TRANSPORTATION RESEARCH RECORD 1037

Transportation Demand Analysis and Issues in Travel Behavior

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1985

Transportation Research Record 1037
Price $14.00
Editor: Naomi Kassabian
Compositor: Harlow A. Bickford
Layout: Marion L. Ross

modes
1 highway transportation
2 public transit

subject areas
12 planning
13 forecasting

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

Printed in the United States of America

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

Transportation demand analysis and issues in travel behavior.

(Transportation research record ; 1037)
1. Transportation–Forecasting–Congresses.
2. Transportation–Planning–Congresses. I. National Research Council (U.S.). Transportation Research Board.
TE7.H5 no. 1037 380.5 s 86-5251
[HE11] [380.5'93]
ISBN 0-309-03953-3 ISSN 0361-1981

Sponsorship of Transportation Research Record 1037

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
William A. Bulley, H. W. Lochner, Inc., chairman

Transportation Data, Economics and Forecasting Section
Joseph L. Schofer, Northwestern University, 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, Lawrence V. Hammel, Susan Hanson, Joel L. Horowitz, Stephen M. Howe, Peter M. Jones, Terry Kraft, David L. Kurth, Steven Richard Lerman, Eugene J. Lessieu, Hani S. Mahmassani, 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
Mary Lynn Tischer, Federal Highway Administration, chairman
Julian M. Benjamin, James C. Echols, David T. Hartgen, David A. Hensher, Joel L. Horowitz, Dennis C. Judycki, Ryuichi Kitamura, Lidia P. Kostyniuk, Steven Richard Lerman, Arnim H. Meyburg, Robert E. Paaswell, Richard H. Pratt, Sandra Rosenbloom, Earl R. Ruiter, Peter R. Stopher, Antti Talvitie, Carina Van Knippenberg, Edward Weiner

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1984.

1037

Contents

Addresses of Authors

Ben-Akiva, Moshe E., Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139
Bertram, D., Department of Civil Engineering, Monash University, Wellington Road, Clayton, Melbourne, Victoria 3168, Australia
Bovy, Piet H. L., Institute for Town Planning Research, Delft University of Technology, P.O. Box 5043, 2600 GA Delft, Netherlands
Bradley, Mark A., Cambridge Systematics Europe, Laan van Meerdervoort 32, 2517 AL The Hague, Netherlands
Chang, Gang-Len, Department of Civil Engineering, University of Texas, Austin, Tex. 78712
Davidson, William A., Barton-Aschman Associates, Inc., 820 Davis Street, Evanston, Ill. 60201
Deaton, Angus, Research Program in Development Studies, Woodrow Wilson School, Princeton University, Princeton, N.J. 08540
Gunn, Hugh F., Cambridge Systematics Europe, Laan van Meerdervoort 32, 2517 AL The Hague, Netherlands
Horowitz, Alan J., Center for Urban Transportation Studies, University of Wisconsin-Milwaukee, P.O. Box 784, Milwaukee, Wisc. 53201
Horowitz, Joel L., Departments of Geography and Economics, University of Iowa, Iowa City, Iowa 52242
Hsu, Poh Ser, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139
Jager, Joke M., Department of Civil Engineering, Delft University of Technology, H. Gorterhof 154, 2624 XL Delft, Netherlands
Koppelman, Frank S., Department of Civil Engineering and Transportation Center, Northwestern University, Evanston, Ill. 60201
Kuah, Geok-Koon, Department of Civil Engineering, University of Maryland, College Park, Md. 20742
Macke, Peter P., Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139
Mahmassani, Hani S., Department of Civil Engineering, University of Texas, Austin, Tex. 78712
Metzger, David N., Center for Urban Transportation Studies, University of Wisconsin-Milwaukee, P.O. Box 784, Milwaukee, Wis. 53201
Pennock, Norma T., TRANSPO Group, 23 148th Avenue, S.E., Bellevue, Wash. 98007
Rutherford, G. Scott, Washington State Transportation Center, 135 More Hall, FX-10, University of Washington, Seattle, Wash. 98195
Scheltes, Wim H., Netherlands Institute of Transport (NVI), Polakweg 13, 2288 GG Rijswijk, Netherlands; formerly with Delft University of Technology
Schoendorfer, David, Department of Civil Engineering, University of Pennsylvania, Philadelphia, Pa. 19104
Spivack, Gary S., Southern California Rapid Transit District, 425 S. Main Street, Los Angeles, Calif. 90017
Stopher, Peter R., Schimpeler-Corradino Associates, 425 South Main Street, Los Angeles, Calif. 90274
Supernak, Janusz, Department of Civil Engineering, San Diego State University, San Diego, Calif. 92182
Tamny, Laiky K., Beverly Enterprises, 873 S. Fair Oaks Avenue, P.O. Box 90130, Pasadena, Calif. 91109
Wilmot, Chester G., National Institute for Transport and Road Research, CSIR, P.O. Box 395, Pretoria 0001, South Africa
Young, W., Department of Civil Engineering, Monash University, Wellington Road, Clayton, Melbourne, Victoria 3168, Australia

Alternative Methods to Estimate Route-Level Trip Tables and Expand On-Board Surveys

MOSHE E. BEN-AKIVA, PETER P. MACKE, and POH SER HSU

ABSTRACT

Several estimation techniques of route-level trip tables are reviewed and tested. Current industry practice of simple expansion of an on-board survey by total boardings is compared with expansion by the iterative proportional fitting (IPF), constrained generalized least-squares, and constrained maximum-likelihood methods. An intervening-opportunity model, which does not use the on-board survey, is also tested. The more complex methods achieve better accuracy and reduced bias by combining the survey data with ride-check data. An empirical case study demonstrates that under the assumption of error-free ride-check data the IPF technique is preferred because of its computational ease without loss of accuracy. The IPF method should enable transit operators to obtain much more accurate trip tables and more reliable on-board survey results for a small additional computational cost.

Ridership data at the level of a single transit route are required for setting headways, for evaluating alternative operating strategies (such as express bus and short turning), and for forecasting revenues. [See, for example, studies by Furth and Wilson (1), Ceder (2), and the U.S. Department of Transportation (3).] For patronage forecasting applications, route-level trip tables serve as the base for pivot-point techniques that use elasticities to predict the effects of changes in headways, travel times, and fares [see, for example, the study by Nickesen et al. (4)].

In this paper the issue is addressed of obtaining cost-effective estimates of route-level origin-destination (OD) trip tables from two data sources--an on-board survey and ride-check data. This could be easily extended to include data from other sources.

There are several reasons for trying to combine ride-check data efficiently with an on-board survey. On-board survey data have several severe limitations: they are expensive to collect, subject to response bias problems, and require a long processing time. This implies that on-board survey data are likely to have large sampling errors because of the small sample sizes and significant biases because of response errors and being out of date. However, they are the only data that contain disaggregate OD information.

Ride-check data, on the other hand, are comparatively less costly to collect and consequently are collected at more frequent intervals. Processing time, particularly in the advent of hand-held electronic data acquisitors, is shorter. They are also not subject to response bias problems. However, they consist of aggregate passenger counts and by themselves could not yield OD information without resort to restrictive assumptions. Thus, the objective of this research is to utilize these more accurate, unbiased, aggregate data to improve the accuracy and reduce the biases of on-board surveys.

Alternative estimation methods are presented of route-level OD trip tables that expand a small sample on-board survey by using ride-check data. This results in much greater accuracy of the available route-level OD information. The aternative methods are illustrated with actual data from the Boston area. The estimation results are compared with those obtained by an intervening-opportunity method based on ride-check information alone and with a matrix obtained by expansion with a single factor of the survey OD matrix to the total boardings counted in the ride checks.

PROBLEM FORMULATION AND NOTATION

Consider a bus route for which the available data consist of a small-sample on-board survey with information on the passengers' boarding and alighting stops and on and off counts from a ride check.

The population of interest is all the passenger trips taking place on the route in one direction and during a fixed time interval. For example, all trips on Route 77 in the outbound direction during the evening peak period would constitute a population of interest. Samples are taken at different points in time (days, months). Therefore, the investigated trip table represents average conditions during the survey period.

For the OD survey a sample of bus runs is drawn and on these runs the survey questionnaire is usually distributed to all passengers.

The ride checks usually consist of counts on all the bus runs spread over the survey period. It is assumed that the counts for a given bus run are stable from day to day during the survey period. The methods developed here treat the ride-check information as error free as a first approximation. In a later stage of this research this restriction will be released and errors in the ride-check data will be considered by using the approach presented by Ben-Akiva et al. (5).

Let i and j denote an origin and a destination stop or group of stops, respectively, where i = 1, ..., I and j = 1, ..., I. Denote the number of passengers boarding transit vehicles at i = 1, ..., I and alighting at j = i, ..., I during a given time period by t_{ij}.

The number of passengers boarding and alighting at every stop is known from the ride check. Denote the number of passengers boarding at i by $t_{i.}$ and the number of passengers alighting at j by $t_{.j}$. Thus,

$$t_{i.} = \sum_{j=i}^{I} t_{ij} \qquad i = 1, \ldots, I \tag{1}$$

$$t_{.j} = \sum_{i=1}^{j} t_{ij} \qquad j = 1, \ldots, I \tag{2}$$

$$t_{..} = \sum_{i=1}^{I} t_{i.} = \sum_{j=1}^{I} t_{.j} \tag{3}$$

where $t_{..}$ denotes the number of total boardings.

Denote the number of observations in the on-board survey by $t^o_{..}$ and the observed cell frequencies by t^o_{ij}, i = 1, ..., I, j = i, ..., I, such that

$$\sum_{i=1}^{I} \sum_{j=i}^{I} t^o_{ij} = t^o_{..}$$

ESTIMATION METHODS

An on-board survey is usually expanded by a single factor, $f = t_{..}/t^o_{..}$, for all the observations during the given time period. In the following section alternative methods are presented that use the ride-check boarding and alighting counts, $t_{i.}$, i = 1, ..., I and $t_{.j}$, j = 1, ..., I, to expand the trip table observed in the on-board survey, t^o_{ij}, i = 1, ..., I, j = i, ..., I, and to obtain combined estimates of the trip table, $\hat{t}_{ij}$, i = 1, ..., I, j = i, ..., I.

The alternative methods investigated in this paper are iterative proportional fitting (IPF), constrained generalized least squares (CGLS), constrained maximum-likelihood estimation (CMLE), and an intervening-opportunity approach. The last method does not use data from an on-board survey.

IPF Method

The IPF method has been widely used in transportation and other fields. It has been referred to as the biproportional method (6), the Furness or Fratar iterative procedure (7), the Kruithof algorithm (8), or Bregman's balancing method (9). In general, IPF estimates for a two-dimensional matrix are proportional to base matrix entries with a constant of proportionality for each row and each column. These multiplicative expansion factors modify the base entries to be consistent with the known row and column totals for the matrix.

The IPF estimator for OD pair (i,j) is given by

$$\hat{t}_{ij} = a_i b_j t^o_{ij} \tag{4}$$

where the proportionality constants or balancing factors a_i and b_j are determined such that the estimated values satisfy the row and column constraints in Equations 1 and 2, as follows:

$$\sum_{j=i}^{I} \hat{t}_{ij} = t_{i.} \qquad i = 1, \ldots, I \tag{5}$$

$$\sum_{i=1}^{j} \hat{t}_{ij} = t_{.j} \qquad j = 1, \ldots, I \tag{6}$$

The IPF estimator in Equation 4 can also be defined as the outcome of the following constrained optimization problem: Minimize

$$\sum_{i=1}^{I} \sum_{j=i}^{I} t_{ij}[\log(t_{ij}/t^o_{ij}) - 1] \tag{7}$$

subject to Equations 1 and 2, where the objective function may be interpreted as a measure of information or as a measure of the discrepancy between the matrix $\{t_{ij}\}$ and the initial matrix $\{t^o_{ij}\}$. Note that for $t^o_{ij} = 0$ the corresponding $\hat{t}_{ij}$ value is zero as well, and the cell is omitted from the objective function. A solution exists if

$$\sum_{i=1}^{I} t_{i.} = \sum_{j=1}^{I} t_{.j}$$

and it is given by the first-order conditions of the Lagrangian:

$$\partial L/\partial t_{ij} = \log(t_{ij}/t^o_{ij}) - \lambda_i - \mu_j = 0 \qquad i = 1, \ldots, I \;\; j = i, \ldots, I \tag{8}$$

where λ_i, i = 1, ..., I and μ_j, j = 1, ..., I are the Lagrangian multipliers of the constraints 1 and 2, respectively. The solution of Equation 8 can be expressed by Equation 4 where

$$a_i = \exp(\lambda_i) \qquad i = 1, \ldots, I \tag{9}$$

$$b_j = \exp(\mu_j) \qquad j = 1, \ldots, I \tag{10}$$

Thus, the balancing factors may be interpreted as exponential transformations of the Lagrangian multipliers of the boarding and alighting constraints.

The IPF estimator in Equation 4 is also defined by the property of constant cross-product ratios, as follows:

$$(t^o_{ij} \cdot t^o_{uv})/(t^o_{uj} \cdot t^o_{iv}) = (\hat{t}_{ij} \cdot \hat{t}_{uv})/(\hat{t}_{uj} \cdot \hat{t}_{iv}) \tag{11}$$

for any OD pairs (i,j) and (u,v) with $t^o_{uj} > 0$ and $t^o_{iv} > 0$.

The variances of the IPF estimates may be calculated by using a linear approximation to obtain

$$\mathrm{Var}(\hat{t}_{ij}) \cong \sum_u \sum_v \sum_k \sum_\ell (\partial \hat{t}_{ij}/\partial t^o_{uv})|_{\hat{t}} (\partial \hat{t}_{ij}/\partial t^o_{k\ell})|_{\hat{t}} \operatorname{cov}(t^o_{uv}, t^o_{k\ell}) \tag{12}$$

The covariances of the observed cell frequencies are unknown. On the assumption that the t^o_{ij}'s are the outcomes of independent random draws from large populations, it is possible to approximate their sampling distributions by independent Poisson variates [see, e.g., the study by Drake (10)] to obtain

$$\mathrm{Var}(\hat{t}_{ij}) \cong \sum_u \sum_v [(\partial t_{ij}/\partial t^o_{uv})|_{\hat{t}}]^2 t^o_{uv} \tag{13}$$

CGLS Method

The CGLS estimation method has been presented by Theil (11) and has recently been used to estimate OD tables by McNeil (12) and by Hendrickson and McNeil (13,14). This method is based on the assumption that the simple expansion by $f = t_{..}/t^o_{..}$ of the entries in the base matrix $\{t^o_{ij}\}$ provides unbiased estimates of the true values $\{t_{ij}\}$. Thus, a base value $(f \cdot t^o_{ij})$ can be expressed as equal to the unknown true value (t_{ij}) plus an error term denoted by ϵ_{ij}, as follows:

$$f \cdot t^o_{ij} = t_{ij} + \epsilon_{ij} \qquad i = 1, \ldots, I \;\; j = i, \ldots, I \tag{14}$$

where $E[\epsilon_{ij}] = 0$.

With matrix notation, the CGLS estimator can be expressed as follows: Minimize

$$(f \cdot t^o - t)' \, V^{-1} (f \cdot t^o - t) \tag{15}$$

subject to

$$Rt = r \tag{16}$$

where

- t^o, t = K x 1 vectors of observed values and unknown parameters, respectively;
- K = number of cells in the base matrix;
- V = K x K variance-covariance matrix of t^o;
- r = C x 1 vector of C linearly independent row and column constraints; and
- R = C x K constraint incidence matrix whose elements are either 0 or 1.

The CGLS estimator is given by

$$\hat{t} = f \cdot t^o + VR'(RVR')^{-1}(r - Rt^o) \tag{17}$$

and the variance-covariance matrix of the estimated values is

$$\mathrm{cov}(\hat{t}) = \sigma^2 \; [V - VR'(RVR')^{-1} RV] \tag{18}$$

with the following estimator for σ^2:

$$s^2 = \left\{[\mathrm{tr}(V)/K]\,(f \cdot t^o - \hat{t})'V^{-1}(f \cdot t^o - \hat{t})\right\}/C \tag{19}$$

where tr(V) is the trace of matrix V. Note that the normalization with tr(V)/K that appears in this formulation is different from the derivations by Theil (11) and McNeil (12) because they assume a normalized covariance matrix V whose trace equals the number of cell entries to be estimated.

This least-squares estimator is not unbiased in situations with small cell values. The bias arises from the fact that the correct model is given by

$$f \cdot t^o_{ij} = \begin{cases} t_{ij} + \epsilon_{ij} & \text{if } \epsilon_{ij} \geq -t_{ij} \\ 0 & \text{if } \epsilon_{ij} < -t_{ij} \end{cases} \tag{20}$$

This implies that

$$E[\epsilon_{ij} \mid \epsilon_{ij} \geq -t_{ij}] \geq 0 \tag{21}$$

where the equality is likely to hold only for large t_{ij}. [See, for example, studies by Judge et al. (15) and by Heckman (16).]

CMLE Method

A maximum-likelihood estimator requires specific assumptions about the sampling distributions of the different data sources. It is often assumed that the base OD matrix is generated by a multinomial sampling process. Other distributions that may be considered are the multivariate hypergeometric, which tends to the multinomial as the population from which the sample is drawn tends to infinity, and independent Poisson distributions of the observed cell values. The sampling distributions of the passenger counts could be approximated by the Poisson distribution. However, in this paper, the ride checks are considered to be observed without errors. On the assumption that all the available counts are compatible, the technique becomes a maximum-likelihood estimation subject to constraints. This method was previously applied to OD matrix estimation by Landau et al. (17) and by Geva et al. (18).

Under the Poisson assumption, the maximum-likelihood estimator is found by solving the following: Maximize

$$L = \prod_{i=1}^{I} \prod_{j=i}^{I} (t_{ij}^{\,t^o_{ij}}/t^o_{ij}!) \exp(-t_{ij}) \tag{22}$$

subject to Equations 1 and 2. The Lagrangian for the logarithm of the likelihood function is as follows: Maximize

$$L^o = \sum_{i=1}^{I} \sum_{j=i}^{I} [-t_{ij} + t^o_{ij}\,\ell n(t_{ij}) - \ell n(t^o_{ij}!)] - a_i\left(\sum_{j=1}^{I} t_{ij} - t_{i.}\right) - b_j\left(\sum_{i=1}^{I} t_{ij} - t_{.j}\right) \tag{23}$$

The first-order conditions are

$$\partial L^o/\partial t_{ij} = -1 + (t^o_{ij}/t_{ij}) - a_i - b_j = 0 \tag{24}$$

and the constrained maximum-likelihood estimator is

$$\hat{t}_{ij} = t^o_{ij}/(1 + a_i + b_j) \qquad i = 1, \ldots, I \;\; j = i, \ldots, I \tag{25}$$

where a_i, i = 1, ..., I and b_j, j = 1, ..., I are the balancing factors as well as the Lagrangian multipliers of the constraints 1 and 2, respectively.

The same result is obtained for the multinomial and, by using an approximation, for the multivariate hypergeometric distributional assumptions:

$$\hat{t}_{ij} = t^o_{ij}/(a_i + b_j) \qquad i = 1, \ldots, I \;\; j = i, \ldots, I \tag{26}$$

Another way of solving the constrained maximum-likelihood problem is to substitute the constraints into the objective function. For the multinomial distribution the kernel of the unconstrained log-likelihood function with a smaller set of unknown parameters can be written as follows: Maximize

$$\sum_{j=2}^{I} \sum_{j=i}^{I-1} t^o_{ij}\,\ell n\, t^o_{ij} + \sum_{j=1}^{I-1} t^o_{1j}\ell n\, t_{1j} + \sum_{i=1}^{I} t^o_{iI}\;\ell n\, t_{iI} \tag{27}$$

Use the constraints 1 and 2 to solve for

$$t_{iI} = t_{i.} - \sum_{j=i}^{I-1} t_{ij} \qquad i = 1, \ldots, I \tag{28}$$

$$t_{1j} = t_{.j} - \sum_{i=2}^{j} t_{ij} \qquad j = 1, \ldots, I \tag{29}$$

and substitute into Equation 27. The first-order conditions of Equation 27 yield the following expression for the maximum-likelihood estimator:

$$\hat{t}_{ij} = t^o_{ij}/[(t^o_{iI}/\hat{t}_{iI}) + (t^o_{1j}/\hat{t}_{1j})] \qquad i = 2, \ldots, I \;\; j = i, \ldots, I-1 \tag{30}$$

Note that Equations 30 and 26 are the same by substituting

$$a_i = t^o_{iI}/\hat{t}_{iI} \qquad b_j = t^o_{ij}/\hat{t}_{1j}$$

For the unconstrained maximum-likelihood formulation estimates of the asymptotic covariance matrix of the t_{ij}'s are given by minus the inverse of the matrix of second derivatives evaluated at $\hat{t}$. Differentiate the first-order conditions with respect to t_{uv}, u = 2, ..., I, v = u, ..., I - 1 to obtain the Hessian of the log-likelihood function as follows:

$$\partial^2 L/\partial t_{uv}\partial t_{ij} = \begin{cases} -(t^o_{ij}/t^2_{ij}) - (t^o_{1j}/t^2_{1j}) - (t^o_{iI}/t^2_{iI}) & \text{for } u = i \text{ and } v = j \\ -(t^o_{ij}/t^2_{1j}) \quad j = i, \ldots, I-1 & \text{for } u \neq i \text{ and } v = j \\ -(t^o_{iI}/t^2_{iI}) \quad i = 2, \ldots, I & \text{for } u = i \text{ and } v \neq j \\ 0 & \text{for } u \neq i \text{ and } v \neq j \end{cases} \tag{31}$$

Minus the inverse of this Hessian evaluation at $\hat{t}$

provides an estimate of the variance-covariance matrix of $\hat{t}$.

With both the MLE and the linear regression approaches it is possible to extend the analysis to include stochastic ride-check data. The MLE approach can also be extended to test for more complex patterns of response biases, particularly in the on-board survey (5).

Problem of Nonstructural Zeros

The IPF and CMLE estimators retain zero values in the expanded matrix for cells that had zero observations in the survey. The problem lies in the structure of those estimators; it is avoided with the CGLS method. (Note, however, that the CGLS estimates may have negative values.)

The problem of nonstructural zeros can be overcome by the grouping of stops in order to reduce the dimensionality of the trip table, by choosing an appropriate sampling strategy, by substituting arbitrary small numbers in the zero cells of the survey matrix, or by a Bayesian technique suggested by Bishop et al. (19, pp. 401-433) and by Kirby and Leese (20).

The last approach (shown in the following paragraph) is a method of smoothing the observations $[t^o_{ij}]$ over the cells of the matrix. Let the probability of observing the sample matrix entries $[t^o_{ij}]$ be expressed by the multinomial distribution and denote for each cell of the matrix the probability that a trip occurs from i to j by $p_{ij}(p_{ij} \geq 0$, for all i,j) where

$$\sum_{i=1}^{I} \sum_{j=i}^{I} p_{ij} = 1$$

The smoothing method estimates nonzero values for cells that are empty by chance but have a nonzero probability that a trip is made. The method also reduces the values of nonempty cells in such a manner that the cell entries continue to add up to $t^o_{..}$.

The Bishop et al. (19) estimator for the smoothed cell values is obtained with the following procedural steps:

1. Select prior probabilities $[p^o_{ij}]$, which may be based on external information or the survey data;
2. Compute the posterior probabilities:

$$\hat{p}_{ij} = (t^o_{ij} + \hat{k}p^o_{ij})/(t^o_{..} + \hat{k}) \quad i = 1, \ldots, I, \; j = i, \ldots, I \tag{32}$$

where the weighting factor is given by

$$\hat{k} = [(t^o_{..})^2 - \sum_{i,j} (t^o_{ij})^2] / \sum_{i,j} (t^o_{ij} - t^o_{..}\, p^o_{ij})^2 \tag{33}$$

3. Compute the "smoothed" cell estimates:

$$\hat{t}^o_{ij} = t^o_{..}\, \hat{p}_{ij} = [t^o_{..}/(t^o_{..} + \hat{k})]\, (t^o_{ij} + \hat{k} \cdot p^o_{ij}) \qquad i = 1, \ldots, I, \; j = i, \ldots, I \tag{34}$$

The prior probabilities p^o_{ij} can be simply assumed to be equal. Thus, $p^o_{ij} = 1/K$ where K is the number of cells. Alternatively, the prior probabilities may be based on the ride-check data in the form

$$p^o_{ij} = (t_{i.} \cdot t_{.j})/(t_{..})^2 \tag{35}$$

By using the latter prior probabilities, the following expression is obtained for the smoothed sample matrix:

$$\hat{t}^o_{ij} = [t^o_{..}/(t^o_{..} + \hat{k})] \left\{t^o_{ij} + [\hat{k}(t_{i.} \cdot t_{.j})/(t_{..})^2]\right\} \tag{36}$$

where

$$\hat{k} = \left[(t^o_{..})^2 - \sum_{i=1}^{I} \sum_{j=i}^{I} (t^o_{ij})^2\right] \Big/ \sum_{i=1}^{I} \sum_{j=i}^{I} \left\{t^o_{ij} - [(t_{i.} \cdot t_{.j})/t_{..}]\right\}^2 \tag{37}$$

Intervening-Opportunity Method

The intervening-opportunity approach is based on the ride-check information alone. The assumption for choosing a particular OD matrix among the many OD patterns that will satisfy the on and off counts is that at a given stop every qualified passenger on the bus is equally likely to alight (21,22). A passenger is qualified if the boarding stop was at least a certain minimum number of stops m before the given stop and if he has not previously alighted. Let V_{ij} be the volume of passengers originating at stop i who are still on board and eligible to alight at stop j; passengers are eligible to alight at j if $j - i \geq m$ and if they have not alighted at a stop before j. Thus, for $j \geq i + m$,

$$V_{ij} = t_{i.} - \sum_{k=i}^{j-1} t_{ik}$$

and for $j < i + m$, $V_{ij} = t_{ij} = 0$. (For $j = i + m$, $V_{ij} = t_{i.}$.) At stop j

$$\sum_{i=1}^{j} V_{ij} = V_{.j}$$

passengers are eligible to alight and $t_{.j}$ passengers actually alight. Thus, a fraction $t_{.j}/V_{.j}$ of all eligible passengers alighted at j. This fraction is applied to every boarding stop i for which there are eligible passengers:

$$\hat{t}_{ij} = (t_{.j}/V_{.j})\, V_{ij} \qquad j = 1, \ldots, I \quad i = 1, \ldots, j-m \tag{38}$$

The tests of this method in this paper are based on a minimum trip length of zero stops because groups of stops are considered in the case study.

CASE STUDY

In the previous section the following four methods were presented to estimate route-level trip tables by expanding an on-board sample with ride-check data:

- Simple expansion by total boardings
- IPF
- CMLE
- CGLS

A trip table can also be synthesized by the intervening-opportunity approach, which is based on ride-check data alone. In this section of the paper the results are reported of applying these five methods to a data set consisting of a small-sample on-board survey and ride checks recently collected for the Massachusetts Bay Transportation Authority (MBTA). The data were collected by Cambridge Systematics during a 4-week period in November and December 1983 for the MBTA bus routes in the Northwest Corridor of the Boston metropolitan area. The methods were applied to Routes 350 and 77.

Description of the Data

Route 350 is 15.2 mi long and runs as an express route from the suburbs of Burlington (1), through Woburn (2), Winchester (3), Arlington (4 and 5) and

Cambridge (6 and 7) to the central business district (CBD) in Boston (8). [The numbers in parentheses refer to the eight groups of stops defined for this route; there is an approximately equal number of stops per group in the suburbs and a finer division (down to single stops) between Arlington and Boston.] The estimation of trip tables for Route 350 was performed for the morning peak (6:00 to 9:00 a.m.) in the inbound direction with 76 observations from the on-board survey and 485 total boardings counted in the ride check and for the afternoon peak (4:00 to 7:00 p.m.) in the outbound direction with 61 observations from the on-board survey and 200 total boardings counted by the ride check. The ride-check information is complete, and total boardings or alightings at each stop of the route for each time period and direction are available. The simple expansion factors for the two matrices for Route 350 are approximately 6 for the a.m. peak and 3 for the p.m. peak.

Route 77 is 5.5 mi long and serves the communities of Arlington and Cambridge. It is a high-density route with scheduled headways of 3 min for the a.m. peak (6:00 to 9:00 a.m.) and 4.5 min for the p.m. peak (3:00 to 6:00 p.m.). The route was divided into seven groups of stops; the most important were Harvard Square in Cambridge (1), Arlington Center (5), and Arlington Heights (7). For the morning peak in the inbound direction, 54 observations were available from the on-board survey and 2,148 passengers were counted in the ride check. For the afternoon peak in the outbound direction, 138 observations were available from the survey and 1,617 passengers were counted by the ride check.

Note that for Route 77 the simple expansion factors are about 40 and 12 for the a.m. and p.m. peaks, respectively, as compared with factors of about 6 and 3 for Route 350. The influence of the on-board sample information on the estimated matrix for Route 77 should therefore be much less significant than the case of Route 350.

Analysis of the Results

The five estimated matrices (described earlier) with total boardings normalized to 100 are given for each route and direction in Tables 1-4. The IPF and the CMLE estimators require iterative solution methods, which in all tested cases converged rapidly.

A comparison of the IPF, CMLE, and CGLS matrices for the outbound direction of Route 350 versus the matrix obtained by the intervening-opportunity model demonstrates the influence of the OD information from the on-board survey on the estimation results. The IPF, CMLE, and CGLS matrices are more concentrated (in accordance with the OD observations) than the intervening-opportunity matrix, which gives more evenly distributed cell entries.

The IPF, CMLE, and CGLS matrices (see Tables 1-4) exhibit the special characteristics of Route 350, which consists of two segments--the first is from the CBD in Boston (8) and Cambridge (6 and 7) to Arlington Center (4) and Winchester (3) (see the first two rows of the matrices for the outbound direction) and the second is from Arlington Center to the outer suburbs in Woburn (2) and Burlington (1) (see row 5 of the matrices). This characteristic is blurred in the intervening-opportunity matrix, as can be seen from a comparison of the respective matrices.

A comparison of Route 350 estimation results with those for Route 77 shows the effect of the on-board survey sample size: with the smaller sample for Route 77 the IPF, CMLE, and CGLS cell estimates are closer to those of the intervening-opportunity model than with the larger sample size of Route 350.

The comparison of the different matrices can be facilitated by applying the following root-mean-square difference measure for a normalized matrix:

$$\left\{(1/K)\sum_{i=1}^{I}\sum_{j=i}^{I}\left[(\hat{t}_{ij}-\hat{t}^b_{ij})/t_{..}\right]^2\right\}^{1/2} \tag{39}$$

or by a measure of average weighted fractional error:

$$\left\{\sum_{i=1}^{I}\sum_{j=i}^{I}(\hat{t}^b_{ij}/t_{..})\left[(\hat{t}_{ij}-\hat{t}^b_{ij})/\hat{t}^b_{ij}\right]^2\right\}^{1/2} \tag{40}$$

where the IPF results denoted by $\hat{t}^b_{ij}$ are used as the basis for comparison and K is the number of nonempty cells in the estimated matrix. The values of these measures for the four cases analyzed are given in Table 5. The comparison clearly shows that the IPF, CMLE, and CGLS estimators yield practically identical results in all four cases. It is interesting to note that the intervening-opportunity matrix, estimated from the ride-check data, is closer to the IPF matrix than the simple expansion based on the on-board survey.

Table 6 presents a comparison of the alternative methods relative to the CMLE method. Under the Poisson distributional assumption the following log-likelihood function is obtained:

$$-C+\sum_i\sum_j t^o_{ij}\,\ell n\, t_{ij} \tag{41}$$

where

$$C = t_{..}+\sum_i\sum_j \ell n\, t^o_{ij}\,! \tag{42}$$

This log-likelihood function is evaluated for the four estimation methods constrained by the ride-check data: CMLE, IPF, CGLS, and intervening opportunity. The values for the non-CMLE methods are subtracted from the maximum value obtained for the CMLE matrix. These differences are given in the top three rows of Table 6. The bottom of this table shows the results of the likelihood ratio test of the intervening-opportunity model. Thus, if the distribution assumption of the CMLE method is true, the intervening-opportunity assumption can be rejected in three of the four cases analyzed. It can also be noted once again that the CMLE, IPF, and CGLS methods produce similar estimates.

Calculation of Standard Errors of the IPF-Estimated Trip Table

A nonparametric approach to standard error estimation known as the "bootstrap" method was employed to check the results obtained by the linear approximation of Equation 13 for the IPF technique. The bootstrap method does not make any distributional assumptions or any approximations and hence should provide a reasonable basis for comparison [see studies by Effron (23-25) for extensive discussions].

To illustrate the application of this method, consider Route 77 in the inbound direction. The full population of 2,147 trips is cast into OD pairs (i,j) according to the matrix cell probabilities (p_{ij}) derived from the Bayesian seeding procedure ($t_{ij} = p_{ij}\cdot 2{,}147$). From this generated population a random sample of the desired size, say 100 trips, is picked. This sample constitutes one on-board survey outcome and is used as input for the IPF procedure. The variance of the IPF results of 1,000 repetitions of this process is an estimate of Var(t_{ij}) for the IPF estimator. The bootstrap variance estimates for the four cases considered are given in Tables 7 and 8.

