Israel the average holding period is 3.5 years). The reason for this phenomenon is that the process of selling and buying a car involves significant transaction costs. A static model assumes a perfectly competitive market with no transaction costs. A dynamic model, on the other hand, allows for the inclusion of variables that measure these costs.
2. Brand loyalty: Brand loyalty is a well-known marketing phenomenon, which is apparent in the car market. It is revealed in the tendency of households to buy a new car of the same make, or even the same model, as that of a previous car. Brand-loyalty reflects lower information acquisition costs and idiosyncratic tastes. Allowing for a brand-loyalty effect in the car type choice model imposes a dynamic structure.
3. Income effect: The money received from selling an old car may be used toward the purchase of another car, so that, all else being equal, a household with a car during time period t - 1 may be able at time t to spend more money on its new car than a household without a car in the previous period. In general, knowledge of the choice made in the last time period provides useful information for the prediction of the choice during the current period. This information can easily be obtained in a household survey. Omitting such information may not only weaken the explanatory power of the model, but also may induce biases in the estimates of the parameters.

On the other hand, a dynamic model introduces econometric difficulties. If the error terms of a model are serially correlated, then the error term will be correlated with lagged explanatory variables. This means that the use of a dynamic model structure requires the assumption of serial independence.

It was assumed that there was no serial correlation, thus allowing for the use of a dynamic model. This decision is justified if the biases caused by violation of serial independence are small compared with the advantages of a dynamic characterization.

Behavioral Framework

The behavioral framework assumed in this study is as follows. Every time period (a year was chosen to avoid the effects of seasonal variations and also because new car models come out yearly), each household evaluates its current car holdings and decides whether to transact in the car market or not. A transaction may mean buying, buying and selling, or just selling. If the decision involves buying a car, then the type of car in terms of make, model, and vintage is also decided on.

The household is assumed to act as a utility maximizer, that is, the household assigns a utility value to each of the alternatives based on the attributes of the alternative and the costs involved, including the transaction costs. The alternatives that the households face are either do-nothing or transact in the car market by buying a specific type of car or selling the existing car or both. The household will decide to transact when the utility of one of the possible alternatives is greater than the utility of the current state.

The decision to transact and the choice of car type are assumed to be based on last year's holdings and current socioeconomic status. This results in a first-order Markov process. This is not an essential assumption to the model, but it is imposed by data-collection limitations and the relative ease in using the model for predictions.

The dependence on only last year's holding may be justified by realizing that, because of the relatively long car-holding period, the most recent car holding has the strongest impact on the current decision. Furthermore, some of the influence of past history is captured by the last holding.

Future expectations probably influence the decision process; for example, a household usually purchases a car with an a priori intention of keeping it for a fixed number of years. Also, expectations about future earnings and use of the car may enter the process. Unfortunately, it is difficult to collect reliable data on future expectations.

In the models presented the possible transactions are either buying (for households that did not own a car last year) or replacing (for households that owned a car). The possibility of buying a second car is omitted because, as previously mentioned, in Israel more than 95 percent of the households have zero or one car.

Another option that is not considered here is selling only. This restriction stems from the nature of the data. In reality, this type of transaction is rare in Israel. However, it can still be allowed in the aggregate, for example, as a function of age of head of household, when the model is used for prediction.

Model Structure

Formally, the model can be stated as follows. Let $j|i$ indicate the transition from owning a car of type i in time period t - 1 to owning a car of type j in time period t, where i = 0 is the state of not owning a car at time t - 1. Assume that

$$U(j|i) = V(j|i) + \epsilon_j \tag{1}$$

where

$U(j|i)$ = random utility of the $(j|i)$ transition,
$V(j|i)$ = average utility of the $(j|i)$ transition, and
ϵ_j = a disturbance that represents unobserved utility of alternative j.

The error terms are assumed to have the following joint probability distribution:

$$F(\epsilon_i, \ldots, \epsilon_j, \ldots, \epsilon_k, \ldots) = \exp - \left\{ \exp(-\mu\epsilon_i) + \left[\sum_{k \neq i} \exp(-\epsilon_k) \right]^{\mu} \right\}, 0 < \mu < 1 \quad (2)$$

This is a special case of the generalized extreme value (GEV) family of distribution functions developed by McFadden (12). The marginal distributions of ϵ_i and ϵ_k for all k are gumbel $(0,\mu)$.

The first term inside the brackets $[-\exp(-\mu\epsilon_i)]$ refers to the alternative of no transaction; the second term $\{[\Sigma \exp(-\epsilon_k)]^{\mu}\}$ refers to all the alternatives that involve a transaction; and μ is a measure of similarity of the unobserved attributes among the transaction alternatives.

It has been shown by McFadden (12) and Ben-Akiva and Francois (13) that the assumption of the GEV distribution on the error terms results in the following choice probabilities:

$$P(j|i) = P(j|tr, i) \cdot P(tr|i) \quad (3)$$

$$P(tr|i) = \begin{cases} 1/\{1 + \exp[-\mu_0(V_{tr|0} + I_i)]\}, & \text{for } i = 0 \\ 1/\{1 + \exp[-\mu_1(V_{tr|i} + I_i)]\}, & \text{for } i \neq 0 \end{cases} \quad 0 < \mu_0 < 1; 0 < \mu_1 < 1 \quad (4)$$

$$P(j|tr, i) = \exp(V_{j|i}) / \sum_{k \neq i} \exp(V_{k|i}), \; j \neq i \quad (5)$$

where the expected maximum utility from available car types is $I_i = \ln \sum_{k \neq i} \exp(V_{k|i})$; and $V_{tr|i}$ is an average utility component of the transition from state i that is equal for all alternative transactions (where tr denotes transaction). The subscript 0 denotes no car owned at time t - 1. The subscript i for $i \neq 0$ indicates that one car of type i was owned at time t - 1.

The choice probabilities describe a dynamic nested logit model with two decision levels. The first level decision is whether to transact or not, and the second (lower) level decision is on the type of car to be purchased, which is conditional on a decision to transact.

The choice probabilities of the first stage have a logit form with scale parameter μ and with I_i (a composite variable from the lower-level model). The choice probabilities of the second level have the logit form with scale parameter normalized to equal one. The particular form of the GEV distribution that was assumed was chosen because it imposes a nested structure in the choice probabilities, which is behaviorally reasonable and computationally feasible.

As mentioned before, two distinct types of transactions are allowed in the model, depending on the level of ownership at time t - 1: buying a first car or replacing an existing car. To enable different specifications for the utilities associated with the two types of transactions, two different transactions models were assumed. The two models can be viewed as one model with all variables specific to either of the two transaction types. Because the two transactions are mutually exclusive (i.e., each household has only one type of transaction in its choice set), the estimation can be done separately.

The utilities of the automobile type choice model are assumed to have the same functional form for all households, regardless of their level of car ownership at time t - 1. However, the specification may include variables that are specific either to first-time buyers or to previous owners. This assumption simplifies the estimation process significantly. It may be justified by realizing that once the decision to transact was made, the household faces the same set of alternatives, whether or not a previous car was owned. The structure of the suggested model is shown in Figure 1.

SPECIFICATION OF THE MODEL

The utility that a household derives from buying or replacing a car is a function of the attributes of the purchased car, the transaction costs, household characteristics, and previous car characteristics. In the following sections the variables in each of these groups that were used to specify the model are described.

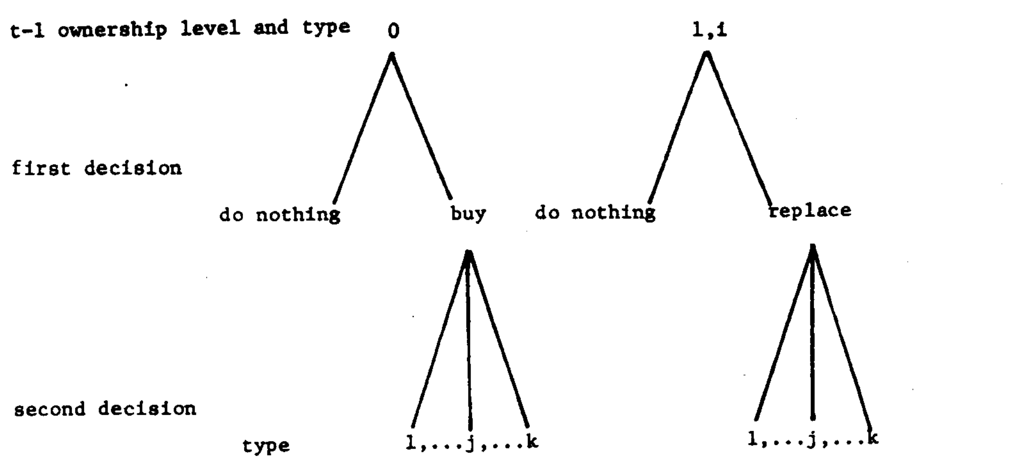
Household Characteristics

Household characteristics affect car purchase decisions in three ways:

1. The income and wealth of the household affect the amount of money it is willing or able to invest in buying a car;
2. Some household characteristics, such as residential and work place locations and household size, influence the need for a car and the type of car suitable for the needs of the household; and
3. There exist household car choice preferences that can be modeled with variables such as age and education.

In this study income was measured in four cate-

Figure 1. Structure of the model.



gories. Also used were proxy variables for income and wealth, such as education, age, and work status.

Household characteristics that affect the need for a vehicle can be divided into two groups: variables that affect the relative utility of owning a car compared with not owning one, and variables that influence the relative utilities of different types of vehicles. The first group consists mainly of accessibility measures for work and other trips. The characteristics that affect the type of car chosen are household size and composition and the need to use a car for work-related purposes. Individual preferences were characterized mainly by interacting age and education of the car user with car attributes such as performance and age.

Attributes of Previous Car

The attributes of the cars owned during the previous time period affect the decision to replace a car and the choice of car type. The replacement model includes the following attributes of the previous car: age, engine size, number of years owned by the household, and average mileage. Engine size serves as a measure of durability, and mileage is a measure of use.

Attributes of the previous car were also assumed to affect the car type choice. Purchase price for each type was defined as its market price minus the sale price of the existing car. Brand loyalty was captured by a dummy variable, which is set to 1 if the type is of the same make as the previous car.

Attributes of Alternative Cars

The car attributes are of special interest because they characterize the alternatives that a household faces when choosing a car, thus enabling predictions of the effects of policy and technological changes on the market. The car attributes were selected according to two criteria: (a) the attribute has to be familiar to prospective car buyers, and (b) a reliable data source for the attribute exists for all the alternative car types. The following attributes were selected: cost--retail price and fuel efficiency; size--dimensions, weight, luggage space, and engine size; performance--acceleration (measured by HP/kg) and maximum speed; and other--age, manufacturing country, and number of cars of the same type in the market.

The alternative-size variable (i.e., the number of cars of the same type in the market) is also of special interest. In the type choice model, each type represents a group of elemental alternatives--all the cars of the same make, model, year, and body type on the market. All of these cars have identical observed attributes. In this case it is necessary to add to the utility of an alternative a normalization term equal to the logarithm of the number of elemental alternatives (ℓnN_j). The coefficient of ℓnN_j is a measure of the variation of the unobserved attributes among cars of the same type. It is reasonable to assume that as cars get older, the heterogeneity among them increases. To enable different coefficients for car types of different ages, the alternative-size variable was introduced in the model as a group of age-specific variables. Four type size variables were used for cars: 2 to 5, 6 to 9, 10 to 14, and 15 or more years old. The type-size variable is not used for new car types. The choice probability for new cars does not depend on ℓnN_j because the buyer faces only one new car that he orders from the dealer. As described in the next section, the type-size variable is also a proxy for search cost so that the coefficients of these variables capture the total affect of alternative size on the utility.

Transaction Costs

Transaction costs arise from two main sources: search costs in terms of time and money that are incurred in the process of searching for a car, and information costs that are caused by incomplete knowledge of the attributes of alternative cars. This model includes a detailed specification of the transaction costs, expressed as functions of the attributes of the previous car, the purchased car, and the nature of the transaction.

In the process of testing alternative model specifications, some simplifying and generalizing assumptions had to be made with respect to the specification of transaction costs (TC). The specification of TC in the car type utility function is as follows:

$$TC(j) = \alpha_1 \delta_{j|i} + \alpha_2 \ell n m_j + \sum_{\ell} \alpha_{3\ell} \cdot \delta_{\nu_j \ell}\, \ell n n_j + \sum_{\ell} \alpha_{4\ell}\, \delta_{\nu_j \ell} \qquad (6)$$

where

$$\delta_{j|i} = \begin{cases} 1, & \text{if make (j) = make (i)} \\ 0, & \text{otherwise} \end{cases}$$

m_j = number of cars of same make as j,

n_j = number of cars of same type as j,

ν_j = age of car j,

$$\delta_{\nu_j \ell} = \begin{cases} 1, & \text{if vintage of car j belongs to vintage group } \ell \\ 0, & \text{otherwise} \end{cases}$$

and α_1, α_2, $\alpha_{3\ell}$ and $\alpha_{4\ell}$ are unknown parameters.

The first term represents brand loyalty. The second term captures the market-size effect of all cars of the same make. The third term captures the effect of the number of cars of the same type on search costs and the basic correction for alternative size of the utility. These effects are allowed to vary with the age of the vehicle. The fourth term represents the effect of the age of the vehicle on the information acquisition costs through a number of age-specific dummy variables. This effect is measured relative to new cars.

The average TC for each of the two types of transactions appears in the respective transaction models as part of the alternative specific constants.

SAMPLING STRATEGY AND ESTIMATION PROCEDURE

The estimation of the models was carried out by using the data from a sample of households that were collected in the Haifa urban area during 1979. The nested structure of the model dictated the sampling strategy. The upper-level (transaction) model requires a sample of buyers and nonbuyers, with and without a previous car. The lower-level (car type) model requires for its estimation the households that bought a car during the study year. A random sample of households would not provide enough such purchases unless it is large, because only 40 percent of the households own cars and less than one-third of these households are expected to purchase a car in a given year.

The sampling strategy used was choice based or endogenously stratified, where the stratification was based on the transaction decision. The total sample consisted of a random sample with one sampling quotient of households that did not transact in the car market during the study year, and another sample with a greater quotient of households that purchased a car during that year. About 500 households that did not buy a car during the study year were surveyed, as well as 700 households that purchased a car in the same year.

The estimation of the model was carried out in two steps. First, the lower-level or conditional model of car type was estimated by using the data of the purchase sample alone. Then the two samples were combined, and the expected maximum utility from all car types was calculated for each household by using the results of the type choice model estimation. The combined choice-based sample was then used to estimate the coefficients of the upper-level models of transaction choice. The alternative specific constants were then corrected to produce consistant estimates.

ESTIMATION RESULTS

Type Choice Model

The type choice model describes the choice of a car type, which is conditional on the decision to buy a car. The alternative cars are defined by make, model, body type, and vintage; the choice set during 1978 consisted of 950 alternatives. To reduce the cost of model estimation, a sample that included the chosen alternative and 19 randomly selected alternatives was selected for each household. This sampling procedure results in consistent estimates of the parameters (12).

The estimation sample had 786 households that purchased a car in 1978 in the Haifa area. The estimation results are given in Table 1.

Table 1. Estimation results—a logit model of car type choice conditional on purchase.

Variables	Coefficient Estimate	Asymptotic t-Statistic
Cost[a]		
Cost when income <10,000 I.L. per month	−0.148	−5.36
Cost when income 10,000-20,000 I.L. per month	−0.137	−8.40
Cost when income 20,000-30,000 I.L. per month	−0.093	−5.22
Cost when income >30,000 I.L. per month	−0.072	−2.70
Cost when income unknown	−0.086	−3.31
Cost when head of household >45 years old	0.0029	0.134
Cost when tax exempt	0.214	6.62
Dummy for high income and expensive car[b]	0.723	1.97
Dummy for age 45 or older and expensive car[b]	1.19	4.38
Fuel efficiency		
Liter per 1000 km for owners who do not get full maintenance and operating cost coverage	−0.0224	−4.69
Liter per 1000 km for owners who get full maintenance and operating cost coverage	−0.0092	−1.15
Size		
Size of car[c] for 5 or more member households	0.0111	1.38
Engine size for receivers of full maintenance costs or self-employed	0.0564	2.69
Dummy for small car[d] and 1- to 2-member households	0.470	2.28
Luggage space when car not used for work	−0.0059	−1.89
Luggage space when car used for work	0.0034	1.05
Performance		
HP/weight when user <30 years old	0.872	1.72
HP/weight when user 30-45 years old	1.89	4.41
Transaction costs and alternative size		
Brand loyalty dummy	1.48	10.6
ℓn number of cars same make x 100	0.248	4.48
ℓn alternative size for cars aged 2-9 years	0.868	15.8
ℓn alternative size for cars aged 10-14 years	0.493	6.79
ℓn alternative size for cars aged 15 or more years	0.904	6.60
Dummy for cars aged 15 or more years	−3.67	−5.80
Dummy for cars aged 10-14 years	−4.76	−10.1
Dummy for cars aged 2-9 years	−6.64	−17.8
Other		
Age of car when main user <30 years old	0.056	2.47
Age of car for first car	0.107	5.40
Dummy for cars made in Israel	0.583	4.15

Note: Number of households = 786, number of observations = 14,834, $\mathcal{L}_{\hat{\beta}} = -1{,}543.79$, and $-2(\mathcal{L}_0 - \mathcal{L}_{\hat{\beta}}) = 1{,}609.7$.

[a] Cost is defined as purchase price or resale price of previous car.
[b] Expensive car = car with purchase value higher than the median (120,000 I.L.).
[c] Car size = length x width (in cm)/1000.
[d] Small car = engine size up to 1300 cc.

The coefficients of the cost variables indicate that, all else being equal, a low price is a desired attribute. The effect of price on the utility decreases as the household's income increases. The cost coefficient for households whose head is 45 years of age or older is not negative, but the overall effect of price and other variables is still negative at all income levels. The only exception is a small group of households of disabled drivers that are exempt from taxes. These households are allowed to sell the car at market price after a few years of ownership, so that a higher cost means for them higher gains. The last two dummy variables in the cost group measure the preference of older people and people with higher incomes for more expensive cars. Age here is probably a proxy for accumulation of wealth.

The effect of fuel efficiency of a car on its choice probability is measured separately for owners who pay for the fuel costs themselves and for owners who are reimbursed by their business or employer. Both groups preferred fuel-efficient vehicles but, as expected, the first group placed higher weight on this attribute.

The data in Table 2 give the marginal rates of substitution between purchase price and fuel economy. To understand these figures, consider an average travel rate of 2000 km per month. Savings of 1 I.L. (Israeli lira) per 1000 km amounts to 24 I.L. a year. For a used car with an average depreciation rate of 10 percent a year, these savings are equal to an increase of 240 I.L. to the purchase price. For a new car that is held for 2 years and depreciates 35 percent during this period, the savings are equal to an increase of 420 I.L. These rough calculations indicate that the marginal rates of substitutions obtained from the model for households that pay for car operating costs are reasonable.

Table 2. Marginal rates of substitution of purchase price versus fuel cost.

	Price Premium (I.L.) to Save 1 I.L. per 1000 km in Fuel Costs			
	Full Car Cost, Not Covered		Full Car Cost, Covered	
Income (I.L. 000s)	<45 Years Old	>45 Years Old	<45 Years Old	>45 Years Old
<10	199	202	83	85
10-20	215	219	90	92
20-30	321	329	132	136
>30	341	426	170	178

Two performance variables were tested during the model estimation process: maximum speed and acceleration (measured by HP/weight). Maximum speed was found to have no effect on choice probabilities. Acceleration was found to be a positive attribute for users younger than 45 years of age, but it had no effect on the decisions of older users. This attribute has a stronger weight for users in the age group 30 to 45 than on younger users. A possible explanation is that the latter group, which has more limited resources, views acceleration as a luxury and puts more emphasis on other, more essential, equalities. It is interesting to note that in the Manski and Sherman model (8), acceleration had negative coefficients for all age groups.

All measures of the transaction costs have highly significant coefficients. The coefficients of the type-size variables are positive and smaller than 1, as expected. The highly significant and negative coefficients of the three vintage dummy variables represent the effect of lower transaction costs involved in purchasing a new car, as well as the effects of other desired qualities of a new car, such as reliability and prestige. The different magnitudes of the three coefficients represent the effects of unmeasured attributes that are related to either age or vintage of the cars.

The other two age variables measure the preference for old cars displayed by young buyers and by first-time buyers. In part this represents more limited resources, but it may also indicate a lesser concern for reliability on the part of inexperienced buyers and young users.

The last dummy variable for locally assembled cars represents a preference for these cars that cannot be explained by their lower prices.

To examine the goodness of fit between the estimated and observed aggregate choice probabilities for the sample, the 950 car types were grouped according to engine size and vintage. Generally, a satisfactory fit between the estimated and observed probabilities was revealed, and none of the differences was greater than 3 percent.

Purchase Model for Households Without a Vehicle

The purchase model for households without a vehicle describes the purchase decision in year t for households that did not own a car in year t - 1. The model was estimated with a sample of 618 households, about half of which came from a random sample of households in the Haifa area and the other half from a random sample of car buyers from the same area. The estimation results are given in Table 3. All coefficients in the model have the expected signs.

Table 3. Estimation results—a logit model of car purchase decision for households without vehicles.

No.	Variable	Coefficient Estimate	Asymptotic t-Statistic
1	One-adult households	−1.04	−2.96
2	Income > 10,000 I.L.	0.498	1.92
3	Income not reported, head of household employed	−0.06	−0.15
4	Head of household self-employed	0.699	1.52
5	Occupation of head of household—academic or managerial	0.624	2.31
6	Education of head of household—more than 12 years	0.416	1.88
7	Age of head of household when older than 50	−0.0093	−2.56
8	Use of car on Saturday (1 = yes and 2 = no)	−0.409	−1.63
9	Travel time to work by bus	0.0185	3.56
10	Travel time to work by car	−0.0131	−1.70
11	Walking distance to bus stop (in minutes)	0.142	4.41
12	Dummy for purchase alternative (corrected)	−6.5	−3.55
13	Expected maximum utility from the car type choice (ℓn sum)	0.512	3.22

Note: Number of observations = 618; $\mathcal{L}_{\hat{\beta}}$ = −339.56; and $-2(\mathcal{L}_0 - \mathcal{L}_{\hat{\beta}})$ = 177.60, d.f. 13.

Attributes 1-6 represent the financial ability to purchase a car; they include income and some sociodemographic descriptors that are correlated with potential earnings. Attributes 8-12 represent the need for a car. They include the following accessibility measures: travel time to work by car and by bus, walking distance to the nearest bus stop, and a proxy to the need for a car for leisure trips, that is, traveling on the Sabbath, when public transportation service is significantly reduced.

The variable age of head of household when older than 50 represents the tendancy to change that is associated with age as well as the increasing difficulty of acquiring a driver's license. The last variable is the expected maximum utility that the household derives from the car choice, given that a decision is made to own a car.

The data in Table 4 demonstrate the influence of socioeconomic and accessibility attributes on purchase probabilities. The effect of the relative accessibility by bus and by car is especially significant; for example, when travel time to work is 10 min by car and 30 min by bus, the purchase probability for a blue-collar employee (column 3 in Table 4) is 0.08. When travel time to work by car increases to 20 min and by bus to 60 min, the probability that the same household will purchase a car more than triples to 0.26 (column 6 in Table 4).

Table 4. Expected purchase probabilities computed for households with various attributes that did not own a car in the previous period.

	Attribute Values by Household Number						
Attribute	1	2	3	4	5	6	7
No. of adults in household	2	2	2	1	2	2	1
Household income	–	–	1	1	–	1	1
Work status of head of household	SE	SE	EMP	EMP	SE	EMP	EMP
Occupation of head of household	AM	AM	BC	BC	AM	BC	BC
Education of head of household	15	15	10	10	15	10	10
Age of head of household	45	45	45	55	45	45	55
Use of car on Saturday	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Travel time to work by bus	0	30	30	30	60	60	60
Travel time to work by car	0	10	10	10	20	20	20
Walking distance to bus stop	2	3	3	3	10	10	10
Inclusive value	6	6	6	6	6	6	6
Purchase probability	0.14	0.22	0.08	0.02	0.54	0.26	0.07

Note: SE = self employed, AM = academic or managerial, EMP = employee, and BC = blue collar.

Purchase Model for Households Already Owning a Vehicle

The purchase model for households already owning a vehicle describes the probability that a household will replace its car during a certain year, given the socioeconomic descriptors of the household and attributes of the existing car. The estimation results are given in Table 5.

An unexpected finding was that high income reduces the probability of replacing a car. This result is strengthened by the negative coefficients of other attributes that are correlated with income, such as education, work status, and coverage of car operating and maintenance costs.

Age of head of household has a negative effect on the probability of replacing the car. So does being a one-adult household, especially if this adult is a woman. It is possible that women and older persons face higher transaction costs. As expected, households that are exempt from taxes have a higher purchase probability. The exemption is generally given to disabled people who need a reliable vehicle and therefore replace it often. Accessibility measures were found to have no effect on the replacement decision. This is expected, because replacing a car will have only a small effect on accessibility.