TABLE 1 Five Estimated Matrices for MBTA Route 350: Inbound Direction

	Boarding Stop by Section								
	8	7	6	5	4	3	2	1	Total
Normalized IPF-Updated Matrix									
Alighting stop Section									
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	15.5	4.1	0.2	0.0	0.0	0.0	0.0	0.0	19.8
4	20.9	0.6	0.2	0.2	0.0	0.0	0.0	0.0	21.9
3	12.0	9.5	0.2	0.1	0.4	0.1	0.0	0.0	22.3
2	2.1	0.3	1.1	0.1	6.4	0.1	0.6	0.0	10.7
1	0.0	1.8	0.0	0.0	23.1	0.2	0.2	0.0	25.4
Total	50.5	16.3	1.6	0.4	29.9	0.4	0.8	0.0	100.0
Normalized CMLE-Updated Matrix									
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	15.9	3.8	0.1	0.0	0.0	0.0	0.0	0.0	
4	20.7	0.9	0.2	0.1	0.0	0.0	0.0	0.0	
3	11.5	9.9	0.2	0.1	0.5	0.1	0.0	0.0	
2	2.4	0.2	1.1	0.1	6.1	0.0	0.7	0.0	
1	0.1	1.4	0.1	0.1	23.3	0.3	0.1	0.0	
Normalized CGLS-Updated Matrix									
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	15.5	4.1	0.2	0.0	0.0	0.0	0.0	0.0	
4	20.9	0.5	0.2	0.2	0.0	0.0	0.0	0.0	
3	12.6	8.5	0.2	0.2	0.4	0.3	0.0	0.0	
2	1.5	0.3	1.1	0.1	7.0	0.2	0.5	0.0	
1	–0.1	2.9	–0.1	–0.1	22.5	–0.1	0.3	0.0	
Normalized Intervening-Opportunity Matrix									
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	14.5	4.7	0.5	0.1	0.0	0.0	0.0	0.0	
4	10.0	3.2	0.3	0.1	8.3	0.0	0.0	0.0	
3	10.1	3.2	0.3	0.1	8.4	0.2	0.0	0.0	
2	4.7	1.5	0.2	0.0	3.9	0.1	0.2	0.0	
1	11.2	3.6	0.4	0.1	9.3	0.2	0.6	0.0	
Normalized Simple-Expansion Matrix									
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	14.5	1.3	0.0	0.0	0.0	0.0	0.0	0.0	
4	14.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	9.2	2.6	0.0	0.0	0.0	0.0	0.0	0.0	
2	2.6	0.0	1.3	0.0	3.9	0.0	0.0	0.0	
1	0.0	2.6	0.0	0.0	46.1	1.3	0.0	0.0	

TABLE 2 Five Estimated Matrices for MBTA Route 350: Outbound Direction

	Boarding Stop by Section								
	8	7	6	5	4	3	2	1	Total
Normalized IPF-Updated Matrix									
Alighting stop Section									
8	0.0	0.0	0.2	18.4	23.9	20.2	4.2	0.1	67.0
7	0.0	0.0	0.4	0.5	4.0	10.8	0.5	3.3	19.5
6	0.0	0.0	0.4	0.6	0.3	0.5	0.6	0.2	2.5
5	0.0	0.0	0.0	0.0	0.3	0.5	0.6	0.2	1.5
4	0.0	0.0	0.0	0.0	0.0	0.3	4.9	3.3	8.5
3	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.1	0.5
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.5
Total	0.0	0.0	1.0	19.5	28.5	32.5	11.0	7.5	100.0
Normalized CMLE-Updated Matrix									
8	0.0	0.0	0.2	18.1	24.9	19.4	4.2	0.1	
7	0.0	0.0	0.4	0.7	3.2	11.6	0.7	2.8	
6	0.0	0.0	0.4	0.6	0.2	0.5	0.6	0.1	
5	0.0	0.0	0.0	0.0	0.2	0.5	0.7	0.1	
4	0.0	0.0	0.0	0.0	0.0	0.3	4.5	3.7	
3	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.1	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	
Normalized CGLS-Updated Matrix									
8	0.0	0.0	0.2	18.5	23.1	21.1	4.0	0.0	
7	0.0	0.0	0.4	0.5	4.7	9.8	0.4	3.7	
6	0.0	0.0	0.4	0.5	0.4	0.5	0.5	0.2	
5	0.0	0.0	0.0	0.0	0.3	0.5	0.5	0.2	
4	0.0	0.0	0.0	0.0	0.0	0.4	5.3	2.9	
3	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.0	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	
Normalized Intervening-Opportunity Matrix									
8	0.0	0.0	0.8	14.4	18.3	21.2	7.2	4.6	
7	0.0	0.0	0.2	4.2	5.5	6.2	2.1	1.3	
6	0.0	0.0	0.0	0.5	0.7	0.8	0.3	0.2	
5	0.0	0.0	0.0	0.3	0.4	0.5	0.2	0.1	
4	0.0	0.0	0.0	0.0	3.1	3.5	1.2	0.8	
3	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.1	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	
Normalized Simple-Expansion Matrix									
8	0.0	0.0	0.0	14.8	39.3	18.0	3.3	0.0	
7	0.0	0.0	0.0	0.0	3.3	4.9	0.0	4.9	
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	0.0	0.0	3.3	8.2	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

TABLE 3 Five Estimated Matrices for MBTA Route 77: Inbound Direction

	Boarding Stop by Section							
	7	6	5	4	3	2	1	Total
Normalized IPF-Updated Matrix								
Alighting stop Section								
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	15.5	3.8	0.0	0.0	0.0	0.0	0.0	19.4
5	6.2	1.2	0.0	0.0	0.0	0.0	0.0	7.4
4	10.1	0.9	0.2	1.0	0.0	0.0	0.0	12.2
3	12.6	0.5	0.1	0.6	0.6	0.0	0.0	14.5
2	19.4	8.7	0.2	1.1	3.4	2.5	0.0	35.3
1	5.9	1.0	0.2	0.4	1.2	2.7	0.0	11.3
Total	69.7	16.1	0.8	3.1	5.2	5.2	0.0	100.0
Normalized CMLE-Updated Matrix								
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	15.5	3.9	0.0	0.0	0.0	0.0	0.0	
5	6.1	1.3	0.0	0.0	0.0	0.0	0.0	
4	10.1	0.9	0.1	1.0	0.0	0.0	0.0	
3	12.6	0.6	0.1	0.6	0.6	0.0	0.0	
2	19.0	8.3	0.1	1.0	3.3	3.5	0.0	
1	6.4	1.2	0.4	0.4	1.3	1.7	0.0	
Normalized CGLS-Updated Matrix								
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	15.5	3.8	0.0	0.0	0.0	0.0	0.0	
5	6.3	1.0	0.0	0.0	0.0	0.0	0.0	
4	9.8	0.9	0.5	1.0	0.0	0.0	0.0	
3	12.7	0.5	0.1	0.6	0.6	0.0	0.0	
2	19.4	9.0	0.6	1.1	3.4	1.8	0.0	
1	6.0	0.8	–0.4	0.4	1.2	3.4	0.0	
Normalized Intervening-Opportunity Matrix								
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	15.7	3.6	0.0	0.0	0.0	0.0	0.0	
5	5.9	1.4	0.1	0.0	0.0	0.0	0.0	
4	9.3	2.1	0.1	0.6	0.0	0.0	0.0	
3	10.1	2.3	0.1	0.6	1.4	0.0	0.0	
2	21.7	5.0	0.3	1.4	2.9	3.9	0.0	
1	6.9	1.6	0.1	0.4	0.9	1.3	0.0	
Normalized Simple-Expansion Matrix								
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	14.8	3.7	0.0	0.0	0.0	0.0	0.0	
5	11.1	1.9	0.0	0.0	0.0	0.0	0.0	
4	7.4	0.0	0.0	0.0	0.0	0.0	0.0	
3	16.7	0.0	0.0	0.0	0.0	0.0	0.0	
2	14.8	7.4	0.0	0.0	1.9	0.0	0.0	
1	13.0	1.9	1.9	0.0	1.9	1.9	0.0	

TABLE 4 Five Estimated Matrices for MBTA Route 77: Outbound Direction

	Boarding Stop by Section							
	7	6	5	4	3	2	1	Total
Normalized IPF-Updated Matrix								
Alighting stop Section								
7	0.0	22.7	6.8	12.2	14.1	18.4	5.7	80.0
6	0.0	2.5	0.6	0.3	2.4	5.2	0.1	11.1
5	0.0	0.0	0.0	0.2	0.1	0.8	0.1	1.3
4	0.0	0.0	0.0	0.4	0.2	0.4	0.2	1.2
3	0.0	0.0	0.0	0.0	0.6	3.5	1.0	5.1
2	0.0	0.0	0.0	0.0	0.0	0.4	0.9	1.3
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	0.0	25.2	7.4	13.2	17.6	28.7	8.0	100.0
Normalized CMLE-Updated Matrix								
7	0.0	22.8	6.7	12.2	14.0	18.7	5.6	
6	0.0	2.5	0.6	0.3	2.6	5.0	0.1	
5	0.0	0.0	0.0	0.2	0.2	0.8	0.1	
4	0.0	0.0	0.0	0.4	0.2	0.4	0.2	
3	0.0	0.0	0.0	0.0	0.7	3.4	1.1	
2	0.0	0.0	0.0	0.0	0.0	0.4	0.9	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Normalized CGLS-Updated Matrix								
7	0.0	22.7	6.8	12.2	14.3	18.1	5.8	
6	0.0	2.5	0.6	0.3	2.3	5.3	0.1	
5	0.0	0.0	0.0	0.3	0.1	0.9	0.0	
4	0.0	0.0	0.0	0.4	0.2	0.4	0.2	
3	0.0	0.0	0.0	0.0	0.6	3.6	0.9	
2	0.0	0.0	0.0	0.0	0.0	0.4	0.9	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Normalized Intervening-Opportunity Matrix								
7	0.0	22.1	6.3	11.1	13.4	21.1	5.9	
6	0.0	3.1	0.9	1.5	1.9	2.9	0.8	
5	0.0	0.0	0.1	0.3	0.3	0.5	0.1	
4	0.0	0.0	0.0	0.3	0.3	0.5	0.1	
3	0.0	0.0	0.0	0.0	1.7	2.7	0.7	
2	0.0	0.0	0.0	0.0	0.0	1.0	0.3	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Normalized Simple-Expansion Matrix								
7	0.0	18.8	9.4	9.4	18.8	13.8	10.9	
6	0.0	2.2	0.7	0.0	3.6	4.3	0.0	
5	0.0	0.0	0.0	0.0	0.0	0.7	0.0	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	0.0	0.7	2.9	2.2	
2	0.0	0.0	0.0	0.0	0.0	0.0	1.4	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

TABLE 5 Normalized Root-Mean-Square Difference and Average Weighted Fractional Error Between the IPF and the Other Methods

	Route 350		Route 77	
Method	Outbound	Inbound	Outbound	Inbound
CMLE				
RMS difference	.0035	.0022	.0007	.0032
Avg weighted fractional error	.8259	.7776	.2082	1.1860
CGLS				
RMS difference	.0035	.0033	.0010	.0028
Avg weighted fractional error	2.2495	1.8458	.3287	1.8165
Simple expansion				
RMS difference	.0334	.0486	.0206	.0236
Avg weighted fractional error	5.4923	6.5511	3.8541	6.1512
Intervening opportunity				
RMS difference	.0232	.0485	.0089	.0122
Avg weighted fractional error	18.4021	59.1127	3.6810	3.7156
$t^o_{..}$	61	76	138	54
$t_{..}$	200	485	1,617	2,148
K	29	27	25	25

TABLE 6 Log-Likelihood Differences Between CMLE and the Other Methods

	Route 350		Route 77	
	Outbound	Inbound	Outbound	Inbound
IPF	.2	.3	4.4	.5
CGLS	.7	2.1	.0	1.9
Intervening opportunity	20.5	44.7	11.1	4.5
Likelihood ratio test of intervening-opportunity matrix				
Test statistic	41.0	89.4	22.2	9.0
K	29	27	25	25
Critical chi-square value at the 0.005 level of significance	13.1	11.8	10.5	10.5
Reject the intervening-opportunity matrix?	Yes	Yes	Yes	No

TABLE 7 Bootstrap Variance Estimates: Route 350, Standard Error Estimation

	Boarding Stop by Section							
	8	7	6	5	4	3	2	1
Inbound: Approximate Standard Error for IPF Matrix								
Alighting stop Section								
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	12.7	12.7	1.9	0.0	0.0	0.0	0.0	0.0
4	7.8	7.5	2.4	1.3	0.0	0.0	0.0	0.0
3	13.6	13.8	2.2	1.2	5.0	1.1	0.0	0.0
2	5.7	4.0	3.1	1.0	6.2	0.8	2.2	0.0
1	0.5	5.5	0.5	0.4	5.9	1.0	2.2	0.0
Inbound: Bootstrap Standard Error Estimate								
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	12.9	12.9	1.2	0.0	0.0	0.0	0.0	0.0
4	6.6	6.4	1.1	0.4	0.0	0.0	0.0	0.0
3	14.5	15.5	1.3	0.4	4.9	0.5	0.0	0.0
2	6.3	3.8	2.1	0.4	6.6	0.4	0.7	0.0
1	0.7	6.0	0.4	0.2	6.1	0.5	0.7	0.0
Outbound: Approximate Standard Error for IPF Matrix								
8	0.0	0.0	0.9	3.4	4.8	5.8	3.6	0.4
7	0.0	0.0	1.2	2.7	4.6	5.6	2.5	2.5
6	0.0	0.0	1.3	2.2	1.4	2.1	2.1	0.8
5	0.0	0.0	0.0	0.0	1.2	1.7	1.8	0.8
4	0.0	0.0	0.0	0.0	0.0	1.5	2.7	2.5
3	0.0	0.0	0.0	0.0	0.0	0.7	0.7	0.3
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Outbound: Bootstrap Standard Error Estimate								
8	0.0	0.0	0.4	2.5	4.7	5.8	3.4	0.6
7	0.0	0.0	0.4	2.4	4.7	5.9	2.0	2.5
6	0.0	0.0	0.3	0.8	0.6	0.8	0.8	0.5
5	0.0	0.0	0.0	0.0	0.4	0.6	0.6	0.4
4	0.0	0.0	0.0	0.0	0.0	1.5	2.8	2.4
3	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.1
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

In the case of Route 350, the results of the bootstrap method and the approximation were close for cells with large OD flows, being within 10 percent for most cases. This holds true for both directions. However, for the Route 77 cases, where the expansion factors are very large, the approximation technique tends to yield estimated standard errors that are greater than those obtained by the bootstrap technique. In the outbound direction, where the total is expanded from 138 to 1,617, results are reasonable, being within 10 percent for most cells. However, in the inbound direction, where expansion is from 54 to 2,148, the approximate standard errors are 12 to 44 percent greater for the large-value cells.

More comparisons are needed before any general conclusions can be made, but the foregoing results would indicate that where expansion is not too large, the proposed standard error estimation technique is quite promising. The results of the proposed technique would appear to give an upper bound where its accuracy is poor, and hence the results are on the conservative side.

Effect of On-Board Sample Size on Accuracy of IPF Estimates

In order to assess the effect of the on-board survey sample size on the accuracy of the IPF-estimated OD flows, the bootstrap technique was applied with different sample sizes for the on-board survey. Sample sizes between 100 and 1,000 observations were used for the case of Route 77, inbound direction; the total number of boardings was 2,147. As a criterion for estimation accuracy the weighted trace of the matrix of variances obtained by the bootstrap simulation method was used for the IPF estimator. The weighted trace is the sum of variances of the cell estimates weighted by the relative mean cell values $t_{ij}/t_{..}$. This is to reflect the greater role taken by larger cell values in route design decisions.

A graph comparing the weighted trace with the on-board sample size is given in Figure 1 for the simple expansion and IPF methods. As may be observed, the addition of the ride-check data reduces the weighted trace significantly, particularly for small on-board surveys. In reality, because of budget con-

TABLE 8 Bootstrap Variance Estimates: Route 77, Standard Error Estimation

	Boarding Stop by Section						
	7	6	5	4	3	2	1
Inbound: Approximate Standard Error for IPF Matrix							
Alighting stop Section							
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	45.3	45.3	0.0	0.0	0.0	0.0	0.0
5	21.2	21.2	0.0	0.0	0.0	0.0	0.0
4	36.6	27.9	5.8	25.0	0.0	0.0	0.0
3	31.0	17.5	4.0	18.4	19.7	0.0	0.0
2	67.9	54.0	5.8	25.4	28.7	40.3	0.0
1	39.2	18.7	4.6	12.1	21.9	40.3	0.0
Inbound: Bootstrap Standard Error Estimate							
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	38.4	38.4	0.0	0.0	0.0	0.0	0.0
5	18.9	18.9	0.0	0.0	0.0	0.0	0.0
4	27.2	22.4	2.7	10.6	0.0	0.0	0.0
3	22.9	13.2	2.0	7.8	13.8	0.0	0.0
2	57.5	48.5	2.7	10.5	20.8	19.8	0.0
1	27.2	15.3	2.2	5.9	16.2	19.8	0.0
Outbound: Approximate Standard Error for IPF Matrix							
7	0.0	18.5	8.0	11.8	17.9	25.9	11.6
6	0.0	18.5	8.0	8.0	15.5	21.6	3.3
5	0.0	0.0	0.0	5.2	3.4	6.1	2.4
4	0.0	0.0	0.0	7.3	5.3	7.4	3.9
3	0.0	0.0	0.0	0.0	8.3	11.9	8.9
2	0.0	0.0	0.0	0.0	0.0	7.0	7.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Inbound: Bootstrap Standard Error Estimate							
7	0.0	17.0	7.9	9.0	16.2	23.8	10.3
6	0.0	17.0	7.9	7.3	14.0	20.4	3.1
5	0.0	0.0	0.0	3.1	2.3	4.1	1.9
4	0.0	0.0	0.0	3.1	2.4	3.1	1.8
3	0.0	0.0	0.0	0.0	7.8	11.9	9.0
2	0.0	0.0	0.0	0.0	0.0	4.1	4.1
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

straints, on-board surveys are likely to be small, so it would appear that including ride-check data is likely to improve the accuracy of the OD flow estimates.

CONCLUSION

This comparison of the numerical estimation results confirms that the IPF, CGLS, and CMLE estimators produce similar trip tables under the assumption that the ride-check data are deterministic. The simple expansion and intervening-opportunity estimates differ significantly from those obtained by the other methods. Criteria for the choice of estimation method for specific applications should include the following:

- Reasonableness of the assumptions,
- Computational burden,
- Flexibility in terms of the information that can be included in the estimation procedure (e.g., different data sources with different levels of accuracy and biases), and
- Ability to provide measures of the accuracy of the estimates.

An evaluation of the presented estimators according to these criteria leads to the following conclusions:

1. The simple expansion procedure retains the biases that are likely to be present in an on-board survey and does not use efficiently the information available from the ride check. On the other hand, the intervening-opportunity approach, based solely on the ride-check data, is also not an appropriate technique at the level of a single bus route.

2. The IPF, CGLS, and CMLE methods, which combine the small-sample OD data from the on-board survey with the more accurate ride-check information, appear to perform equally well in the four cases that were tested. Among these three methods the IPF method is the simplest. For the CGLS and CMLE methods it is possible with relative ease to obtain measures of reliability of the estimated matrix entries. A simple estimation technique has been tested for the standard errors of the IPF method. As indicated, initial results are promising. Thus, the only major drawback of the IPF method is that the assumption of deterministic constraints cannot be relaxed.

3. The key assumptions of the CGLS estimator are that the base matrix entries are unbiased estimates of the true matrix values and that the error terms are normally distributed. Therefore, the CGLS estimator does not have optimal properties for small-sample on-board surveys and matrices with small cell entries.

4. The CMLE method requires assumptions on the sampling distributions of the base matrix values.

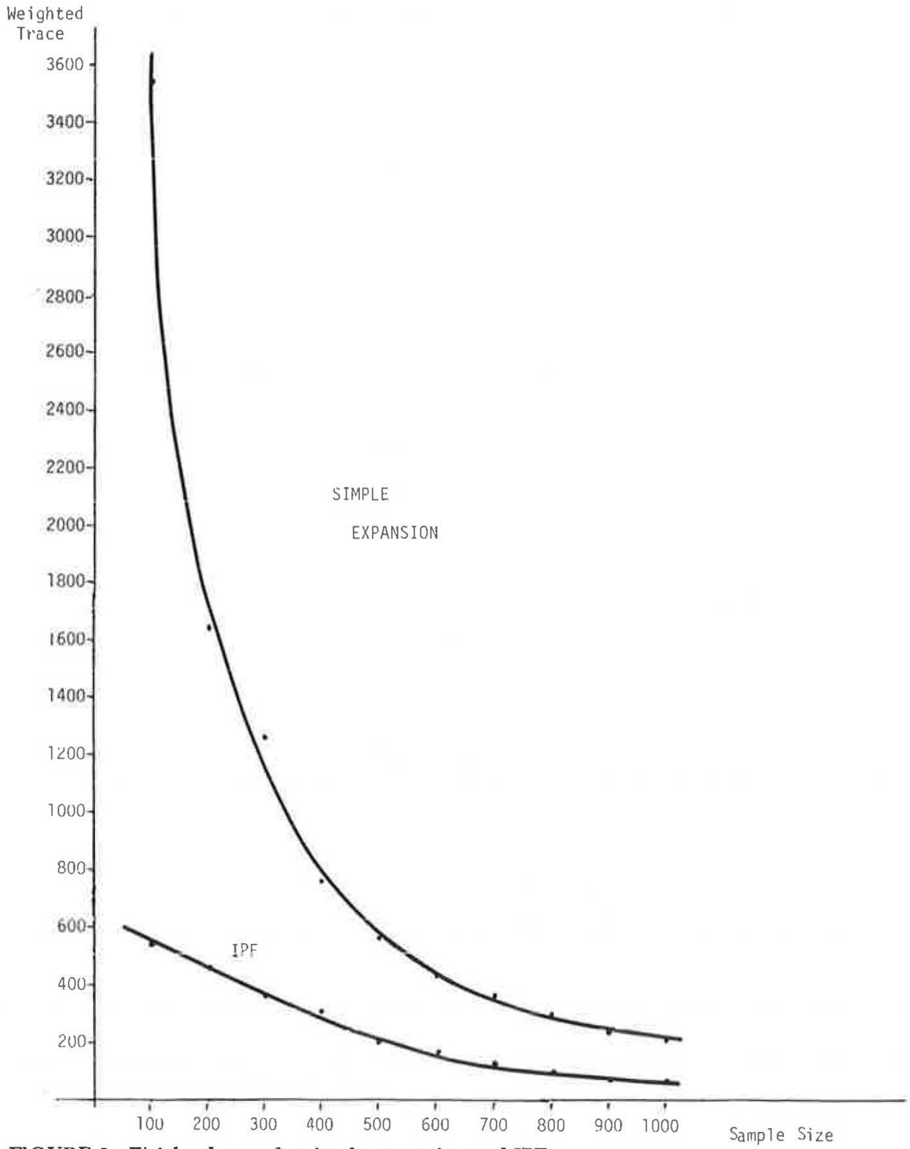

FIGURE 1 Weighted trace for simple expansion and IPF.

The solution technique is computationally more complex than that for the IPF method. Accuracy measures for the matrix cell entries are readily available.

5. With both the CMLE and the least-squares approaches it is possible to extend the analysis to include stochastic ride-check data. This is a more realistic assumption because the ride-check data are also derived from a sample of bus trips and therefore are not free of sampling errors.

6. In terms of flexibility (e.g., consideration of multiple sources of information with different levels of accuracy), the CMLE approach is the most favorable among those considered.

Thus, for the case of an on-board survey and deterministic ride-check data the IPF technique with Bayesian seeding of nonstructural zero cells appears to be the most effective method, although the CMLE and CGLS methods also produce satisfactory results.

REFERENCES

1. P.G. Furth and N.H.M. Wilson. Setting Frequencies on Bus Routes: Theory and Practice. In Transportation Research Record 818, TRB, National Research Council, Washington, D.C., 1981, pp. 1-7.
2. A. Ceder. Bus Frequency Determination Using Passenger Count Data. Transportation Research, Vol. 18A, Nos. 5/6, 1984.
3. Microcomputers in Transportation: Software and Source Book. Report UMTA-URT-83-1. U.S. Department of Transportation, 1983.
4. A.H. Nickesen, A.H. Meyburg, and M.A. Turnquist. Ridership Estimation for Short-Range Transit Planning. Transportation Research, Vol. 17B, No. 3, 1983.
5. M. Ben-Akiva, H. Gunn, and H. Pol. Expansion of Data from Mixed Random and Choice-Based Survey Designs. Presented at the International Conference on New Survey Methods in Transport, Australia, Sept. 12-16, 1983.
6. M. Bacharach. Biproportional Matrices and Input-Output Change. Cambridge University Press, London, 1970.
7. S.P. Evans and H.R. Kirby. A Three-Dimensional Furness Procedure for Calibrating Gravity Models. Transportation Research, Vol. 8, 1974, pp. 105-122.

8. J. Kruithof. Calculation of Telephone Traffic. De Ingenieur (E. Elektrotechnik), Vol. 52, 1927, pp. E15-E25.
9. B. Lamond and N.F. Stewart. Bregman's Balancing Method. Transportation Research, Vol. 15B, No. 4, 1981, pp. 239-248.
10. A.W. Drake. Fundamentals of Applied Probability Theory. McGraw-Hill, New York, 1967.
11. H. Theil. Principles of Econometrics. Wiley, New York, 1971.
12. S. McNeil. Quadratic Matrix Entry Estimation Methods. Ph.D. thesis. Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, Pa., Sept. 1983.
13. C. Hendrickson and S. McNeil. Estimation of Origin/Destination Matrices with Constrained Regression. In Transportation Research Record 976, TRB, National Research Council, Washington, D.C., 1984, pp. 25-32.
14. C. Hendrickson and S. McNeil. Matrix Entry Estimation Errors. Presented at the Ninth International Symposium on Transportation and Traffic Theory, The Hague, 1984.
15. G.G. Judge, W.E. Griffiths, R.C. Hill, and T.C. Lee. The Theory and Practice of Econometrics. Wiley, New York, 1980.
16. J.J. Heckman. The Common Structure of Statistical Models of Truncation Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. Annals of Economic and Social Measurement, May 4, 1976, pp. 475-492.
17. U. Landau, E. Hauer, and I. Geva. Estimation of Cross-Cordon Origin-Destination Flows from Cordon Studies. In Transportation Research Record 891, TRB, National Research Council, Washington, D.C., 1982, pp. 5-10.
18. I. Geva, E. Hauer, and U.Landau. Maximum Likelihood and Bayesian Methods for the Estimation of O-D Flows. In Transportation Research Record 944, TRB, National Research Council, Washington, D.C., 1983, pp. 101-105.
19. Y.M.M. Bishop, S.E. Feinberg, and P.W. Holland. Pseudo-Bayes Estimates of Cell Probabilities. In Discrete Multivariate Analysis, M.I.T. Press, Cambridge, Mass., 1975.
20. H.R. Kirby and M.N. Leese. Trip-Distribution Calculations and Sampling Error: Some Theoretical Aspects. Environment and Planning A, Vol. 10, 1978, pp. 837-851.
21. S. Tsygalnitzky. Simplified Methods for Transportation Planning. M.S. thesis. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, 1977.
22. J. Simon and P.G. Furth. Generating a Bus Route Origin-Destination Matrix from On-Off Counts. Working Paper. Northeastern University, Boston, Mass., 1983.
23. B. Effron. Nonparametric Standard Errors and Confidence Intervals. Canadian Journal of Statistics, Vol. 9, No. 2, 1981, pp. 139-172.
24. B. Effron. A Leisurely Look at the Bootstrap, the Jackknife and Cross-Validation. American Statistician, Vol. 37, No. 1, 1983, pp. 36-48.
25. B. Effron. Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods. Biometrika, Vol. 68, No. 3, 1981, pp. 589-599.

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

Route Choice Analyzed with Stated-Preference Approaches

PIET H. L. BOVY and MARK A. BRADLEY

ABSTRACT

An application of scenario-based, or stated-preference, survey and analysis techniques is described in the context of cyclists' route choice. Route choice modeling with observed choice data is hampered by the cost of processing network data and by the difficulty of assessing the alternative routes and the perceived attributes of the routes considered by individual travelers. An alternative approach is to obtain stated evaluations of well-defined hypothetical routes. Such data were collected from commuting cyclists in the city of Delft in the Netherlands and analyzed by using functional measurement to estimate the relative importance placed on such route attributes as time, traffic level, and surface quality. Though the techniques used are well founded in the marketing and psychology literature, the route choice context raises issues that are particularly important for their application in transport analysis. A case study of the application of stated-preference techniques to route choice is discussed and empirical results obtained for urban bicycle trips are presented.

An understanding of the relative influence of time and cost versus qualitative factors on route choice is valuable in several types of transport system analysis. Alternative plans for new roads or cycling facilities may involve changes in travel times that must be weighed against costs or benefits in terms of other choice factors. These trade-offs are important in predicting the use of new facilities and assessing the benefits to the users. Such information may also be useful for large, network-based studies to define the paths that best represent travel alternatives and to assign the predicted flows for those alternatives to the network. Despite these clear needs for a more thorough understanding of the trade-off process in route choice behavior, theory building as well as modeling efforts in this field are still in need of development. There may be many reasons for this. For example, a route in a network is a difficult concept to deal with in quantitative and statistical analyses. Also, the choice situation with routes is relatively complex, being composed of many alternatives, which are not all clearly distinguishable and overlap slightly. Tackling the route choice problem with revealed-preference random-utility modeling approaches, therefore, poses serious difficulties to the researcher. One way to overcome a number of these problems is to collect preference data by offering hypothetical travel options to individuals in survey form, each option defined in terms of the attributes assumed to be most important. This general method is what is termed the "stated-preference" approach as opposed to revealed preferences inferred from choices among real options.

An application of stated-preference survey techniques in modeling route choice of bicyclists in the city of Delft in the Netherlands is described.

The relative importance of factors such as travel time, surface quality, traffic level, and cycling facility type was studied by varying them experimentally across sets of hypothetical routes. A secondary focus of this study was a practical one, that is, to assess the relative performance of various techniques for route description, grouping of alternatives, measuring preferences, and estimating preference functions in a hypothetical route choice context.

The main approach used in this study to assess cyclists' trade-offs is called functional measurement. This technique originated in the field of mathematical psychology (1) and has been developed in applications to many choice contexts, some within the transport modeling field (2). Thus, the purpose here is not to argue or extend the theoretical validity of the techniques, although the behavioral assumptions are made clear. Rather, the objective of this paper is to provide empirical evidence of the utility and efficiency of standard, established techniques of the stated-preference approach in route choice analysis and application.

APPROACHES TO ROUTE CHOICE ANALYSIS

In the past, route choice analysis has followed two main approaches. In the motivational-attitudinal approach, travelers are asked to state their reasons or motivations in selecting routes in a network. The results are often in the form of qualitative evaluations of the adequacy and importance of individual route attributes and of overall beliefs about alternative routes (3,4). From such studies it is generally agreed that the single most important influence on a driver's choice of route is travel time but that there are other important factors as well. Such studies provide useful input to other approaches by identifying choice factors, helping to define market segments, providing operational definitions for qualitative factors, and giving insight into the relevant choice sets and choice constraints facing individuals. Although it is possible to relate intended or actual choices to reported perceptions and attitudes (5), it is difficult to apply such relationships in assessing policies that can be characterized only as changes in observable route attributes.

With the revealed-preference approach, observable route characteristics are related directly to observed route choices, often by using an individual random-utility maximization framework, such as that of the logit model (6,7,pp.299-330). This approach requires knowledge or an estimate of the set of alternative routes considered by each person, as well as objective data for all salient attributes for each route in the set. Collection of such data is generally difficult and costly and does not ensure that the route characteristics used in modeling consistently represent the subjective measurements of those attributes made by travelers. This problem is especially relevant to route choice, where travelers perceive many route characteristics continuously, and these perceptions may vary a great deal over the course of travel. A variation on the revealed-preference approach requires travelers to state the choice alternatives considered for a familiar choice context and to provide their perceived values along a given set of attributes for each alternative. With this information, it is possible to estimate the relative importance and interactions of perceived route attributes in behavior.

STATED-PREFERENCE STRATEGY

The stated-preference approach is similar to this latter form of revealed-preference analysis, but with the set of choice alternatives and their attributes given in a hypothetical context and behavior measured as a rating, ranking, or stated choice among the alternatives rather than as an actual choice. Although the responses are not subject to the perception processes and choice constraints of actual choice contexts, Louvière et al. (8) found that the two approaches, carried out on the same sample, resulted in similar trade-offs among important mode-choice factors. In a larger mode-choice study (9), which included bicycle commuting, models from scenario-based data validated well against the sample's actual choice among their self-described mode alternatives.

Though stated-preference methods are still evolving rapidly, applications in the transportation field are already numerous. Theoretical background as well as practical aspects of application may be found in numerous reports (10-12). The relative advantage of the stated-preference approach, in most cases, is the controlled nature of the choice scenarios. This feature allows greater freedom in defining choice contexts, alternatives, and attributes as well as direct comparison with the responses across individuals. The ability to obtain multiple responses from each individual reduces sample size requirements and also enables the estimation of truly individual models. With these advantages comes the liability that the success of the approach depends largely on the consistency of the hypothetical alternatives and the corresponding sets of attributes with their perception in actual choice situations.

The study by Morisugi et al. (13) is the only application of a stated-preference analysis to route choice known to the authors. The study uses hypothetical route attributes to estimate values of time for qualitative route factors such as reliability, comfort, and safety by trading off between two factors at a time. In contrast, the authors designed a study by using the full-profile approach in which each hypothetical route was defined completely in terms of a selected set of variables.

Because of the peculiarities of a route, a simple transfer of experiences from other travel-related choices does not appear justified: spatial perception and visual impressions play an important role in the identification by the traveler. Unlike modes, routes cannot be readily labeled or classified into easily understandable categories and because a route is in fact a chain of different links, it is intrinsically heterogeneous. These aspects of routes require special consideration in a stated-preference study.

To examine the usefulness of stated-preference techniques in route choice analysis, a study was designed to serve both substantive and methodological aims: to analyze trade-offs of cyclists for route characteristics of their regular home-to-work trip (this trip purpose generates a large proportion of urban bicycle travel in the Netherlands) and to study the importance of various experimental design features and presentation techniques in the performance of the stated-preference approach.

CONTEXT OF THE SURVEY

In order to keep comprehension problems to a minimum, the study was directed toward frequent and experienced bicycle commuters. Hypothetical route choice situations were arranged and presented to them in questionnaire form. In the questionnaire, options for stimulus presented (the descriptions of choice contexts and alternative routes) and response measurement (the way of expressing preferences toward route alternatives) were varied methodically. In addition, the questionnaire asked subjects to rate the importance of the route choice factors directly and evaluate the survey in terms of ease of response and similarity to actual choice situations. The sample selected for this study consisted of 134 employees of Delft University, who lived in Delft and commuted by bicycle at least twice a week. As a result of the selection, most members of the sample faced similar traffic conditions and comparable commuting distances and used similar cycling facilities for at least part of their journey.

An attempt was made, however, to achieve a wide cross section across characteristics such as age, sex, profession, other modes available, and cycling frequency (more than twice a week) for sake of segmentation analysis and to explain variations in individuals' preferences.

The basic definition of the term "route" had to match the experience of the respondents yet be simple enough to represent with a small set of attributes. To this end a verbal description together with a pictorial form using a map were applied. A route was defined that consisted of a trip from home to work. Routes were further defined along a single set of attributes, each of which was assumed to be homogeneous along each alternative. People were asked to assess various routes as if they were single links. Conceptually, if the utility of a route is assumed to be a linear sum of link utilities, then the relative preference between two routes can be modeled as a function of the important differences in noncommon links. This conceptual solution, however, does not preclude difficulties in perceiving routes as homogeneous. Various presentational approaches were used to confront this problem.