The attributes that describe the previous car have the expected effect on replacement choice. The replacement probability increases with age and use

Table 5. Estimation results—a logit model of car purchase decision for households with a vehicle.

Variable	Coefficient Estimates	Asymptotic t-Statistic
One-adult household	-1.52	-2.35
Woman head of household	-1.23	-1.77
Income < 10,000 I.L., head of household employed	0.455	0.816
Income > 20,000 I.L., head of household employed	-0.550	-2.00
Income not reported, head of household employed	-0.919	-2.15
Head of household self-employed	-0.0654	-0.14
Full car maintenance cost covered	-0.150	-0.30
Exemption from car taxes	1.51	2.03
Education of head of household—more than 12 years	-0.494	-2.18
Age of head of household 35 years of younger	0.816	2.71
Age of head of household 50 years or older	-0.721	-2.65
Monthly kilometerage exceeds 2000 km	0.960	2.37
Previous car characteristics		
Year	-0.0907	-2.55
Engine size smaller than 1300 cc	0.699	2.30
Engine size larger than 1800 cc	-0.829	-1.59
No. of years car was owned by household	0.076	1.49
Dummy for purchase alternative (corrected)	6.5	3.67
Expected maximum utility from car type choice	-0.092	-0.625

Note: Number of observations = 582; $\mathcal{L}_{\hat{\beta}} = -267$; and $-2(\mathcal{L}_0 - \mathcal{L}_{\hat{\beta}}) = 271$, d.f. 17.

of the vehicle and with duration of holding, and decreases with engine size (a proxy for durability). The positive effect of duration of holding on replacement reflects the decrease in relative utility of the existing vehicle with time.

The coefficient of the expected maximum utility in this model is negative but small. It is not significantly different from zero. This means that given the socioeconomic characteristics of the household and the attributes of the existing car, the utility from the car type choice was not found to have an effect on the replacement decision. In other words, the decision to replace a car is independent of the type choice.

This result deserves an explanation because it is expected that the replacement decision would be dependent on the characteristics of available cars. Two reasons may account for this effect, and both reasons are related to the fact that this model is based on cross-sectional data. First, exogenous variables such as fuel costs and microeconomic conditions do not vary across the sample. Thus any possible effect of the variables is contained in the transaction-specific constant. Second, technological changes in the car industry are not so dramatic as to cause a major shift in choice from one period to another. Thus the expected maximum utility from all car types is well represented by the attributes of the existing car and household characteristics, and those attributes are represented in the replacement model.

Forecasting Results

The model developed in this study was used to evaluate the automobile-demand effects of the following two scenarios:

1. A change in tax rates that will result in an increase of 20 percent in the price of new large cars with an engine size larger than 1900 cc, and a decrease of 20 percent in the prices of new small cars with an engine size smaller than 1300 cc (it is assumed that the changes in the prices of new cars will cause price changes of similar proportions in the used car market), and
2. An increase of 100 percent in fuel prices.

As expected, the price changes created a relative advantage for small cars, which resulted in an increase in purchases of these cars and a corresponding decrease in purchases of larger cars. An interesting finding is that the increase occurs mainly in the purchases of new cars in the engine-size category that benefited most from the tax changes, namely, cars with engine sizes of 1100 to 1300 cc. Apparently, for households that intended to buy a small car, the price reduction enabled the purchase of a newer or a bigger car within the same category. On the other hand, the price changes caused some potential buyers of intermediate or large cars to choose a smaller car, and the savings thus gained could be used to purchase a newer car. The net effect of these shifts is an increase of 38 percent in the category of new cars with engine sizes between 1100 and 1300 cc, and a 10 percent decrease in the demand for cars with engine size larger than 1400 cc.

The assumed changes in car prices did not affect the purchase decision. This result is obvious for households that already own a car and consider whether to replace it, because according to the model the replacement decision is independent of the car type decision. For households without cars, the purchase decision is affected by the expected maximum utility from all car types. However, these values are not changed much by the proposed price changes, and the predicted aggregate effect is negligible.

The increase of 100 percent in fuel price increased the demand for fuel-efficient vehicles (with engine volumes of up to 1000 cc) by 16 percent and caused a corresponding decrease in the demand for bigger cars, especially for gas-guzzelers in the 1900 cc or larger category. The choice of a smaller car enabled the buyer to purchase a newer car with the same budget, so that the overall demand for new cars increased by 7 percent.

The 100 percent increase in fuel price was found to have a strong impact on the probability of buying a first car. The estimated number of purchases decreased by 47 percent as a result of the fuel price hike. However, the implied assumption that the car prices will remain unchanged in the face of such changes in demand is unrealistic, and the real effect of a large increase in fuel price on purchase probabilities of first cars is expected to be smaller.

SUMMARY AND CONCLUSIONS

In this work dynamic transaction models for car ownership were defined and estimated. The main advantage of these models is their dynamic structure--Markovian of the first order--which provides for a direct account of transaction costs and characteristics of the previous ownership level as well as the attributes of cars in the choice set. The nested structure of the logit model used in this study provides for an efficient data-collection effort and eases the estimation process.

Specification of the models includes policy-sensitive variables such as characteristics of car types, socioeconomic variables, and accessibility variables.

The use of the models to support policy decisions was demonstrated for two scenarios involving changes in purchase and operating costs. For a more comprehensive policy analysis, the models developed here can be easily supplemented to include all the transactions that are possible in the car market (e.g., transactions in multicar households and transactions that result in a reduction in level of car ownership). A full set of such models can be incorporated into a system of equations to represent equilibrium conditions in a car market. A general structure of such equations is described by Manski (14). It is believed that the use of the dynamic models developed in this work, in the framework of equilibrium equations, can provide a useful system for the analysis of policies that affect the car market.

ACKNOWLEDGMENT

The research reported here was performed at the Transportation Research Institute, Technion--Israel Institute of Technology, Haifa. The study was sponsored by the Israeli Ministry of Transport, and we would like to thank Nurit Straus, the project monitor, for her continuing support. Partial support was also provided by a grant from the National Science Foundation to the Massachusetts Institute of Technology to study the demand for automobiles. We also greatly benefited from the enlightening discussions we had with Charles Manski.

REFERENCES

1. C. Lave and K. Train. A Behavioral Disaggregate Model of Auto Type Choice. Transportation Research, Vol. 13A, 1979, pp. 1-9.
2. Wharton Econometric Forecasting Associates. An Analysis of the Automobile Market: Modeling the Long Run Determinants of the Demand for Automobiles. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1977.
3. M. Ben-Akiva and S. Lerman. Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work. Transportation, Vol. 3, 1976, pp. 357-376.
4. S.R. Lerman and M. Ben-Akiva. Disaggregate Behavioral Model of Automobile Ownership. TRB, Transportation Research Record 569, 1976, pp. 34-55.
5. L.D. Burns, T.F. Golob, and G.C. Nicolaidis. Theory of Urban Household Automobile Ownership Decisions. TRB, Transportation Research Record 569, 1976, pp 56-75.
6. J. Kain and G. Fauth. Forecasting Auto Ownership and Mode Choice for U.S. Metropolitan Areas. Department of City and Regional Planning, Harvard Univ., Cambridge, Mass., Working Paper, 1977.
7. R.R. Johnson. Aggregation and the Demand for New and Used Automobiles. Review of Economic Studies, Vol. 45, 1978, pp. 311-327.
8. C.F. Manski and L. Sherman. An Empirical Analysis of Household Choice Among Motor Vehicles. Transportation Research, Vol. 14A, 1980, pp. 349-366.
9. D.A. Hensher and T. Manefield. A Structured Logit Model of Automobile Acquisition and Type Choice. Proc., Australian Transport Research Forum, Hobart, Tasmania, Australia, 1982.
10. N.S. Cardel and F. Dunbar. Measuring the Societal Impact of Automobile Downsizing. Transportation Research, Vol. 14A, 1980, pp. 423-434.
11. J.H. Boyd and R.E. Mellman. The Effect of Fuel Economy Standards on the U.S. Automobile Market: An Hedonic Demand Analysis. Transportation Research, Vol. 14A, 1980, pp. 367-378.
12. D. McFadden. Modelling the Choice of Residential Location. In Spatial Interaction Theory and Residential Location (A. Karlquist et al., eds.), North-Holland, Amsterdam, 1978.
13. M. Ben-Akiva and B. Francois. μ Homogeneous Generalized Extreme Value Model. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Working Paper, 1981.
14. C.F. Manski. Short Run Equilibrium in the Automobile Market. The Maurice Falk Institute for Economic Research in Israel, Jerusalem, Discussion Paper 8018, 1980.

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

Experiments with Optimal Sampling for Multinomial Logit Models

YOSEF SHEFFI AND ZVI TAREM

In this paper a recently published method for optimizing the sample used in estimating discrete-choice models is tested. The work is intended to identify and explore the elements that influence the effectiveness of this methodology in designing sampling procedures for estimating logit models. The investigation includes both analytical and numerical tests. The results indicate that the sample optimization method can improve the accuracy of the resulting estimates, as compared with random sample.

Data collection is, in many cases, the major cost item in studies that involve the estimation of econometric models. Techniques for sample design have therefore been developed for many econometric and statistical models (1). In this paper discrete choice models, which are extensively used in travel-demand analysis, are examined, and, in particular, the multinomial logit (MNL) model is discussed. The focus here is on a method for optimizing the sample used to estimate discrete-choice models. The applicability of this sample optimization approach to the collection of the sample points (the data) used to estimate MNL models is examined. Also examined is the appropriate amount of effort that should be invested in the sample optimization process.

The original development of the sample optimization method, which is the subject of this paper, is from Daganzo (2). Daganzo's method is a stratified sampling technique. It assumes that the population to be sampled from can be partitioned into separate groups (or strata) and that observations can be sampled independently from each group. The objective of the sampling method is to determine how many observations should be drawn from each group so that the total estimation error is minimized. The estimation error is a composite measure of the error in all the model parameters. Naturally, this minimization is subject to a budget constraint. This sampling method attempts to determine the best allocation of the sampling budget. (The companion problem, that of determining the minimum budget required to achieve a certain accuracy, is somewhat more difficult. Its solution, however, can be inferred from the solution of the problem under consideration.)

Three main points are discussed in this paper. The first point is the applicability of the approach in terms of potential. The question examined in this context is the sensitivity of the sampling error to different sample designs. The second point is that the solution of the sample optimization (SO) problem requires prior estimates (or guesses) of the values of the parameters of the model to be estimated. The applicability of the whole concept depends, naturally, on the required accuracy of these prior estimates. The tests described in this paper explore this point in some detail. The third point is related to the first point. It has to do with the question of the amount of effort that should be invested in obtaining these prior estimates. Such an effort should be judged in comparison to the level of effort of the entire study, which means that the relevant question is the allocation of effort between obtaining the prior estimates and the estimation itself.

This paper is organized as follows. First, Daganzo's SO method is outlined. Then the application of this method to the MNL model is reviewed. Next, the question of the applicability of the SO method is explored by looking at a simple one-parameter model and a two-parameter model. Then the aforementioned issue of resource allocation in the framework of a small case study is discussed, and finally conclusions are given.

It should be noted that the conclusions of this paper are based on numerical experiments, which means that not all the results can be generalized in all circumstances. The experiments are described in further detail by Sheffi and Tarem (3).

SAMPLE OPTIMIZATION PROGRAM

Daganzo's SO method attempts to minimize the error associated with the estimation of the parameters of a discrete-choice model. The optimization problem is formulated as a mathematical minimization program, where a composite measure of the estimation error serves as the objective function and the sample group sizes are the decision variables. This approach assumes that the model under consideration is estimated by using the maximum likelihood (ML) method. It also assumes that the distribution of explanatory variables in each group is known. [This information may not be available, in which case the methods discussed by Lerman and Manski (4) may be used.]

The objective function of the SO program relates the sampling error to the sample group sizes. This expression can be derived from the Kramer-Rao lower bound on the covariance matrix of ML estimators. Letting $\underline{x}$ be a vector of explanatory variables, y be the dependent variable, and $\underline{\theta}$ be the vector of parameters for some model, this bound ($\Sigma_{\hat{\theta}}$) is given by

$$\Sigma_{\hat{\theta}} = \left\{-E[\nabla_\theta^2 L(\underline{\theta}|y,\underline{x})]\right\}^{-1} \qquad (1)$$

where $L(\cdot|\cdot,\cdot)$ is the log-likelihood of the sample $(y,\underline{x})$ evaluated at $\underline{\theta}$, $\nabla_\theta^2 L(\cdot)$ is the θ-Hessian of $L(\cdot)$, and $E[\cdot]$ denotes the expectation operator that, in Equation 1, is carried out with respect to both y and $\underline{x}$.

For stratified sampling, where all observations are independent, the sample log-likelihood is given by the sum

$$L(\underline{\theta}|y,\underline{x}) = \sum_{k=1}^{K} \sum_{n=1}^{N_k} L(\underline{\theta}|y_n^k, \underline{x}_n^k) \qquad (2)$$

where

$L(\underline{\theta}|y_n^k,\underline{x}_n^k)$ = log-likelihood of sample point n from group k,
$(y_n^k,\underline{x}_n^k)$ = observed values at this point,
N_k = number of observations in group k, and
K = number of groups in the sample.

The Hessian of this function is

$$\nabla_\theta^2 L(\underline{\theta}|y,\underline{x}) = \sum_{k=1}^{K} \sum_{n=1}^{N_k} \nabla_\theta^2 L(\underline{\theta}|y_n^k, \underline{x}_n^k) \qquad (3)$$

In stratified sampling it is assumed that all observations from a given group (k) are realizations of some underlying distribution $f_{y,\underline{x}}^{(k)}(y,\underline{x})$ that characterizes the group. Thus all these observations have the same expectation. The expectation of Equation 3 is therefore

$$E[\nabla_\theta^2 L(\underline{\theta}|y,\underline{x})] = \sum_{k=1}^{K} N_k E^{(k)}[\nabla_\theta^2 L(\underline{\theta}|y,\underline{x})] \qquad (4)$$

where $E^{(k)}[\cdot]$ denotes the expectation taken over the distribution $f_{y,\underline{x}}^{(k)}(y,\underline{x})$, and the designations n and k are omitted from the notation of the likelihood function in order to clarify the presentation. The final expression for the bound on the parameter covariance matrix is obtained by combining Equation 4 with Equation 1, i.e.,

$$\Sigma_{\hat{\theta}}(\underline{N}) = \left\{-\sum_{k=1}^{K} N_k E^{(k)}[\nabla_\theta^2 L(\underline{\theta}|y,\underline{x})]\right\}^{-1} \qquad (5)$$

To minimize the estimation error, a scalar measure of the size of the parameter covariance matrix has to be defined. A family of such measures can be defined by using a quadratic form of the covariance matrix with a (column) vector of constants, $\underline{z}$, i.e.,

$$F = \underline{z}^T \Sigma_\theta \underline{z} \qquad (6)$$

where F is the estimation error, Σ_θ is the true parameter covariance matrix, and the superscript T denotes the transposition operation. Because the true covariance matrix is not known, the approximation in Equation 5, which holds asymptotically for maximum likelihood estimators, is used instead. Thus $F(\underline{N}) = \underline{z}^T \Sigma_{\hat{\theta}}(\underline{N})\underline{z}$, where $\underline{N} = (\ldots,N_k,\ldots)$. The form of the error measure used in this paper uses a vector $\underline{z} = (1,1,\ldots,1)$, i.e., $F(\underline{N})$ is the sum of the elements of the parameter covariance matrix.

The optimal sample composition is derived by minimizing $F(\underline{N})$ with respect to the N_k's. The unconstrained solution to the minimization is, obviously, to sample an infinitely large number of observations from each group. The estimation error then approaches zero. The sample size, however, is bounded by the budget available for sampling, and possibly by some other physical size constraints. The total budget constraint may be expressed by

$$\sum_{k=1}^{K} c_k N_k < B \qquad (7)$$

where c_k is the cost of sampling one unit from group k, and B is the total budget available. Physical group size constraints may be expressed as

$$N_k < N_k^{max} \quad \text{for some groups } k \tag{8}$$

In addition, the constraint set should always include nonnegativity of the group sizes.

The SO program can be summarized as follows:

$$\underset{N_k}{\text{Min}} \quad F = \underline{z}^T \left\{ -\sum_{k=1}^{K} N_k E^{(k)} [\nabla_\theta^2 L(\underline{\theta} | y, \underline{x})] \right\}^{-1} \underline{z} \tag{9a}$$

Subject to

$$\sum_{k=1}^{K} c_k N_k < B \tag{9b}$$

$$0 < N_k < N_k^{max} \quad \text{for all } k \tag{9c}$$

Daganzo (5) indicates that this program has a unique local minimum for any constant vector $\underline{z}$ and any form of the log-likelihood function $L(\underline{\theta}|y,\underline{x})$. This means that the problem can be solved by using standard nonlinear, constrained optimization methods. The algorithm used in this work is based on the gradient projection (6) method.

The exact form of the objective function depends on the specific model for which the sample is designed. Sheffi and Tarem (3) formulate and solve this program for several model forms. In the next section the derivation of this expression for MNL models is reviewed. The remainder of the paper is aimed at evaluating the usefulness and applicability of the approach.

SAMPLE OPTIMIZATION FOR LOGIT MODELS

The logit formula is the most widely used discrete-choice model because of the simplicity of its form. A detailed description of the model can be found in Domeneich and McFadden (7).

The logit model can be used to quantify some aspects of individuals' choice among a set of alternatives. The model can be interpreted in the framework of random utility maximization by assuming that each decision maker attaches a measure of utility to each alternative and chooses the one with the largest utility. The utility of alternative j to an individual randomly drawn from the population (u_j) is modeled as the sum of a systematic utility term (v_j) and an error term that is assumed to be randomly distributed across the population. The systematic utility captures the model specification in terms of the relationships between the utility and the explanatory variables; thus $v_j = v_j(\underline{\theta},\underline{x})$. The specification of the random part determines the family of models to be used. If these random terms are assumed identically and independently Gumbel distributed, the resulting model is the MNL mode. The MNL model gives the probability that each available alternative is chosen (i.e., it has the highest utility)--$P_j(\underline{\theta},\underline{x})$--as

$$P_j(\underline{\theta}, \underline{x}) = \exp[v_j(\underline{\theta}, \underline{x})] / \sum_{i \in I} \exp[v_i(\underline{\theta}, \underline{x})] \tag{10}$$

where I is the index set of the available alternatives. In most cases the systematic utility is assumed to be linear in the parameters, and thus $v_j(\underline{\theta},\underline{x}) = \underline{\theta}^T\underline{x}$.

To develop the SO objective function for the MNL model, the θ-Hessian of the log-likelihood function has to be derived for such models. The likelihood of a sample point n can be written as

$$L^*(\underline{\theta}|\underline{a}_n, \underline{x}_n) = \prod_{i \in I} P_i(\underline{\theta}, \underline{x}_n)^{a\,ni} \tag{11}$$

where $\underline{a}_n$ is an indicator variable vector that contains the observed choice, i.e., $a_{nj} = 1$ if alternative j is chosen by the nth decision maker in the sample, and $a_{nj} = 0$ otherwise. The vector $\underline{x}_n$ includes the explanatory variables for the nth observation. The choice probabilities are given by Equation 10. The logarithm of Equation 11 is simply

$$L(\underline{\theta}|\underline{a}_n, \underline{x}_n) = \sum_{i \in I} a_{nj} \log P_{nj} \tag{12}$$

where $P_{nj} = P_j(\underline{\theta},\underline{x}_n)$ for ease of notation. The sample log-likelihood includes the sum over n of $L(\underline{\theta}|\underline{a}_n,\underline{x}_n)$, i.e.,

$$L(\underline{\theta}|\underline{a}, \underline{x}) = \sum_{n=1}^{N} \sum_{j \in I} a_{nj} \log P_{nj} \tag{13}$$

where N is the total sample size.

The derivation of the θ-Hessian of the sample log-likelihood function is simple but somewhat lengthy (3). The final result of applying the Hessian operator to the log-likelihood function is

$$\nabla^2 L(\underline{\theta}|\underline{a}, \underline{x}) = -W^T Q W \tag{14}$$

where W is the matrix of attribute differences for an individual randomly drawn from the population, i.e., row j of W is the difference $\underline{x}_j - \underline{x}_I$, where I is the index of the last alternative (any other alternative can be chosen as a base). Q is a square matrix with the elements,

$$[Q]_{ij} = P_i(\delta_{ij} - P_j) \quad \text{for } i, j = 1, 2, \ldots, I-1 \tag{15}$$

where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. After inserting Equation 14 into the objective function of the sample optimization program (Equation 9a), this function becomes

$$F = \underline{z}^T \left\{ -\sum_{k=1}^{K} N_k E^{(k)} [-W^T Q W] \right\}^{-1} \underline{z} \tag{16}$$

Computing the expectations of $E^{(k)}[\cdot]$ in Equation 16 requires prior knowledge of both the distribution of the attributes in all groups and the values of the unknown parameter vector ($\underline{\theta}$). The latter is required for computing the choice probabilities that appear in the elements of Q. As previously mentioned, it is assumed in this paper that the attribute distributions are known before sample optimization. The main concern of this paper is with the required accuracy of the initial parameter guesses.

Because the function under the expectation operator is complicated, a numerical Monte Carlo approach for computing these expectations was adopted. With this approach, M observations were drawn from the distribution of the attributes and the average, where

$$(1/M) \sum_{m=1}^{M} [-W_m^T Q_m W_m] \tag{17}$$

was used as an approximation of the true expectations.

INACCURACIES IN INITIAL GUESSES: ONE-PARAMETER MODEL

In this section two of the issues that determine the applicability of the SO approach are examined. These questions are addressed in the context of a simple logit model that includes only two alternatives and a single parameter.

The usefulness of the SO method depends on two separate questions. The first is whether SO actually improves the accuracy of the resulting parameter estimates. Although SO assures minimum error in estimation, the improvement relative to other sample designs may be insignificant. In this case the optimization process is not cost effective. The second question is the dependence of the optimization results on the accuracy of the initial parameter guesses used in the optimization. If the optimization process requires accurate parameter values to yield satisfactory sample composition, its usefulness will be limited because having such accurate parameter values obviates the need for the estimation process.

Thus, for the SO method to be useful, it is necessary that the estimation error will decrease when an optimal sample is used, but also that this optimal composition may be obtained without an accurate initial parameter guess.

The tests described in the following sections are designed to determine if and when these conditions can be met for a simple logit model, where the issue can be addressed analytically. The simple logit model chosen for this analysis includes two alternatives and one parameter. The systematic utilities of these alternatives are $x_1\theta$ and $x_2\theta$, respectively. The choice probabilities have the form

$$\begin{aligned} P_1 &= \exp[(x_1 - x_2)\theta]/\{1 + \exp[(x_1 - x_2)\theta]\} \\ &= \exp(W\theta)/[1 + \exp(W\theta)];\ P_2 = 1 - P_1 \\ &= 1/[1 + \exp(W\theta)] \end{aligned} \tag{18}$$

and the optimization objective function (Equation 16) for this model is given by

$$\begin{aligned} F(\underline{N}) &= 1/\sum_{k=1}^{K} N_k E^{(k)}[W^2Q] = 1/\sum_{k=1}^{K} N_k E^{(k)}\{W^2\exp(W\theta) \\ &\quad \div [1 + \exp(W\theta)]^2\} \end{aligned} \tag{19}$$

The minimization of $F(\underline{N})$ in Equation 19 is equivalent to the maximization of the reciprocal of $F(\underline{N})$, i.e.,

$$\underset{N_k}{\text{Min}}\ F(\underline{N}) = \underset{N_k}{\text{Max}}\ F'(\underline{N}) = \sum_{k=1}^{K} N_k E^{(k)}[W^2Q] \tag{20}$$

For a problem with a simple budget constraint (such as Equation 9b), the solution of this SO program is to sample all observations from the group (k) with the largest value of

$$\alpha^{(k)} = E^{(k)}[W^2Q]/c_k \tag{21}$$

The total sample size will, of course, be B/c_ℓ, where ℓ is the group sampled. From Equation 21 it is clear that if the expectations $E^{(k)}[\cdot]$ are similar in all groups, the sample should include observations from the group with the lowest sampling cost. If the expectations differ considerably, however, a group with higher sampling cost may contribute more to the estimation accuracy and should therefore be chosen for sampling.

The accuracy of the initial parameter guess, denoted by θ_0, needed in computing the $\alpha^{(k)}$'s is important only if it can cause the sampling from the wrong group. In other words, as long as the values of $\alpha^{(k)}$ computed by using θ_0 suggest the same choice of group as would happen with the true parameter (θ), the optimal sample composition is not affected by inaccuracies in θ_0.