ROUTE ATTRIBUTES AND LEVELS OF THE EXPERIMENTS

The set of route attributes was chosen after the results of previous research had been assessed (14,2). On the basis of such work and previous research in Delft, travel time, surface quality, traffic level, and cycle facility type were selected. Descriptions of the factors and levels for the survey are given in Table 1. Three levels were chosen for each attribute to allow the estimation of nonlinear (quadratic) effects. For the quantitative variable, time, a base value of 12 min was chosen, roughly the median of the respondents' reported one-way travel times. The high and low levels were defined as 15 and 9 min, a range equal to half the base level and encompassing the majority of self-reported times. This range was considered large enough to be perceivable in actual choice situations but not so great as to overshadow the influence of changes in the qualitative attributes.

The levels for facility type and surface quality were defined as commonly encountered types of cycle network construction. Though the definitions were made as mutually independent as possible, there is bound to be some correlation in the perception of these two attributes (i.e., separate bicycle paths are most likely to have an adequate surface). The traffic-level factor was the most difficult to define, as can be seen from Table 1 (a translation from the Dutch survey).

Two variations were included in the survey to test the influence of factor and level presentation on the perception of the qualitative variables and on the ease of comparing hypothetical routes. One-half of the sample was given photographs portraying each of the levels for these factors in addition to the normal verbal descriptions. An overlapping half of the sample was asked to classify their own home-to-work route according to the factor levels that best characterized its major portions, using the given verbal and (in some cases) pictorial descriptions.

PRIMARY EXPERIMENTAL DESIGN

In contrast to the trade-off matrices approach used by Morisugi et al. (13), which presents combinations of pairs of attributes, all others held constant, full profiles were used here; that is, choice sets were presented with alternatives varying across all attributes. The full-profile approach has proven more understandable in practice and more stable if there are significant interactions between variables (15).

To allow the estimation of the independent effect of each attribute, the factors and levels in the route-choice scenarios were arranged in an orthogonal design. To estimate all main effects and all interactions, evaluations of all 3^4 = 81 possible route configurations (treatments) would be required. To limit the size and difficulty of the experiment, the analysis of the trade-offs between route factors was performed at two levels of aggregation. At the individual level a simple piecewise linear main-effects model was assumed where the unobserved error term has the same distribution across all routes:

$$U_i(x) = \sum_{k=1}^{K} \sum_{j=1}^{m_k} (\alpha_{ikj} \cdot x_{kj}) + \varepsilon \quad (1)$$

where

$U_i(x)$ = overall utility or preference measure given to an alternative by individual i,

TABLE 1 Descriptions of Factors and Levels Given

FACTOR	LEVEL	GIVEN VERBAL DESCRIPTION
Facility Type	Physically Separated	This portion of the roadway is meant only for bicycles and mopeds, and is totally separated from other traffic by curbing or plantings. Pedestrians are also provided with a separate walkway.
	Reserved On-Street	This is a full lane of the roadway, reserved for bicyclists and marked with a white stripe on the surface. Now and then there are autos parked on this lane.
	Non-Existent	There is no separate space for bicyclists on the street. They must ride in the same lanes as other traffic.
Surface Quality	Smooth	The surface over the whole route is good asphalt with no large cracks.
	Moderate	In this case, about half of the route is asphalt and the other half brick. There are occasional bumps and cracks.
	Rough	This route has a brick surface from beginning to end with bumps and cracks which one must try to avoid.
Traffic Level	Light	A car or bicycle comes along this route from time to time. There are few crossing pedestrians. Cyclists can easily ride next to each other. You never have to stop for others.
	Moderate	Quite a few cars use this route, but this is not a hinderance to cyclists. You can still ride along with other cyclists when it does not present a problem.
	Heavy	Many cyclists and autos ride along this route. It is very busy here. It is often difficult to enter or cross. You must also wait for other traffic and crossing pedestrians. It is not possible to ride next to other cyclists.
Travel Time		This refers to the total time that you are travelling, including all of the delays encountered on the route. The average travel times used in this survey are:
	Short	- 9 minutes,
	Medium	-12 minutes, or
	Long	-15 minutes.

K = number of attributes of the alternative,
m_k = number of levels of attribute k,
α_{ikj} = the partworth contribution of level j of attribute k to individual i,
x_{kj} = presence or absence of level j of attribute k, and
ε = error term.

With this type of model only a one-ninth fractional factorial design (nine orthogonal alternatives) must be included in each survey to estimate the main effects for each individual. Because individual-level estimates were desired mainly for market segmentation rather than for strict tests of functional form, the simple design was deemed adequate.

Because certain interactions were thought to be potentially important for the qualitative factors, a more extensive block design was used to allow their estimation. To this end, three blocks of nine routes were designed, each block being internally orthogonal. The blocks were distributed evenly across the sample and across other survey variations. Together, the three main-effects designs form a one-third fractional design [see Master Plan 8 by Kocur (10)], allowing aggregate estimation of the two-way interaction between each pair of qualitative variables (although certain interactions may not be separable from other two-way and higher-order terms).

At the aggregate level, therefore, models were specified consisting of main effects and selected two-way interaction terms:

$$U(x) = \sum_{k=1}^{K} \sum_{j=1}^{m_k} [(\alpha_{kj} \cdot x_{kj}) + \sum_{h=1}^{K-1} \sum_{l=1}^{m_k} (\beta_{h,kj} \cdot x_{kj} \cdot x_{hl})] + \varepsilon \qquad (2)$$

for specified kj and hl.

In order to minimize the loss of variation through aggregation, sample segments analyzed with aggregate models were kept as homogeneous as possible in terms of preferences. The individual models were used to guide segmentation, particularly where the aggregate model included only the same main effects. With respect to the statistical specification of the aggregate model, it should be noted that in contrast to the individual model, the error terms presumably will not be identically and independently distributed across all observations. Especially because of the repeated-measurement type of observations, the error terms will tend to be more highly correlated for repeated observations within individuals than for observations across individuals. Assuming independence within individuals will not bias the estimates but may lead to an underestimate of their standard errors, as discussed later.

RESPONSE MEASUREMENT SCALES

In the survey, three response scales were attempted:

- A verbally defined seven-point scale assigning a value to the strength of preference (e.g., "always choose option A," "slightly prefer option B," "no preference"),
- An extension of this scale to a continuous one on which the percentage probability of choosing each option can be indicated, and
- Ranking of the options in order of preference.

For the presentation of the rating scale and percentage score methods, the set of alternatives was transformed into sets of route pairs with every alternative placed against a common base route. This base route was chosen to have the middle level of all four factors. Use of this pairwise format rests on an assumption that responses from partial choice sets are transitive across the remaining choice pairs.

Each respondent was asked to complete each of the three response tasks, which, for sake of comparison, were offered in a fixed order. The first nine routes to be rated were presented one at a time, all against a common base route. The second set of routes, presented on separate cards, had to be ranked. This method uses the full choice set of nine alternatives. The final set was presented in the same pairwise manner as the first set, but now the respondent had to indicate the percentage of times he would choose each alternative route.

To ensure that the three blocks of routes would be well distributed across all three response methods and that subjects would not simply transfer preferences from one scoring method to another, each subject was given all three blocks of nine routes, randomly assigned to one of the six possible permutations of the three sets.

GRAPHICAL ANALYSIS OF AVERAGE SCORES

An overall picture of the relative importance of the choice factors and factor levels and of nonlinearities and interactions can be gained from a graphical plot of the average responses. In Figure 1, the average (utility) scores for the rating scale are plotted according to the attributes of the routes they represent. Because the chosen design accounted for interactions between facility type and the other factors, three separate plots are presented for each of these combinations of two factors against the average scores. If these interactions were not present, each plot would show parallel lines for the three facility types. The relative importance of each of the factors is shown by the slopes of the lines connecting the factor levels or can be approximated from the range of the average scores of the extreme levels. Changes in travel time in the chosen range appear to have the greatest linear effects, directly followed by surface quality, which appears slightly less important. Changes in traffic level and facility type clearly have less influence on the average score values. All the main effects have the expected direction (sign) and are more or less linear, with the exception of traffic level. Certain interaction terms appear to be present: a smooth surface has the least effect when a separate bicycle path is concerned and a rough surface has the smallest effect when no cycling facility is present. These results suggest that the facility type itself has implications regarding surface quality. The same type of interaction is suggested by the reduced influence of light traffic when combined with physically separated bikeways. It is probable that these are not true interactions in the behavioral sense but are due to an interaction in the perception of these attributes from the survey presentation. At this level of average scores, roughly the same relationships were discovered with both of the other response methods (ranks and percentages).

ESTIMATION OF PREFERENCE FUNCTIONS

Estimation Approaches

The metrically scaled data were analyzed by using ordinary (dummy) least-squares regression. To estimate the partworth utilities in Model 1 only $\sum_{k=1}^{K}(m_k - 1)$ linearly independent variables are needed to completely specify the preference model. Therefore, each attribute with m_k-levels is converted into $(m_k - 1)$ dummy variables, where the omitted level serves as a reference (16). In this case, the base route in the paired comparisons, which combines the middle levels of all four factors, was taken for reference. The original Model 1 was then estimated as follows:

$$U(x) = \beta_0 + \sum_{k=1}^{K} \sum_{j=1}^{m_k-1} (\beta_{kj} \cdot x_{kj}) + \varepsilon \quad (3)$$

where

β_0 = the utility for the choice alternative, which has been coded zero for all attributes (base route);
β_{kj} = the differential partworth utility of level j of attribute k, which is the difference in utility between each attribute and the reference; and
x_{kj} = 1 if level j of the kth attribute is present in a choice alternative; 0 otherwise.

At the individual level the main-effects model could be estimated algebraically because there are nine orthogonal equations (observations) to determine the same number of unknown parameters.

At the aggregate level ordinary least-squares regression (OLS) was applied. For efficient regression estimates, the error terms ε must be assumed to be independent and identically distributed across route alternatives and individuals. Because all alternatives represent the same generic choice (a route), this common assumption was considered reasonable for the application-oriented nature of the study. Yet it must be recognized that the error terms will tend to be more highly correlated for repeated observations within individuals than for observations across individuals and that each individual's taste for certain types of routes may vary along factors not controlled for in the design. As mentioned previously, this will cause the standard errors given by the OLS procedure to be biased downward. A conservative adjustment is to assume that the error terms are completely correlated within individuals, adjust the standard errors, and check for significance. If doubt remains, one can use more complex generalized linear regression (GLS) methods.

The ranked data were exploited for estimation by using a so-called "exploded" logit analysis (17). This is a procedure for exploiting the information of ranked choice sets to estimate the parameters of

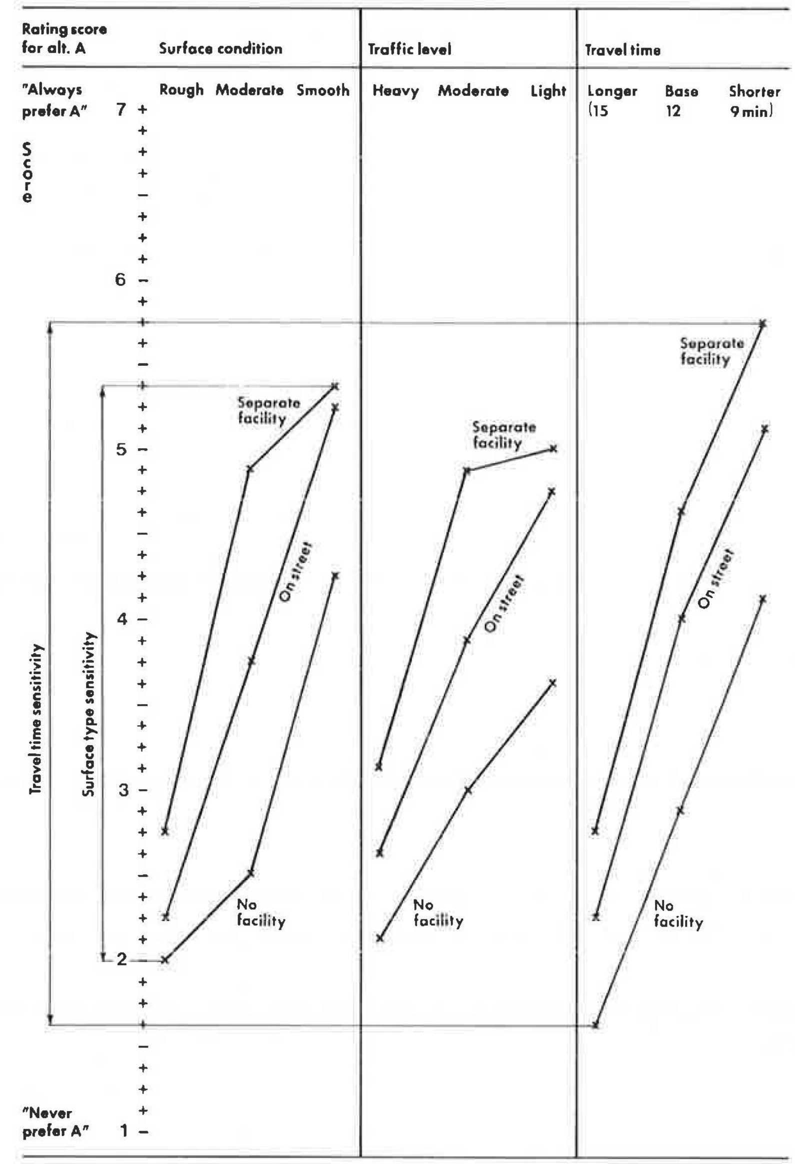

FIGURE 1 Graphical analysis of average rating scores.

the utility function in the multinominal logit model. The "explosion" means the decomposition of a single ranked choice set into a series of unranked and statistically independent choice sets. Each separate ranking can be treated as being chosen over all alternatives that rank equally below it. The choice model to be estimated thus has the following form:

$$P_i \{n | n \varepsilon NCM\} = \exp\{U_i^n(x)\} / \sum_{n=1}^{N} \exp\{U_i^n(x)\} \quad (4)$$

where $P_i(n)$ is the probability that individual i ranks alternative n ε N highest in subset M of the choice set L and U_i^n are the utility functions as, for example, in Equation 3.

The parameters are estimated by using maximum-likelihood techniques. This exploded logit approach requires that two important assumptions be made: first, the validity of the I.I.A. property, which means that the utility of an alternative is not influenced by whatever other alternatives are in the choice set, and second, as with the metrically scaled data regression analysis, the independence and identical distribution across all alternatives and individuals of the error terms. The logit approach was used as a comparison to the OLS approach, though only as a first step toward a more appropriate but more complicated procedure, including tests of the I.I.A. property.

Estimation Results with Different Approaches

Table 2 contains the results of the most aggregate main-effects model estimated for the entire sample by using each type of response data and the maximum number of observations possible (only the top six ranks were chosen for the logit model, however). To facilitate comparison of the parameter values of the different models as well as for better assessment of the trade-offs, all parameters are also expressed relative to the travel-time values as minute equivalents. Overall, the estimation results look plausible and consistent. For the regression models, the constant terms (grand mean) were not significant and were very small, and they are not reported. All

TABLE 2 Aggregate Models for Each Response Data Type

response method	scores (1-7)		percent (0-100)		ranked (1-6)	
individuals	119		114		121	
observations	1071		1026		741	
R-squared (RHO-squared)	.56		.56		(.23)	
	OLS coefficients		OLS coefficients		LOGIT coefficients	
	raw (T-stat)[b]	minutes[a] (range)	raw (T-stat)[b]	minutes[a] (range)	raw (T-stat)[b]	minutes[a] (range)
1. no facility	-0.96 (7.9)	-2.0	12.1 (6.2)	-1.5	-0.65 (6.0)	-1.8
		(.51)		(.35)		(.52)
2. separate path	0.48 (4.0)	1.0	5.0 (2.5)	0.6	0.53 (5.2)	1.4
3. rough surface	-1.44 (11.8)	-3.1	-21.4 (10.7)	-2.7	-1.22 (10.5)	-3.3
		(.92)		(.87)		(.90)
4. smooth surface	1.15 (9.5)	2.4	20.6 (10.3)	2.6	0.80 (7.9)	2.2
5. heavy traffic	-1.23 (10.1)	-2.6	-16.1 (7.9)	-2.0	-0.95 (8.2)	-2.6
		(.62)		(.49)		(.51)
6. light traffic	0.52 (4.2)	1.1	7.8 (3.8)	1.0	0.20 (1.9)	0.6
7. longer time	-1.66 (13.4)	--- 6.0	-25.7 (12.5)	--- 6.0	-1.42 (11.8)	--- 6.0
		(1.0)		(1.0)		(1.0)
8. shorter time	1.15 (9.3)	---	22.6 (11.1)	---	0.82 (8.0)	---

a) Coefficients per minute travel time are normalized using the design range of 6 minutes between the time levels, and the estimated time coefficients. The range is the difference between the coefficients for the extreme levels of each factor, normalized to the range of the travel time factor.

b) t-values based on independence across all observations. Very conservative estimates result by dividing with $\sqrt{9}$, 9 being the number of designed responses per subjects (see also (18)).

other estimates appear statistically significant; longer travel time shows the most precise and separate facility the least precise coefficients. [Even when complete dependence between observations within subjects is assumed, by dividing t-values with $(9)^{1/2}$, most parameter estimates remain significant.] All models indicate, by the range between the normalized effects of the extreme levels, that travel time is most important (in the given time ranges), directly followed by surface type. From the models it can be observed, for example, that for an average trip length of 9 min, an improvement from "no facility" to "separate path" will compensate for a travel time loss of about 3 min. An improvement from a rough to a middle-quality surface also can compensate for a travel time detour of 3 min. These tradeoff values suggest a fairly high sensitivity of bicycle travelers toward changes in qualitative route factors.

The three data types yield similar models in terms of the explained variation and the relative importance of the factors. Models that included the interactions identified in Figure 1 showed them to be marginally significant [for details see report by Bovy et al. (19)].

Estimation Results with Various Segmentations

Various segmentations were attempted to improve the explanatory and predictive power of the models. First, the respondents were grouped according to their individual coefficients for time, as estimated from their scores on the rating scale. Table 3 provides evidence that this type of clustering is effective, because the rho-squared values for internal segmentation increase noticeably. The coefficients for both time levels are significantly higher for the time-sensitive group. The other group shows significantly higher disincentives from rough surface, heavy traffic, and no facility. This appears to be the comfort-sensitive segment. Whereas the time-sensitive bicyclists are willing to spend only 1.5 min of extra time to use a route that has a separate bicycle path instead of no facility, the comfort-sensitive segment appears willing to accept a detour in such a case of more than 6 min (on an average trip length of 9 min).

A second segmentation was done compositionally, according to age. The travel time coefficients for the respondents under 40 are higher than those of the over-40 segment, whereas the coefficients of the other variables are not significantly different between both segments.

The corresponding time-valued figures show that the older cyclists are willing to sacrifice much more travel time for better route quality than the younger riders. Although this particular external segmentation is less effective in improving the model, groupings of this type are more useful in application, and the composition of the internal segments can help to define more effective groupings.

TABLE 3 Estimated Coefficients and Validation Results for Segmentations (Logit Model, Top Three Ranks)

TYPE OF SEGMENTATION	NONE		INTERNAL				EXTERNAL			
			HIGH WEIGHT ON TIME		LOW WEIGHT ON TIME		UNDER 40 YEARS		OVER 40 YEARS	
OBSERVATIONS	374		208		166		179		195	
RHO-SQUARED	.29		.40		.34		.34		.27	
	RAW (T-STAT)[b]	NORM[a] (*RANGE*)	RAW (T-STAT)[b]	NORM[a] (*RANGE*)	RAW (T-STAT)[b]	NORM[a] (*RANGE*)	RAW (T-STAT)[b]	NORM[a] (*RANGE*)	RAW (T-STAT)[b]	NORM[a] (*RANGE*)
1. NO FACILITY	-0.64 (4.0)	*-1.27* *(0.49)*	-0.36 (1.5)	*-0.44* *(0.24)*	-1.07 (4.2)	*-3.07* *(1.09)*	-0.52 (2.1)	*-0.77* *(0.32)*	-0.79 (3.6)	*-1.87* *(0.66)*
2. SEPARATE PATH	0.83 (5.6)	*1.64*	0.82 (3.2)	*1.00*	1.21 (5.7)	*3.48*	0.79 (3.4)	*1.17*	0.89 (4.6)	*2.11*
3. ROUGH SURFACE	-1.72 (6.6)	*-3.41* *(0.85)*	-1.02 (2.2)	*-1.25* *(0.52)*	-2.31 (6.0)	*-6.64* *(1.71)*	-1.67 (3.5)	*-2.48* *(0.66)*	-1.80 (5.5)	*-4.27* *(1.08)*
4. SMOOTH SURFACE	0.94 (6.6)	*1.86*	1.54 (4.6)	*1.89*	1.27 (5.7)	*3.65*	0.99 (3.3)	*1.47*	0.93 (4.6)	*2.20*
5. HEAVY TRAFFIC	-1.15 (6.1)	*-2.28* *(0.55)*	-0.87 (2.6)	*-1.07* *(0.20)*	-1.57 (5.8)	*-4.51* *(1.11)*	-1.23 (3.8)	*-1.83* *(0.43)*	-1.13 (4.7)	*-2.68* *(0.67)*
6. LIGHT TRAFFIC	0.52 (2.9)	*1.03*	0.12 (0.4)	*0.15*	0.76 (3.2)	*2.18*	0.49 (1.6)	*0.73*	0.56 (2.5)	*1.33*
7. LONGER TIME	-1.30 (6.0)	- 6.0 *(1.0)*	-2.01 (4.6)	- 6.0 *(1.0)*	-1.25 (4.6)	- 6.0 *(1.0)*	-1.96 (4.4)	- 6.0 *(1.0)*	-1.04 (4.0)	- 6.0 *(1.0)*
8. SHORTER TIME	1.73 (9.8)	-	2.87 (9.0)	-	0.84 (3.5)	-	2.08 (6.8)	-	1.49 (6.6)	-

VALIDATION RESULTS	NONE	INTERNAL	EXTERNAL
ORIGINAL SAMPLE : 1107 RANKINGS (123 SUBJECTS)			
SPEARMAN RANK CORRELATION :	.745	.778	.750
HOLD BACK SAMPLE: 99 RANKINGS (11 SUBJECTS)			
SPEARMAN RANK CORRELATION :	.746	.768	.753

The predictive improvement from the segmentations can be checked by examining the correlation between the actual and predicted rankings. Table 3 shows a small increase in the Spearman correlation from the clustering segmentation but only a marginal improvement from segmentation by age. Most of this improvement is likely in the top few rankings, on which the models are based and which are most important for forecasting.

Validity and Reliability Tests

The true validity of these models in explaining behavior can be judged only with respect to independent data on observed choices. There were, however, several steps taken to increase confidence in the results. In the preceding sections the robustness of the results to differences in composition of choice sets, response method, and type of estimation has been shown. On the other hand, the modeling results appeared to be clearly sensitive to segmentation of the subjects.

The internal validity of the models was tested by using the unsegmented and segmented models to predict rankings for a small hold-back sample of 11 individuals. Even with this limited sample, the recovered rankings are at a level almost identical to that for the original sample (bottom of Table 3). The external validity of the model can only be tested with an independent source of data. Unfortunately, limited data exist for bicycle route choice.

External validity can be judged somewhat through comparison with the results of similar studies. In a concurrent stated-preference analysis in Wisconsin, Axhausen (2) identified similar trade-offs among the same four attributes (by using distance instead of time). He also found slope, land use, and cycling experience to be influential. These factors, however, exhibit much less variation in the Netherlands and were assumed not to affect the route choice of most cyclists. In another stated-preference experiment (9), also in Wisconsin but this time for automobile versus bicycle choice, time, surface, traffic, and facility were shown to have roughly the same relative influence as that reported in the preceding discussion.

A further test dealt with the method of presenting the qualitative factor levels, an aspect that appears of crucial importance in the context of route choice. Four different methods of presentation were included in the experimental design and were randomly distributed among the subjects: a purely verbal description of each factor level, the same verbal description illustrated with photographs, the verbal description including an exercise to use this for a categorization of the actual home-to-work route, and all three elements combined. Table 4 shows the relative factor utilities normalized with respect to time.

The breakdown indicates that because the qualitative factors were portrayed more clearly, they were generally given more importance relative to travel time and that the inclusion of photographs appears to have helped clarify these factors to a greater extent than asking subjects to categorize their actual route.

The results also indicate that using all presentation approaches in combination was generally no more effective than using either one separately. If the "all methods" category can be assumed to contain the most informed responses, it appears that the use of photographs (column 2) without some relation to the actual routes (columns 3 and 4) may make qualitative factors appear more homogeneous and more important than in actual choice contexts.

Finally, after the scenario comparisons, each respondent was asked to assess the importance of the four attributes in his daily route choice, the clarity of these attributes in the survey presentation, and the overall difficulty of relating the scenarios to his own choice situations. A preliminary analysis of these evaluations showed that the choice factors were generally well understood and considered important in choosing a route. This information also supported the relative importance of the factors found in the analysis.

Over three-fourths of the subjects reported little or no difficulty in comparing hypothetical choices. Of the individual factors, only traffic level appeared to present some difficulty in comprehension. This is the factor that encompasses the widest variety of physical attributes and is likely to be quite variable over an actual route. Interestingly, those who were asked to classify their own route in these terms reported a less clear understanding of this factor. Perhaps the task brought out the inconsistency between simplified and actual routes. Those who were given photographs, on the other hand, reported less trouble understanding the attributes.

As for the different scoring methods, the respondents found the ranking of routes with cards to be easiest and the verbal scale slightly more difficult. The percentage scale was reported as decidedly the most difficult.

SUMMARY AND CONCLUSIONS

By itself, the stated-preference approach appears to give stable estimates of the trade-offs among specific choice factors in a specific context at a cost usually much lower than that of alternative methods. The information on trade-offs within different market segments is useful in ranking alternative policies and in identifying advantaged and disadvantaged groups of users affected by those policies. The differential weights placed by bicyclists on various route factors can help planners in designing bicycle facilities. The outcomes also suggest that extending current minimum-time traffic assignment models with other route factors should be seriously considered for applications in bicycle traffic.

The magnitude of the estimated coefficients suggests a fairly high sensitivity of bicyclists toward the chosen route factors, a result that might be partly due to the specific nature of a stated-preference survey. By showing that the differences in attributes between alternatives vary explicitly, the subject's responses presumably are far more differential than when he is confronted with changes in real alternatives. For more detailed policy analyses it is suggested therefore that the stated-preference models be validated (and probably scaled down) by calibrating them with actual choice data as far as possible. The findings also indicate that the method of presenting the choice context and attributes can have a significant effect on estimates of individuals' trade-offs. This difficulty is essentially what distinguishes this analysis from most other contexts in which scenario-based analyses have been performed successfully. The use of photographs and maps appears to be a useful aid in understanding the experiment, but further research into methods of defining and presenting qualitative route factors should have priority in extending these techniques to a wider range of route choice contexts.

TABLE 4 Relative Linear Utilities of the Qualitative Factors Normalized to Time by Presentation Subgroup[a]

Factor	verbal alone n=22 1	verbal + photo's n=35 2	verbal + own route n=37 3	all methods n=34 4	average n=128 5
Facility type	2,8	7,1	3,2	3,5	4,5
Surface type	4,0	8,5	6,4	6,8	6,6
Traffic level	3,5	6,4	5,9	3,5	5,0

[a]These utilities are based on averages from the individual-level models using the rating scale data

In terms of measuring preferences, ranking of the routes appears to be easiest for the respondents and most comparable to revealed-preference data. Metric scales, on the other hand, allow simpler regression analysis and provide more information at an individual level. Before such scales are used extensively in mail-out experiments, however, it appears that improvement and simplification of the scale presentation and grouping of alternatives are necessary. The percentage scale is not recommended: it is the most difficult for the subjects and can easily lead to response errors. The estimation of the models appears robust with both regression and discrete choice methods. Apart from these methodological issues, further work should be done to study additional route choice factors. For example, qualitative route factors such as safety, variability, and signposting are important for many policy areas and could be incorporated into a stated-preference analysis.

ACKNOWLEDGMENTS

The research presented in this paper was partially funded by the Dutch Ministry of Transport and Public Works. Thanks are due to Dick den Adel of the Delft Institute for Town Planning Research for doing all of the survey production and administration and to Hugh Gunn and Ben Jansen for their constructive criticisms. The authors wish to thank the anonymous referees for their very helpful comments and criticisms.

REFERENCES

1. N.H. Anderson. Foundations of Information Integration Theory. Academic Press, New York, 1981.
2. K.W. Axhausen. Bicyclists Evaluate Their Environment: Some Results. M.Sc. thesis. Department of Civil and Environmental Engineering, University of Wisconsin, Madison, 1984.
3. S.M. Carpenter. Drivers' Route Choice Project: Pilot Study. Paper 085/WP42. Transport Studies Unit, Oxford University, Oxford, England, 1977.
4. R.D. Huchingson et al. Survey of Motorist Route-Selection Criteria. In Transportation Research Record 643, TRB, National Research Council, Washington, D.C., 1977, pp. 45-48.
5. M. Wachs. Relationships Between Drivers' Attitudes Toward Alternate Routes and Driver and Route Characteristics. In Highway Research Record 197, HRB, National Research Council, Washington, D.C., 1967, pp. 770-787.
6. R. Hamerslag. Investigation into Factors Affecting the Route Choice in Rijnstreek-West with the Aid of a Disaggregate Logit Model. Transportation, Vol. 10, 1981, pp. 373-391.
7. M. Ben Akiva et al. Modelling Interurban Route Choice Behavior. In Proceedings of the Ninth International Symposium on Transportation and Traffic Theory (J. Volmuller and R. Hamerslag, eds.), VNU Science Press, Utrecht, Netherlands, 1984.
8. J.J. Louviere et al. Laboratory-Simulation Versus Revealed-Preference Methods for Estimating Travel Demand Models. In Transportation Research Record 794, TRB, National Research Council, Washington, D.C., 1981, pp. 42-51.
9. Development of Wisconsin Work Trip Models for Forecasting Modal Choice. Division of Planning and Budget, Wisconsin Department of Transportation, Madison, 1981.
10. G. Kocur et al. Guide for Forecasting Travel Demand with Direct Utility Assessment. UMTA, U.S. Department of Transportation, 1982.
11. J. Benjamin and L. Sen. A Manual of Procedures to Analyze Attitudes Towards Transportation. U.S. Department of Transportation, 1982.
12. S.R. Lerman and J.J. Louviere. Using Functional Measurement to Identify the Form of Utility Functions in Travel Demand Models. In Transportation Research Record 673, TRB, National Research Council, Washington, D.C., 1978, pp. 78-86.
13. H. Morisugi et al. Measurement of Road User Benefits by Means of Multiattribute Utility Function. Papers, Regional Science Association, Vol. 46, 1981, pp. 31-43.
14. W.F. Upcott. Cyclists' Route Choice: A Behavioural Analysis of the Journey to School. M.Sc. thesis. Cranfield Institute of Technology, Cranfield, Bedfordshire, England, 1976.
15. M.N. Segal. Reliability of Conjoint Analysis: Contrasting Data Collection Procedures. Journal of Marketing Research, Vol. 19, 1982, pp. 139-144.
16. H. Timmermans. Decomposition Multiattribute Preference Models in Spatial Choice Analysis: A Review of Some Recent Developments. Progress in Human Geography, Vol. 8, No. 2, 1984, pp. 189-221.
17. R.G. Chapman and R. Staelin. Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model. Journal of Marketing Research, Vol. 19, 1982, pp. 288-301.
18. J.J. Louviere and G. Woodworth. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. Journal of Marketing Research, Vol. 20, Nov. 1983, pp. 350-367.
19. P.H.L. Bovy et al. Toepassing van Stated Preference Methoden Voor de Analyse van Fietsroute-Keuzegedrag (in Dutch). Institute for Town Planning Research, Delft University of Technology, Delft, Netherlands, 1984.

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

Tests of the Scaling Approach to Transferring Disaggregate Travel Demand Models

HUGH F. GUNN, MOSHE E. BEN-AKIVA, and MARK A. BRADLEY

ABSTRACT

The transferability of disaggregate travel demand models is viewed as a problem of predicting the values of model parameters in a new context by using the estimated values from an existing model. The approach tested in this paper is to use survey data from the new context to estimate transfer-scale parameters that correct for context-specific effects. The empirical tests investigate a variety of transfer models with different numbers of transfer-scale parameters ranging from complete reestimation of the entire model specification to none at all. The results support the partial transfer model in which scale parameters are estimated for a limited number of functional subgroups of variables such as alternative specific constants, level-of-service attributes, and personal characteristics.

During the past decade a substantial body of literature has emerged with empirical evidence on the variability of the estimated parameters of disaggregate travel demand models between urbanized areas and over time. [See, for example, the studies by Watson and Westin (1,pp.227-249), Atherton and Ben-Akiva (2), Parody (3), Talvitie and Kirshner (4), Train (5), McCarthy (6), Silman (7), Koppelman and Wilmot (8), McCoomb (9), Tzieropoulos (10), Rose and Koppelman (11,pp.471-491), and Supernak (12,pp.533-559).] Evidence of stable values of estimated parameters provides a direct indication of the range of validity of a model. A model that is not stable over time is likely to produce inaccurate predictions and a "transferable model" would permit more cost-effective analyses of transportation plans and policies. Therefore, many of the transferability studies have tended to focus their attention on statistical tests that reject or accept equality of parameter values between populations from different areas and time periods. The conclusions of these studies are mixed. Tests of entire sets of parameters have usually shown significant differences between urban areas (4), whereas comparisons of subsets of parameters, in particular coefficients of travel times and cost variables, have found some similar values in models estimated for very different populations. [See, for example, the review by Ben-Akiva (13).]

Because it is unrealistic to expect an operational travel demand model to be perfectly specified, an estimated model is in principle context dependent. This means that it would not be useful to define transferability in terms of equality of parameter values in different contexts. The proper role of statistical transferability tests is to help identify those aspects of a model that need to be further investigated and potentially respecified to obtain an improved model (13).

Thus, a constructive definition of transferability must be based on pragmatic considerations. It is assumed a priori that model parameters have different values in different contexts and consideration is given to the more general issue of whether an existing model provides information that can be used in some way to improve forecasting in a new context.

The value of transferring an existing model depends on the magnitude of the differences in parameter values between the two contexts. This is the transfer bias, which is generally unknown. In addition, the value of a transfer would depend on the accuracy of the estimated model in the base context and on the level of the available information in the transfer context (e.g., the size of the new survey).

A variety of methods to update an existing model with information from a new context were suggested by Atherton and Ben-Akiva (2). For example, if it is believed that the transfer bias is negligible and a travel survey is available in the new context, then a Bayesian model-updating approach can be used. However, the available evidence suggests that transfer biases do play an important role for at least some subgroups of model parameters.