For example, assume that there are only two groups, and that the sampling costs are the same in both. If the true $E^{(1)}[W^2Q]$ is 10 times larger than the true $E^{(2)}[W^2Q]$, computing $\alpha^{(k)}$ with even a bad guess of θ will still probably suggest sampling from group 1. If, on the other hand, the true $E^{(1)}[W^2Q]$ is only 10 percent larger than the true $E^{(2)}[W^2Q]$, a slight inaccuracy in θ_0 may reverse the choice initiated by Equation 21. In this case, however, the contribution of both groups to the estimation accuracy is similar, and sampling from the wrong group would not introduce a large increase in the estimation error (F).

In summary, Equation 21 indicates that if the group attribute distributions (and hence the group expectations) are considerably dissimilar, sampling from the wrong group may cause a large estimation error, but the correct group for sampling may be relatively easy to determine. In cases when this determination is more difficult (i.e., when the groups are similar), the cost of an error is not high. Thus this analysis leads to the conclusion that SO should be useful in this case, even with questionable prior estimates of θ.

TWO-PARAMETER MODEL

A similar analysis can be applied to a slightly more complicated model, which includes two alternatives and two parameters. In this case the choice probabilities have the form

$$\begin{aligned} P_1 &= \exp(W_1\theta_1 + W_2\theta_2)/[1 + \exp(W_1\theta_1 + W_2\theta_2)]; \\ P_2 &= 1/[1 + \exp(W_1\theta_1 + W_2\theta_2)] \end{aligned} \tag{22}$$

where W_1 and W_2 are the two elements of attribute differences vector $\underline{W} = (W_1, W_2)$. The SO objective function (Equation 16) in this case is given by

$$F(\underline{N}) = \underline{z}^T \left\{\sum_{k=1}^{K} N_k E^{(k)}[Q\ W^T W]\right\}^{-1} \underline{z} \tag{23}$$

where the single element of the matrix Q is

$$(Q)_{1,1} = \exp(W_1\theta_1 + W_2\theta_2)/[1 + \exp(W_1\theta_1 + W_2\theta_2)]^2 \tag{24}$$

The general analysis of this case cannot be carried out analytically because of the complexity of Equation 24. The approach followed here was to analyze a specific sample design case with known true parameters. The problem setup included two groups with the following attribute distributions:

$$W_1^{(1)} = W_1^{(2)} = 1;\ W_2^{(1)} \sim N(0.5, 0.25);\ W_2^{(2)} \sim N(-0.5, 0.25)$$

The true parameters (see Equation 22) were set to $\theta_1 = \theta_2 = 1.0$. The true group expectations can be calculated by using the simulation method, explained by Equation 17, as follows:

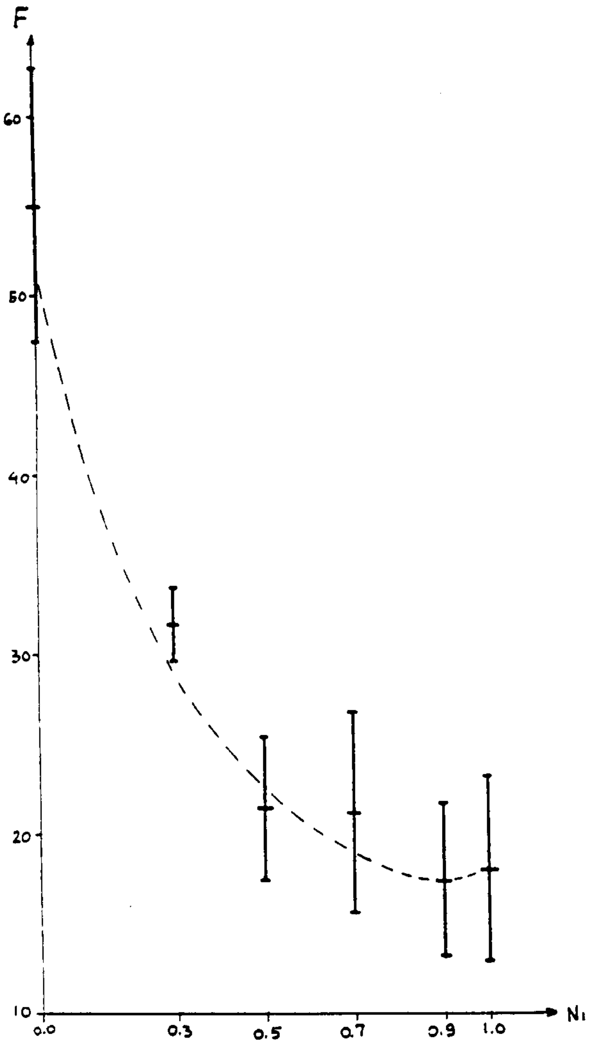
$$E^{(1)}[QW^TW] = \begin{bmatrix} 0.1496 & 0.0549 \\ 0.0549 & 0.0545 \end{bmatrix};\ E^{(2)}[QW^TW] = \begin{bmatrix} 0.2234 & -0.122 \\ -0.122 & 0.1176 \end{bmatrix}$$

The budget constraint was set to $N_1 + N_2 \leq 1$, which implies that $c_1 = c_2 = 1$ and that the N_k's can be looked on as sample shares rather than number of observations. Because the budget constraint is always binding in these problems, the sample composition can be represented by the single variable N_1, and N_2 can be replaced by $1 - N_1$.

The dependence of the estimation error on the sample composition was determined by evaluating the objective function (Equation 23) at different values of N_1. The resulting curve is shown as the dotted line on Figure 1. The estimation error has a distinct minimum at $N_1 = 0.908$, which corresponds to the value $F^* = 17.567$. It rises sharply for values of N_1 less than 0.69 (the 10 percent deviation mark).

Each sample composition is associated with a

Figure 1. Intervals $F \pm \sigma_F$ plotted versus the analytical curve.



unique value of the objective function in Figure 1. The sampling process, however, introduces a randomness that may cause the actual estimation error to deviate from the one indicated in Figure 1. This is because once the group size is determined, the actual observations are still randomly sampled within each group. Thus different samples with the same composition may result in different estimation errors. To verify the relationships shown in Figure 1, a simulated data set was generated. Attribute observations were generated from the previously mentioned distribution of the explanatory variables within each group. The chosen alternative was determined by simulating the total utilities of the alternatives to each individual and recording the alternative with the largest utility as the chosen one. This simulation was carried out by generating a Gumbel-distributed random variable (by using the cumulative distribution inversion method) and adding it to the observed utility.

The logit estimation routine computes, apart from the parameter estimates, an estimate of the parameter covariance matrix based on the sample. An approximate estimation error may be computed by summing the elements of this matrix (see Equation 6). Five different samples were generated for each selected composition, and the estimation error was computed for each one by using that procedure. An interval of probable values for the estimation error was derived from the mean and standard deviation of the five measurements (i.e., $F = \bar{F} \pm \sigma_F$, where $\bar{F}$ is the average and σ_F is the standard deviation of the five values). These intervals are also plotted in Figure 1. As demonstrated in the figure, the sampling results depict the same relation between the estimation error and the sample composition as shown by the analytical curve.

In the particular example solved here, Figure 1 demonstrates that the SO is worthwhile even when the randomness of the sampling procedure is accounted for. In general, however, this may not be the case if the variance of the attribute distribution is large. Such a case means that the groups are, statistically, quite similar. As in the one-parameter case, this means that SO is not cost effective because the (expected) cost of an error in the groups' composition is not large.

The dependence of the optimal solution on the accuracy of the initial guesses was determined by solving the SO problem by using different values of the initial parameter guesses ($\underline{\theta}_0$) around the true parameters ($\underline{\theta}$). Figure 2 shows contours of equal composition over a range of values of $\underline{\theta}_0$ around the true value of $\underline{\theta} = (1.0,1.0)$. The figure shows that in most of the region, except for the upper right corner, the optimal composition is within 10 percent of the best composition. The best composition is given by $N_1 = 0.908$, which was computed by using the true parameter values.

Figure 3 demonstrates the same point from a different angle. The relationships between the estimation error and the initial guesses used in the optimization process can be derived by reading, from Figure 1, the values of F that correspond to the sample compositions shown in Figure 2. These values can then be transformed to percentage differences from the minimum error, $F^* = 17.567$. Figure 3 depicts contours of equal percentage differences over the same range of $\underline{\theta}_0$ used in Figure 2. As shown in Figure 3, most of the region analyzed lies within 10 percent of the minimum error. In summary, it can be concluded that although arbitrary sample compositions may yield large estimation errors (as seen in Figure 1), the use of SO, even with a wide range of possible initial parameter guesses, limits the error to small deviations from the minimum error obtained by using the true parameter values.

Figure 2. Contours of equal N_1 over the range of the initial parameter guesses θ_0.

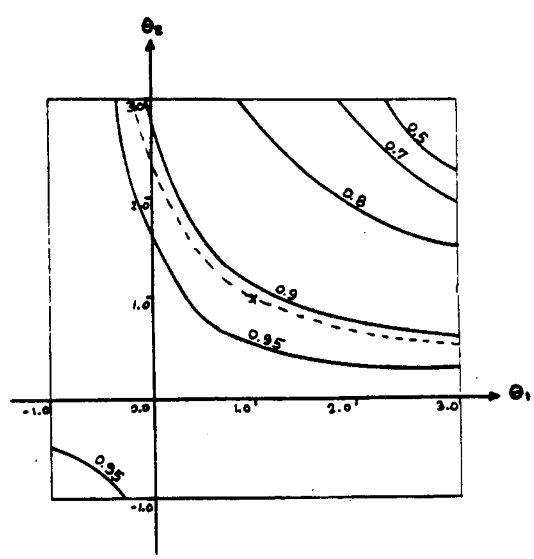
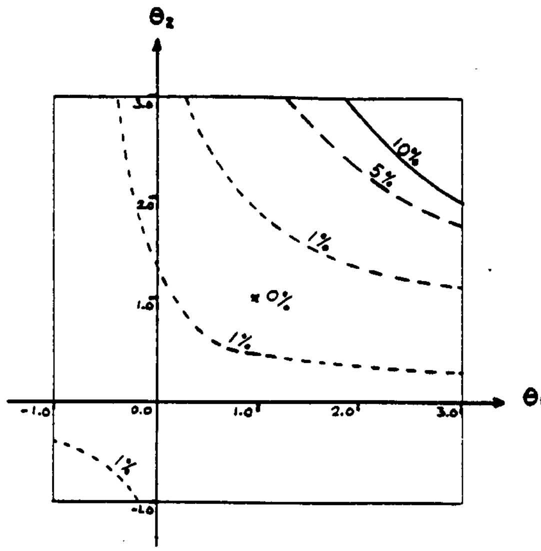


Figure 3. Contours of equal error (F) over the range of initial parameter guesses θ_0.



OPTIMAL BUDGET ALLOCATION

The initial parameter guesses used in the optimization process may come from two distinct sources. The first one is an external source, such as another study or a set of studies conducted elsewhere or in the past. The second one is an internal source, such as a pilot study conducted on the current population. In this case a small presample may be randomly drawn in order to estimate $\underline{\theta}_0$. The final parameter estimation will be based on a combined sample, including the observations of the presample and the main sample. The relevant question here is what is the appropriate relative investment in the initial sample that will yield the best accuracy of estimation when using the combined sample.

The procedure followed in this research for determining the optimal allocation of the sampling budget was to first allocate some prescribed amount (B_1) to an initial random sample. The parameter estimates based on this sample were used as initial guesses in determining the optimal sampling scheme for the main sample, subject to the remaining budget B_2. The main sample was then drawn and combined with the initial one and used to estimate the model. The estimation error was computed from the estimated parameter covariance matrix of this model. The optimal allocation was determined by parametrically varying the amount spent on the initial sample.

The existence of an optimal allocation stems from the fact that when the budget (B_1) spent on the initial random sample is small, the resulting estimates of the parameters are not accurate. Thus the main sample will not be close to optimality, and the estimation error can be expected to be large. On the other hand, when most of the budget is spent on the initial sample, the resulting initial estimates will be accurate, and the small main sample is close to being optimal. The combined sample, however, will include primarily the random, nonoptimal sample, and the estimation error is again expected to be large. Therefore, there may be some optimal allocation of the budget such that the size of the random sample is sufficient to provide relatively accurate estimates, but the remaining optimized sample is sufficiently large to reduce the error measure.

This procedure was carried out by using a large data set as a population. The data were extracted from the 1977 National Personal Travel Study (NPTS) data base. A simple model of automobile ownership levels was used as an example model in these tests. The model included three alternatives: owning two or more cars, owning one car, and owning no car. The systematic utilities of the alternatives were specified as

$$u_1 = \theta_1 + \theta_3 \cdot \text{INCOME} + \theta_4 \cdot \text{HHSIZE}$$

$$u_2 = \theta_2 + \theta_3 \cdot \text{INCOME}$$

$$u_3 = 0.0$$

where INCOME is measured in $10,000 units, and HHSIZE is the number of members in the household. The population data set contained 7,393 observations partitioned into three groups along the income dimension, according to the following ranges:

Group	Income Range ($)	Observations
1	0-7,500	2,565
2	7,500-20,000	3,331
3	>20,000	1,497
Total		7,393

The distributions of the attributes (INCOME and HHSIZE) were estimated from the data.

A budget size of B_1, varying between 40 and 200, was allocated to the initial random sample (assuming a cost of one unit for all observations). The composition of the main sample was determined by solving the optimization problem with the constraint $N_1 + N_2 + N_3 \leq B_2$, where $B_2 = 200 - B_1$. The two samples were then combined to yield a sample of size 200, and the estimation error was computed from the combined sample. This procedure was repeated five times for each value of B_1. The interval $\bar{F} \pm \sigma_F$ of the five measurements is plotted versus B_1 in Figure 4. A shallow minimum can be observed around $B_1 = 80$, which means that 80 observations should be sampled at random. The results of this estimation should be used to optimize the composition of the remaining 120 observations. The shape of the relationship shown in Figure 4 suggests, however, two hypotheses.

1. The optimal size of the initial sample is fixed, probably because it corresponds to the minimum sample size that yields reasonable initial estimates for the optimization. In this case the optimal initial sample size (B_1) is independent of the total sample size (B).
2. Optimizing a larger sample requires more accurate initial guesses, which implies a larger initial sample. In this case the optimal initial sample size (B_1) is a fixed proportion of the total sample size (B).

To test these hypotheses in the context of the examples analyzed in this section, the test procedure used in this case study was repeated for total sample sizes of B = 400 and 600 observations. The means of the five estimation error measures computed for each selected value of B_1 are plotted in Figure 5. The horizontal axis of the graph is the ratio B_1/B, and the vertical axis represents the estimation error. The measurements obtained from each value of B (i.e., 200, 400, and 600) were normalized for comparison purposes. The figure shows that for all total sample size values, the estimation error does not have a distinct minimum but is flat over the region up to $B_1/B = 0.5$ and rises thereafter.

Thus it can only be concluded that the initial

Figure 4. Estimation error intervals plotted versus the budget spent on the initial sample for total sample size of 200.

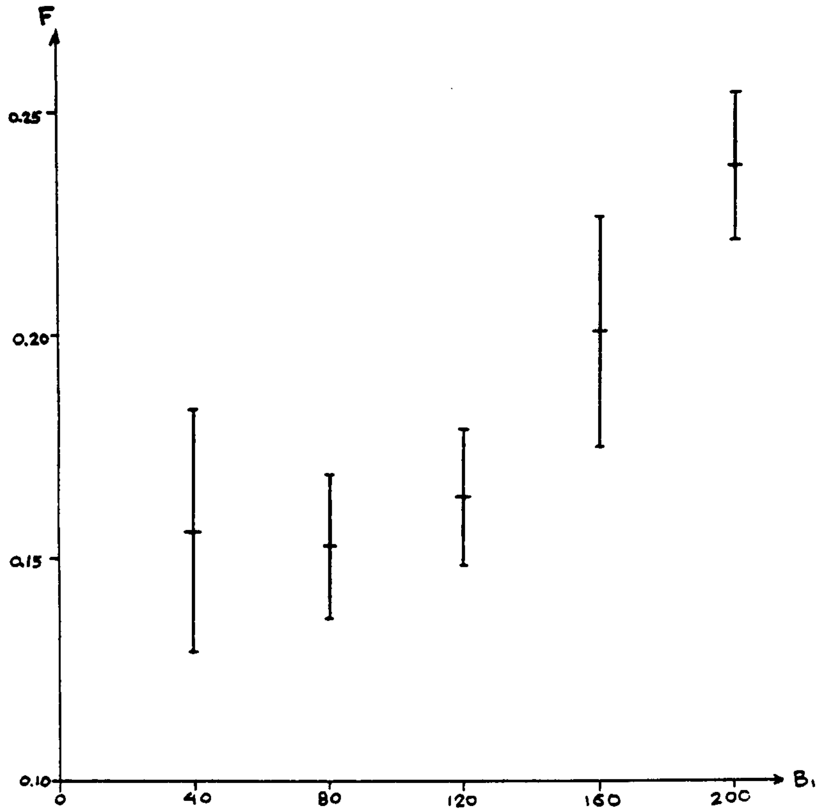
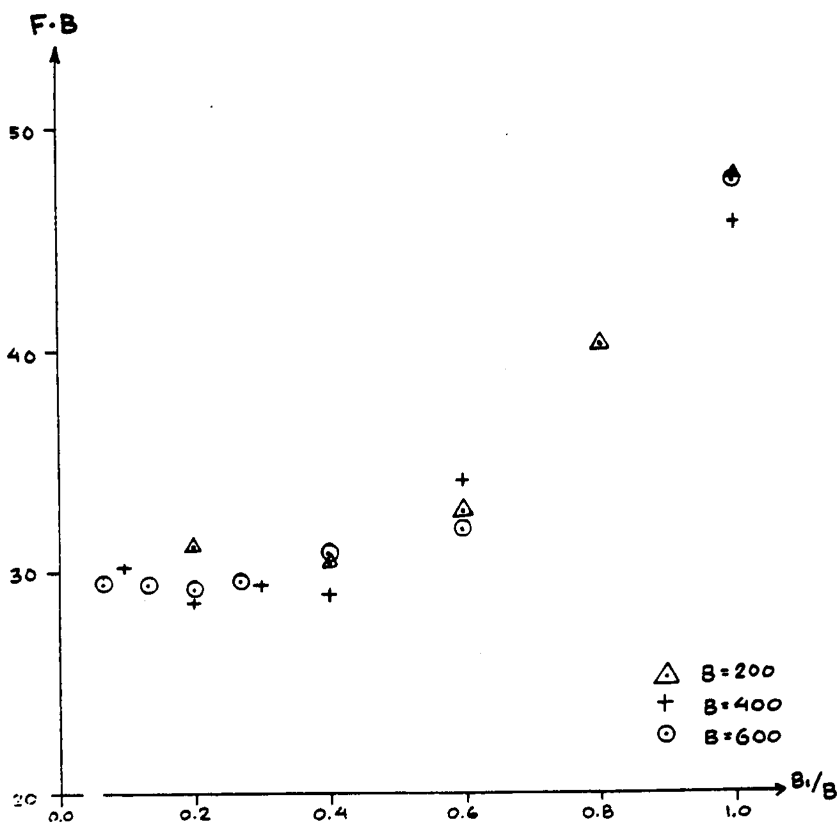


Figure 5. Mean estimation error versus ratio of initial sample size to total sample size.



sample size for this example should be less than one-half of the total sample size. This appears to suggest that, in general, the initial random sample can be small, regardless of the total sample size. The size of this sample may in fact be dictated by the requirements on the estimation of the distribution of the explanatory variables in all the groups. This point was not addressed in this paper, which assumed that this distribution is known.

CONCLUSIONS

The two major conclusions from the work described here may be stated as follows:

1. The SO procedure can introduce a significant increase in parameter estimation accuracy, and
2. This optimization need not be based on accurate initial parameter guesses; only a small pilot sample is needed to produce sufficiently accurate guesses.

It should be emphasized, however, that these conclusions result from a specific set of tests performed on prespecified models. Even though these models were chosen without any regard to the final results, these results can be generalized only with caution. The results are, however, encouraging in that the SO procedure appears to be worthwhile in cases where it can be applied. It requires nonlinear optimization software, which may not be easily used in many environments.

REFERENCES

1. W.G. Cochran. Sampling Techniques. Wiley, New York, 1963.
2. C.F. Daganzo. Optimal Sampling Strategies for Statistical Models with Discrete Dependent Variables. Transportation Science, Vol. 14, No. 4, Nov. 1980.
3. Y. Sheffi and Z. Tarem. Experiments with Optimal Sampling for Discrete Choice Models. Office of University Research, U.S. Department of Transportation, Aug. 1982.
4. S.R. Lerman and C.F. Manski. Sample Design for Discrete Choice Analysis of Travel Behavior: The State of the Art. Transportation Research, Part A, Vol. 13A, No. 1, Feb. 1979.
5. C.F. Daganzo. On the Uniqueness and Globality of Optimal Data Gathering Strategies. Transportation Science, Vol. 16, No. 2, May 1982.
6. E. Polak. Theory of Optimal Control and Mathematical Programming. McGraw-Hill, New York, 1979.
7. T. Domeneich and D. McFadden. Urban Travel Demand: A Behavioral Approach. North-Holland, Amsterdam, 1975.

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

Procedure for Predicting Queues and Delays on Expressways in Urban Core Areas

THOMAS E. LISCO

A procedure that predicts morning inbound and evening outbound queuing delays on express highway facilities in downtown areas is discussed. The procedure is based on the relationships among hourly traffic capacities at bottleneck points, daily volumes at those points, and associated queues and delays. The need for such a procedure arose from difficulties in using traffic assignment or other existing analysis techniques to predict queues and delays associated with alternative highway plans. Empirical delay data for developing the procedure came from nearly 600 speed runs conducted on the express highway system in and near downtown Boston. Fourteen queuing and potential queuing situations were analyzed. The relationships derived appear to be generalizable, and the specific results from the Boston area should apply to other urban areas of comparable size.

A procedure that predicts peak-period queuing and delays on express highway facilities in downtown areas is discussed. The procedure is based on the relationships among hourly traffic capacities at bottleneck points, daily volumes at those points, and associated peak-period queues and delays. (In this paper the term daily volume refers to average weekday traffic.) The procedure was developed by comparing observed bottleneck capacities with empirical delay data for traffic upstream of the bottlenecks. Capacities were derived from traffic counts at bottleneck locations. The delay data were from almost 600 speed runs conducted on express highway facilities in and near downtown Boston, mostly during 1978 and 1979. The procedure was developed for use in detailed evaluations of potential traffic impacts and benefits of alternative highway investments in downtown areas.

The need for such a procedure arises initially from difficulties in using the output from traffic assignment models to predict peak-period operating conditions and cost-benefit statistics associated with alternative highway plans. The basic problem is that the regional traffic assignment process derives speeds for individual links separately based on their individual volume/capacity (v/c) ratios and does not consider the queuing effects of bottleneck locations. Thus in typical downtown area queuing situations, where one bottleneck highway segment can create queues stretching into many other segments, traffic assignments cannot indicate the locations and extents of queues or the delays associated with them. Because queuing can be of major importance in peak-period expressway operations in downtown areas, the assignments can be grossly inaccurate in predicting peak-period operating speeds. Similarly, the associated cost-benefit statistics can miss much of the phenomenon they are intended to measure.

A potential solution to this problem would be to attempt a queuing analysis based on peak-period traffic assignment results. Such an analysis would fail for two reasons. First, by its very nature a traffic assignment is balanced, with all highway links clearing all traffic assigned to them for the time period of analysis. Thus there is no possibility of an assignment producing for a bottleneck link the different vehicle arrival and service rates necessary to perform a queuing analysis. Second, a well-calibrated traffic assignment will indicate all bottleneck links operating exactly at capacity during peak periods, with no indication of which are major and minor bottlenecks. In some cases the assignments will indicate volumes greater than actual capacities at bottlenecks, but the degree to which such volumes are indicated is related more to the nature of the capacity constraint in the assignment program than to the queuing phenomenon. Therefore, these greater-than-capacity volumes are not particularly helpful in predicting the extents of potential queues.

An alternative solution would be to perform a queuing analysis based on daily traffic assignment volumes with given fractions of daily traffic assigned to peak hours. The traffic assigned to peak hours would be compared with capacities at bottleneck points. Again there would be severe problems. One problem is that different bottlenecks process different fractions of daily traffic during the peak periods, with lower fractions being handled by severe bottlenecks. Thus a given fraction applied to all bottlenecks would underestimate the effects of small bottlenecks and overestimate the effects of large ones. A more important consideration is that queues rarely contain more than several hundred vehicles at one time. Thus any procedure that attempts to predict queues through calculating differences between arrival and service rates must project flows with a great deal of accuracy. Certainly, this cannot be done by allocating fractions of daily traffic to hourly flows at bottleneck points. As before, the delays calculated will relate far more to the assumptions used in the allocation than to the queuing phenomenon.

Because of the difficulties involved in predicting vehicle arrival and service rates from traffic assignments and, more generally, the problems of accurately predicting these rates by any method (1-3), the procedure documented in this paper follows an approach that predicts queuing delays directly without calculating the difference between arrival and service rates. Specifically, the analysis approach assumes that there is a consistent relationship between daily traffic volume at a bottleneck point compared with capacity, and typical peak-period delays upstream of the bottleneck.