In the following section of this paper, the transfer-scaling approach designed to correct transfer biases is presented, followed by a description of the tests that were performed to find the most appropriate approximation for the transferred model. The empirical evidence does not support complete equivalence of parameter values. It suggests that a partial transfer by using the scaling approach is an effective method of transferring disaggregate travel demand models. Additional practical aspects and data requirements for implementing this approach are discussed by Gunn and Pol (14).

THE SCALING APPROACH TO MODEL TRANSFER

In the previous section transferability was characterized in the context of predictive models as being a property of general similarity between members of a family of models. It was argued that, to be a useful property, this similarity need not be a complete equivalence.

First, for practical purposes there will be a range of parameter values within which the response characteristics of the model are effectively equivalent. Models within this range are, to the practitioner, transferable regardless of the statistical evidence for the nonequivalence of the parameters.

More generally, there may exist simple transformations of one model that will bring it within such a range of another. Thus the transfer process may itself involve estimating parameters from data collected in the area to which the transfer is being made. Such a property will be of practical value if it is easier (cheaper) to estimate the transfer coefficients than to estimate the model coefficients in the new context.

Some empirical evidence will be presented here about a family of simple transformations that could provide such a link. The basic idea is not new; it has been proposed by Atherton and Ben-Akiva (2) and by Ben-Akiva (13) and has also been used by other researchers in this area [Rose and Koppelman (11), Supernak (12), and Koppelman et al. (paper in this Record)]. The family of transformations consists of a set of transfer-scaling factors for each of a number of subgroups of explanatory variables; the variables in each subgroup have already been scaled by their respective coefficients in the base model.

Thus a completely reestimated model is included as the special case when each subgroup contains only one member, and the "naive" transfer is also included, which is the case in which all variables are contained in one group and the transfer-scaling factor is set at unity.

The behavioral rationale behind the approach rests on the hypothesis that the (measurable) variables in the models can be collected into subgroups on the basis of the function they serve and that that function may be more or less important in different contexts.

For example, it may be expected that the level-of-service variables will be jointly influenced by real income levels and individually influenced by the specific characteristics of each travel mode in that context. Thus traffic conditions, service reliability, the standard of comfort of transit vehicles, and the quality of pedestrian and cycle facilities specific to the context will also affect the relative importance of the level-of-service variables and hence the scale of the coefficients in the models.

Other variables entering travel demand models are mostly concerned with effects that can be broadly categorized as personal characteristics, household characteristics, time of day, area type, and local effects.

Personal characteristics affect the absolute levels of modal attractiveness in mode-split models, and they are the main explanatory variables in travel frequency models. In mode-split models, these variables primarily reflect the availability of the alternative travel modes. However, they also account for contextual variations in reliability, safety, and other effects not proportional to travel time and the differing importance of these variations to different types of individuals. Once again, their absolute scale is potentially context-related.

The most important household variables are income levels and the level of competition for household automobiles. The former might be expected to be related to price levels (potentially context-specific) and the latter to patterns of automobile use and specifically to the use of the automobile for the journey to work (also potentially context-related through the availability of parking space and the competitive position of the other modes).

Time-of-day effects may be influenced in certain cases by local work practices (e.g., flex time for work, opening hours for shopping). Inasmuch as demand varies between hours of darkness and light, these effects will also be dependent on the time of year of the survey.

Area-type variables are probably the most obviously context-specific, even when the measured variable can be defined quite rigorously. For example, zones may have exactly the same population density but entirely different land use patterns.

Local-effects variables are included in destination choice models to allow for perceived differences in the attractiveness of near and distant destinations due to differences in knowledge or familiarity. Measured distance has been included in some models to represent this effect; intrazonal dummy variables are also used. Once again, these effects could be expected to vary with context and the choice of zoning system.

Finally, there is the question of the variables and effects that are not specifically included in the model specification. The means and variances of these residual effects determine both the levels of the alternative-specific constants and the absolute scale of the representative utility function. Here too the size of these effects may be context-specific.

In conclusion, the inclusion of transfer-scale parameters in the manner that has been outlined allows the empirical investigation of a variety of possible context-specific effects ranging from one per variable (complete reestimation) through one per functional subgroup (partial transfer) and only for omitted effects (complete transfer) to none at all (naive transfer).

CONTEXT OF THE EXPERIMENT

The transferability experiment was based on two adjacent regions of the Netherlands, one centered at Rotterdam and The Hague and the other at the city of Utrecht, each having respective populations of approximately 3 million and 1 million inhabitants. The data collection exercises took place 5 years apart and at different times of the year. In this section some statistics are presented to identify those variations in travel patterns and in major background variables that should be attributed to the differences in geographical area, year, and time of year of the survey. These statistics are derived from a nationwide travel survey conducted annually by the Netherlands Central Bureau of Statistics, which includes some 10,000 households a year. This survey has been used throughout the transferability experiment to provide a control data set against which to judge both base and transfer surveys.

Average levels of household size, income, driving-license possession, and automobile ownership in the base and transfer areas for the control data set are as follows:

Variable	Base Area	Transfer Area
Household size	2.68	3.00
Income (guilders)	28,300	30,300
Driving licenses per household	1.17	1.22
Automobiles per household	0.83	0.89
Average no. of trips per day	1.29	1.34
Mean trip distance (km)	16.7	17.9
No. of households in sample	5,439	1,680
No. of trips in sample	24,461	8,283

It can be seen that the two areas are very similar in all these respects; the transfer area has somewhat larger households, higher income per household, higher average license possession, and higher automobile ownership per household.

In terms of travel characteristics, the levels displayed in Table 1 for the control data set indicate once again that there is very little difference between the two areas. There is a general tendency for more trips to be made in the transfer area,

TABLE 1 Average Level of Trips

	Base Area		Transfer Area	
Age Group	Percentage of Population in Age Group	Trips per Day	Percentage of Population in Age Group	Trips per Day
12-17	11	1.47	14	1.50
18-26	15	1.38	15	1.45
27-46	36	1.47	37	1.47
46-65	25	1.15	23	1.26
Over 65	14	0.84	11	0.88

regardless of age group. However, the effect is small; around 5 percent more trips were reported.

When the tables were compiled from the control data set, surveys conducted on weekends were excluded, as were surveys on public holidays or days adjacent to public holidays. This was done to maximize the correspondence of the control data with the experimental data.

Trip lengths are also slightly higher in the transfer area; self-reported distances suggest an increase of around 7 percent in the transfer area from the level in the base. These figures have been corrected to adjust for a systematic reporting bias established by Moning (15).

Overall mode use differs little between the areas, despite the absence in the transfer area of the tram and metro modes that are available in the base:

	Percentage of All Trips Reported	
Mode	Base Area	Transfer Area
Automobile		
Driver	28	28
Passenger	12	8
Transit	3	3
Walk	31	26
Two-wheeled vehicle	24	33
Other	2	2

A feature of travel in the Netherlands is the intensive use of bicycles and mopeds. The absence of gradients and the high density of population and services typically combine to make the slow modes (walk, cycle, or moped) account for in excess of 50 percent of all trips. The only noticeable difference between the areas is the more intensive use made of two-wheeled vehicles in the transfer area, which draws travel from both the walk and the automobile (passenger) modes.

In conclusion, in all major aspects the two areas are very similar and appear to be ideal candidates for a model transfer.

Because the experimental surveys were conducted at different times of the year, further analysis was performed to establish the significance of variations in trip rates and mode use over the different months of the year. The results are given in Table 2, which suggests that there are no extreme variations for which allowance need be made. Note that this data set excludes holiday travel, work-based business travel, and travel by commercial travelers, truck drivers, and other professional travelers. Thus, to some extent, the factors that cause seasonal fluctuations have been deliberately suppressed.

Last, some analysis was performed to establish the changes that had taken place over time during the 6 years between the base and the transfer survey periods. The statistics are not presented here, but the conclusions were that household incomes had grown initially and then leveled off or even declined somewhat during the period of general economic stagnation in the early 1980s. License ownership had continued to increase and so too had automobile ownership, although to a lesser extent and with some suggestion of a decline in the year of the transfer survey, 1982. More information about these trends has been given by Daly (16,17).

TABLE 2 Seasonality in Travel by Mode: Netherlands 1982

	Avg No. of Trips per Person			
Month	Automobile	Transit	Slow	Total
January	0.53	0.09	0.75	1.37
February	0.54	0.08	0.69	1.31
March	0.57	0.10	0.72	1.39
April	0.49	0.07	0.67	1.23
May	0.51	0.08	0.68	1.27
June	0.49	0.06	0.86	1.41
July	0.55	0.06	0.71	1.32
August	0.58	0.07	0.75	1.40
September	0.52	0.09	0.84	1.45
October	0.53	0.09	0.72	1.34
November	0.57	0.09	0.76	1.42
December	0.60	0.10	0.68	1.38

COMPARISON OF THE BASE AND TRANSFER SURVEYS

The base sample consisted of data from some 3,000 households collected in the last months of 1977 for a regional transportation study in the Netherlands (the Zuidvleugel study), an overview of which is given by Daly et al. (18). The transfer sample was collected in the late spring of 1982 and consisted of some 1,000 household interviews. In this later survey, it was ensured that the content of the questionnaire and the definitions of mode and purpose that were used were entirely consistent with those in the earlier study. In terms of questionnaire layout and presentation, however, no attempt was made to replicate the earlier study. Rather, every effort was made to improve the clarity of the questionnaire and generally to minimize the difficulty of completion in an attempt to maximize response.

The household interview was later augmented with several thousand interviews with travelers. However, only the analysis of the household survey data will be reported in this paper.

In terms of characteristics of households and persons (number of members, workers, automobiles, income levels, etc.), after expansion both base and transfer surveys compared well with the control data. In keeping with the trends over time and the area-to-area differences apparent from the control data, automobile ownership by household was nearly 40 percent higher in the transfer data, and driving licenses by household were around 50 percent higher. Although the transfer area is generally a more prosperous area, the overall economic decline has left it only some 2 percent better off in terms of real household income than was the base area 6 years earlier.

In Table 3 trip rates per person by travel mode are compared for the base and transfer surveys. As might be expected, the large increase in automobile availability has resulted in a comparable increase (50 to 60 percent) in automobile trips. What is more surprising is the large increase in the slow modes, walking and two-wheeled vehicles (bicycle and moped).

Table 4 shows the same comparison by person type and trip purpose, and it is clear that the increased trip rate in the transfer area reflects general increases in travel irrespective of person type or travel purpose. By far the largest increases are in travel for discretionary activities; this is also

TABLE 3 Frequency of Trips per Person: Experimental Data

Mode	Base Survey	Transfer Survey	Percent Change
Automobile			
Driver	0.28	0.44	+57
Passenger	0.11	0.17	+55
Transit	0.08	0.09	+12
Walking	0.42	0.62	+48
Two-wheeled vehicle	0.40	0.73	+82
Other	0.02	0.05	-
Total	1.31	2.10	+60

TABLE 4 Trip Rates: Experimental Data

	Base Survey	Transfer Survey	Percent Difference
Person type			
Worker	1.38	2.01	+46
Unemployed	1.09	1.88	+72
Housewife	1.25	2.14	+71
Student	1.39	2.39	+72
Other	0.98	1.79	+83
Trip purpose			
Work	0.28	0.44	+57
Shopping	0.23	0.32	+39
Education	0.32	0.39	+22
Social and recreational	0.26	0.53	+100
Other	0.21	0.40	+90

reflected in the differences in mobility increases for workers and nonworkers.

A detailed comparison of the transfer survey with data from the same area in the same year extracted from the control data set suggests that the transfer survey recorded some 25 percent more trips of less than 15 min than would be expected on an average day. This result, peculiar to the survey instrument or to the days of the survey, would go some way to explaining the large differences shown in Tables 3 and 4 but would not of itself account for the entire effect.

Another result that should be noted is that in the base survey some 23 percent of the interviewees reported no travel at all on the survey day. This proportion of nontravelers is nearly four times as high as was found in the transfer survey, where only 6 percent of interviewees reported no travel at all. The levels in the control surveys were around 12 percent in all years, and levels of 10 percent for nontravelers have been found in large British travel surveys (19,pp.49-64). It appears that the large increase in mobility can be explained only by a combination of an increased number of trips per traveler and an increasing proportion of the population reporting travel.

A variety of factors could have contributed to produce this effect. The questionnaire layout and administration of the transfer survey may well have been significantly improved over the base survey. Another factor concerns the definition of the survey day; the base survey used an arbitrary 24-hr period and a portion of travel was "lost" when trips were started but not completed in that period. In the transfer survey, the interviewees were requested to complete travel diaries for a 24-hr period commencing at the time that they first left their home on the survey day, a measure specifically introduced to minimize lost travel. There is also the observation that the population in the transfer area appears to travel more than their counterparts in the base area, albeit by only a small amount. However, the weather probably did play a part, despite the evidence of Table 2 that travel patterns change little over the months of the year. The base survey was conducted during a particularly unpleasant period of winter weather, the transfer survey in an unusually hot early summer.

In summary, the principal differences between base and transfer data sets are ones of automobile and transit availability, the apparent generation of an extra number of short, slow-mode trips, and generally higher mobility rates.

Overall, the experimental populations are similar in all respects but private vehicle ownership and driver's license possession. However, their patterns of travel demand are substantially different, at least before the influences of the range of background variables incorporated in the models are accounted for. It appears that the two surveys offer quite a stern test of the transferability of the base system of travel demand models.

OUTLINE OF THE MODEL SYSTEM UNDER INVESTIGATION

A full description of the structure and component models in the Zuidvleugel model system may be obtained from the series of project reports, and an overview is presented by Daly et al. (18). In brief, the system is an application of discrete choice theory to the range of aspects of travel demand from license possession, automobile ownership, and personal travel frequency by the household to travel mode and choice of destination. It has been estimated by using disaggregate data.

The standard simplifications are made to reduce the number of alternatives to an analytically tractable level; potential elemental destinations are grouped into zones, and trips are assigned to one of a small number of categories of travel mode and destination activity purpose.

Perhaps the most significant simplification is the reduction of complex trips involving visits to multiple out-of-home destinations to a categorization according to the activity carried out at a single selected primary destination. The advantages of this simplification are primarily to ensure consistency in mode use throughout the trip and to allow network conditions on both outward and return trips to influence the overall choice of mode. The principal disadvantage is that for the 16 percent or so of trips that involve more than one destination there is some systematic underestimation of the travel generated. These points are discussed more fully in the reports of the Zuidvleugel study.

Given these approximations, travel demand is characterized as a choice of the mode, destination, and frequency of trips for each distinct purpose. The structure of the models is hierarchical, allowing for the possibility of unmodeled similarities between clusters of alternatives.

A distinctive feature of the models is their interconnection into a linked system. This linkage is extended to the longer-term decisions about household automobile ownership, inasmuch as the representation of private vehicle accessibility gains is through an expected utility (logsum) term from the travel models.

This system of models has been used intensively for the last 4 years to appraise both short-term and long-term policy proposals. Its response characteristics have been examined in detail and have been found to be reliable, at least in applications in the region from which the data were drawn.

TRANSFERABILITY TESTS

Given a thoroughly tested and apparently suitable model system and data specifically collected for the purpose, the broad aim of this experiment was to assess the practical worth of the transfer-scaling method. This approach to transferability offers the possibility of trading off model precision against the number of parameters that must be estimated; the question to be tackled was whether an advantageous trade could be made.

The experimental approach taken here is aimed at casting light on the manner in which model precision deteriorates in the progression from the complete reestimation of each parameter toward the naive transfer of the entire base model specification. Essentially, the performance of transfers involving different subsets of variables has been examined by estimating the corresponding transfer-scale parameters from the transfer survey data.

A crucial question concerns the way in which model performance is to be assessed. The usual considerations of model fitting apply here; the two most important might be characterized as validity and adequacy. In the context of the transfer, a third consideration could be broadly defined as relevance.

For the purpose of this investigation, the validity of the various models that have been fitted has been judged in terms of their performance in predicting response over a series of predefined prediction tables for various market segmentation schemes. The model predictions of the probabilities with which certain options are selected are summed over groups of individuals and compared with the proportion of times that the option actually was selected in the data base. A second consideration here is that the signs of all subgroups of variables should be in accord with common-sense a priori reasoning.

The assessment of the adequacy of the model here has been formed primarily on the evidence that a more detailed model could be supported by the data; that is, a much better model could be achieved by estimating more coefficients.

The relevance of the transfer of a particular subgroup of parameters is perhaps the simplest of all to characterize; it was deemed that if the transfer-scale parameter was estimated with reasonable precision but was insignificantly different from zero, that subgroup of variables was irrelevant in the transfer context.

Formal statistical tests can be associated with each of these three basic checks on model performance. For the prediction tables, a range of tests has been proposed by Horowitz (20) to assess the likelihood that the distribution of observations across each table is consistent with a well-fitting model. For the assessment of the support that the data provided for more complex model specifications, the familiar likelihood ratio test is available, providing that it is ensured that the successive specifications are nested. Last, the t-ratio of the transfer-scale factor, taken together with the estimated standard error, can indicate both the accuracy of measurement and the evidence for a significant difference from zero for that factor.

Not all of these and other possible statistical summaries will be presented here, quite simply because the authors do not wish to investigate the hypothesis that a transferred model is "well-fitting" in the sense of being identically equivalent to a best local model. Rather, accepting that it will be an approximation, they would like to know the extent to which it differs from the "best available" model and eventually to be able to form a judgment as to whether the decrease in model precision could be acceptable given the savings that could be made in model estimation.

Accordingly, rather than summarizing the prediction tables by a single statistic, an attempt has been made to preserve more information to indicate the overall performance of the model. In the following section, each table is reduced to three numbers. These are constructed as follows; first, for each cell in the table and conditional on the fitted model, an estimate is formed of the variability to be expected, for the given sample size, for the data points falling in that cell. For each data item contributing an element of probability, say p, to the prediction of the number of times that a particular cell is chosen, an approximate variance of that contribution is estimated as $p(1 - p)$, assuming the model to be given and correct. The variance of the overall prediction is then given as the sum of the variances of the constituent elements. The cell is then categorized into one of three groups according to whether the model prediction is (a) less than one, (b) between one and two, or (c) more than two standard errors from the observed data. In an informal sense, the more cells that can be assigned to category (a) rather than (b) and the more that can be assigned to (b) rather than (c), the better the model. In principle, if the model were well-fitting, the distribution of cells among these categories should tend to that expected for a standard normal distribution (i.e., approximately 70, 25, and 5 percent, respectively, in each of the categories). In this case correlations between the model and the data, together with misspecification in the model, will lead to departures from that pattern. Despite this, the three numbers are a simple and useful summary of the prediction tables in which the discrepancies between predictions and observations are viewed in relation to the expected accuracy of the observed data.

In practice, a number of relevant background variables were selected and for each variable, two prediction tables were formed that stratified alternatively by mode and destination area chosen. The triplet of numbers summarizing model performance was then accumulated over both tables.

SOME EMPIRICAL RESULTS

The data collected for the transfer study area have now been used to reestimate the model system fitted to the base survey. For most of the models, transfer-factor models have also been estimated. In this section, results will be presented for only four of the models in the system, although these will be discussed in the wider context of the larger number of exploratory estimations that have been performed. The models that will be presented address, respectively, the joint choice of mode and destination for shopping trips and for personal business trips and the frequency of trips for the same two purposes.

In Table 5 a sequence of transfer-factor tests for the personal business joint mode and destination model is given. This trip-purpose category contained business travel for private reasons, for example, visits to banks and insurance offices.

The explanatory variables that enter the model are listed in the first column. These have been organized into groups. The first group contains mode-specific constants. Three modes are distinguished--automobile (driver or passenger), transit (bus or train), and slow (walking or two-wheeled vehicle). The second overall group has been labeled "overall utility" and is composed of the three subgroups "level of service," "other variables," and "size variables." Within the subgroups are the elementary explanatory variables that constitute the base model. Most of these are self-explanatory; the variables "cars/license" and "female driver" are probably the least transparent. They refer respectively to a constructed variable proxying automobile competition by means of the number of licenses and

TABLE 5 Transfer of the Personal Business Joint Mode-Destination Model

					Transfer Model			
	Base Model		Reestimated Model		Partial		Complete	
Variable	Coefficient	T(0)	Coefficient	T(0)	Coefficient	T(0)	Coefficient	T(0)
Constants								
Car	-4.79	–	-5.07	–	-4.74	–	-4.39	–
Transit	-1.09	–	-0.37	–	-0.72	–	-0.67	–
Overall utility							0.83	28.7
Level of service[a]					0.82	26.3		
In-vehicle time	-.046	3.9	0.59	2.5				
Walk access time	-.067	2.8	0.96	2.7				
Waiting time	-.024	2.0	2.02	2.3				
Travel cost	-1.02	8.6	0.87	9.7				
Slow-mode distance	-.539	10.9	0.83	14.2				
Other variables[a]					0.93	10.0		
Cars/license	2.00	4.9	1.24	5.5				
Female driver	1.52	3.6	0.14	0.8				
CBD destination	0.55	2.7	2.42	4.1				
Intrazonal dummy	0.58	3.5	1.33	4.4				
Logsum car	.793	7.2	0.97	7.6				
Size variables					0.00[b]		0.00[b]	
Population	-3.39	.45	-2.56	.42				
Service employment	-2.23	.74	-1.45	.50				
Retail employment	0.00	–[b]	0.00	–[b]				
No. of observations	352		405		405		405	
Likelihood at zero	-1424		-1883		-1883		-1883	
Final likelihood	-770		-784		-804		-804	
Rho-squared (0)	.459		.584		.573		.573	

[a]See text; transfer results given as scale factors on base survey estimates.
[b]The coefficient of the last size variable (or group of size variables) is constrained to be exp(0), i.e., 1.0. Coefficients are unscaled.

automobiles in the household and a dummy variable identifying trips made by women who possess driver's licenses and are members of automobile-owning households.

The second column sets out the fitted coefficients and t-statistics for the base survey. The third, fourth, and fifth columns contain results from the transfer-scale experiment. Column 3 sets out the completely reestimated model, in which variables are entered already scaled by their coefficients in the base model. Thus, for example, the factor 0.59 for in-vehicle time should be interpreted as a coefficient of (0.59 x -0.046) on the same variable; that is, the transfer value is just some 59 percent of the base value for that variable. There are two exceptions to this rule; both the size variables and the alternative specific constants are given as unscaled values.

Looking through the level-of-service variables first, it may be seen that the transfer scales are generally less than unity, with the exception of waiting time, for which the transfer survey value actually appears rather more plausible than the base value. Among the other variables, levels are generally around 1.0, with the exceptions of the central business district (CBD) dummy (twice as highly weighted in the transfer survey) and the "female driver" effect, which appears to be absent in the transfer data.

Size variables are within one standard error or so of each other, and levels of explanation provided by the model are apparently rather higher than those in the base survey.

Overall, then, the variables that appear in the base model specification have proved, with a single exception, to have a statistically significant role in the transfer model. Transfer-scale factors are generally around unity; the highest is 2.42 and the lowest, 0.59. All signs are correct.

It should be noted that the specification of the base model contained a number of other variables that were not included in the transfer model; these were variables specific to the base region, such as constants for destination zones in specific cities in the base area. Nationally these variables may be deemed to be included in the specification, but to take zero values in the data. Their absence does not affect the other results.

A transfer-scale model is given in column 4; it involves the three major subgroups of variables in addition to the mode-specific constants. The size variable subgroup is constrained in this estimation [the use of size variables in the estimation of destination choice models has been discussed by Daly (16)].

The compound scale factors for the subgroups reflect the trends within subgroups seen in the previous model. Overall the level-of-service subgroup acquires a scale factor of 0.82, extremely well determined in statistical terms, and the "other variables" subgroup acquires a factor of 0.93, also very well determined.

There is a small decrease in the explanatory power of the model, as measured by the rho-square index (calculated with respect to zero here), but the model still appears to provide a relatively high degree of explanation. In the tables given here, the statistic rho-squared (0) is the commonly used measure of model fit computed by calculating the difference between the log-likelihoods of the fitted model and of a "null" model in which all coefficients are set to zero and dividing this by the log-likelihood of the null model. Rho-squared (C) denotes a similar statistic, but with the null model replaced by a model in which alternative specific constants are estimated from the data set, all other coefficients being set at zero. Note that the last size variable is retained with a coefficient set at 1.0 in all of these models.

Comparison of the log-likelihood values of this model with those of the fully specified model shows that a reduction of 10 scale parameters has resulted in a decrease of 20 units of log-likelihood. The chi-square 95 percent probability value with 10 degrees of freedom is 18.3. Given that (-2 x log-likelihood decrease) = 40, there is clear statistical

TABLE 6 Personal Business Joint Mode-Destination Model: Prediction Table

	No. of Standard Deviations								
	Reestimated Model			Partial Transfer			Complete Transfer		
Category	0-1	1-2	2+	0-1	1-2	2+	0-1	1-2	2+
Distance to destination	32	9	13	30	10	14	31	9	14
Arrival time	49	14	1	46	14	4	46	14	4
Age group	54	14	3	56	9	6	55	10	6
Car competition	42	6	2	40	5	5	40	5	5
Education level	49	13	1	46	15	2	47	15	1
Sex and license	36	6	0	30	8	4	31	7	4
Occupation	47	14	2	46	10	7	46	10	7
Population density at destination	26	9	5	25	9	6	25	9	6
Total	335	85	27	319	80	48	321	79	47

evidence that the more detailed model can indeed be justified on these data.

However, looking at the accompanying prediction table (Table 6) and the summarized results from the validation tests puts the decrease in model precision into perspective. Although every single prediction table has been made worse, the broad picture has not been changed greatly.

Finally, the complete transfer is given in column 5 in which all variables other than the mode-specific constants and the size variable are grouped into a single compound and given a single transfer-scale factor. Because the levels of the two separate transfer-scale factors were so similar in the previous model, this factor also turns out to be around 0.8, and to be extremely well determined in statistical terms.

Log-likelihood values and prediction tables confirm that the partial and the complete transfer models are virtually identical.

In Table 7 the same results are given for the model of joint mode and destination choice for shopping trips. The broad conclusions are similar to those drawn for the previous table.

First, looking at the completely reestimated model in column 3 where the transfer-scale factors are statistically significant (t-ratios more than 2.0), they vary around unity with highest values being 3.5 and the lowest value being 0.93. There is much more scatter in these transfer-scale factors than in the corresponding ones for the personal business model.

There are 5 statistically insignificant factors out of 11, which in this case suggests that some of the effects are absent or not well measured in the transfer context. However, comparison with column 2 shows that two of these five were also statistically insignificant in the base model, and none of the others was particularly well measured. The distance variable was included in the base specification primarily to proxy prejudices toward local facilities unconnected with relative level-of-service attributes; in the transfer survey, the effect appears to be in the opposite direction, although the evidence is not statistically significant.

The log-likelihood decrease on moving from the fully specified model to the partial transfer is 19 units, in exchange for 10 parameters as in the previous table, so that once again there is clear statistical evidence that the full specification can be justified on these data.

TABLE 7 Transfer of the Daily Shopping Trip Joint Mode-Destination Model

					Transfer Model			
	Base Model		Reestimated Model		Partial		Complete	
Variable	Coefficient	T(0)	Coefficient	T(0)	Coefficient	T(0)	Coefficient	T(0)
Constants								
Car	–6.49	–	–6.07	–	–5.36	–	–5.80	–
Transit	–3.66	–	–3.23	–	–3.03	–	–3.08	–
Overall utility							1.09	27.0
Level of service[a]					1.11	26.0		
In-vehicle time	–.055	3.7	0.93	2.5				
Walk access time	–.033	2.0	0.59	0.7				
Waiting time	–.013	1.4	3.06	1.0				
Travel cost	–.265	1.3	3.50	5.5				
Slow-mode distance	–.700	11.5	1.37	14.5				
Slow distance (8 km)	.352	2.2	2.19	9.4				
Other variables[a]					0.94	8.4		
Cars/license	1.34	4.0	1.69	5.2				
Female driver	0.77	2.9	0.47	1.4				
CBD destination	0.35	1.7	2.01	0.4				
Distance	–.103	3.0	–5.13	1.6				
Logsum car	.751	6.5	0.98	6.7				
Size variables					0.00[b]		0.00[b]	
Population	–3.08	.43	–2.74	.29				
Retail employment	0.00	–[b]	0.00	–[b]				
No. of observations	441		566		566		566	
Likelihood at zero	–1806		–2403		–2403		–2403	
Final likelihood	–896		–805		–824		–825	
Rho-squared (0)	.504		.665		.657		.657	

[a]See text; transfer results given as scale factors on base survey estimates.
[b]The coefficient of the last size variable (or group of size variables) is constrained to be exp(0), i.e., 1.0. Coefficients are unscaled.

TABLE 8 Daily Shopping Trip Mode-Destination Model: Prediction Table

	No. of Standard Deviations								
	Reestimated Model			Partial Transfer			Complete Transfer		
Category	0-1	1-2	2+	0-1	1-2	2+	0-1	1-2	2+
Distance to destination	29	4	6	23	5	11	24	5	10
Arrival time	36	9	1	35	10	1	35	10	1
Age group	48	8	3	49	6	4	51	4	4
Car completion	39	2	3	35	6	3	36	5	3
Education level	45	6	3	44	7	3	44	6	4
Sex and license	30	4	2	29	4	3	30	3	3
Occupation	42	7	3	41	8	3	41	9	2
Population density at destination	29	3	7	29	3	7	29	3	7
Total	298	43	28	285	49	35	290	45	34

The evidence of the validation tables (Table 8) is that the approximation introduced by the partial transfer does not overly distort the predictions within the various groupings of the data; only the distance segmentation shows a serious deterioration, as might be expected from the discovery that the effect in the base model is wrongly signed in the transfer. The car competition effect, proxied by the "cars/license" variable, is also somewhat distorted by the overall scale in the partial transfer, but the worsening is less serious. Other than these two effects, the reduction of the specification to the partial transfer system shows only slight deteriorations in the predictions over the different tables.

The values of the transfer-scale parameters, as in the personal business model, are both around 1.0 and highly statistically significant.

Another similarity with the personal business model is that the complete transfer in column 5 is virtually indistinguishable from the partial transfer.

Tables 9-12 set out the same form of models for the prediction of trip frequencies. Rather fewer variables are involved, and only one comparison is made--full reestimation against complete transfer.

One difference is that logsum variables were poorly identified in the base models and have been estimated separately in the transfer models to avoid distortion.

Table 9 gives the results for the personal business trip frequency model. The full set of transfer-scale factors is given in column 3; only the influence of sex on trip frequency failed to attract a statistically significant transfer-scale parameter, although the logsum variable was also statistically insignificant.

The full transfer in column 4 has exchanged three fitted coefficients for four units of log-likelihood. The 95 percent value of chi-square with 3 degrees of freedom is 7.8, so that the more elaborate model can probably be justified. However, the prediction tables indicate that there is very little difference in the model predictions. The overall scale factor, at 0.68, is quite well measured in a statistical sense.

TABLE 9 Transfer of the Personal Business Trip Generation Model

	Base Model		Reestimated Model		Complete Transfer	
Variable	Coefficient	T(0)	Coefficient	T(0)	Coefficient	T(0)
Trip	−2.25	10.6	−2.29	5.8	−2.22	6.5
Joint logsum	.064	1.6	.030	0.7	.025	0.6
Overall utility					0.69	6.6
Age 16 and under	−1.42	7.8	0.74	5.9		
Age 65 and over	−0.23	1.6	1.64	1.8		
Nonworker	0.75	5.1	1.16	5.5		
Female	0.28	2.2	−0.41	0.8		
No. of observations	3817		2439		2439	
Likelihood at zero	−2646		−1691		−1691	
Final likelihood	−1290		−974		-974	
Rho-squared (0)	.512		.424		.421	
Rho-squared (C)	.044		.027		.023	

TABLE 10 Personal Business Trip Generation Model: Prediction Table

	No. of Standard Deviations								
	Base Model			Reestimated Model			Complete Transfer		
Category	0-1	1-2	2+	0-1	1-2	2+	0-1	1-2	2+
Age of person	7	1	1	7	1	1	7	1	1
Sex/car available	4	0	0	4	1	0	3	2	0
Education level	8	0	0	5	4	0	5	4	0
Occupation/profession	6	3	0	6	3	0	5	4	0
Household size	5	3	0	4	4	0	4	4	0
Household income	5	3	0	7	2	0	8	1	0
Household cars/license	3	1	0	6	1	0	6	1	0
Total	38	11	1	39	16	1	40	15	1

TABLE 11 Transfer of the Daily Shopping Trip Generation Model

Variable	Base Model Coefficient	Base Model T(0)	Reestimated Model Coefficient	Reestimated Model T(0)	Complete Transfer Coefficient	Complete Transfer T(0)
Trip constant	−2.12	7.4	−2.61	6.9	−2.41	8.7
Joint logsum	.005	0.3	.089	2.2	.066	1.8
Overall utility					0.86	16.1
Age 16 and under	−1.85	12.1	0.81	8.4		
Age 65 and over	−0.40	3.0	0.97	2.0		
Nonworker	1.56	11.7	0.97	10.2		
Female	0.75	6.6	0.82	4.8		
No. of adults	−0.17	2.1	1.22	2.5		
Primary education	−0.46	4.3	0.58	1.8		
No. of observations	3817		2439		2439	
Likelihood at zero	−2646		−1691		−1691	
Final likelihood	−1525		−1084		−1086	
Rho-squared (0)	.424		.359		.358	
Rho-squared (C)	.148		.122		.120	

TABLE 12 Daily Shopping Trip Generation Model: Prediction Table

	No. of Standard Deviations								
	Basic Model			Reestimated Model			Complete Transfer		
Category	0-1	1-2	2+	0-1	1-2	2+	0-1	1-2	2+
Age of person	8	1	0	4	3	2	3	3	3
Sex/car available	2	0	2	4	1	0	5	0	0
Education level	5	3	0	7	2	0	5	4	0
Occupation/profession	4	2	3	6	3	0	6	3	0
Household size	6	2	0	4	3	1	4	3	1
Household income	5	3	0	6	2	1	6	2	1
Household cars/license	2	2	0	4	3	0	4	3	0
Total	32	13	5	35	17	4	33	18	5

Table 11 presents the same comparisons for the shopping-trip frequency model. All the variables included attract statistically significant coefficients, with the exception of the "primary education" variable (a dummy identifying those whose highest completed education was at the primary level). The simplification of the specification to the complete transfer exchanged only two units of log-likelihood for five parameters; in this case the simplest model appears quite adequate to explain the variation in the data. The single transfer-scale parameter, 0.86, is once again very well measured in a statistical sense.