To search for such a relationship, an extensive analysis was conducted of the complex expressway queuing phenomenon in and near downtown Boston. Delay data were compared with volumes and capacities at bottleneck points, and a set of rules was developed that operates in the formation of queues and appears to explain the interrelationships among them. Ultimately, a procedure was developed that predicts morning inbound queues and evening outbound queues for downtown area expressways. The procedure is in two parts. In the first part the average maximum peak-period delays are predicted by using a comparison of daily bottleneck volumes with hourly capacities. In the second part queue speeds are derived from hourly v/c ratios of queue sections, and queue lengths are calculated from queue speeds and delays.

In this analysis no attempt has been made to predict outbound morning delays or inbound evening delays, or delays on highways that are not downtown oriented. Also, no consideration has been given to predicting delays caused by heavy stop-and-go traffic with no explicit bottleneck points. Such circumstances were not adequately represented in the data. Further, the procedure as presented does not include any consideration of the variation of queue lengths during the peak period. Patterns of within-peak variations tend to be similar among queues and can be adjusted as circumstances require.

BASIC RELATIONSHIPS GOVERNING MORNING AND EVENING PEAK-PERIOD QUEUING DELAYS

The basic relationships between average maximum peak-period delays and daily traffic related to hourly bottleneck capacity are shown in Figures 1 and 2 for morning and evening peak periods. The relationships shown are manually fitted curves from the Boston speed-run data. Six data points are for the morning peak period, and eight data points are for the evening peak period. The data in the figures indicate that peak-period queues and delays begin to materialize when daily traffic volumes reach the vicinity of 8 to 10 times the hourly capacity at bottleneck points. Evening peak-period delays are greater than morning delays for any given daily volume relative to hourly bottleneck capacity because evening peak-period traffic tends to be heavier than morning peak-period traffic. Similarly, evening delays increase more quickly than morning delays for given increases in daily volumes relative to bottleneck capacities.

In evaluating the curves shown in Figures 1 and 2, it can be seen that their shapes are quite regular and sensible. Also, the relationships between the fitted curves and the data points are close. In no case does the predicted delay from the curves

Figure 1. Daily traffic volume as a multiple of hourly capacity at bottleneck versus average maximum morning peak-period delay.

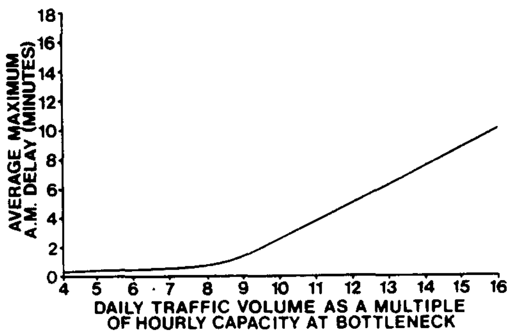
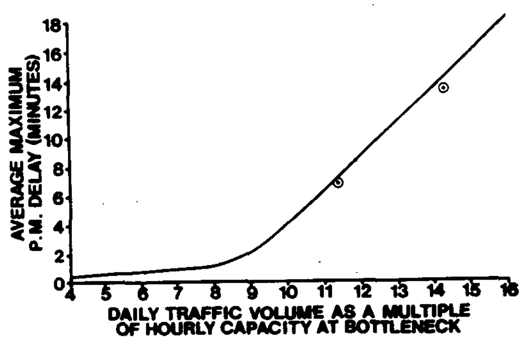


Figure 2. Daily traffic volume as a multiple of hourly capacity at bottleneck versus average maximum evening peak-period delay.



differ from the experienced average delay of the speed runs by more than 1 min. This difference represents less than 15 vehicles per lane in a typical queue.

Of the 14 data points, only 2 are irregular in their derivation. These data points are circled in Figure 2. The circled point with the greater delay is for travel from Logan Airport to downtown Boston through the Sumner Tunnel, an inbound rather than an outbound route. This data point was included in the evening outbound statistics because Logan Airport is a major traffic generator in the Boston core area, and because evening peak-period traffic from Logan Airport can be considered to be outbound, regardless of its direction.

The second irregular data point, which shows less delay, is the data point for I-93 and the Boston Central Artery southbound during the evening peak period. In the derivation of delay data, the segment of this route considered is assumed to have one long queue, even though it has an intermediate section that does not become solidly queued every evening. Because this section is quite short, it was not considered to substantially affect the validity of the data point.

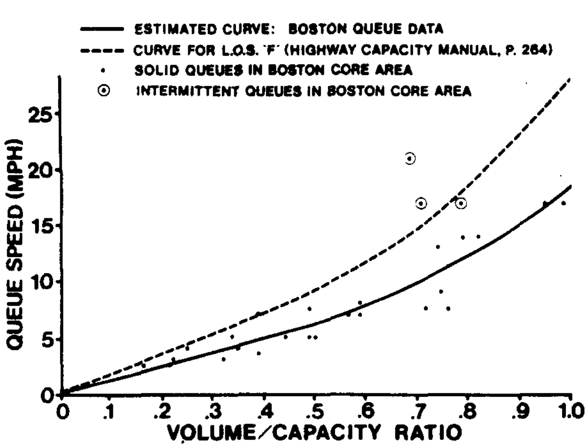
There is one major drawback in the data: there are so few data points; i.e., a total of 14 to fit two curves. Boston has only a few explicit bottleneck points on its express highway system in and near the downtown area; thus data were taken for all of them.

CALCULATING QUEUE LENGTHS FROM DELAYS

To calculate queue lengths from the delay curves shown in Figures 1 and 2, it is necessary to compare queue speeds on highway segments with speeds on the same segments under uncongested conditions. When the delay per unit distance that the difference between queue speed and congested speed implies is known, as well as the total delay in the queue, queue length can be determined by calculating the distance of travel necessary to accumulate the total delay.

Information on queue speeds is shown in Figure 3. The data in this figure relate queue speeds to conventional hourly v/c ratios and also indicate what is, in effect, a level-of-service F curve for queues. The input speed data for the figure were actual speeds from speed runs for all segments of all morning and evening queues on the highway system in the downtown Boston area. As the data in this figure reveal, almost all of the observed speeds are

Figure 3. Relationship between hourly v/c ratio and queue speed: morning and evening queues.



within 1 or 2 mph of what would be predicted by the estimated curve.

Also shown in Figure 3 is the level-of-service F curve from the 1965 Highway Capacity Manual (4, p. 264). It is interesting to note that the estimated curve for queues in the downtown Boston area has speeds less than those of the curve in the Highway Capacity Manual. Although the reason for this is not clear, it appears that the level-of-service F curve in the Highway Capacity Manual was derived from statistics for stop-and-go conditions, with no explicit bottleneck points and no explicit queues. There is support for this notion because the only three Boston data points that are near the curve in the Highway Capacity Manual (the points circled in Figure 3) are those for I-93 and the Central Artery (southbound) in the evening. As noted previously, this section of highway has a segment that is not solidly queued every evening. Thus average speeds are higher. In any case, the fitted Boston curve is appropriate for estimating existing and future queue speeds and lengths.

The following is a hypothetical delay and queue-length calculation. Suppose an expressway has three travel lanes inbound, each of which has a capacity of 2,000 vehicles per hour. Total inbound capacity of the highway is 6,000 vehicles per hour. At one point there is the constriction of a lane being dropped. Beyond this point two lanes remain with a total capacity of 4,000 vehicles per hour. Suppose also that the average weekday traffic inbound at the bottleneck is 50,000 vehicles, or 12.5 times the hourly capacity at that point. Finally, suppose that the highway operates at 55 mph during uncongested periods.

The questions to be answered are as follows: (a) What will be the average maximum morning delay upstream of the bottleneck? and (b) How long will the average maximum morning queue be in which that delay will be experienced? The answer to the first question comes directly from Figure 1. With an average daily traffic volume 12.5 times the hourly bottleneck capacity, the average maximum morning delay will be about 5.7 min.

The calculation of queue length is a little more complicated. In the queue area the v/c ratio is 0.67 (4,000 vehicles per hour traveling on three lanes that could handle 6,000 vehicles if it were not for the bottleneck). This corresponds with a queue speed of 9.5 mph (as shown in Figure 3). At this speed it takes 6.316 min to travel a mile (1/9.5 x 60). In uncongested conditions it takes 1.091 min to travel a mile (1/55 x 60). Thus a vehicle traveling 1 mile in the queue will incur 5.225 min of delay (6.316 - 1.091). Because the total delay in the queue was calculated to be 5.7 min, the average maximum queue length will be 1.091 miles (5.700/5.225), or 5,760 ft.

Clearly, the procedure for calculating queue delays and lengths is quite simple. A little more work is required if there are on-ramps and off-ramps or variations in capacity within the queued section. In such cases v/c ratios and speeds must be calculated separately by segments of the highway section (moving upstream from the bottleneck) and delays added up by segment until the total queue delay is achieved.

One final note is appropriate concerning the application of the model. In determining hourly bottleneck capacity for the determination of delay, the actual peak-period capacity of the bottleneck should be used, including vehicle mix, weaves, and geometrics. Alternatively, counts may be used. However, for determining queue length, capacities should be considered to be approximately 2,000 vehicles per lane per hour because vehicle mix, weaves, and geometrics become largely irrelevant when vehicles are waiting in line.

DETERMINING LOCATIONS OF BOTTLENECK POINTS

Before the procedure for predicting queue delays and lengths can be carried out, the exact locations of the bottleneck points relevant to the given queues must be identified. This task can be more complex than the application of the procedure. During the course of the development of the basic model in this study, a number of methods of selecting bottleneck points were tested in an attempt to develop consistent relationships between peak-period delays and daily volumes relative to hourly capacities at bottlenecks. Ultimately, the best relationships were established by using data that resulted from defining and selecting bottlenecks according to the rules set forth in the following sections. In performing the queuing analysis, the same rules should be used for determining the locations of the bottleneck points.

Simple Queue

When a queue forms on an express highway with heavy traffic, the location of the queue will be upstream of the point with the highest daily volume relative to capacity, which point is the bottleneck point. Such a point may be at a constriction, such as a bridge or a lane drop, or at a merge or diverge of a major flow of traffic.

A simple queue is shown in Figure 4, which shows a bottleneck point and the queue upstream of it. Also shown in Figure 4 are areas upstream of the queue and downstream from the bottleneck where free flows of traffic are maintained.

Two Queues in Succession

In some circumstances a highway may have two bottleneck points in succession. Such a circumstance is shown in Figure 5, which depicts an upstream bottleneck A and a downstream bottleneck B. Here queues and delays depend primarily on which is the greater bottleneck (higher daily volume relative to capacity). If bottleneck A is the greater bottleneck, a queue will develop upstream of bottleneck A but no queue will develop at bottleneck B, because bottleneck A will meter traffic to bottleneck B, so that no queue can develop there. Similarly, if bottleneck B is the greater bottleneck, a queue will form there but none will form at bottleneck A, because traffic will meter itself in anticipation of the queue downstream.

The only circumstance in which queues will develop at both locations will be where the bottlenecks are relatively far apart and substantial volumes of traffic enter and leave the highway between them. In this circumstance traffic at the two bottlenecks is mostly composed of different vehicles, and delays at the two bottlenecks should be predicted separately by using the volume relative to capacity at each.

Split at Head of Queue

Where a highway divides at the head of a queue, three potential bottleneck points may be considered for predicting queue length and delay. This circumstance is shown in Figure 6, which shows bottleneck A before the diverge point and bottlenecks B and C to the left and right after the diverge point. Hypothetical queues predicted from the bottlenecks are shown in the figure, where each queue is based on the daily volume relative to capacity of the given bottleneck.

In the case shown, bottleneck A would generate the smallest queues and delays, bottleneck B would generate the largest queues and delays, and bottleneck C would generate queues and delays of intermediate length and duration. Because it produces the largest queues and delays, bottleneck B should be used for prediction. Potential queues formed by bottlenecks A and C would simply be submerged in the bottleneck B queue.

Split Near Head of Queue

A somewhat similar circumstance to that of a split at the head of a queue is that of a major diverge point near the head of a queue, with the diverging traffic entering a bottleneck itself shortly after the diverge point. This circumstance is shown in Figure 7, which again shows the potential bottlenecks for use in queue and delay prediction. As shown in the figure, bottleneck A is on the main line just before the diverge point, bottleneck B is

Figure 4. Simple queue.

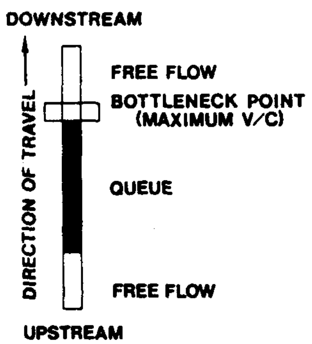


Figure 5. Two queues in succession.

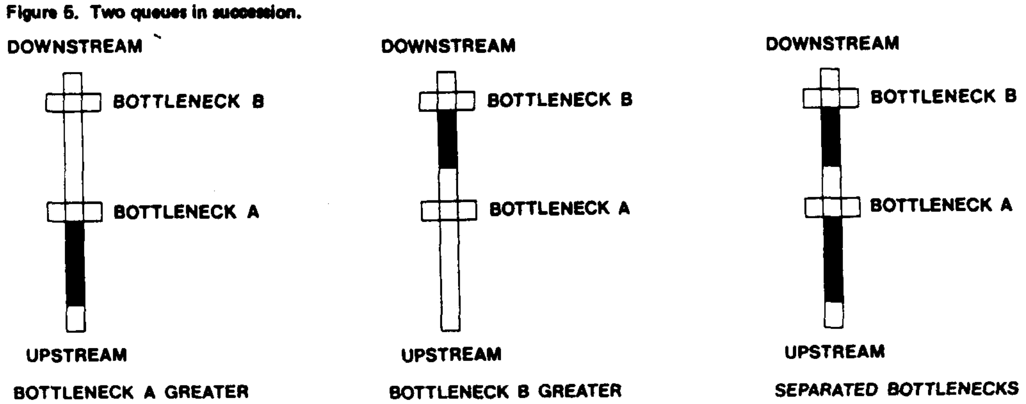


Figure 6. Split at head of queue.

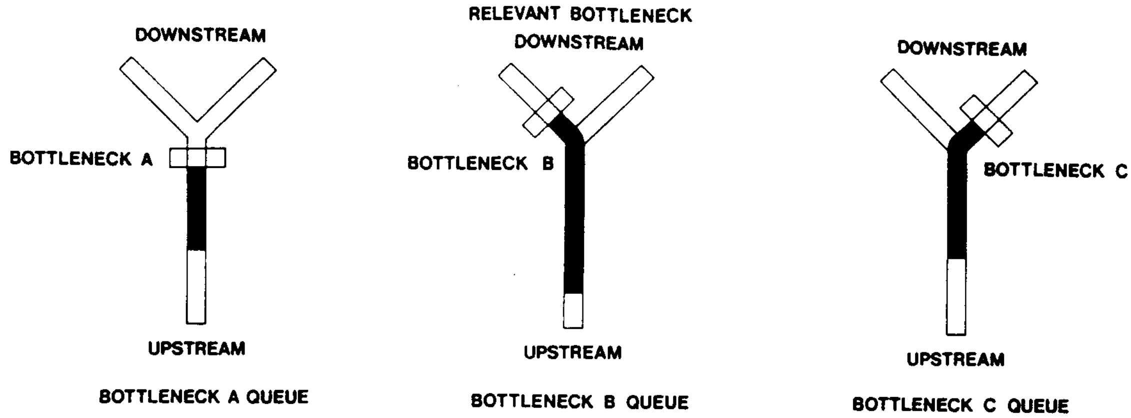
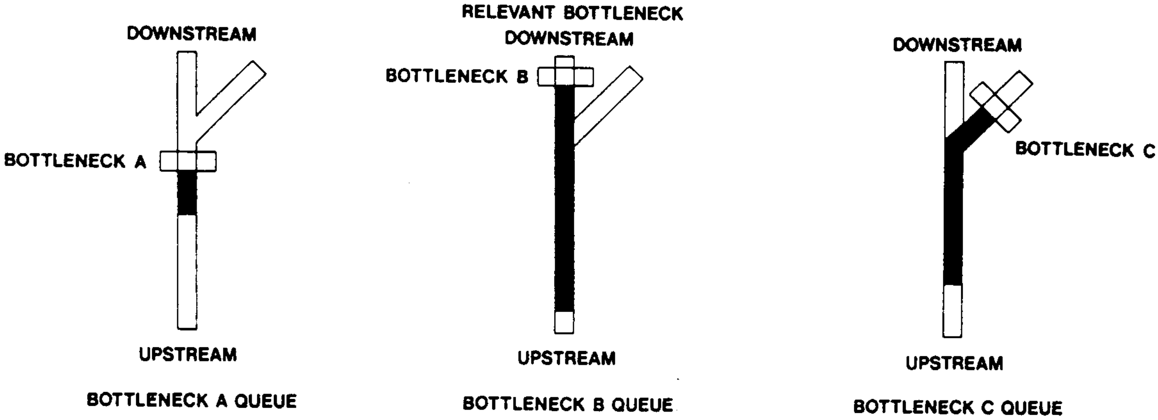


Figure 7. Split near head of queue.



on the main line downstream of the diverge point, and bottleneck C is on the route used by the diverging traffic.

Also shown in Figure 7 are hypothetical queue lengths implied by the three bottlenecks individually. Bottleneck A would generate the shortest queue, bottleneck B the longest queue, and bottleneck C a queue of intermediate length. In this case it is the bottleneck that produces the queue that stretches to the point farthest upstream that should be used for prediction. In the example the relevant queue is from bottleneck B. As before, potential queues from the other bottlenecks would simply be submerged in the bottleneck B queue.

Two Queues Joining at Bottleneck

Yet another circumstance is that of two major highway flows joining and encountering a bottleneck at the merge point. Such a situation is shown in Figure 8. In this case the question is whether the daily volume relative to capacity of the joined flow at bottleneck A should be used to predict equivalent queues and delays for the two merging flows of traffic, or whether the two flows at bottlenecks B and C should be considered separately. In this circumstance the flows should be considered separately. The daily volumes to be used are those at bottlenecks B and C. The capacities to be used, however, are not those at bottlenecks B and C, but the fractions of the capacity at bottleneck A available through channelization to the traffic flows from bottlenecks B and C.

Queue Joining Queue Near Bottleneck

A final circumstance is that of two major highway flows joining upstream of a bottleneck on one of them. This circumstance is shown in Figure 9 by three hypothetical cases. In all three cases a main line queue is generated from bottleneck A. Three different possible queues are illustrated from bottleneck B, which is upstream from bottleneck A and applies to the merging traffic where it enters the main flow.

In case 1 bottleneck B creates a small queue for the entering traffic. This is the circumstance in which the entering traffic is a relatively small fraction of the traffic on the main line and can merge into the main flow without difficulty. Presumably, the relationship between daily traffic and potential merge capacity at bottleneck B would create only a minor queue. In case 2 a queue is formed upstream of bottleneck B equal in length to that on the main line. Here both flows are determined effectively by bottleneck A, and there is really one queue with two equivalent tails. In case 3 bottleneck B creates a queue longer than that of the main line upstream of the merge point. Here the queues are probably separate in cause and operation.

Which of these three cases applies in any given situation is difficult to determine because the general circumstance is, in part, equivalent to two queues in succession. The following guidelines, however, may help determine which case applies. If the traffic flows through bottlenecks A and B are largely composed of different vehicles, the queuing

Figure 8. Two queues joining at bottleneck.

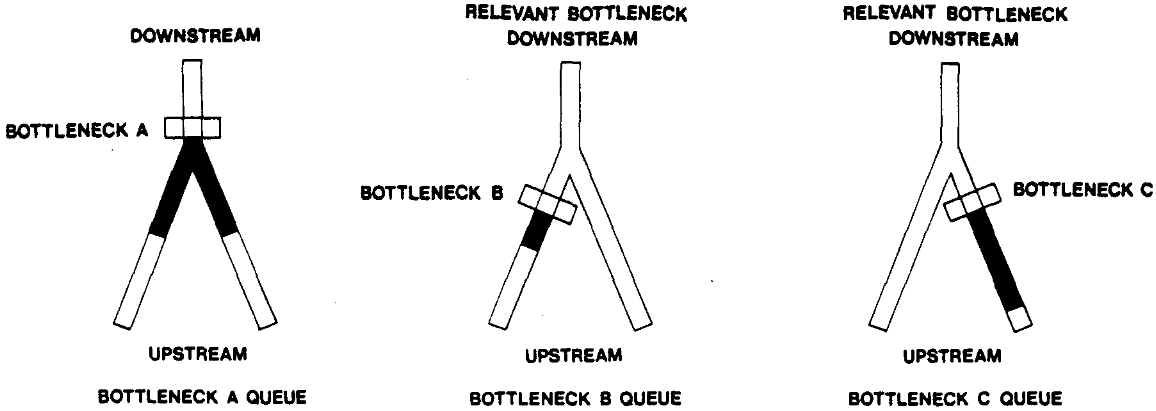
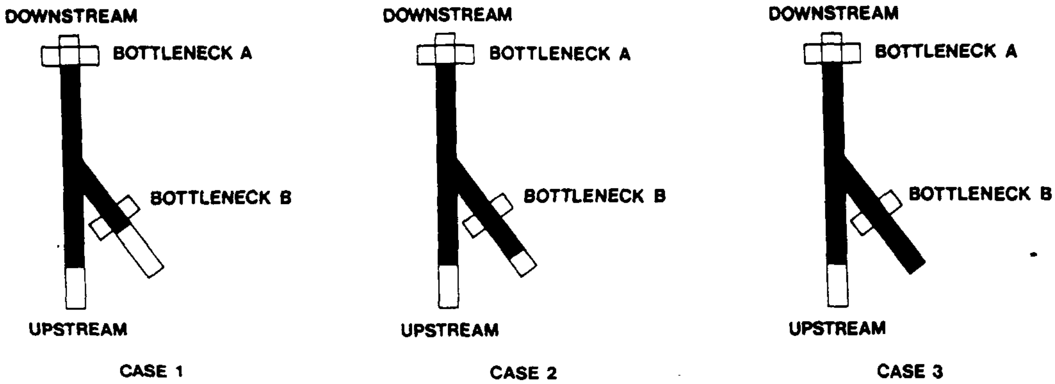


Figure 9. Queue joining queue near bottleneck.



prediction can probably be accomplished separately for the two bottlenecks, as in cases 1 and 3. If most vehicles from both routes are destined for bottleneck A, however, the queuing should probably be predicted by assuming one queue with equivalent tails from bottleneck A, as in case 2.

Summary

The rules just discussed for bottlenecks would indicate that

1. The relationship between queue delays and daily volumes compared with hourly capacities pertains only to unbroken stretches of congested traffic;
2. The ratio to apply is that of the point with the highest daily volume compared with hourly capacity, the point of which will be at the head of the queue; and
3. The delay to apply is that to the most distant end of the queue.

There are qualifications, and the rules need to be applied with careful attention paid to actual circumstances. But with adequate consideration of geometrics and traffic flows, following the rules previously described yields clear relationships between queue delays and daily volumes relative to hourly capacities at bottleneck points.

STRENGTHS AND LIMITATIONS OF THE PROCEDURE

The procedure described in this paper has a number of strengths. Primary among these are its ability to use traffic assignment data as input, its simplicity, and its generally reasonable and consistent results. The procedure appears to solve successfully the extremely difficult problem of predicting vehicle arrival and service rates. At the same time, however, the relationships developed for the procedure are based on data collected for only a few queues. Only six data points for morning inbound queues and eight data points for evening outbound queues could be derived from observations of traffic in the Boston core area. Further, some of these data points are subject to question.

An additional limitation of the procedure is its narrow range of applicability: morning inbound and evening outbound queues in the cores of urban areas about the same size as Boston. No attempt was made to calibrate procedures for queues in reverse flows or in nondirectional flows (such as on circumferential routes), for temporary queues where construction projects are under way, or for queues in urban areas of different sizes. Nevertheless, the basic approach appears to be applicable to these circumstances, and analogous procedures could be derived for them with further data collection and analysis.

Certainly, addressing problems of queuing is central to improving the operations of many urban expressway systems. To the extent that the basic approach can be applied to other cities and circumstances, the prediction of queuing from relationships between daily traffic and bottleneck capacities may provide a powerful analysis tool. It could enhance considerably the analyst's ability to predict and evaluate the potential impacts of urban expressway projects.

ACKNOWLEDGMENT

The preparation of this paper was financed through a contract with the Massachusetts Department of Public Works and with the cooperation of the FHWA, U.S. Department of Transportation.

REFERENCES

1. M. Wohl. Notes on Transient Queuing Behavior, Capacity Constraint Functions, and Their Relationship to Travel Forecasting. In The Regional Science Association Papers (M.D. Thomas, ed.), Regional Science Association, The Wharton School, Philadelphia, 1968, pp. 191-202.
2. D.R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
3. W.A. Stock, R.C. Blankenhorn, and A.D. May. The "FREQ3" Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Rept. 73-1, 1973.
4. Highway Capacity Manual. HRB, Special Rept. 87, 1965, 411 pp.

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