TABLE 13 Log-Likelihood Values for Various Mode-Destination Models

	Model				
Trip	Zero	Naive	Complete	Partial	Final
Work	−3381	−2408	−2226	−2221	−2169
Social	−3092	−1523	−1499	−1482	−1354
Recreation	−2376	−949	−919	−910	−765
Shopping	−2403	−858	−824	−824	−782
Personal business	−1883	−828	−804	−803	−770
Education	−4667	−1717	−1504	−1483	−1398
Other	−3476	−1566	−1413	−1392	−1336

SUMMARY OF RESULTS FOR MODE-DESTINATION CHOICE

In the previous section attention has been restricted to the various models associated with shopping and personal business trips. Table 13 presents a summary of results for the models of joint choice of mode and destination for seven activity purposes, listing the various log-likelihood values associated with the following models:

1. Zero coefficient,
2. Naive transfer (all coefficients including the constants set at the base values),
3. Complete transfer (updating of scale and constants),
4. Partial transfer (involving two or three different functional subgroups), and
5. Final, in which all coefficients have been reestimated and the model specification has been extended to include area-specific effects.

Two points should be made about these models. First, the likelihood associated with the zero-coefficient model is somewhat approximate, having been calculated over a different sample of destinations than that used for the other models; it also differs from the null model of Tables 5, 7, and 9 in that the size variables have been excluded altogether. Second, the final model has been developed on a data base enriched with observations of travelers crossing screen lines; however, the likelihood reported in the table is calculated only for the household interview data and is comparable with the other entries.

The move from the naive transfer to the complete transfer involves the estimation of 3 parameters, from complete to partial generally involves an extra 1 (2 for work and education), and from the partial to the final an extra 20 or so. Each successive refinement of the specification is clearly justified in terms of statistical fit (as judged by the likelihood ratio test), although there is an obvious trend of diminishing returns in terms of likelihood gain per extra parameter fitted.

CONCLUSIONS

The conclusions based on the empirical evidence can be summarized under the same broad headings identified for the transferability tests.

First, in terms of the adequacy of the transfer-factor models, it was found that the performance of the model can usually be improved by reestimating on the local data, that is, by using more variables and more parameters. The transfer-factor model should be regarded as an approximation, although the nature of the two data sets (and in particular the scarcity of transit trips in the transfer survey) may actually make this model more reliable than the completely reestimated model for some variables.

Second, in terms of validity, the transfer-factor models perform worse in the prediction tests applied here, but the deterioration in performance over the reestimated model is not dramatic. For most purposes, the predictions of the transfer-factor models are quite similar to those of the most detailed specifications.

Last, in terms of relevance, all the transfer factors here were very well estimated and statistically different from zero (when compared with standard errors returned by the maximum-likelihood estimation routine). Interestingly, they were not very different from unity in most cases.

Overall, then, this experiment has provided some evidence of a stability in model specification and coefficient levels between the base and transfer surveys. This stability falls short of complete equivalence, and it would be unwise to rely too heavily on such a transfer without some corroboration or updating from local data sources. The transfer-factor approach allows variables to be scaled in blocks and appears to be a constructive way in which to adapt models to different environments.

ACKNOWLEDGMENTS

The work reported in this paper is part of a transferability study commissioned by the Transportation and Traffic Engineering Division of the Dutch Ministry of Transport and Public Works. The study was carried out by Cambridge Systematics Europe. Thanks are due to several anonymous referees for suggestions on presentation.

REFERENCES

1. P.L. Watson and R.B. Westin. Transferability of Disaggregate Mode Choice Models. Regional Science and Urban Economics, Vol. 5, 1975, pp. 227-249.
2. T.J. Atherton and M. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. In Transportation Research Record 610, TRB, National Research Council, Washington, D.C., 1976, pp. 12-18.
3. T.E. Parody. An Analysis of Disaggregate Mode Choice Models in Prediction. In Transportation Research Record 637, TRB, National Research Council, Washington, D.C., 1977, pp. 51-57.
4. A. Talvitie and D. Kirshner. Specification, Transferability and the Effect of Data Outliers in Modelling the Choice of Mode in Urban Travel. Transportation, Vol. 7, No. 3, 1978.
5. K. Train. A Validation Test of a Disaggregate Mode Choice Model. Transportation Research, Vol. 12, No. 3, 1978.
6. P.S. McCarthy. Further Evidence on the Temporal Stability of Disaggregate Travel Demand Models. Department of Economics, Purdue University, Lafayette, Ind., 1981.
7. L.A. Silman. The Time Stability of a Disaggregate Modal-Split Model. Environment and Planning A, Vol. 13, 1982.
8. F.S. Koppelman and C.G. Wilmot. Transferability Analyses of Disaggregate Choice Models. In Transportation Research Record 895, TRB, National Research Council, Washington, D.C., 1982, pp. 18-26.
9. L.A. McCoomb. Analysis of the Transferability of Disaggregate Demand Models Among Ten Canadian Cities. Transportation Development Centre, Transport Canada, Montreal, Quebec, Canada, 1983.
10. P. Tzieropoulos. Transfer Spatial et Temporel de Modeles Desagrégés du Choix Modal. In Proceedings of the World Conference on Transport Research, Studium Gesellschaft Nahverkehr mbH, Hamburg, West Germany, 1983.
11. G. Rose and F.S. Koppelman. Transferability of Disaggregate Trip Generation Models. In Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Science Press, Utrecht, Netherlands, 1984.
12. J. Supernak. Disaggregate Models of Mode Choice: An Assessment of Performance and Suggestions for Improvement. In Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Science Press, Utrecht, Netherlands, 1984.
13. M. Ben-Akiva. Issues in Transferring and Updating Travel-Behaviour Models. In New Horizons in Travel Behaviour (P.R. Stopher et al., eds.), Lexington Books, Lexington, Mass., 1981.
14. H.F. Gunn and H. Pol. Model Transferability: The Potential for Increasing Cost-Effectiveness. Presented at International Conference on Travel Behavior, Noordwijk, Netherlands, 1985.
15. H.J. Moning. The National Travel Survey in the Netherlands. Central Bureau of Statistics, Heerlen, Netherlands, 1983.
16. A.J. Daly. Estimating Choice Models Containing Attraction Variables. Transportation Research, Vol. 16B, No. 1, 1983, pp. 5-15.
17. A.J. Daly. Recent Developments in Car-Ownership and License Holding in the Netherlands. Presented at Conference of the Colloquium on Traffic Studies (CVS), Netherlands, 1983.
18. A.J. Daly, J. van der Valk, and H.H.P. van Zwam. Application of Disaggregate Models for a Regional Transportation Study in the Netherlands. In Proceedings of the World Conference on Transport Research, Studium Gesellschaft Nahverkehr mbH, Hamburg, West Germany, 1983.
19. H.F. Gunn. An Analysis of Travel Budgets into Mandatory and Discretionary Components. Proc., PTRC Summer Annual Meeting, University of Warwick, England, PTRC Education and Research Services, Ltd., London, 1981.
20. J.L. Horowitz. Testing Probabilistic Discrete Choice Models of Travel Demand by Comparing Predicted and Observed Aggregate Choice Shares. University of Iowa, Iowa City, 1983.

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

Implementation of Service-Area Concepts in Single-Route Ridership Forecasting

ALAN J. HOROWITZ and DAVID N. METZGER

ABSTRACT

Transit service area is the basis of several direct-demand models of single-route ridership. This service-area concept has not been integrated into the more recent single-route models that have evolved from the Urban Transportation Planning System and similar four-step simulation procedures. How the service-area concept was incorporated into the Transit Ridership Forecasting Model is explained and the application of the concept on routes selected from the Milwaukee urban area is illustrated. It is shown that use of the service-area concept removes a serious tendency for four-step models to underestimate ridership on marginal routes. It is also shown that proper application of the service-area concept can reduce both computation time and data requirements.

Extensive work has been done on modeling ridership for entire transit systems. Those models, notably those of the Urban Transportation Planning System (UTPS), have been in use for nearly a decade and are particularly good for long-range forecasts where there are major system changes. Parallel, but much less recognized work has been done on estimating ridership for single routes. These models are designed for short-term and mid-term planning where changes to the system are confined to only a few routes.

Early single-route ridership models were basically of two types: direct demand and special-purpose simulation. The direct-demand models (1-5) were essentially statistical or elasticity equations that estimated total route ridership on the basis of service area, dwelling-unit density, and aggregated service characteristics. They had the advantages of simplicity and explicit consideration of the existing state of the system and service areas of routes. Simulations (6) applied the same four-step approach (i.e., trip generation, trip distribution, mode split, and trip assignment) as found in the full-system models to special single-route cases, such as many-to-one and park-and-ride routes. It appeared that the four-step approach had particular advantages in terms of flexibility and accuracy across a variety of route types, so development of a new single-route model proceeded in that direction. The Transit Ridership Forecasting Model (TRFM) (7-9), the model that forms the basis of the research described here, illustrates one way of applying the four-step approach.

TRFM, like the direct-demand models, maintains consistency with service areas used for other route-planning activities. Implementation of the service-area concept permitted significant simplifications in the four-step approach. The program size is smaller than that of the Transit Operations Planning (TOP) Model (10), which is a highly sophisticated direct-demand model, yet through the windowing and focusing concept (11), it can handle routes in systems of nearly any size--even on a small microcomputer. The windowing and focusing concept was implemented with two simplifying assumptions: (a) the possibility of multiple transfer trips can be ignored and (b) small travel zones are required only on the route of interest. These assumptions were subjected to sensitivity tests (8) and were not found to introduce significant error. Implementation of the windowing and focusing concept has accuracy advantages as well. It is possible to represent the route of interest in far more detail than would be practical with full system models. Earlier tests of TRFM showed that once the model has been calibrated on a single route in a radial network, it can accurately (within 5 percent) forecast ridership on other existing routes in the same network. A detailed description of TRFM may be found in the references cited earlier.

The standard definition of a service area is the region within 0.25 mi of the route or set of routes in a system. Furthermore, the service-area concept has frequently been extended to account for natural and man-made barriers to travel (such as school district boundaries) and temporal constraints on service (such as service only during rush hours). Any such service area is arbitrary in that some of those who live outside the service areas still use transit and others refuse to walk even a short distance to a route. Nonetheless, the service-area concept persists as one of the most useful and used tools for transit route planning.

Implementation of the service-area concept is relatively easy in radial systems with little circuity. In these systems, almost all possible transit trips can be satisfied with at most one transfer. Thus it is only necessary to first determine which trips occur exclusively within the service area of the route of interest and then determine which trips go between that route's service area and the service areas of immediately connecting routes. All other possible transit trips can be ignored.

Serious problems with the service-area concept develop when there is substantial circuity in the network. Then it becomes difficult to allocate some land parcels exclusively to service areas of single routes, especially parcels near transfer points. In gridded networks, where there is substantial circuity, even single-transfer trips have alternative paths through the transit network. The foregoing simple procedure for handling service areas in radial systems fails to adequately describe transit trip making in gridded systems.

TRFM was not originally designed for large, gridded transit systems. When the model was upgraded for this type of system, it was deemed necessary to retain TRFM's simplicity of operation, to not increase the amount of data preparation, and to maintain its implementation on a small microcomputer. Much of this was accomplished by exploiting to the maximum extent the interactive graphics features of the program (9).

OVERVIEW OF THE SERVICE-AREA ALLOCATION PROCEDURE

To properly allocate land to service areas in gridded networks it was found necessary to implement a version of multiple-path trip assignment. It is interesting to note that multiple-path trip assignment has not been implemented on UTPS. Such an implementation would be difficult without placing severe restrictions on the manner in which zones can be defined, because transit riders' choice of routes depends heavily on walking distance. So that the highway and transit networks can be consistent, UTPS permits zones of arbitrary shape, that is, having parts both inside and outside service areas of routes. But for a multiple-path assignment to work well the algorithm must know where service areas of connecting routes overlap, that is, where people have a true choice between routes. This would not be possible unless zonal boundaries closely corresponded to service-area boundaries. Consequently, UTPS does not check for overlapped service areas but instead permits planners to adjust walking times to maintain a reasonable split between routes. As will be seen later in this paper, such ad hoc methods of handling multiple paths have serious consequences when ridership on routes with poor service characteristics is estimated.

On the other hand, TRFM forces planners to construct zones by following service-area boundaries and to construct networks according to a rigid set of rules. The rules are so specific that the eventual location of nodes and links in a TRFM network, as displayed on a CRT screen, contains all the information necessary to completely regenerate the underlying zone structure. If necessary, the computer can be directed to find every overlapped service area without explicitly knowing the zonal boundaries. The following example helps illustrate how TRFM networks are constructed and how the program interprets them.

Figure 1 shows a hypothetical gridded transit network with 13 routes and 42 transfer points. Route B is the route of interest, that is, the route for which ridership is estimated. A TRFM network focuses on the route of interest; immediately connecting routes are shown in far less detail. If it is assumed that riders do not make multiple transfers, the TRFM network and zone system look like those in Figure 2. The routes that do not immediately connect to the route of interest are shown as dashed lines. All trips that use the route of interest must be on this network.

There is a possibility of two paths between a point on a connecting route and a point near a transfer node along the route of interest. For example, a trip from x to y could be satisfied by following Route I and then Route B or by following Route F and then Route K. This latter path does not use the route of interest. The same is true for the reverse trip. The only people who have these two potential paths are persons traveling between the overlapped service areas of Routes B and K and the overlapped service areas of Routes F and I. This choice between two paths should be presented to only these few riders, not all persons traveling to or from zone x.

Multiple-path procedures for highway traffic assignment have been available for some time, rising from heuristic procedures of the 1960s to efficient, precise stochastic assignment procedures in the

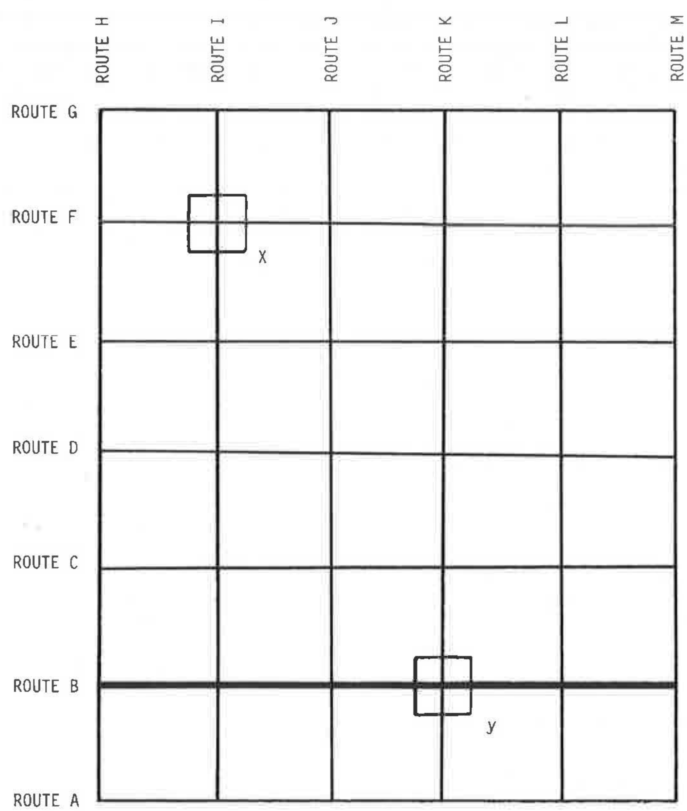

FIGURE 1 Hypothetical gridded network.

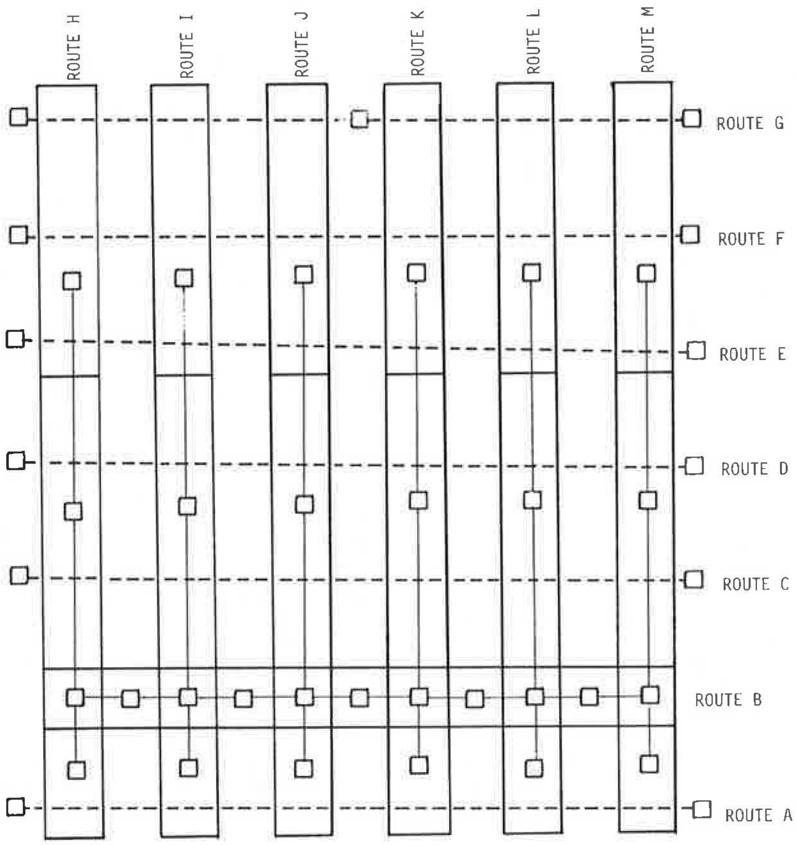

FIGURE 2 TRFM network and zone system for the hypothetical network.

1970s (12,13). But, contrary to conventional wisdom, trip assignment on transit routes is not the same problem as traffic assignment. In highway networks, the number of possible paths is huge and many of the paths have nearly identical disutilities. For transit trips, the reluctance to transfer between routes is so strong that optional paths are few. In well-laid-out transit systems, where overlaps between service areas of different routes have been minimized, the optional paths are obvious to most riders and should be obvious to a properly constructed assignment algorithm. For example, most bus trips are made without transferring, and these riders simply do not consider possible paths that involve transfers, even in rare cases where there may be small travel time savings (14-16). Many of the remaining riders find that only one path can satisfy their trip within one transfer. Very few transit riders have a choice of more than two paths. All that is really needed is a bipath algorithm. Thus, multiple-path trip assignment for transit networks must be handled differently from that for highway networks. Not only are there fewer path choices, but as will be seen, the way in which zone boundaries are defined and used becomes critical.

The bipath algorithm, in theory, is quite straightforward. It has to determine where multiple paths exist, compute the overlapped service areas, determine the fraction of trips that have a choice of two paths, and assign the trips on the basis of relative service characteristics of the two paths. However, in practice such an algorithm can become complex while dealing with all the possible shapes of zones and overlaps of service areas.

In order to easily discuss the various aspects of how the service-area concept is implemented with a bipath algorithm, it is necessary to first define a few terms. The route of interest is here called the primary network (Route B in Figure 2). The immediately connecting routes and the primary network together form the secondary network (Routes B, H, I, J, K, L, and M in Figure 2). The secondary network must be a tree; that is, it must not contain circuits. Routes that intersect the secondary network but not the primary network are called tertiary networks (any of Routes A, C, D, E, F, and G). Each tertiary network is also a tree. There may be several tertiary networks associated with a single secondary network. If all tertiary networks were to be overlaid on the secondary network and all intersections between them were to be explicitly designated as nodes, the combined network would look very much like those created for UTPS and other full-system models. However, intersections between tertiary networks and the secondary network are not explicitly indicated, so they are called virtual nodes. Sections of routes between adjacent virtual nodes on tertiary networks are called virtual links.

Tertiary networks are drawn on the CRT display in the same manner as secondary networks. Explicit nodes are drawn first and then explicit links are connected to pairs of explicit nodes. A sample CRT display with both secondary and tertiary networks is shown in Figure 3. Only the explicit links are important to the bipath algorithm. They are of three types: (a) links that always form virtual nodes at intersections with the secondary network, (b) links with a one-way characteristic that form virtual nodes but only allow these nodes to be used by trips going in one particular direction, and (c) links that do not form virtual nodes at intersections with the secondary network. Most tertiary networks are composed of links of the first type, and many of these networks have only a single explicit link. However,

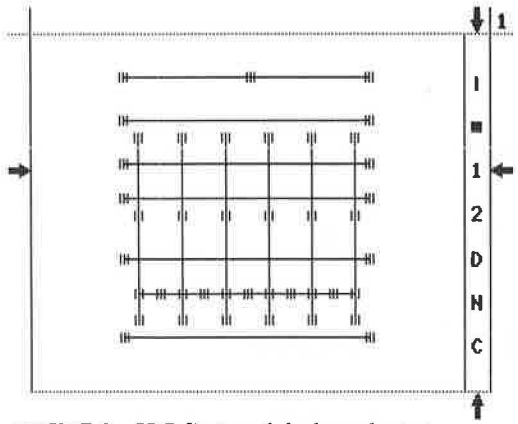

FIGURE 3 CRT display of the hypothetical network.

the latter two types of links can be used where it is clear that certain paths should be excluded from choice sets of riders.

TRFM requires that its various networks be trees for four reasons. First, a more general network configuration adds little or nothing to the accuracy of forecasts and requires substantially more data preparation time. Second, trees can be much more compactly stored in the computer's memory. Third, algorithms for analyzing trees are far faster than those for general networks. And fourth, it is much easier to regenerate the underlying zone structure of a tree.

The decomposition of general networks into trees is not an unusual practice, given that it is a standard technique of computerized network analysis. TRFM, however, makes planners aware of this process, giving them complete control over how the decomposition is to be accomplished and allowing them to optimize their data requirements around it. Because of the tree representation, TRFM is able to dispense (without sacrificing generality of network representation) with most zonal centroids, walking links, transfer links, and other artificial network elements that are required by full-system models. Equivalent network elements are generated internally; the planner need not be aware of their existence.

TRFM's capacity gives an idea of the compactness of network representation that can be achieved when networks are created by superimposing trees and exploiting service areas. TRFM runs on a 64 K microcomputer (an Apple II+/IIe), which is small by current standards. It can easily handle networks with 160 explicit nodes, 320 explicit one-way links, 800 virtual nodes, and 1,600 virtual links without memory swapping. Because TRFM requires few artificial network elements, its effective size is even larger. It can handle single-route (and even some multiple-route) problems that previously could only be analyzed on main-frame computers or large minicomputers. An implementation of TRFM on a slightly larger microcomputer would permit analysis of systems in the largest of cities.

The bipath algorithm is not applied to the roughly two-thirds of trips that occur exclusively on the primary network (i.e., within the service area of the route of interest). Of the remaining third or so, considerably less than half will be assigned to the tertiary networks and thus discarded from ridership estimates on the route of interest. This is done by (a) determining which trips can feasibly use tertiary networks by looking for overlapped service areas and (b) splitting them between the secondary and tertiary networks on the basis of service characteristics. The mathematical aspects of the algorithm will not be described in detail. Rather, the remaining portions of this paper will concentrate on discussing how well the concept works and how it may be applied to full-system simulation.

TRFM'S IMPLEMENTATION OF THE BIPATH ALGORITHM

In TRFM's specific implementation of the bipath algorithm the following six assumptions are made:

1. Boundaries of service areas along routes parallel those routes at a fixed distance, typically 0.25 mi.
2. The CRT drawing of the network is to scale.
3. Bus running time between two points following a tertiary network is identical to the bus running time between the same points on the secondary network. Consequently, the only difference between the level of service on these two paths relates to out-of-vehicle time. Of course, paths along tertiary networks that are obviously poor choices can be selectively deleted by the planner.
4. Transfer coordination does not exist at virtual nodes. The transfer time is taken to be the mean waiting time of the appropriate intersecting route.
5. Choices between paths exist only where the service areas of two paths overlap.
6. The probability of choosing a particular path is provided by a logit model as suggested by Dial (12).

The second through sixth assumptions were made to save data preparation time, and they could easily be made more rigorous if it were found necessary. The analysis described later in this paper suggests that more rigor is not necessary. In fact, a good argument can be made for further simplifications to the bipath algorithm in the single-route case.

In order to perform the bipath algorithm, the program must go through the following steps:

1. Identify link segments associated with each zone (the program must generate its own links not explicitly designated by the planner at the ends of routes);
2. Find the areas of zones;
3. Identify the tertiary networks;
4. For each virtual node, determine its location, intersecting links, directionality of intersecting links, and the closest transfer node on the primary network;
5. For each origin-destination pair, determine by using overlapped service areas the fraction of trips that can possibly use a tertiary network;
6. Split these trips between secondary and tertiary networks by comparing the various components of out-of-vehicle time, thereby determining the fraction that does not use the secondary network;
7. Compute the number of trips that use the secondary network from the fractions determined in the previous two steps and the number of trips for each origin-destination pair;
8. Resolve any double counting of trips due to multiple overlaps of service areas; and
9. Assign the trips to the secondary network and count them in total ridership.

Because of the underlying tree structure of the networks, this algorithm can be efficiently executed. Steps 1 to 8 take about 140 sec on a typical network with 200 virtual nodes.

An earlier version of the algorithm further subdivided overlapped service areas into 400 smaller parcels in order to better account for differences in walking distances to the intersecting routes. It was found that provision of this level of detail added almost nothing to model accuracy but much to computation time, so the step was abandoned.

COMPARISON WITH OTHER AVAILABLE MODELS

A direct comparison with UTPS is not possible, because multiple-path assignment has not been implemented on it. At best it can be discussed why an all-or-nothing trip assignment, which ignores service-area considerations, is inadequate for projections of ridership on individual routes. This will be done by example.

In 1981 the Milwaukee County Transit System, in cooperation with the Southeastern Wisconsin Regional Planning Commission, ran several projections of system ridership by using UTPS. They had originally hoped to use the projections for route-level planning but backed off when, in spite of all best efforts, it was found that the current year projections failed to consistently match actual passenger loads. The root-mean-square (RMS) error in route ridership was 43 percent of the average ridership, even though projected total ridership for the system was in good agreement with actual figures.

It stands to reason that multiple-path trip assignment would have somewhat improved the route-level results had such an algorithm been available. But any multiple-path trip assignment algorithm would have been greatly handicapped by a zone structure that was originally designed for highway planning. Zones were based on quarter sections and many zones were as large as full sections (i.e., 1 mi^2). Major arterials, and consequently bus routes, follow quarter-section boundaries. Thus, zonal centroids were as much as 0.50 mi from transit routes. Under these conditions it would be nearly impossible for an algorithm to determine which portions of zones are within the service areas of connecting routes.

Figure 4 shows a portion of the UTPS network in Milwaukee. The origin zone is a quarter section and is bounded by three routes: Route 60 running east and west and Routes 27 and 35 running north and south. Those who wish to reach Area E on Morgan Avenue (Route 50) have an apparent choice of three paths, even though at most two choices are available to anyone. Route 50, as will be discussed later, provides poor service, so an all-or-nothing assignment would throw all trips onto Path I. However, it is highly unlikely that persons in the hatched areas would use this path. Rather, those at A will use Path II, those at B will use Path III, and those at C and D will choose their path on the basis of relative service characteristics. Similar mistakes in allocating trips to paths are repeated at hundreds of other zones. For a route like Route 50, the mistakes are cumulative, leading to a systematic under-

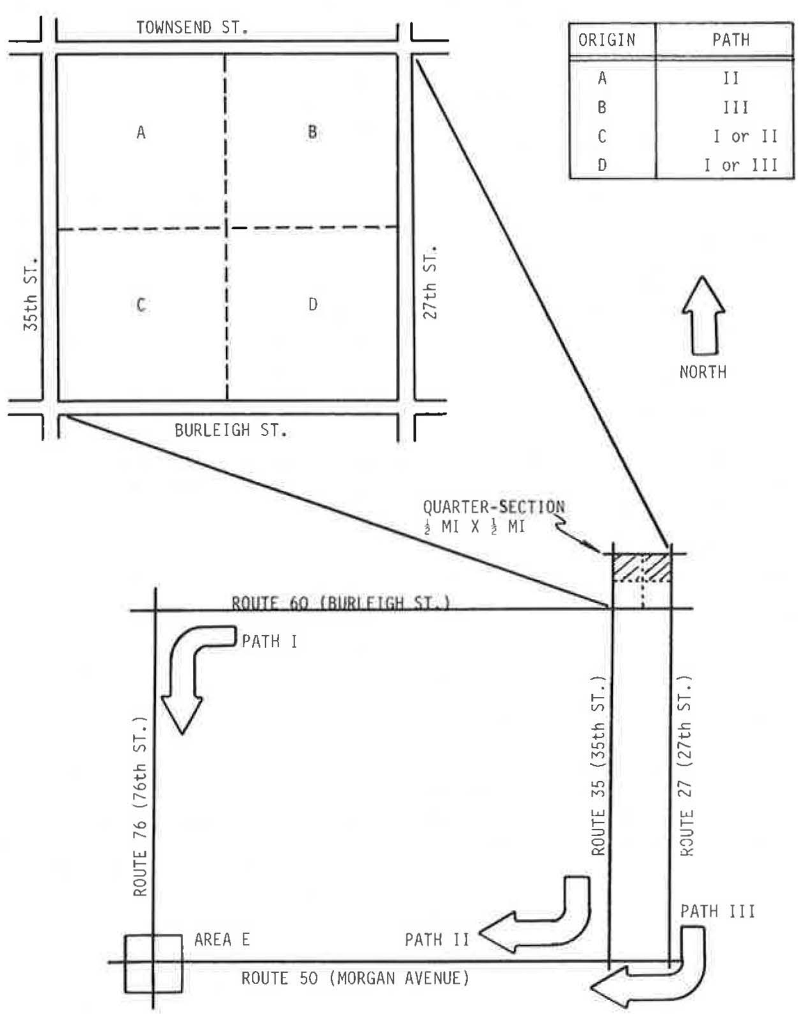

FIGURE 4 Partial UTPS network, Milwaukee, Wisconsin.

counting of ridership and a redistribution of link loads. This tendency for all-or-nothing assignment to be particularly unkind to routes with relatively poor service characteristics was repeatedly seen in the Milwaukee case study. The 14 routes with the lowest actual volumes had an RMS error of 75 percent of average ridership. Most of this error was due to underprediction. Specifically, UTPS allocated to Route 50 only 9 percent of its actual ridership and to Route 15 (a considerably larger route to be discussed later) 59 percent of its actual ridership. The magnitude of this error is such that a comparison of predicted link loads to actual link loads would be meaningless.

From the viewpoint of total system ridership these problems are of little importance. In this case the planner would only be interested in the fraction of all trips that use transit, and a misallocation of transit trips away from poorer routes would affect total ridership only if levels of service varied greatly among alternative paths. However, at the route level these misallocations are costly. For the previously cited example, because it has lower headways, Path I is allocated all of the trips, although it should be allocated at most 50 percent of possible trips.

These problems could be mitigated by a much finer division of zones, but the issue persists of what to do about overlapping service areas. An improper placement of an important trip generator could have major implications for predicted riderships on nearby routes. It is desirable that multiple-path trip assignment be implemented in full-system models, but the algorithm will not be effective unless the zone system closely matches service areas of routes. This is not as easy as it sounds because a minor realignment of a single route could necessitate a complete reconstruction of the entire zone system. TRFM avoids these problems by requiring a custom zone system for each route that is analyzed. Such a procedure would be impractical when a full-system model was used in a large city.

A more practical, multiple-path algorithm for a full-system model could be implemented by following essentially the same steps as those of TRFM. First, there would necessarily be a hierarchy of three zone systems: very fine, coarse, and medium. The very fine zone system would be used for organizing demographic data and perhaps for calculating trip generation. The coarse zone system would be needed for trip distribution and mode split and would be necessarily consistent with the highway network. The medium zone system would be used for trip assignment. The size of the medium zones would be small by current standards, because each zone has to be unambiguously assignable to the service area of a single route or assignable to the overlapped service areas of intersecting routes. On the basis of experience with TRFM, the zone would be approximately 0.25 mi^2 and would be centered on routes.

Second, trips would be assigned by order of the number of transfers required. Trips that can be satisfied without transferring would be assigned first. Almost all these trips (e.g., those with both ends exclusively within the service area of a single route) have only one possible path. The few remaining zero-transfer trips can then be split among the routes that share the two trip ends. Trips that can be satisfied with a single transfer, having at least one end exclusively within the service area of a single path, would be assigned to that path. The remaining single-transfer trips can be split among the small number (usually two) of alternative single-transfer paths. Trips that can only be satisfied with more than one transfer can be assigned to the shortest path, because there are so few multiple-transfer trips.

ILLUSTRATION OF THE SERVICE-AREA CONCEPT

The first transit system analyzed by using TRFM was that located in Racine, Wisconsin (8). The bus system in this small community was of the radial type, with little consequential overlapping of service areas. All but one route met in the downtown at a common transfer point. Because of the structure of the Racine system, all-or-nothing trip assignment was more than adequate. Calibration of the model was accomplished by varying a single parameter in the logit mode-split equation. This single parameter is a coefficient on total trip disutility that has been converted to units of in-vehicle time. It would have the same value as the coefficient on the in-vehicle time variable in a more traditionally calibrated logit equation. All other parameters were taken from other published studies that were conducted in or around southeastern Wisconsin. Once this minimal calibration exercise was performed, TRFM produced consistently good estimates of actual ridership in all the test networks. Details of the Racine tests may be found elsewhere (8).

The Racine networks could only serve as a partial test of the service-area concept. Ridership on all the routes was of roughly the same magnitude and there was too little circuity. A far more complex test for the concept, to be reported here, was judged necessary. The Milwaukee County Transit System (MCTS) provided such a test. MCTS is a gridded system with a few radial routes, emanating from the central business district (CBD). It serves a large, heterogeneous metropolitan area, with performance of its various regular routes ranging from excellent (30,000 riders per day) to poor (less than 500 riders per day). The MCTS network is as difficult a test as could be devised for a ridership forecasting model.

Two routes on opposite sides of the performance spectrum were randomly selected from the Milwaukee system. The first, Route 50 along Morgan Avenue, is at the fringe of the urban area and largely serves residences, a large high school, and a subregional shopping center. Its total ridership is low (about 1,200 riders per day), barely meeting MCTS's minimum performance standard of 22 passengers per bus hour. As might be expected, Route 50 operates on long headways, typically 30 min. The second route, Route 15, is a larger-than-average route (12,000 riders per day). It serves the heart of the CBD, the University of Wisconsin--Milwaukee, a regional shopping mall, major industrial areas, strip commercial areas, and residences. Route 15 has 19 connecting routes. Being able to estimate ridership on both Routes 15 and 50 by using exactly the same set of parameters and with a minimum of recalibration would lend strong support to the service-area concept.

To the extent possible, data were prepared according to the procedures established in the Racine case study. Only once was it necessary to deviate from those procedures--in handling trips arriving at the high school on Route 50. Milwaukee has court-ordered busing to implement desegregation in its public schools. As a result, many of Route 50's patrons are captive riders, whose distribution of origins is dictated by artificially contrived boundaries. To accommodate this situation in TRFM, a separate subnetwork was created to specifically carry these riders. The appropriate number of captive riders was forced onto this subnetwork, yielding an estimate of the number of captive riders on each link of Route 50. Consequently, the trip generation, trip distribution, and mode-split steps of the model were overridden for these riders.

Ridership was estimated for both routes by using the original Racine parameters, producing results that were approximately 35 percent too low in both cases. This underestimation was expected, having been seen in other tests on the Milwaukee system. Racine is a much smaller city, has relatively fewer individuals of the type that normally patronize transit, places less dependence on its transit systems for busing students, and has a newer transit system with a lower overall level of service. In short, people in Racine have less of a predisposition to travel by transit. Nonetheless, this completely hands-off application of TRFM to Milwaukee by using another city's parameters did significantly better on the sample routes than did UTPS, which was specifically calibrated for Milwaukee.

In order to provide a better fit to the actual data in Milwaukee, the single mode-split parameter was adjusted until the percentage error in total ridership on both routes was minimized. The other 29 parameters in the model were left alone. Because of this adjustment, there is less interest in how well TRFM predicts total ridership than in seeing how well the model represents link loads and in observing consistencies between the two routes. A good match between actual and estimated link loads can only be achieved when boardings and alightings are accurately predicted everywhere along the route.

The actual and estimated load profiles for Routes 50 and 15 are shown in Figures 5 and 6, respectively. These link loads are the average of the two directions of travel. Overall, the model matched the load profiles well. The peak load points are properly located, and the peak link loads are accurate. The RMS error in link load for Route 50 was 21 percent; the RMS error in link loads for Route 15 was 9 percent. These are small errors, considering the low level of data aggregation represented by link loads. The largest errors in link loads occur near the ends of Route 50. These errors can be attributed to the peculiar behavior of riders on routes with very long headways in cities with cold weather. Many riders who arrive at their stop early catch their bus going in the opposite direction and ride through the layover. Available on-off counts revealed the number of riders behaving in this fashion, so the double counting of these riders could be eliminated from

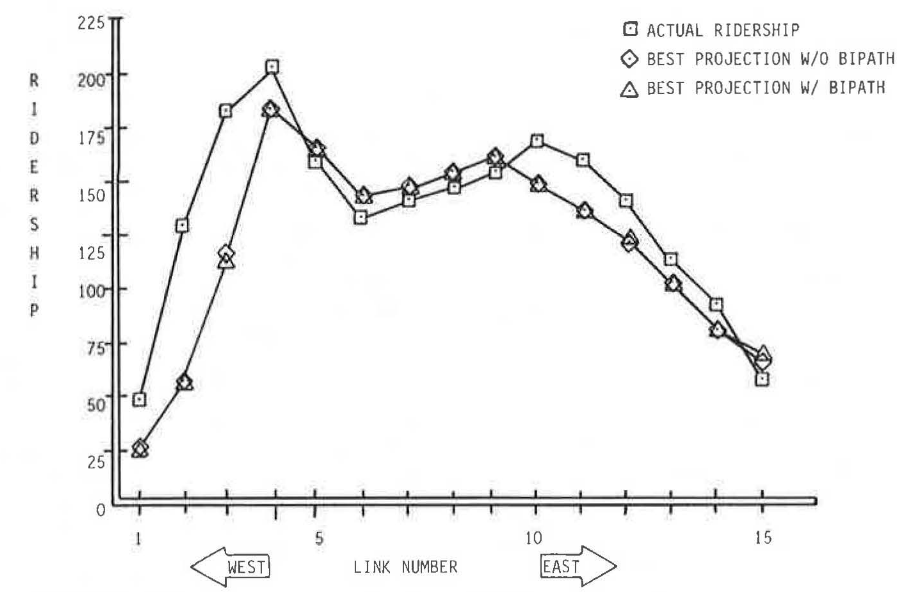

FIGURE 5 Ridership comparisons for Route 50 (Morgan Avenue).

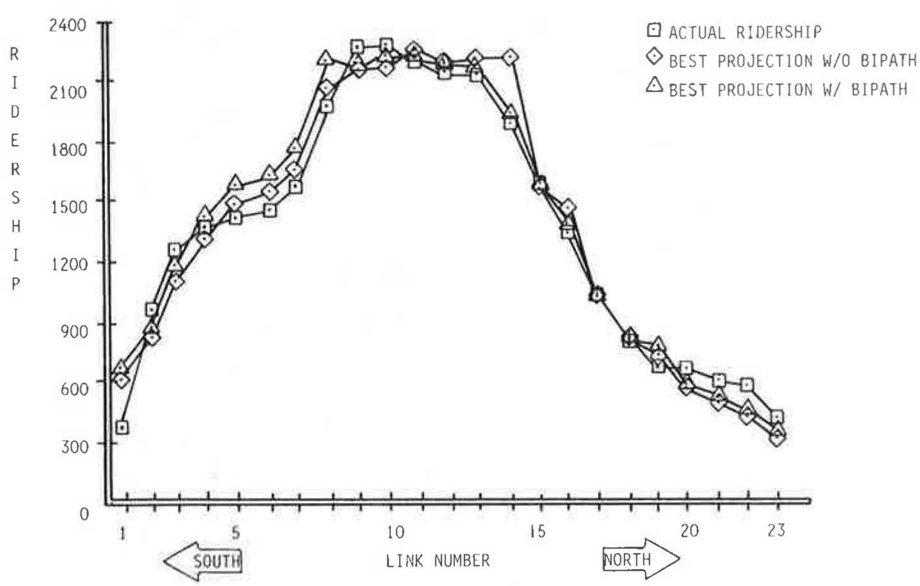

FIGURE 6 Ridership comparisons for Route 15 (Oakland Avenue).

total ridership, but it was impossible to determine where these riders actually boarded the buses. Consequently, it was not possible to properly simulate their behavior.

Of particular importance is the good fit to data on both routes by using an identical set of parameters. The model is unbiased with respect to the performance of route, demonstrating that the service-area concept is working well. Although only two routes were rigorously tested, these routes were so dissimilar that any such bias should have been obvious. As expected, the model is free from the misallocations that were so evident with all-or-nothing assignment in UTPS.

Much of the complexity in implementing the service-area concept stemmed from the need for multiple-path trip assignment. It is only logical, then, to question the importance of having this type of assignment. One way to do this is to eliminate tertiary routes from the networks and recalibrate the model by again adjusting the mode-split parameter. When this was done, the mode-split parameter increased by only 9 percent, far less than the change of 35 percent necessary to recalibrate the Racine model for Milwaukee. The load profiles (Figures 5 and 6) were virtually unchanged. On the basis of this exercise, two observations can be made. First, only about 20 percent of trips are affected by the bipath algorithm, so eliminating it does not obviously distort the results. Second, the bipath algorithm affects results in nearly the same way as an important parameter in the logit equation of the mode-split step. Thus, it would be difficult to determine the need for the bipath algorithm simply by observing how well it reproduces total ridership and link loads. The strongest arguments for the bipath algorithm are that (a) it is a logical extension of the service-area concept, (b) it is consistent with the way riders behave on transit networks, and (c) there is little extra data preparation required. The bipath assignment could be eliminated or downgraded, but at the expense of a bias in an important parameter, which could later have an unfavorable effect on forecast validity.

Routes 50 and 15 provided one additional test of the service-area concept: ease of data preparation. Both networks contained nearly every route in the Milwaukee system. Route 50 had 120 explicit nodes and 208 explicit links; Route 15 had 149 explicit nodes and 266 explicit links. Total data preparation time was approximately 2 person months. However, almost all of this time was spent reducing socioeconomic and demographic data for the entire Milwaukee urban area to a compatible level of spatial aggregation. Of course, the data previously prepared for the UTPS runs were totally useless. Each route took only about 3 days to complete once this initial preparation had been done. The portion of data specifically needed for the bipath algorithm required only about 2 hr of preparation time. This amount of data preparation is considerably less than that needed for a UTPS run, but TRFM certainly does not qualify as a quick-response technique. Data preparation time is more on the level of what would be needed for careful application of a direct-demand model.

DISCUSSION AND CONCLUSION

There has long been an inconsistency between the way in which transit planners designed service and the way in which the more sophisticated simulation models predicted ridership. This inconsistency can be attributed largely to historical accident; the simulation models were pioneered by highway engineers who found little use for the concepts of service area or transferring. In highway networks, either access existed or it did not. Where highway access did exist, highways were ubiquitous. Motorists had little impediment to switching streets in order to reach destinations as quickly as possible. The simulation models represented this situation rather well. Zones of nearly any size could be created at nearly any location and collapsed into a single point. Nodes could be placed at all arterial intersections to permit turning. And highly efficient traffic assignment algorithms could be written to find alternative paths through the network and estimate link volumes. However, transit networks behave in an entirely different way. They are neither ubiquitous nor encourage switching between routes, facts fully appreciated by transit planners.

The large number of attempts to forecast transit ridership with direct-demand models stands as a testament to transit planners' dissatisfaction with models developed for highway planning. The research reported in this paper demonstrates that there is a middle ground, at least for the single-route case. It is possible to retain the sophistication of the four-step models while giving full respect to the peculiarities of transit networks. In addition, it appears that this middle ground can exist for full-system models, too.

Unfortunately, the specific test of the bipath algorithm was inconclusive. Because of the small fraction of trips affected by the bipath algorithm in the single-route case, the algorithm would be extremely difficult to validate unambiguously in any system, including those with extensive origin-destination data. Even though an algorithm at this level of sophistication may not be essential to ridership forecasts on single routes, it is hard to imagine how a less sophisticated multiple-path algorithm could be successfully implemented in the full-system case. An accurate full-system model must at least be capable of determining which land parcels are in the service area of any given route, which land parcels are shared by routes, and what alternative paths exist for trips between parcels that are shared. This determination is made in the single-route case largely through the decomposition of the system into a hierarchy of trees; the bipath algorithm plays only a minor role in that determination.

The authors' experience in forecasting ridership by employing the service area concept confirms its utility and flexibility. Not only is it useful for simulating conventional fixed-route service, it can be extended to handle park-and-ride service, skip-stop service, downtown shuttles, and, as seen earlier, forced busing. However, implementing the concept may require a complete restructuring of existing data bases, as well as major enhancements to existing simulation models.

ACKNOWLEDGMENTS

Raw ridership data and UTPS results were provided by the Milwaukee County Transit System. Milwaukee Public Schools provided information on enrollments and district boundaries. The authors wish to thank Sam Seward, Senior Planner at MCTS, for his many insightful interpretations of the data, UTPS results, and TRFM results. They would also like to thank Ron Jensen-Fisher at UMTA for help in clarifying the role of UTPS in forecasting ridership on individual routes. This research was supported by UMTA through its University Research and Training Program and by the University of Wisconsin Extension.

REFERENCES

1. H.S. Chadda and T.E. Mulinazzi. A Transit Planning Methodology for Small Cities. Transit Journal, Vol. 3, No. 2, Spring 1977, pp. 19-40.
2. J. Hsu and V.H. Surti. Framework of Route Selection in Bus Network Design. In Transportation Research Record 546, TRB, National Research Council, Washington, D.C., 1975, pp. 44-57.
3. Peat, Marwick, Mitchell and Co. Analyzing Transit Options for Small Urban Communities--Analysis Methods. Report PPM.76.1.1. UMTA, U.S. Department of Transportation, Jan. 1978.
4. Peak, Marwick, Mitchell and Co. Simplified Aids for Transportation Analysis--Estimating Annual Ridership and Operating Expense for Fixed Route Bus Systems in Small Urban Areas. Report UMTA-IT-06-9020-79-3. UMTA, U.S. Department of Transportation, Jan. 1979.
5. J.H. Shortreed. Bus Transit Route Demand Model. In Transportation Research Record 625, TRB, National Research Council, Washington, D.C., 1977, pp. 31-33.
6. M.H. Rapp and G.D. Gehner. Criteria for Bus Rapid Transit Systems in Urban Corridors: Some Experiments with an Interactive Graphic Design System. In Highway Research Record 455, HRB, National Research Council, Washington, D.C., 1973, pp. 36-48.
7. A.J. Horowitz. Transit Ridership Forecasting: A Microcomputer Application. Transportation Quarterly, Vol. 38, No. 2, April 1984, pp. 297-310.
8. A.J. Horowitz. Simplifications for Single-Route Transit Ridership Forecasting Models. Transportation, Vol. 12, 1984, pp. 261-275.
9. A.J. Horowitz. The Design of a Single-Route Ridership Forecasting Model. In Transportation Research Record 980, TRB, National Research Council, Washington, D.C., 1984, pp. 112-117.
10. M.A. Turnquist, A.H. Meyburg, and S.G. Ritchie. Innovative Transit Service Planning Model that Uses a Microcomputer. In Transportation Research Record 854, TRB, National Research Council, Washington, D.C., 1982, pp. 1-6.
11. R.B. Dial. Demand Forecasting for New Options and Technology: Resource Paper. In HRB Special Report 143: Urban Travel Demand Forecasting, HRB, National Research Council, Washington, D.C., 1973.
12. R.B. Dial. A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration. Transportation Research, Vol. 5, 1971, pp. 83-11.
13. R.L. Tobin. An Extension of Dial's Algorithm Utilizing a Model of Tripmakers' Perception. Transportation Research, Vol. 11, 1977, pp. 337-342.
14. A.J. Horowitz. Subjective Value of Time in Bus Transit Travel. Transportation, Vol. 10, 1981, pp. 149-164.
15. A.J. Horowitz and D.J. Zlosel. Transfer Penalties: Another Look at Transit Rider's Reluctance to Transfer. Transportation, Vol. 10, 1981, pp. 279-282.
16. E. Stern. A Cognitive Scalar for the Public Transport Connection Matrix. The Professional Geographer, Vol. 32, No. 3, Aug. 1980, pp. 326-334.

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

The Usefulness of Prediction Success Tables for Discriminating Among Random Utility Travel Demand Models

JOEL L. HOROWITZ

ABSTRACT

The development of an empirical random utility travel demand model, like the development of most other statistically based models, typically includes testing and comparing several different functional specifications of the model to determine which specification best explains the available data. This paper is concerned with comparisons based on prediction success tables and indices. It is shown by example that prediction success tables and indices can lead to selection of the incorrect model when a correctly specified model is compared with an incorrectly specified one. This can happen even with data sets sufficiently large to make the effects of random sampling errors negligibly small. Accordingly, it is concluded that prediction success tables and indices should not be used for model selection. Alternative selection procedures that are both reliable and easy to use are described.

The development of an empirical random utility travel demand model (e.g., a logit or probit model), like the development of most other statistically based models, typically includes testing and comparing several different functional specifications of the model to determine which specification best explains the available data. For example, in developing a logit mode choice model, a specification in which the utility function is linear in the travel time might be compared with a specification in which the utility function is linear in the logarithm of the travel time. A variety of formal statistical procedures for testing and comparing alternative specifications of models is available (1-4). The discussion in this paper is concerned with comparison procedures based on prediction success tables and indices (3). These procedures have no formal justification, but they have greater intuitive appeal than do many of the formal procedures and therefore are attractive in practice.

A prediction success table for a model of choice among J alternatives contains J rows and J columns The entry in the (i,j) cell of the table is the number (or proportion) of individuals in the available data set who are observed to choose alternative i and predicted by the model under consideration to choose alternative j. Intuition suggests that a model with relatively large diagonal elements in its prediction success table is likely to provide a better explanation of the available data than is a model with relatively small diagonal elements because the former model gives a higher proportion of correct predictions of choice than does the latter. A single indicator of a model's prediction success can be obtained by forming a suitable average of the diagonal elements of its prediction success table. The resulting prediction success index provides an unambiguous criterion for discriminating among several models when no model dominates the others in terms of all of the diagonal elements of the models' prediction success tables.

The purpose of this paper is to show by means of examples that an erroneously specified model can have larger diagonal elements in its prediction success table and a larger prediction success index than does a correctly specified model. This can happen even with data sets sufficiently large to make the effects of random sampling errors negligibly small. Thus, contrary to intuition, prediction success tables and indices do not provide reliable means for comparing models with different specifications. Alternative comparison techniques that are both reliable and easy to use are described in the final section of the paper.

DEFINITIONS OF PREDICTION SUCCESS TABLES AND INDICES

Prediction success tables and indices were proposed originally by McFadden (3) as goodness-of-fit indicators for random utility models. A prediction success table for a model can be developed as follows. Let the available data consist of observations of N individuals who choose among J alternatives. Let P_{jn} denote the probability that individual n in the data set (n = 1, ..., N) chooses alternative j (j = 1, ..., J) according to the model under consideration. Let S_{jn} equal 1 if individual n is observed to choose alternative j and 0 otherwise. For each pair of alternatives (i,j) (i,j = 1, ..., J) define N_{ij} as

$$N_{ij} = \sum_{n=1}^{N} S_{in}P_{jn} \tag{1}$$

and define $\hat{\pi}_{ij}$ by

$$\hat{\pi}_{ij} = N_{ij}/N \tag{2}$$

Then N_{ij} and $\hat{\pi}_{ij}$ respectively represent the number and proportion of individuals in the data set who are observed to choose alternative i and predicted by the model to choose alternative j. N_{ii} and $\hat{\pi}_{ii}$ respectively represent the number and proportion of individuals who are correctly predicted to choose alternative i. A prediction success table for the model is the J x J array whose (i,j) element is either N_{ij} or $\hat{\pi}_{ij}$. Either form of the table contains the same diagnostic information, and it is a matter of convenience which is used. In this paper, it will be convenient to use the form based on $\hat{\pi}_{ij}$.

The total proportion of choices successfully predicted by the model under consideration is

$$\hat{\pi} = \sum_{j=1}^{J} \hat{\pi}_{jj} \tag{3}$$

This constitutes a goodness-of-fit index for the model. However, a better index can be achieved by averaging the differences between the proportions of correct predictions for each alternative obtained from the model and the proportions of correct predictions that would be obtained if each alternative were assumed to be chosen by each individual with a probability equal to the alternative's observed aggregate share. The resulting prediction success index is

$$\hat{\sigma} = \sum_{j=1}^{J} [\hat{\pi}_{jj} - (\hat{\pi}_{j\cdot})^2] \tag{4}$$

where

$$\hat{\pi}_{j\cdot} = \sum_{k=1}^{J} \hat{\pi}_{jk} \tag{5}$$

[Equation 4 corrects a typographical error in a previous study (3) that has the effect of exchanging the order of the subscripts on the right-hand side of Equation 5.] When $\hat{\pi}$ (or $\hat{\sigma}$) is used to compare models, the model with the largest $\hat{\pi}$ (or $\hat{\sigma}$) value is preferred to the others because this model yields the largest proportion of correct predictions in the case of $\hat{\pi}$ or, in the case of $\hat{\sigma}$, the largest increase in the proportion of correct predictions relative to the proportion implied by the observed aggregate shares.

CRITERION FOR EVALUATING USEFULNESS OF PREDICTION SUCCESS TABLES AND INDICES

In this paper, prediction success tables and indices will be evaluated according to their abilities to distinguish between correctly and incorrectly specified models. Before this can be done, it is necessary to consider the effects of random sampling errors on the ability of any statistical procedure to distinguish between correct and incorrect models and to identify a method for dealing with these effects. Random sampling error arises because different individuals with the same observable characteristics (i.e., the same values of a model's explanatory variables) and the same sets of alternatives may make different choices because of the effects of unobserved factors. As a result, the estimated parameter values, choice probabilities, and goodness-of-fit statistics for a model tend to have different values in different finite samples of individuals. These random fluctuations in estimation results can cause a goodness-of-fit statistic for an incorrectly specified model to be more favorable than that for a correctly specified model on occasion, even if the statistic usually or on the average favors the correct model. Random sampling error therefore constitutes a "noise factor" that impairs the ability of test statistics to distinguish correct models from incorrect ones.

Random sampling error always can be made negligibly small by making the sample used for estimating and testing models sufficiently large. Moreover, if the sample is large enough to make the effects of sampling error negligible, then it always is possible to determine unambiguously whether a model is correct by comparing the values of its choice probabilities for each set of values of the explanatory variables with the observed choices of individuals with the same values of the explanatory variables. A model whose choice probabilities for the available alternatives differ from the observed proportions of individuals choosing these alternatives is incorrect. Accordingly, it is reasonable to demand for comparison statistics, such as prediction success tables and indices, that they be capable of distinguishing without error between correct and incorrect models in the absence of random sampling error. In formal statistical terms, this property of a test is called consistency. Statistical test procedures that are not consistent usually are considered to be unacceptable.

In the next section, it will be shown by example that prediction success tables and indices are not consistent when used to discriminate among models. In other words, prediction success tables and indices can result in the selection of an incorrect model in a comparison with a correct one, even if the sample used for estimating and testing the models is large enough to make random sampling errors negligibly small. To show this, it is necessary to be able to evaluate the limits of the entries in a prediction success table and of $\hat{\pi}$ and $\hat{\sigma}$ as the sample size approaches infinity (large-sample limits). It follows from the strong law of large numbers that as the sample size N approaches infinity, the entries $\hat{\pi}_{ij}$ in a prediction success table approach

$$\pi_{ij} = E[Q_i(X)P_j(X)] \tag{6}$$

where $Q_i(X)$ denotes the true probability that a randomly selected individual for whom the values of the explanatory variables are X chooses alternative i (i.e., the probability according to the correctly specified model and the true parameter values), $P_j(X)$ denotes the large-sample limit of the probability according to the model under consideration that a randomly selected individual for whom the values of the explanatory variables are X chooses alternative j, and E denotes the expectation over the distribution of explanatory variables X in the population being sampled. The large-sample limits of $\hat{\pi}$ and $\hat{\sigma}$ are obtained by substituting Equation 6 into Equations 3 and 4. These limits will be denoted by π and σ, respectively.

TWO EXAMPLES OF INCONSISTENCY

Suppose that a model of choice among two alternatives (e.g., mode choice between automobile and transit) is being developed. Then J = 2, and

$$\pi_{11} = E(Q_1P_1) \tag{7}$$

$$\pi_{22} = \pi_{11} - E(P_1) - E(Q_1) + 1 \tag{8}$$

where the argument X of P_1 and Q_1 has been suppressed to simplify the notation. If $P_1(X) = Q_1(X)$ for all X (i.e., the model under consideration is correctly specified), Equations 7 and 8 become

$$\pi_{11} = E(Q_1^2) \tag{9}$$

$$\pi_{22} = \pi_{11} - 2E(Q_1) + 1 \quad (10)$$

By subtracting Equation 9 from Equation 7 and Equation 10 from Equation 8, one obtains the large-sample limits of the differences between the diagonal elements of the prediction success tables of an arbitrary model P and the correctly specified model Q. Denote the limits of these differences by $\Delta\pi_{jj}(j = 1,2)$. Then

$$\Delta\pi_{11} = E\, Q_1(P_1 - Q_1) \quad (11)$$

$$\Delta\pi_{22} = \Delta\pi_{11} - E(P_1 - Q_1) \quad (12)$$

Now suppose that models P and Q yield the same predictions of the aggregate shares of alternatives 1 and 2. Then $E(P_1 - Q_1)$ and

$$\Delta\pi_{11} = \Delta\pi_{22} = E\, Q_1(P_1 - Q_1) \quad (13)$$

Equivalently,

$$\Delta\pi_{11} = \Delta\pi_{22} - E\{[Q_1 - E(Q_1)](P_1 - Q_1)\} \quad (14)$$

Finally, suppose that in addition to satisfying $E(P_1 - Q_1) = 0$, P_1 has the property that

$$P_1(X) = \begin{cases} 1 \text{ if } Q_1(X) > E(Q_1) \\ 0 \text{ otherwise} \end{cases} \quad (15)$$

In other words, model P assigns individuals deterministically to alternative 1 if $Q_1(X) > E(Q_1)$ and deterministically to alternative 2 otherwise. Model P is misspecified because $P_1(X) \neq Q_1(X)$ whenever $Q_1(X)$ differs from 1 or 0. However, it can be seen from Equation 14 that $\Delta\pi_{11} > 0$ and $\Delta\pi_{22} > 0$. Therefore, if the sample size is large enough to make random sampling errors negligibly small, the diagonal elements of the prediction success table of the erroneous model P will exceed the corresponding elements of the prediction success table of the correct model Q. Similarly, the goodness-of-fit indices $\hat{\pi}$ and $\hat{\sigma}$ will be larger for model P than for model Q when the sample size is sufficiently large. Thus, the prediction success tables and indices will lead to selection of the wrong model in large samples and are inconsistent. The following example illustrates this result numerically.

Example 1

In a model of mode choice between automobile and transit let mode 1 be automobile and mode 2 be transit. Let the correctly specified model be

$$Q_1(T) = 1/[1 + \exp(-0.1T)] \quad (16)$$

where T denotes transit travel time minus automobile travel time in minutes. Let the distribution of T in the sampled population be uniform on the interval [-10,10]. Then $E(Q_1) = E(Q_2) = 0.5$ in this population, and $E(Q_1^2) = E(Q_2^2) = 0.27$. It follows from setting $P_i = Q_i$ in Equation 6 that the large-sample limit of model Q's prediction success table is

$$\text{Table}(Q) = \begin{bmatrix} 0.27 & 0.23 \\ 0.23 & 0.27 \end{bmatrix} \quad (17)$$

The values of π and σ for model Q are $\pi(Q) = 0.54$ and $\sigma(Q) = 0.04$.

Now define the misspecified model P by

$$P_1(T) = \begin{cases} 1 \text{ if } T > 0 \\ 0 \text{ otherwise} \end{cases} \quad (18)$$

Then $E(P_1) = E(Q_1)$, and $E(P_1Q_1) = 0.31$. It follows from Equation 6 that the large-sample limit of model P's prediction success table is

$$\text{Table}(P) = \begin{bmatrix} 0.31 & 0.19 \\ 0.19 & 0.31 \end{bmatrix} \quad (19)$$

The values of π and σ for model P are $\pi(P) = 0.62$ and $\sigma(P) = 0.12$. Thus, if the sample size is sufficiently large and models Q and P are compared by using their prediction success tables or their π- or σ-values, the erroneous model P will be accepted and the correct model Q rejected. This is true despite the fact that model P yields predictions that can be both unreasonable and highly erroneous. For example, suppose that T = 1 for a certain population group (i.e., transit travel time exceeds automobile travel time by 1 min). Then model Q yields the result that 48 percent of the members of this group use transit, whereas model P yields the unreasonable and erroneous result that no members of the group use transit.

Example 1 shows that the use of prediction success tables and indices for model selection can lead to selection of an erroneously specified model and rejection of a correctly specified one. However, the erroneous model P used in this example cannot be obtained through maximum-likelihood estimation, which is the standard method for estimating empirical choice models. This suggests the possibility that prediction success tables and indices may discriminate correctly among models when the sample size is large if consideration is restricted to models that can be obtained through maximum-likelihood estimation. The next example shows that even when this restriction is imposed, prediction success tables and indices can select the wrong model.

Example 2

As in Example 1, let individuals choose between the modes automobile (mode 1) and transit (mode 2). Let the correctly specified model be given by Equation 16. Assume that the values of T in the sampled population are restricted to those shown in Table 1 (e.g., because the sample is stratified) and that each of these values occurs with probability 1/9. Let the erroneous model be specified as

$$P_1 = 1/[1 + \exp(-\alpha C)] \quad (20)$$

where α is a positive constant and C is the cost of transit travel minus the cost of automobile travel in dollars. Assume that in the sampled population, there is a unique value of C associated with each value of T (e.g., because of the stratification procedure that is used) and that the C-values corresponding to the T-values are as shown in Table 1.

TABLE 1 Values of Explanatory Variables for Example 2

T (min)	C ($)	T (min)	C ($)
−80.0	−1.00	10.0	0.29
−60.0	−0.97	20.0	0.52
−20.0	−0.52	60.0	0.97
−10.0	−0.29	80.0	1.00
0.0	0.0		

The large-sample limit of the prediction success table of the correct model Q is

$$\text{Table (Q)} = \begin{bmatrix} 0.4046 & 0.0954 \\ 0.0954 & 0.4046 \end{bmatrix} \qquad (21)$$

The π- and σ-values of model Q are π (Q) = 0.8092 and σ (Q) = 0.3092. The large-sample limit of the maximum-likelihood estimate of α, which can be computed by using methods described elsewhere (4), is 4.2877. The large-sample limit of the prediction success table of the erroneous model P can be obtained from Equation 6 by using Equation 16 to evaluate the Q probabilities and Equation 20 with α = 0.2332 to evaluate the P probabilities. The result is

$$\text{Table (P)} = \begin{bmatrix} 0.4060 & 0.0940 \\ 0.0940 & 0.4060 \end{bmatrix} \qquad (22)$$

The π- and σ-values of model P are π (P) = 0.8120 and σ (P) = 0.3120. It can be seen that the prediction success tables and π- and σ-values all favor the erroneous model P. Although the differences between the prediction success tables and π- and σ-values of the two models are small, a comparison of the models based on any of these criteria will lead to acceptance of the erroneous model and rejection of the correct one if the sample size is large enough to make random sampling errors negligible. As an example of the prediction errors that can result from selection of the incorrect model, suppose that transit improvements cause T to decrease from 20.0 to 10.0 for a certain population group while C remains unchanged. Then model Q yields the result that transit ridership in this group increases by 126 percent, whereas model P yields the result that there is no change in transit ridership.

DISCUSSION

The examples presented here show that prediction success tables and indices are unreliable means for discriminating among models. They can result in acceptance of an incorrect model and rejection of a correct one, even when the sample used for estimation and testing is large enough to make random sampling errors negligible. Because, as will now be discussed, comparison procedures that do not have this deficiency are readily available, prediction success tables and indices should not be used for model selection.

The appropriate procedure to use for comparing two models depends on whether the models are nested or nonnested. Two models are nested if one model can be obtained from the other by assigning appropriate values to the latter model's parameters. In nonnested models, this cannot be done; given the values of either model's parameters, it is not possible to choose values of the other model's parameters so that the two models become identical. Models P and Q in Example 1 are nested because P can be obtained from Q by setting the coefficient of T in Q equal to $-\infty$. Models P and Q in Example 2 are nonnested. See the discussion by Horowitz (1) for further examples of nested and nonnested models.

Comparisons of nested models can best be carried out by using likelihood ratio or t-tests (3). In a comparison of a correctly specified model with an incorrectly specified one, these tests always select the correct model in the absence of random sampling error (i.e., they are consistent). When random sampling error is present (as it always is in practice), likelihood ratio and t-tests have high probabilities of selecting the correct model when a correctly specified model is compared with one that is seriously erroneous (4). Likelihood ratio and t-tests are easily implemented because they rely on information that is virtually always included in the outputs of computer programs used for estimating random utility travel demand models.

Nonnested models can be compared easily by using the likelihood ratio index statistic modified to account for the effects of any differences in the numbers of estimated parameters in the models being compared (1,2). Like the likelihood ratio and t-tests for nested models, comparisons based on the modified likelihood ratio index are consistent and with samples of practical size, where random sampling error is present, have high probabilities of selecting the correct model when a correctly specified model is compared with a seriously erroneous one (1,2). Comparisons based on the modified likelihood ratio index can be implemented by using information that is included in the outputs of existing computer programs for estimating random utility travel demand models.

ACKNOWLEDGMENT

The research reported in this paper was supported in part by an assistance agreement between the Urban Mass Transportation Administration and the University of Iowa.

REFERENCES

1. J.L. Horowitz. Evaluation of Usefulness of Two Standard Goodness-of-Fit Indicators for Comparing Non-Nested Random Utility Models. In Transportation Research Record 874, TRB, National Research Council, Washington, D.C., 1982, pp. 19-25.
2. J.L. Horowitz. Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models. Transportation Science, Vol. 17, 1983, pp. 319-350.
3. D. McFadden. Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. In Behavioural Travel Modelling, (D.A. Hensher and P.R. Stopher, eds.), Croom Helm, London, 1979.
4. J.L. Horowitz. Identification and Diagnosis of Specification Errors in the Multinomial Logit Model. Transportation Research, Vol. 14B, 1980, pp. 331-34.

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

Travel Demand Forecasting with the Quick-Response Microcomputer System: Application and Evaluation of Use

G. SCOTT RUTHERFORD and NORMA T. PENNOCK

ABSTRACT

A description, application, and evaluation of the quick-response microcomputer system (QRS) are presented. QRS is found to be a well-programmed computer version of the manual techniques presented in NCHRP Report 187. The system is, with few exceptions, easy to understand and operate. Observations about QRS operation include the following: (a) Some data files must be laboriously reentered every time the model is run; (b) screen prompts and written documentation sometimes fail to give sufficient guidance; (c) the gravity model output is never transformed into an origin-destination matrix, although it is labeled as such; (d) the mode-choice model has several undesirable features, the most troublesome being the lack of an explicit transit penalty, making calibration difficult; (e) the software may be so easy to apply that it can be used without much thought; and (f) the best application of QRS might be for local traffic analysis and not for corridor or regional studies.

In 1978 a two-phase research effort of NCHRP culminated with publication of two volumes in the NCHRP Report series. Phase I consisted of identifying travel-related urban policy issues and assessing existing methods and procedures that could be used to respond to these issues quickly. The results of Phase I were presented in NCHRP Report 186 (1). Phase II included developing a User's Guide "to describe transferable parameters, factors, manual techniques, and the like, to enable the user to carry out a simplified [travel demand] analysis without the necessity of reference to other sources" (2). Phase II results were published in NCHRP Report 187 (2). These reports will henceforth be referred to as NCHRP Reports 186 and 187.

In the years following its publication, NCHRP Report 187 has become a popular reference and planning tool (3) along with such standards as Characteristics of Urban Transportation Demand (4), Characteristics of Urban Transportation Systems (5), and the Institute of Transportation Engineers' trip generation manual (6).

Some years after publication of NCHRP Report 187, FHWA established funds for the creation of microcomputer software that would incorporate the report's quick-response techniques for travel demand forecasting. The resulting software, called the quick-response microcomputer system (QRS), was released into the public domain in February 1984.

The following description of QRS is not intended as a substitute for the user's manual; rather, it is meant to be a supplement. This paper provides a summary of QRS format and application; however, readers who wish to use QRS should begin with the study of NCHRP Report 187 (2) and the QRS user's manual (7).

HARDWARE AND SOFTWARE REQUIREMENTS

QRS has been written and compiled in two versions, one for the Apple microcomputer and the other for the IBM PC. In either case, the computer must be configured with a minimum of 64K random access memory (RAM), two disk drives (for 5 1/4-in. floppy disks), and a video monitor capable of displaying 24 lines and 80 columns. A printer is optional because users can copy output from the monitor. However, copying is a tiresome procedure, so a printer is strongly recommended. Users of the Apple II+ or Apple IIe also need a language card.

QRS is written in UCSD PASCAL, and its programs run within the UCSD p-system. Users must purchase the UCSD p-system and use the file management facilities of the p-system to run QRS. Purchase of the full p-system represents a substantial software investment; a "run-time" version may be purchased more cheaply, and it has all the capabilities required to operate QRS. However, users of the run-time version will not be able to modify files or programs.

All the modules within QRS are menu-driven; that is, a list of possible responses is presented for the user's choice. The following paragraph is excerpted from the user's manual (7):

> The QRS has been designed to be user friendly. Users direct the system by selection of functions from a menu. Ease of data entry has been incorporated in the system's basic design. After review of this manual and NCHRP Report 187, the system should provide sufficient prompting information at the screen to allow operation with minimal reference to printed material.

In its present form, application of QRS is subject to upper bounds on certain parameters. These include

- Maximum number of zones, 50;
- Maximum number of assignment links, 800; and
- Maximum travel time, 40 min.

GENERAL DESCRIPTION

Trip Generation

QRS calculates trip productions based on the number of dwelling units in each zone. Users are free to use their own trip rates or to use QRS default values. Attractions are calculated for each zone according to default equations or user-specified values.

After users have established production and attraction rates, they must provide zonal data. For each zone, the following data are required:

- Average income or average automobile ownership per household,
- Retail employment,
- Nonretail employment, and
- Total dwelling units.

Trip Distribution

The gravity model in QRS is the traditional formulation. Required input for each of the three trip purposes includes the following:

- Productions and attractions for each zone,
- Travel time/friction factor relationship,
- Intrazonal travel times, and
- Interzonal travel times for each ij interchange.

Users may input productions and attractions directly or may recall the file saved after trip generation. Travel times may be entered directly for each interchange. Alternatively, QRS will calculate interzonal travel times, given the following data:

- Zone type for each zone [central business district (CBD) or suburb],
- X and Y coordinates for each zone centroid (measured in inches),
- Map scale (miles per inch),
- Circuity factor to convert airline distance to over-the-road distance, and
- For each interchange, the following percentages: (a) distance in CBD, (b) distance (a) on arterials, (c) distance in central city, (d) distance (c) on arterials, (e) distance in suburbs, and (f) distance (e) on arterials.

Users may elect to use default values for friction factors, which are available for each of the four population groups. If users have area-specific factors, these may be directly entered into the gravity model. In either case, the program forces users to begin with a travel time of 1 min and to increase by 1-min increments up to a maximum of 40 min. Corresponding friction factors are needed for each of three trip purposes.

Mode Choice

QRS calculates mode split based on previously saved files of person trips and travel times and new data required for calculation of impedances.

QRS employs travel times calculated for trip distribution as in-vehicle times for automobile impedance. In-vehicle time for transit, excess time for both modes, and travel cost for both modes are calculated based on the following new data:

- For each origin zone the following: automobile occupancy, income, automobile access time, and walk or drive time for transit access;
- For each destination zone the following: parking cost and time to walk to destination after leaving automobile or transit; and
- For each origin-destination (OD) interchange the following: transit speed, circuity, headway, fare, and transfer time.

These values may be held constant for all interchanges, may be unique for each interchange, or may be constant for all but a selected group of interchanges. The user should know that travel times calculated for trip distribution include OD terminal times. If the travel time file is used without modification for mode choice, special care must be given to definition of automobile access times.

Traffic Assignment

The QRS traffic assignment model is basically a bookkeeping function. Users supply trip tables that were created in the trip-distribution or mode-choice models. Any number of trip tables may be combined so long as the tables are all of equal size. Normally users combine the tables for the three trip purposes and make a single assignment of total daily traffic.

Summary

Figure 1 shows the required data for and output from the four basic modules: trip generation, trip distribution, mode choice, and traffic assignment.

APPLICATION OF QRS

The authors used QRS to forecast the demand for light rail transit (LRT) in Spokane, a medium-sized city in Washington State. Figure 2 shows the study corridor. In keeping with the time and budget constraints of the feasibility study, and in view of the fact that the study was one of only feasibility versus a complete-alternatives analysis, the authors decided that the LRT forecast could best be accomplished with QRS.

The Study Area

The maximum number of zones that QRS will accommodate is 50; therefore it was necessary to aggregate the region's 286 zones into new groupings. The results of aggregation are shown in Figure 3. Forty-six districts were formed; the four extra ones were intended to be used as external zones.

Trip Generation Results

QRS default trip rates and default trip purpose shares were used as a starting point for calculation of productions and attractions. The results were compared with local totals for the three trip purposes in each of the 46 zones. An iterative process was then used to adjust the production rates for each income category until the resulting production and attraction zonal totals were acceptably close to the local totals.

Trip Distribution

One major input to the gravity model was a travel time matrix, representing minutes of in-vehicle and out-of-vehicle time for every automobile interchange

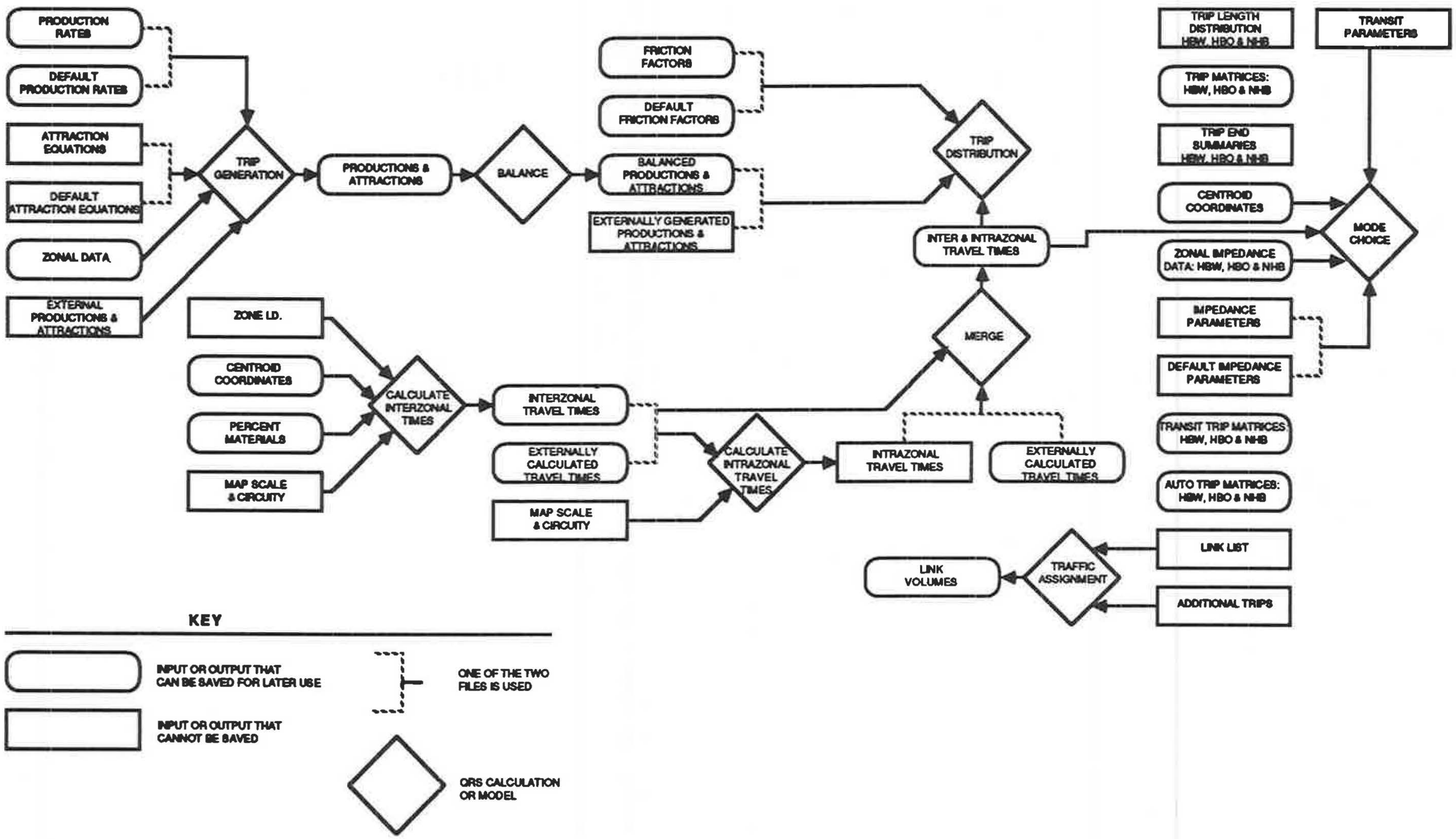

FIGURE 1 QRS: flowchart.

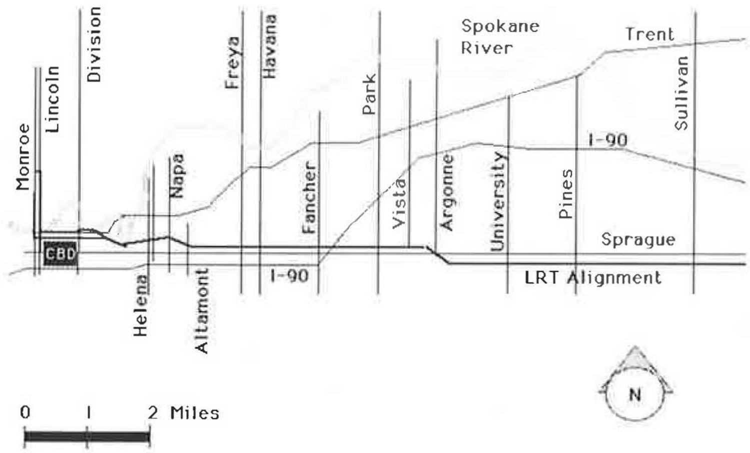

FIGURE 2 Spokane, Washington, East Valley Corridor.

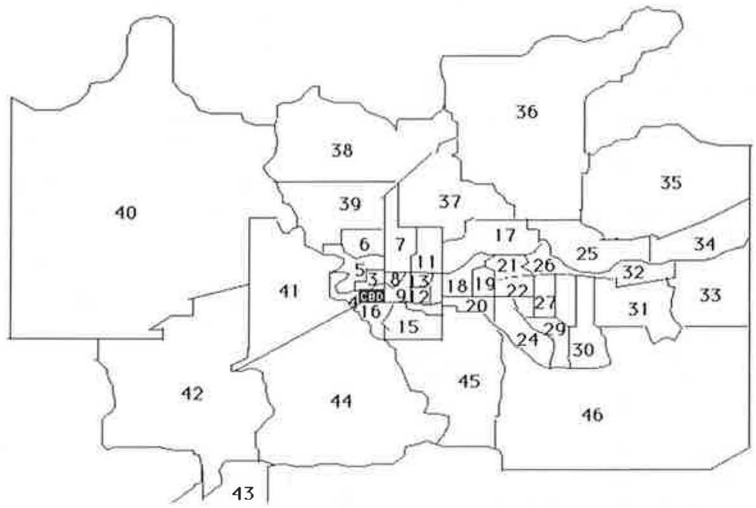

FIGURE 3 Spokane region: zones for LRT study.

in the 46-zone region. Centroids were assigned to each zone on the basis of demographic density; a land use map was used to estimate the most likely center of activity for each zone. Each centroid was assigned X-Y coordinates (in inches) so that QRS could use centroid coordinates and a user-supplied circuity factor to convert airline distance to over-the-road distance. Actual driving distances for several interchanges were compared with the distances that resulted with a circuity factor of 1.22 (the QRS default factor), and the comparison indicated that the default factor was acceptable.

In addition, each zone was assigned to one of three zone-type categories: CBD, central city, or suburban. Zones 1 and 2 were CBD, zones 3 through 16 were central city, and the remaining zones were suburban.

In a region of 46 zones, there are 2,116 possible interchanges. This number can be reduced to 1,058 because travel time in one direction is assumed equal to travel time in the opposite direction. The number can be further reduced to 1,012, because QRS calculates the 46 intrazonal travel times on the basis of interzonal times. For each of the 1,012 interchanges remaining, it was necessary to estimate both the portion of the trip in the CBD, central city, and suburbs and the portion of the trip on arterials and freeways.

These estimates were made by using Spokane street maps. Consideration was given to Spokane's unique topography in which there is a limited number of river crossings. Consideration was also given to the location of freeway interchanges.

Friction factors for the gravity model were adapted from factors used by local agencies. Plots of travel time versus friction factors were made for each of three trip purposes, and smooth curves were drawn through the data points. The factors used for the LRT study were taken from the curves.

Trip length distributions for 1980 were not locally available, but 1980 census data included work trip information. Census records indicated that the 1980 home-to-work trip in the Spokane region averaged 18.4 min, whereas the QRS average home-to-work trip was 18.0 min. This was considered to be an acceptable match, given the difference in calculation methods.

Mode Choice

Three major data sets had to be prepared as input to the mode-choice model. The first data set was OD data for each of the 46 zones. The format of the resulting file is given in Table 1. A similar file

TABLE 1 Format of QRS Mode-Choice OD Data File

Zone	Origin Zone Data: Automobile Occupancy	Income ($000s)	Automobile Walk Time (min)	Transit Time (min): Walk	Transit Time (min): Drive	Destination Zone Data: Parking Cost (cents)	Automobile Walk Time (min)	Transit Walk Time (min)
1	1.3	3.0	3	10	0	100	3	10
2	1.3	3.0	3	10	0	100	3	10
3	1.3	4.0	1	7	0	0	1	7
4	1.3	5.0	1	10	0	0	1	10
5	1.3	7.0	1	7	0	0	1	7
6	1.3	8.0	1	7	0	0	1	7
7	1.3	7.0	1	7	0	0	1	7
8	1.3	6.0	1	10	0	0	1	10
9	1.3	3.0	1	10	0	0	1	10
10	1.3	5.0	1	10	0	0	1	10

had to be created for each of the three trip purposes.

Transit walk times were estimated by measuring over-the-road distances on a base map, and the distances measured were from zone centroids to the nearest bus line. In very large zones, these transit walk times were increased to allow for the fact that a small portion of the zone's residents actually lived within practical walking distance of a bus line.

The second major data set required was a matrix of automobile travel times. The travel time matrix used in trip distribution was created based on certain assumed trip end conditions, including automobile access times.

The third major data set was one that described transit parameters for each interchange. Bus schedules and a route map from the Spokane Transit Authority (STA) were used to establish transit speed, transit circuity, transit fare ($0.50 for all interchanges), transit headway, and transfer time for each interchange in the 46-zone region.

Transit speeds were calculated for individual routes by measuring distances between time points on a transit route map and using route schedules to determine time between the points. These calculations showed an average system speed of 14 mph, which agreed with average speeds reported by STA (8). In the QRS application, routes with higher speeds were thus credited.

Transit headways were examined for individual routes during the morning peak period [for home-based work (HBW) trips] and during midday [for home-based nonwork (HBO) and non-home-based (NHB) trips]. It was determined that a 30-min headway was typical of the system as a whole.

A file of centroid coordinates was required as mode-choice input. Unfortunately, the coordinate file used in trip distribution could not be used for mode choice because mode-choice coordinates must be expressed in miles rather than in inches.

The final input to mode choice is a set of parameters used by QRS to convert time and cost to impedance units. These parameters, and the values used in the Spokane forecast, include the following:

- Weight for excess time, 2.50;
- Income to value-of-time factor, 0.33;
- Automobile operating cost, $0.08 per mile (which represents a behavioral cost);
- Model exponent for HBW trips, 1.55;
- Model exponent for HBO trips, 2.40; and
- Model exponent for NHB trips, 2.15.

Calibration of the mode-choice model was achieved through successive iterations. Adjustments were made to the model exponents, transfer penalties, and transit access times until the results were acceptably close to 1980 STA ridership figures. Two ridership measures were used to judge the QRS results:

- Total STA weekday ridership in 1980, which was 24,360; and
- Average 1980 weekday ridership on Valley routes, which was 3,450 (estimated as 97 percent of 1982 ridership, because 1980 daily ridership was 97 percent of that in 1982) (8).

The QRS mode-choice model was considered to be calibrated when the following results were achieved:

- Region weekday ridership equalled 24,400 (desired total, 24,360), and
- Valley route weekday ridership equalled 3,490 (desired total, 3,450).

Forecast

Future-year demographic forecasts were made available by local agencies. Year 2000 was the target for the LRT feasibility study; for each of the 286 zones, data on total households, total employment, and total group quarters population were available for the year 2000. These data were aggregated to correspond to the 46 zones used in the LRT study, just as had been done with 1980 data.

For the year 2000, total employment was split between retail and nonretail categories in the same proportion as had been observed in 1980. Exceptions were made for zones where total employment was forecast to change significantly. The zonal data were used in the calibrated trip generation model, and the output productions and attractions were entered into the calibrated gravity model.

Mode choice for the forecast year involved creating a rough design of the proposed LRT line with an attendant feeder bus service. A base map was prepared that included the LRT route as approved by the feasibility study steering committee. North-south bus routes in conceptual form were added in the Valley, but Valley bus routes with east-west orientation were assumed defunct except for one express route. Bus routes outside the Valley were assumed to be unchanged, both in coverage and in scheduling.

The same data files that had been prepared for mode-choice calibration had to be prepared for the 2000 forecast. Each of 1,012 interchanges was examined separately to identify the operating characteristics that were appropriate. Transit speed for each interchange was calculated by means of measuring distances on a base map and assuming transit speed to be a function of the share of the trip made by LRT versus the share made by bus. Determination of the interchanges that could fairly be assigned to LRT was necessarily a matter of judgment.

In the final analysis, ridership forecast for the proposed LRT fell short of the minimum criterion of 900 passengers per peak hour and peak direction established by the feasibility study's steering committee. The forecast resulted in a figure of 6,579 daily LRT riders, which was converted to 610 peak-hour peak-direction passengers according to local transit information (12.3 percent of daily total trips = peak-hour total; 75 percent of peak-hour total = peak direction total) (8).

EVALUATION OF THE USE OF QRS

QRS is not, and was not intended to be, a scaled-down version of the Urban Transportation Planning System (UTPS). It was designed to be a computerized application of the techniques presented in NCHRP Report 187. Those techniques were established for manual travel demand forecasting; QRS has therefore inherited all the methodological shortcuts and shortcomings that were included in NCHRP Report 187. QRS was intended to allow the user to make a forecast faster than manual procedures would permit, but it was not intended that the QRS forecast would be better than a manual one. Therefore, it is not the authors' intention to criticize the inherent methodological structure of QRS. Instead, this paper concentrates on the authors' opinions about the QRS goal to be a "user-friendly" system with "ease of data entry" (7). However, some additional comments are provided.

The following criteria were applied in evaluating QRS:

- Is QRS user friendly?
- Does QRS incorporate ease of data entry?
- Is QRS quick?
- Is QRS responsive?
- When is QRS appropriate?

The User-Friendly System

All the programs within QRS are menu-driven. That is, the user is always presented with a list of numbered alternatives from which to choose. A single keystroke corresponds to each alternative, and QRS instructs the user to press the appropriate key. When the program begins (i.e., when the disk is "booted"), the QRS insignia appears along with a request to hit the ENTER key to proceed. Next appears a menu of all the QRS forecast modules: trip generation, trip distribution, and so on. The user (with a single keystroke) chooses a module, and a menu for that module appears. This menu contains a numbered list of the steps within the program.

Additional menus appear within each element of a program as data are required. The numbered options allow the user to recall previously saved files, to enter data directly, or to exit.

Finally, when a data file is being viewed, a menu of file management options appears. This menu provides single-keystroke options for editing, printing, saving, and scrolling. Many of the data files are structured so that a row of data corresponds to each analysis zone, and each row often has five or more entries. A convenient feature of QRS is the option to edit these files one entry at a time or a whole row at a time. Users may proceed more efficiently than if the option were not available.

A criticism of the QRS format relates to recall of previously saved files. The only ways that users can view a listing of files stored on the data disk are to view them before booting QRS or to exit QRS and reenter the operating system. If users are unfortunate enough to forget the name of a file needed in the middle of a program run, the only way they can view the necessary file directory is to reboot the disk. This is annoying because booting with the UCSD p-system is incredibly slow. However, worse than losing time is the fact that any portions of the module already executed will be lost; booting requires that the whole process begin anew.

In general, QRS fulfills its promise to be user friendly. The only fault the authors find in this regard is the lack of a more convenient way to view file directories. In addition, although most of the QRS menus include HELP as an option, very few of the HELP message files can be found. Most requests for help remain unanswered.

Ease of Data Entry

In one sense, users will find data entry easy. At most points in the program where data must be supplied a menu directs users to create a new file or to recall a previously saved file. Usually, on-screen instructions are worded so that users will have no doubt about how to enter the required data. For example, within the mode-choice part of the program users will be required to create a file of OD data. QRS prompts will be the following:

- Enter zone number (enter 999 to exit),
- Input values for zone separated by spaces.

In cases where on-screen prompts are not clear, the user's manual (7) usually provides the necessary information. However, there are exceptions worth noting. First, in the mode-choice model, destination parking cost must be provided. The user should know from NCHRP Report 187 that one-half a trip's parking cost should be charged to each half of the trip. The user probably will not know, because neither the on-screen prompt nor the documentation explains, that the data will be divided by 2 when impedance is calculated.

Second, and also within the mode-choice module, a file on X-Y zonal centroid coordinates must be supplied. The documentation and the prompts do not point out that these coordinate measurements must be in miles, not inches. Users will be tempted to reuse the file of coordinates created for the trip distribution module, but that file is in inches and should not be used in mode choice (unless the base map's scale was 1 in. = 1 mi).

Third, within the trip distribution module, the user must supply intrazonal travel times. The on-screen prompt leads users to believe that these times must be directly entered each time the gravity model is run. If a file of travel times was created and saved during the first gravity model run, the intrazonal times were saved as a part of the same file. On subsequent gravity runs when users are asked to input intrazonal travel times, they should elect option 2 (direct input). QRS will retrieve and display the intrazonal times from the previously saved file.

Fourth, the user is never told that circuity supplied as a transit parameter within mode choice will also be applied to the X-Y zonal coordinates for calculation of automobile travel times. This means that alteration of transit circuity cannot be used as a means of testing improved transit service.

Most of the required data files may be saved and recalled as needed for iterative applications of QRS. This is essential because calibration of the models will no doubt require several, if not many, trials, especially with the gravity model and the mode-choice model. It is extremely unfortunate that certain data files must be directly input for each iteration.

A minor example is the trip generation module, in which users must always enter the attraction equations, because QRS reverts to the default values when users exit the program. This does not constitute a significant problem, but users should be careful to keep records of the equations.

A more serious example is the mode-choice module, in which users must input for every interchange in the region five transit parameters: speed, circuity, headway, transfer time, and fare. If users are working with a region that has uniform transit coverage and performance, they may enter these parameters once and they will be held constant for all interchanges. However, most likely a truer representation of the transit system will be achieved only if some of these parameters are varied for some of the interchanges. The designers of QRS realized that this would be the case and built in an option that allows the user to intervene in impedance calculations for any or all interchanges.

To intervene, users must first enter a list of all OD pairs requiring intervention. If users are working with a sizable number of zones, the list of interchanges requiring intervention may be long. The problem is that users cannot save this list; it must be entered every time. Furthermore, not only must users enter it for every iteration of the model, but for every trip purpose (HBW, HBO, and NHB trip tables are analyzed separately for mode choice).

After users supply this list of interchanges, QRS will calculate impedances for each interchange. QRS progresses in typical OD matrix order, beginning with 1-1, 1-2, 1-3, ..., 1-n and ending with ...n-n-1, n-n. Whenever QRS comes to one of the interchanges requiring intervention, it pauses and prompts users to enter the five transit parameters. These data also cannot be stored; users must enter them directly for each trip purpose and for each iteration.

The inability to save these data represents a serious flaw because mode-choice calibration will no doubt require several iterations, and the time required to directly enter the list of interchanges and transit parameters can be extensive. For the Spokane LRT forecast, 41 zones were accessible to transit, creating 1,681 interchanges (41 x 41). Of these 1,681, intervention was necessary for several hundred. A single run of the mode-choice model took 4 1/2 hr at the terminal (1 1/2 hr for each of three trip purposes).

In summary, QRS provides ease of data entry in most cases. The exceptions to the rule are not numerous, but they are significant.

Speed and Responsiveness

The QRS documentation states, "The basic approach to quick response does not rely on coded transportation networks . . . . Considerable time is saved by not coding networks" (7). QRS does not use coded networks because NCHRP Report 187 did not use them and the QRS designers merely programmed the manual techniques. This means that users will not necessarily save "considerable time." Assembling the necessary data for impedance and travel time calculation is an onerous burden requiring that each interchange be examined individually so that the parameters may be identified.

A fair estimation of the time required to make a forecast with QRS under normal circumstances is difficult. The authors' best estimate is that a person already familiar with QRS and general forecasting procedures and with ready access to the required data could accomplish a regional or corridor analysis in 4 to 6 weeks.

An evaluation of QRS responsiveness is related to the other criteria: user friendliness, ease of data entry, and speed. QRS is responsive to users' needs if it rapidly evaluates alternative scenarios, and indeed once the QRS models are calibrated, they fairly rapidly examine forecast results on the basis of varying data. For instance, transit assignment resulting from a fare decrease could be calculated in just a few hours. However, more complex scenarios, perhaps involving major alterations in transit network structures, would require more time; determining impedance parameters would take several person days.

Appropriateness

Given the burden of continued entry and reentry of data and the nature of models that may rely on default values, the authors believe that QRS is most appropriate for local traffic studies. For example, a regional shopping center traffic study or major residential development could easily be handled with QRS. However, for regional or corridor studies with greater than 25 zones, a network-based system would be preferred.

The mode-choice model has no way to add a transit penalty separate from other parameters such as access time. This capability is needed for the calibration process. At last report, FHWA was developing a new mode-choice model, and until it is available the current one should not be used for transit patronage estimation because it is so difficult to calibrate.

SUMMARY

In general, the QRS designers have succeeded in producing software that computerizes the techniques presented in NCHRP Report 187. The program's most serious deficiencies involve its inability to save extensive files of data, so that a great deal of time is required to enter them. These deficiencies could be easily remedied, and it would also be easy to add a protocol to total trips in the mode-choice output matrices. Other elements of QRS that are perhaps undesirable concern the lengthy process of establishing parameters for transit impedance and travel time calculation. These elements cannot properly be considered flaws exclusive to QRS; they were inherited from the NCHRP Report 187.

QRS, in its present form, is a tool that can be useful for rapid calculation of zonal productions and attractions. If users already have access to appropriate travel times, as would be the case in any regional planning conference, the gravity model could also be quickly applied. However, the authors believe that the mode-choice model, in its present form, is not a particularly useful tool. It is cumbersome and may require that input be less than ideally logical. Personnel at the Transportation Systems Center are reportedly planning to incorporate a different mode-choice model in the future.

In summary, the authors established four criteria with which to judge application of QRS. The results of this evaluation are as follows:

1. Is QRS user friendly? Yes. With few exceptions, users will have no trouble understanding and applying the software.
2. Does QRS incorporate ease of data entry? Sometimes. Users are generally well-informed about

the actual procedures to follow during data entry, but the fact that some lengthy files cannot be saved and reused is a significant failing.

3. Is QRS quick? Yes and no. The authors believe that users could complete a regionwide forecast in 20 to 30 person days, given readily available data and prior knowledge of QRS. Hence, QRS is faster than a mainframe UTPS-type forecast. On the other hand, estimating some of the QRS input parameters requires days of person effort, and trip matrices must be manually summed. In general, QRS is quick, but not as quick as it should be.

4. Is QRS responsive? Sometimes. A calibrated QRS can respond quickly (in hours) to some program or policy data changes. Examination of other alternatives might require days or weeks of effort.

Potential users of QRS should be familiar with the techniques presented in NCHRP Report 187 before deciding to use QRS for a travel demand forecast. The short-cut methods contained therein--and in QRS--are not applicable to all situations.

REFERENCES

1. Comsis Corporation. Travel Estimation Procedures for Quick Response to Urban Policy Issues. NCHRP Report 186. TRB, National Research Council, Washington, D.C., 1978.
2. Comsis Corporation. Quick-Response Urban Travel Estimation Techniques and Transferable Parameters: User's Guide. NCHRP Report 187. TRB, National Research Council, Washington, D.C., 1978.
3. J.W. Dickey. Metropolitan Transportation Planning, 2nd ed. McGraw-Hill Book Company, New York, 1983.
4. Wilbur Smith and Associates and DeLeuw, Cather and Company. Characteristics of Urban Transportation Demand. U.S. Department of Transportation, April 1978.
5. DeLeuw, Cather and Company. Characteristics of Urban Transportation Systems. U.S. Department of Transportation, June 1979.
6. Trip Generation: An Informational Report, 3rd ed. Institute of Transportation Engineers, Washington, D.C., 1983.
7. Comsis Corporation. Quick Response System (QRS) Documentation. U.S. Department of Transportation, Jan. 1984.
8. Transit Monitoring Program: Summary Report for June-December 1982. Spokane Regional Planning Conference, Spokane, Wash., Jan. 28, 1983.

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

Using Regional Forecasting Models of the Urban Transportation Planning System for Detailed Bus Route Analysis

PETER R. STOPHER, WILLIAM A. DAVIDSON, LAIKY K. TAMNY, and GARY S. SPIVACK

ABSTRACT

Planning bus system operations has traditionally relied heavily on the acquired knowledge of bus system planners and has been one of the last areas of transportation planning to be computerized. There are a number of programs available to design the allocation of drivers and vehicles to a bus system, but the planning process still lacks a detailed capability to determine the desired headways and short lines based on the levels of demand for service. Conventional wisdom has held that the regional forecasting models, based on zonal-level analyses of trip generation, trip distribution, mode split, and assignment, are too aggregate and too coarse to permit them to be used to assist in such planning efforts. There is no question that these models are coarse and aggregate. However, this paper demonstrates that they are still sufficiently realistic and accurate to be used for bus route planning at a line-by-line level and that for large bus systems they may be much better suited to the planning issues involved than any other available methodology. Some specific requirements that the models must meet to be used in this manner are described. A procedure is detailed for producing bus system statistics from the standard planning models of the Urban Transportation Planning System (UTPS) and it is shown how this procedure can be used in conjunction with the UTPS procedures to undertake detailed long-range planning of a bus system. The capability of the procedure to produce data that accurately reflect the base year is shown to be considerably greater than that normally associated with aggregate travel-forecasting models. The capability of using the procedure to refine a long-range bus system is demonstrated in a case study from the Los Angeles area, and this shows that the procedure has the capability to provide clear indications of a variety of improvements to the efficiency of the planned bus network.

Planning bus system operations has traditionally relied heavily on the acquired knowledge of bus system planners and has been one of the last areas of transportation planning to be computerized. In the present state of the art, there are a number of programs available to design the allocation of drivers and vehicles to a bus system [e.g., RUCUS and HASTUS (1-4)], each of which works on a line-by-line basis and is capable of determining an efficient, although probably not optimal, allocation of both drivers and vehicles. These tools allow a system to put into practice the service configuration that has been determined from other considerations, for example, changes in service levels to meet demand and changes that may be indicated to reduce operating costs.

This planning process still lacks a detailed capability to determine the desired headways and short lines based on the levels of demand for service. Perhaps more important, bus system planning has been undertaken only on a short-range basis with any degree of detail. Long-range planning of bus system configurations has not been attempted to a large extent, even though part of the planning of future long-range capital investment in transit should consider the implications for fleet size and system operation. Conventional wisdom has held that the long-range regional forecasting models, based on zonal-level analyses of trip generation, trip distribution, mode split, and assignment, are too aggregate and too coarse to permit them to be used to assist in such planning efforts. There is no question that these models are coarse and aggregate. However, the authors believe that it can be demonstrated that they are still sufficiently realistic and accurate to be used for bus route planning at a line-by-line level and that for long-range future planning they may be much better suited to the planning issues involved than any other available methodology. There are some specific requirements that the models must meet to be used in this manner, and there will remain a need for a significant level of professional judgment to be applied to the final results. Nevertheless the basic conclusion is that the models are capable of assisting the planning process at this level of detail and particularly for long-range planning applications.

In the remainder of this paper the goals of this procedure, the steps required to develop models that are of sufficient accuracy to be used in this manner, and a computer program (URAP) that works with the Urban Transportation Planning System (UTPS) (5) models to produce the information needed for route-level planning are described. A case study of the application of the procedures is given to demonstrate how the procedure can be used to refine service levels that would feed that next step of the process--the development of run cuts and schedules.

GOALS OF THE PROCEDURE

The primary goals of this procedure are to be able to develop changes in bus route service levels that are consistent with demand for bus service and known elasticities of demand and to provide a basis for estimating changes in service levels required for long-range planning purposes well beyond the time frame usually associated with detailed bus route planning. It is also intended to systematize the trial-and-error procedures that are more likely to be used in long-range planning for bus service changes and to provide a bus system design that represents one possible system to meet demand at a reasonable level of efficiency and with prespecified policy requirements for service.

DEVELOPMENT OF THE TRANSIT NETWORK

Fundamentally, the route-analysis procedure consists of refining the transit network description in terms of both the transit lines themselves (deletion of transit lines and definition of short-line operations) and changes in headways. Therefore, it is important that the transit network be built to provide as realistic a simulation of the actual transit system as possible. If the transit network is not a careful, realistic simulation, planning of bus route revisions will necessarily be too inaccurate to be useful, and one would also need to question the degree of inherent accuracy in any individual line loadings. However, in case it should be construed that the levels of accuracy indicated here are required only to enable the route-analysis procedure described in this paper, it should be stressed that most transit networks are not built with adequate attention to detailed realism and are likely to provide misleading results for any long-range planning application. The level of accuracy described here is necessary to the route-analysis application, but it should be achieved in any case for realistic transit planning of any description.

Ideally, two network definitions are needed--one for the morning or evening peak and one for the midday period. For each of these, the network description will show the appropriate average headways for the period, the average speed or travel time on each link of the network in each period, and line descriptions over the network for each bus line and significant subline operation. A number of aspects of the development of the transit network are worth reviewing, because they have a marked effect on the accuracy and realism of the network and because they include a number of judgmental aspects of transferring actual bus lines into state-of-the-art transit network methodology.

A careful study of existing bus system operations is required initially, ideally focused on defining the headways. It is recommended that average headways be developed by counting the actual numbers of trips made on a bus route during each of the peak periods and in the base (midday) period. Average headways are defined by dividing the length of the period (peak or base) by the number of trips in that period. Counts of trips should be made from published schedules or the schedules used to generate work pieces for driver assignments. In either case, a control point should be defined for each bus route and the count of trips made at that point. In the process, and by looking at the beginning and ending points of each bus trip, the analyst can gain a rapid definition of the alternative short-line operations that are scheduled. These can then be developed into a definition of the sublines. The authors have found that it is preferable to define the count in the direction for each peak for which the trip count is maximized and to define the subline operations that occur in that direction only.

It is often suggested that the limitations in state-of-the-art networks make it best to define each bus route as two one-way routes. However, many problems are generated by such a definition, including a lack of ability to build in layover required to maintain headways on round trips and overestimation of the total number of vehicles required to provide the service. Because of this, the authors have found it preferable to define routes as two-way unless this makes it absolutely impossible to produce a reasonable simulation of a route. Even when a route traverses some segments of one-way streets, generally in the central business district (CBD), it is preferable to define the route as a two-way route and define a special two-way transit link on the one-way streets. Provided that care is taken in connecting walk links to nodes on such a two-way street segment, the resulting bus line will usually provide the most realistic simulation. Of course, genuine one-way routes, such as express bus services that operate only in the peak direction, should be defined as one-way routes. These should generally be the only such routes in the network, however.

Care is needed in defining the ends of bus lines that may make a loop, because state-of-the-art transit networks will generally prohibit a line from crossing itself. Coding to the midpoint of the loop, along one side of the loop, or around the loop to end at the division point are each possible strategies that should depend on the size of the loop and the extent to which appropriate connections from zone centroids can be provided along the nodes on the loop.

Subline operations often present serious coding problems, particularly when there are a large number of such operations on some lines. As a general rule, and bearing in mind that the goal of the network building is to provide a basis for demand forecasting, the sublines should be defined from the viewpoint of the bus user rather than from the viewpoint of the operator. Thus, suppose that A, B, C, and D are four points along a bus line with A being the beginning of the line and D the end of the line. The bus operator may run the following operations: a bus starting at A, driving to C, returning to B, proceeding to D, and then returning to A. The mirror image of this operation may also be scheduled, beginning initially at D with short runs to B, C, and finally A. This scheme may potentially define as many as eight distinct bus trips (A to C, A to D, B to C, and B to D and all the return trips). At a minimum, given the limitations of state-of-the-art transit networks, four two-way lines would need to be defined for this. However, most bus riders will perceive that the route offers three different headways: one between A and B, one between B and C, and one between C and D. Because relatively few riders will ride from one end to the other of the line, most bus riders will be unaffected by the fact that some buses do not offer service on the entire length of the line. Therefore, the line may, if headways are identical along AB and CD, be defined as two lines: one from A to D and one from B to C. If all operations are of the form described earlier, the line from A to D will have the base (lowest) headway, and the line from B to C will have half the base headway, because all buses traverse this portion of the line. Similar reductions in complexity of sublining need to be made for demand modeling to present an effective simulation of the bus system.

The speeds and distances on the network must also be scrutinized carefully. For a number of reasons, it is rarely possible for an aggregate coded network to produce line descriptions that exactly match the times and distances known to occur in reality. A more serious problem often arises because there is a lack of good information on the true times and distances on specific lines. In transit properties that have made a substantial number of recent route modifications, there may be fairly reliable information on the actual running times from beginning to end of a line, but no reliable information on the length of the line. Nevertheless, replication of true distances and times in the network description is most important to use of aggregate network tools to assist in defining service needs. The authors suggest that the line-by-line distances and times be checked carefully and that any falling outside a 10 percent error margin be reexamined carefully. This reexamination should consist of checking the network output reports to see if there is any link in evidence on which there is either a disproportionately long or short travel time compared with the length of the link. Second, careful examination is required of the line description to make sure that there are no "tunnels" or airline links that violate the geometry of the underlying street system.

Once a sufficiently accurate network description has been achieved, it is important to examine the results of a base loading of the network. For this, the entire travel-forecasting procedure must be run with base-year data. In examining the results, there can be no substitute for the person in most transit properties who has an encyclopedic knowledge of the system and its current loadings. Maximum loadings on each line should approximate fairly well the known maximum load points, both in location and volume. Further, the pattern of loading along the line should be reasonably close to reality. A long initial or final segment of line that runs empty in the assignment when buses are actually running with 20 to 40 passengers per bus would be an obvious indication of problems in the line description or the connections from the zone centroids to the transit network. There is considerable potential for error in transit assignments resulting from a poor choice of walk and automobile connectors to the transit network, so these assignments must be considered prime candidates for modification if loads are not found to match reality. In addition, it is most important that lines that share a common segment of street be described identically in terms of the network nodes traversed. In standard assignment procedures that share patronage on a common street segment among all the lines on that segment, a common street segment can be recognized only by absolute identity in segment descriptions. If the option to have loads split between common line segments in proportion to service frequency is used and common segments are not found to have proportional loadings, the fault is almost certainly in the lack of identity in the coding. The same applies for any other proportioning of the loads.

It cannot be overemphasized that a significant amount of time is needed to ensure that all such errors and problems are resolved and removed from the network. Any one of these errors will compromise the use of the network to assist in line planning. However, the accuracy achieved in this process is also necessary for many other aspects of long-range planning.

THE UTPS-COMPATIBLE ROUTE-ANALYSIS PROGRAM

In UTPS, there are two alternative transit network procedures. Although new applications are encouraged to use the newer INET procedure, which is built from the highway network and reflects existing highway loadings in determining bus speeds on shared right-of-way, many existing planning agencies use the older UNET procedure. UNET is based on an independent transit network, and congestion on the streets can be reflected only by a manual adjustment to the transit running speeds or times. The route-analysis program described in this section was developed principally to work with UNET networks. It can be used with INET networks, but some of its features are unnecessary in that context, because INET contains some capabilities that the route-analysis program was designed to add to UNET.

The UTPS-compatible route-analysis program (URAP) provides four primary features:

- Addition of several elements of line descriptions that add to the realism of the descriptions,
- Determination and reporting from the assignment of a number of statistics that are not available from standard reports,
- Computation of several alternative estimates of line-by-line service levels, and
- A capability to impose some service modifications and determine their effect on the system requirements.

It is beyond the scope of this paper to provide a detailed description of each of these features, details of which may be found elsewhere (6). A list of the tables provided by URAP is provided in Table 1. A brief description of the features is provided in this section. Computationally, URAP is extremely simple and involves primarily only the organization of data already available from the ULOAD assignment of trips to the transit network. In addition, there are no assumptions involved in the URAP program that are a function of size of the region, percentage of trips on transit, or size of the transit property or properties operating in the region.

TABLE 1 URAP Output Reports

Report No.	Contents	Optional?
1	Input global parameters	No
2	Input annual parameters	No
3	Input policy headway values	No
4	Line record information	Yes
5	Maximum load point summary	Yes
6	Operating statistics	Yes
7	Compressed operating statistics	Yes
8	Summary of total operating statistics	Yes
9	Annual statistics	No
10	Undefined headway values	No
11	Excess passenger demand summary	Yes
12	Operating cost model statistics	Yes

Primarily, URAP operates by taking certain user-provided inputs and using these to modify ULOAD data or compute additional statistics from the ULOAD and user-provided information. ULOAD assigns transit trips to the transit network and generates loadings by line and by link for the transit network. If steps preceding the use of ULOAD split transit trips into peak and base average hourly loads and ULOAD is run for each of the time periods with relevant trip tables, the output information available to URAP consists of assignments of transit person trips by link and line for each of an average peak hour and an average base hour.

URAP allows the user to specify vehicle capacities that can vary by up to 10 types and where each vehicle type can have a different peak and base capacity (allowing for standees in the peak but not in the base). Factors can be used that indicate the number of hours of peak and off-peak service in a weekday, a Saturday, and a Sunday and the number of weekdays, Saturdays, and Sundays operated in the year (thus allowing specification of Saturday or Sunday service on certain public holidays). Other factors can be used to apply to the trip tables loaded on the transit network to convert results to one peak hour and one midday hour. URAP also requests the user to specify the layover at each end of the line, which can be input as a percentage of the one-way trip time or as an absolute number of minutes. In subsequent computations, URAP adds to this the number of minutes required to increase the sum of the one-way trip time and layover to an integer multiple of the headway (to allow the bus to return on the route at the same headway). Thus, if a line operates on a 20-min headway and has a one-way trip time of 57.6 min with layover specified as 10 percent, URAP redefines the one-way trip time as 63.4 min (57.6 + 5.8) and then requires the bus to lay over for a further 16.6 min to reach a multiple of the 20-min headway. Circuity factors are also available.

Among the special reports that URAP offers are the four highest links on each line listed by the node pair (in a directional sense) and provided for each of the selected peaks (a.m. or p.m.) and the midday. These are obtained by reading the loaded legs files from ULOAD and involve no computation. In addition, URAP reports the daily and annual vehicle miles and vehicle hours of travel for the transit system by company; this involves using the line miles and hours from ULOAD, circuity factors (if any) input by the user, the factors for expansion from input trip tables to annual data, and (for vehicle hours only) the amount of layover specified by the user. Because URAP has no information on deadhead time and distance, these are revenue vehicle hours and miles.

The most useful aspect of URAP is the set of service-level alternatives provided. Four scenarios are described: coded, loaded, nominal, and modified. The four scenarios are each accompanied by similar information. Under the coded scenario, the program lists the headways as coded into the network and shows the maximum load, the vehicle requirements for each of the peak and base periods, and the daily vehicle miles and vehicle hours implied by the headway and trip information. If the network was built in UNET, the trip time and distance information will necessarily vary from that produced by the network and the assignment, because of the addition of layover, and any user-specified local circuity. If the network was built by using INET, there may be little or no difference. The numbers of vehicles are calculated taking into account the length of each period. Thus, if there is a bus route that takes 3 hr and 17 min in the peak for a round trip, including layover, and the peak is defined as 3 hr, the vehicle requirements will be 3 hr divided by the headway (because no vehicles can run a second trip). On the other hand, if another route has a round-trip time in the peak of 2 hr and 43 min and a headway of less than 17 min, at least one bus can run a second trip. This is taken into account in determining vehicle requirements. On each line, the vehicle miles and vehicle hours are estimated for each of the two periods selected. The program also prints out the maximum load on the line and indicates whether this represents an overload.

In the loaded scenario, URAP calculates the headways needed to provide sufficient service to fill the buses by using capacities provided by the user for each vehicle type. In this case, five vehicle types can be used, corresponding to the five transit modes allowed in the coding of UTPS networks. Thus, vehicle type 1 corresponds to mode 4, vehicle type 2 to mode 5, and so on. If mode 4 is specified as local bus with a peak capacity (including standees) of 65 passengers, the maximum hourly peak load is divided by 65. The result of this calculation is the number of buses per hour required to service the peak demand. Dividing this number into 60 min provides the peak loaded headway, which is adjusted to the next lowest headway included in the input list by the user. Identical calculations can be performed for the base period or the user can specify that the peak-to-base ratio in the coded headways is to be maintained, irrespective of base loadings. Vehicle requirements and all other statistics are then calculated by using these loaded headways as the basis.

For example, suppose line 17 has a coded peak headway of 10 min, a coded base headway of 20 min, a peak hourly load of 492 passengers, and an off-peak hourly maximum load of 164 passengers, with peak capacity of 65 passengers and base of 48. The peak load requires 7.57 buses per hour, which is approximately an 8-min headway. In the base period, the need is for 3.42 buses per hour, which is an 18-min headway. Therefore, this line will be recomputed to an 8-min peak and 18-min base headway, with all statistics recomputed accordingly. If input headways were only 5, 7, 10, 12, 15, and 20 min, among others, then the 8 and 18 min would instead be replaced by 7 and 15 min as the nearest headways that would provide sufficient capacity to meet the demand. If the maximum loads generate headways that are longer than those coded, the longer headways that are adequate for demand are listed. Suppose line 20 has a maximum peak load of 102 passengers and a maximum base load of 41 passengers, with coded headways of 20 min in the peak and 40 min in the base. URAP will determine that the demand service level is 38 min peak and 65 min base. Assuming that these are the nearest input headways, the line would be recorded to 35 and 60 min.

The nominal scenario changes the vehicle requirements by imposing policy headways whenever a loaded headway is longer than a maximum headway for a line input by the user; the format shown in Table 2 is used. Thus, the user specifies each coded headway and a maximum policy peak headway and maximum and minimum base headways corresponding to each. Typical input information for this is shown in Table 2. Thus, for all lines on which the loaded headways do not violate the policy headways set out by the user, the nominal and loaded data are identical. However, if the loaded data represent a violation of policy constraints, the policy headway is substituted and line statistics are recomputed for the policy headways. Thus, if the 20-min peak headway has a maximum

TABLE 2 Typical Inputs of Policy Headways

Coded Headway (min)	Policy Headways (min)		
	Maximum Peak	Maximum Base	Minimum Base
2	10	10	2
3	10	10	2
4	12	12	2
5	15	15	2
6	15	15	2
7	20	20	2
8	20	20	3
10	20	20	5
12	25	25	5
15	30	30	5
.	.	.	.
.	.	.	.
.	.	.	.
60	60	90	20

base headway specified of 45 min, line 20 in the foregoing example would be reset to 35 min peak and 45 min base, with vehicle requirements, vehicle hours, and vehicle miles recalculated accordingly.

Last, the modified scenario represents changes that can be input by the user on a line-by-line basis. The parameters that the user may override include the vehicle capacity (allowing specification of special vehicle types, such as articulated buses, to serve one or more lines), deadhead time, maximum passenger load, layover time, and circulation time. Also, the user can input short lines and new lines and obtain system statistics for such added lines without the necessity of returning through the simulation process. This addition of new lines does require, however, that the user estimate the maximum passenger load that such a new line would carry without the benefit of a simulation to establish this. The primary benefits of this estimation are to determine the likely effects on fleet requirements of such changes as well as to override some of the automatic recalculations of URAP. For example, a few bus lines may exhibit a spuriously high load that cannot be corrected through a more realistic network description. The recomputation of vehicle requirements, headways, and so on can be overridden by specifying a maximum passenger load on the line that is more realistic and having URAP recompute headways based on this. Special layover times can also be input for short lines, where the percentage layover may result in violations of union rules because of the shortness of the trip time.

URAP has also assembled the necessary data to run a UTPS-compatible operating cost model and is designed to output a disk file that can be used as input to a cost model that uses variables such as vehicle miles, vehicle hours, and peak vehicles. These values are provided on a line-by-line basis, so that line-by-line cost estimates are possible.

USING THE ROUTE-ANALYSIS PROGRAM

The route-analysis program is not a model in the conventional sense, so there is no calibration step involved in its use. The underlying calibration is that of the transit network. The route-analysis program assists the process of network calibration by providing statistics such as peak and base vehicle requirements, daily and annual vehicle miles and vehicle hours of travel, and specific line statistics such as round-trip time and distance, peak passenger load, and location of the peak load point, all of which can be checked against actual system statistics. The detection of significant departures between the actual system and the route-analysis program outputs should lead to identification of problems in the transit network that require correction.

Once the network has been calibrated satisfactorily, simulation runs can be made and URAP can be used to analyze the performance of the system. First, URAP can be used to generate system statistics for the simulation situation that provide a guide to the performance of the system. Loaded bus requirements that are significantly higher than the coded ones signify that the bus system is overloaded, a situation not easily determined from a single figure in standard network assignments. If nominal bus requirements are significantly higher than loaded-bus requirements, this signifies that the policy headways that override demand headways result in a need to provide an excessively high level of service compared with demand. Similarly, if loaded and nominal bus requirements are below those coded, it is indicative that too many buses are being provided for the level of demand or to maintain policy service levels.

On a line-by-line basis, the statistic of peak load on a line is asterisked if the load exceeds capacity. An examination of the table of line statistics reveals quickly on which lines there is an overload, and when the accompanying headways and vehicle requirements for the loaded condition are compared with the coded line, the order of magnitude of the overload is also revealed.

Interpretation of these statistics should, however, be made in conjunction with the transit assignment outputs of line loadings (Report 2 in ULOAD) or in conjunction with URAP Report 5 giving the four maximum load links. Because of the aggregate nature of transit networks and the incidence of connections from zone centroids, it is possible for the loaded network to generate a one-link extraordinary peak load. This load is probably not a real peak and represents rather the result of the aggregation process. Therefore, the loads on the line around the peak-load point should be investigated to determine whether there is a sustained overload (indicating a genuine need to increase capacity) or only a one- or two-link overload that drops off rapidly on either side of the peak-load point. A further use of Report 2 from ULOAD is to investigate the incidence of heavy loadings on the line with a view to defining sublines to take care of the overload situations or to make an underloaded line more efficient without reducing service as drastically as might appear necessary otherwise. These sublines can be tested initially in URAP alone; the proportion of the peak load on the new subline can be defined by using the relative headways of the subline and its parent line to split the load. This will provide some information on the likely savings to be achieved by the short lining, although it will not reveal new transfer patterns and other potential path changes. Subsequently, the sublines can be coded into the network and the entire simulation of transit ridership reiterated. This analysis provides the basis for refining the bus network to provide a more efficient service pattern.

There is a pitfall in the process, however. As in highway-network, capacity-constrained iterations, changes in the service levels to provide more appropriate service will generate changes in the demand levels. Specifically, if URAP indicates that a bus line is overloaded and requires more frequent service, coding of the loaded or nominal headways will result in reduced waiting times for the network paths served by the route. This, in turn, will lead to an increased patronage on the line and will generate a further increase in the peak load. Hence, the loaded or nominal headway will be insufficient to carry the enhanced demand. In the reverse case, a line that is underloaded will have a longer loaded or nominal headway than the coded one. Replacing the coded headway with this demand headway will lengthen the waiting time and reduce demand still further. Stable convergence is unlikely if the path building used in the network is all or nothing, because the subsequent path building will drop paths out of the long-headway lines and add paths that use short-headway lines. Thus, if one continually adjusted headways to match demand levels, all lines that were under capacity to start would theoretically end with the maximum policy headways on them, and all lines that started with overloads would end with high frequencies, probably on the order of 1-min headways or less. Clearly, this is neither logical nor desirable. Use of URAP outputs to adjust the headways is more appropriately to apply about half of the change indicated by the loaded scenario and continue to readjust in smaller increments from this. Such a procedure produces a relatively rapid convergence to an acceptable service level on each line.

CASE STUDY APPLICATION

First, a base-year transit network was constructed to cover all bus service in the six-county Los Angeles metropolitan area. Of necessity, the network was constructed by using the UNET program in UTPS, because when this project began, there was no suitable regional highway network available from which to construct an INET network. The network involved using practically the limiting values of nodes and links in the network and necessitated development of several FORTRAN programs to seek out errors in a systematic fashion and to find and delete unused links and nodes in the network. The final base-year network involved use of over 7,500 nodes and 30,000 links and described a bus network with nearly 2,500 peak-period buses. Over 470 individual bus lines and sublines were needed to describe the network. Detailed statistics were available only from the largest operator in the region the Southern California Rapid Transit District (SCRTD). Therefore, all calibration was performed against the SCRTD portion of the regional network, which consists of 305 of the lines and 2,000 of the peak-time buses.

A number of tests were performed to check the base-year network. Briefly, these checks revealed that no coded lines differed by more than 10 percent from actual values of round-trip travel time and one-way distance, except for a few routes that were identified as having incorrect actual values. Chi-square and Kolmogorov-Smirnov tests were performed between actual and network values of times and distances, and no statistically significant differences were found. Finally, with URAP, the network produced a coded peak-vehicle requirement (PVR) of 1,619 buses for SCRTD and a nominal PVR of 1,871 compared with the actual base-year PVR of 1,848 buses. This discrepancy of 23 buses, or 1.2 percent, was considered to be satisfactory. A summary of the statistics for the base-year network under coded and nominal conditions is shown in Table 3.

The primary purpose of the use of URAP in conjunction with the standard UTPS models in this case study was to refine the background bus network for a proposed long-range future systemwide network that included an initial rail line of 18.6 mi. Initial simulations of the rail and bus network with year 2000 trip estimates from the Southern California Association of Governments (SCAG) provided the statistics shown in Table 4 for the original bus network. The bus operating cost for this network of $435.7 million in 1983 dollars was estimated by using a UTPS-based operating cost model (7) that provided an estimate of $398.5 million for the base-year network. Thus, the increase in the peak-vehicle requirement from 1,871 to 1,895 for SCRTD together with increases in revenue-vehicle miles and revenue-vehicle hours and transit ridership growth from 1,170,000 to 1,924,000 daily trips resulted in an estimated increase in cost of $37.2 million. The issue to be determined was whether the future network could be operated more economically without significant loss of patronage.

TABLE 3 Statistics of the Base-Year Network

Statistical Measure	Value	
	Coded	Nominal
Peak-vehicle requirement (PVR)		
SCRTD local buses	1,228	1,120
All operators' express buses	420	793
Total SCRTD buses	1,619	1,871
Total systemwide buses	2,278	2,447
Total daily revenue-vehicle hours	32,900	33,000
Total daily revenue-vehicle miles	420,000	444,000
Daily linked passenger trips	1,170,000	1,170,000

TABLE 4 Summary of Changes in PVR by Iteration (SCRTD Lines Only)

Variable	Network Iteration		
	Original	Second	Third
Peak coded vehicles	1,775	1,820	1,919
Peak nominal vehicles	1,895	1,858	1,907
Base coded vehicles	1,111	905	977
Base nominal vehicles	985	936	942
Revenue-vehicle miles	97,350,000	95,260,000	96,540,000
Revenue-vehicle hours	7,610,000	7,260,000	7,360,000
Linked passenger trips	1,924,000	1,817,000	1,863,000
Bus operating cost[a] ($)	435,697,000	384,522,000	421,563,000

[a] 1983 dollars.

A series of iterative adjustments was made in the network on the basis of the URAP and ULOAD reports. First, all lines were identified from the URAP outputs that were overloaded in either the peak or the base period. Each such line was examined in the ULOAD reports to determine whether the peak load was of extremely limited duration or was spread over a significant portion of the line. In the latter case, consideration was given to defining a new subline. This resulted in the definition of 64 new sublines, the suggested deletion of 43 lines or sublines, and replacement of the current operation of 23 lines or sublines with one of the new sublines. This was a net decrease of two coded sublines in the entire network, but represented some significant shifts of service. All the proposed changes were submitted for review by the planning staff of SCRTD, after which some of the lines recommended for deletion were restored for policy reasons and 10 of the sublines were either removed or assigned different end points, where it is feasible to turn back buses. Even from this first round of network revisions, the majority of routing changes was greeted with no surprise by the SCRTD staff. Most identified changes were logical in light of present loadings, and the extent of the overloaded line segments that were used to identify sublines corresponded well with known segments of heaviest loading. This provides a further indication that the network possessed a high degree of realism and accuracy.

A measure of the degree to which the network is unable to satisfy demand is the difference between nominal and coded vehicle requirements. Table 4 shows the results of the second and third iterations. (The first iteration is not reported here, because a number of errors were found subsequently in it, and the second iteration provided corrections to this.) From the initial network, it can be seen that the nominal PVR was 120 buses greater than the coded one, whereas the base coded network was 136 buses too great. By the second iteration, not only were the nominal vehicle requirements for both peak and base lower than in the original network, but the differences between coded and nominal had decreased markedly, being 38 buses in the peak and 31 buses in the base period. The third iteration shows maintenance of this improvement, although the differences here are that some lightly loaded lines were returned to a higher service level because the previous adjustments had reduced patronage too far. This shows clearly in the annual cost per daily linked trip (not all of which are on the SCRTD buses). For the original network, this cost is

$226.45, for the second iteration it is $211.62, and for the third iteration it is back up to $226.28. However, the difference between coded and nominal PVRs is now 12 buses, with the nominal PVR being slightly lower than the coded PVR. At the same time, the base-vehicle requirement is overestimated by 35 buses. These changes show a slight overcorrection of the deficiencies in the second iteration and auger well for a stable system at the fourth iteration.

Adjustments were made in these iterations by accepting the URAP-generated nominal headways in those instances where lines were severely under-loaded or overloaded and coding about half the change between URAP and the original coded headways in all other cases. The use of about half of the headway change is necessary to dampen out the cyclical shift of patronage from lightly loaded lines and into heavily loaded lines. Even so, the second iteration shows too large a swing to the heavy lines and away from the light lines, and this was modified in the third iteration.

At time of writing, a fourth iteration was being developed to complete the redesign of the background bus system. Even before this, it could be seen that the redesign that was enabled by URAP allowed three iterations to reduce the PVRs, the base-vehicle requirements, and the operating cost, all with a relatively small loss of transit riders. More important, the revisions to the network were easily identified and were lengthy to input only because of the large size of this particular case-study network. This procedure would be efficient for a small or medium-sized network.

The most important results of this analysis are that information is provided to allow a systematic adjustment to be made to the bus system for a long-range planning situation; such an adjustment is not usually feasible. Furthermore, all indications are that the final results will be an increase in the efficiency of the resulting bus system and a reorientation of service to where the greatest demands are.

CONCLUSIONS

A procedure for producing bus system statistics from standard UTPS planning models and how this procedure can be used in conjunction with the UTPS procedures to undertake detailed long-range planning of a bus system have been described. It has been shown that the capability of the procedure to produce data that accurately reflect the base year is considerably greater than that normally associated with aggregate travel-forecasting models. Given the increasing importance of developing planning strategies to contain operating costs for transit systems and comprehensive operating-cost plans for regions contemplating major capital investments in transit, this procedure is an important one that adds a needed dimension to the battery of UTPS models.

A number of subsequent improvements are contemplated for the program, including the development of graphical displays of the ULOAD reports that are used in conjunction with URAP and addition of a capability for URAP to output a modified network file by using the nominal headways or some predetermined fraction of the change between coded and nominal headways. Such enhancements will remove much of the time-consuming portion of the current procedure. This capability to refine bus systems through a long-range route analysis also permits a fairly extensive capability to simulate alternative futures and determine probable directions of service changes that should be planned.

ACKNOWLEDGMENTS

The authors would like to thank SCRTD for permission to use the results described in this paper and for their active support on its presentation. The authors wish to thank particularly Keith Killough of SCRTD for his suggestions and help both in the original work and in the documentation of the results in this paper.

REFERENCES

1. H.V. Wallis. Experiences in Computer-Assisted Scheduling Installations in North America. Proceedings of Third International Workshop on Transit Vehicle and Crew Scheduling, Montreal, Canada, June 1983.
2. L. Luedtke. RUCUS II: A Review of the System. Proceedings of Third International Workshop on Transit Vehicle and Crew Scheduling, Montreal, Canada, June 1983.
3. J-Y. Blais and J-M. Rousseau. HASTUS for Bus Operator Scheduling: The Man-Machine Interface. Publication 231. Centre de Recherche sur les Transports, Université de Montreal, Quebec, Canada, Nov. 1981.
4. J-Y. Blais. HASTUS System: An Interactive System for Vehicle and Bus Crew Scheduling. Proceedings of Third International Workshop on Transit Vehicle and Crew Scheduling, Montreal, Canada, June 1983.
5. Urban Transportation Planning System Introduction. UMTA and FHWA, U.S. Department of Transportation, Jan. 1980.
6. Barton-Aschman Associates, Inc. Route Analysis Program Documentation (URAP2). Technical Memorandum 8.1. Transportation Planning and Modeling Services Project, Southern California Rapid Transit District, Los Angeles, Calif., Nov. 1983.
7. K.G. McDonald, R. Alvarez, and P.R. Stopher. The Development of a Bus Operating-Cost Model for Long-Range Cost Forecasting. Presented at 64th Annual Meeting of the Transportation Research Board, Washington, D.C., 1985.

The opinions expressed in this paper are those of the authors and do not necessarily reflect the position or policies of SCRTD. The authors are solely responsible for any errors or omissions in the paper.

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

The Demand for Personal Travel in Developing Countries: An Empirical Analysis

ANGUS DEATON

ABSTRACT

Conventional travel survey data, whether household- or traveler-based, are scarce in developing countries, and it is suggested that household expenditure surveys, which are relatively common, be used instead. Although many of the traditional topics of travel surveys cannot be addressed with expenditure data, many others can, and there are compensating advantages of coverage and scope. This concept is illustrated with data from household surveys of Sri Lanka, Hong Kong, Thailand, and Tunisia. Broad regularities in travel expenditures are discussed, both within and across countries, and for the Tunisian survey, a more detailed regression analysis is presented that focuses on the interrelations between travel expenditures and vehicle ownership.

A preliminary empirical analysis of the demand for personal travel in a number of developing countries is presented. The data come from household expenditure surveys and typically were not collected with the specific purpose of analyzing travel behavior. They are therefore inferior in many respects to standard travel surveys, which typically contain a great deal of "physical" (as opposed to financial) information about travel, for example, the number, length, frequency, and timing of trips for various purposes. However, by giving up this detail, several advantages are gained. Household expenditure surveys are relatively plentiful around the world. Most countries publish cost-of-living or other price index statistics, and household surveys are the standard way of obtaining the weights for their construction. The quality of design and data processing is often excellent; response rates are high by western standards, as is the quality of the interviewing and coding staff. The surveys are typically large and are representative of the country as a whole. Their size means that there is a great deal of useful data, and quantity is to some extent a substitute for lack of detail. However, representativeness is even more important. The choice-based sampling problem that arises from surveys confined to travelers is avoided, and information is given not only for urban areas, where a good deal is known about travel patterns in less-developed countries (LDCs), but also for rural areas for which there is much less information.

Studies such as those by Maunder (1-3), Eastman and Pickering (4), Heraty (5), and Thobani (6) have revealed much about trip patterns of household members in a number of urban contexts in developing countries, but research on personal travel in the countryside has been hampered by the lack of an adequate framework for description. In towns, trips can be conveniently classified by purpose; in LDCs they are usually to work or to school. They are regular, usually occurring 6 days a week, and they are predominantly single-purpose. None of these descriptions fits rural travel patterns, so that even the taxonomic framework around which a questionnaire could be designed is lacking. Expenditure surveys avoid these problems by asking much simpler and more limited questions, so although they cannot answer many of the questions to which answers are desired, they can tell a great deal about such matters as what fraction of households spends anything on travel at all, whether travel expenditures form a significant enough share of the budget to be worth investigating, and which modes account for the expenditures that do exist. This paper is concerned with presenting some of such information and demonstrating its usefulness. Many of the standard patterns that appear in the studies of trip patterns in urban areas can be discerned in the expenditure data here, so the approach of this paper can be seen as complementary to that of the standard studies.

There are two main empirical sections to the paper. In the first, the broad patterns of travel behavior are characterized by using data from five surveys or sets of surveys. These come from Hong Kong in 1979-1980; Sri Lanka, 1980-1981; Tunisia, 1979-1980; Thailand, 1975-1976; and from several suburbs of Delhi, India, in 1979-1982. In the second section, the Tunisian data are subjected to a more detailed analysis of the influence on household travel patterns of regional, occupational, and family composition variables. The analysis is, at this stage, primarily designed to be descriptive but is also guided by its potential relevance to questions of how transportation services should best be priced. Modern pricing or tax and subsidy theory [see, for example, an overview by Atkinson and Stiglitz (7)] has evolved a set of pricing rules that depends both on the distribution of commodity demands and on their sensitivity to price. In particular, if the pricing authority has an interest in improving the distribution of real income and if this interest cannot be met by direct systems of taxes and subsidies (and in LDCs this is usually thought to be very difficult), then it will generally be desirable to tax highly those expenditures that are more heavily consumed by richer consumers and subsidize those that figure most prominently in the budgets of the poor. However, if different goods are differentially elastic to price changes, this must be allowed for, too, and it is typically undesirable on efficiency grounds to tax heavily those commodities that consumers will readily find substitutes for in response to price increases. Clearly then, a good deal of empirical evidence is required to assess alternative pricing schemes; much of this evidence can be provided by household survey data. In particular, the surveys provide an excellent picture of who gets what, so that the distributional effects of price policy can readily be seen. Descriptions of the relationship between household income and expenditures on the various travel modes give a national picture of the consequences for real income distribution of changes in transport pricing policies, and the relationships with other variables such as region and household demographics reveal the differential effects of policy in different regions or over different household types.

TRAVEL PATTERNS FROM HOUSEHOLD SURVEYS

Evidence from four household surveys is discussed in this section; all are standard socioeconomic surveys specializing in the detail of household expenditures. The surveys are the 1980-1981 socioeconomic and labor force survey of Sri Lanka, with a sample size of nearly 10,000 households; the 1975-1976 socioeconomic survey of the whole kingdom of Thailand, with 11,000 households (here a 10 percent random sample is used in order to minimize processing costs); the 1979-1980 household expenditure survey of Hong Kong, with over 4,000 households; and the 1979-1980 expenditure survey of Tunisia, with almost 6,000 households. These particular surveys were chosen because of their current accessibility, diversity, and relatively detailed information on travel expenditure by modes.

Table 1 lists the evidence on travel expenditures as a component in total consumer expenditure. In principle, the denominator is total expenditure on nondurable goods and services, and shares are calculated for each household and then averaged by using weights as necessary to reflect the sample design. Note that this is quite different from total consumer expenditure on travel expressed as a share of all consumer expenditure. This alternative concept is effectively a weighted average of individual shares with weights proportional to household total expenditure; the rich are therefore effectively overrepresented. As an example, the aggregate figures for Sri Lanka corresponding to those in the table are 5.0, 6.7, 4.5 and 2.0; the rich spend a higher proportion of their outlay on travel.

Delhi apart, there is a good deal of uniformity in these figures. Shares tend to be higher the higher the level of development and are higher in urban than in rural areas. Nevertheless, travel expenditures still exist in the countryside; even if a high proportion of trips are made by foot, and even in the (presumed) absence of regular commutation trips, the rural shares of expenditure are only 20 to 30 percent lower than those in the cities. Even in the extremely low-income tea estates of Sri Lanka, where incomes are barely above subsistence by most criteria, more than 1 1/2 percent of the budget is devoted to personal travel. The very high Indian figures, taken from work by Maunder (1), may be peculiar to Delhi where urban resettlement to relatively distant suburbs enforces high travel costs on even poor consumers. There may also be understatement of income, which would artificially inflate the ratios.

An alternative way of assessing the importance of travel expenditure is to examine the fraction of households spending nothing on travel. These ratios, corresponding to those in Table 1, are given in Table 2. Because surveys have a finite reporting period, these figures will be somewhat of an overestimate because occasional trips will only show up for a fraction of households. Nevertheless, the pattern is consistent with that in Table 1. Travel expenditure is a part of the budget for the vast majority of urban dwellers, and even in the rural areas, only a minority of households show no such expenditures. If Tables 1 and 2 are taken together, shares of the budget devoted to travel for households that spend something on travel can be calculated. By this measure, travel composes much the same share of travelers' budgets in the rural areas as it does in the cities and towns.

These patterns must be disaggregated by rich and poor households if the distributional effects of transport policy are to be assessed. Because expenditure groups are not comparable across countries,

TABLE 1 Personal Travel Expenditure as Share of Total Outlay

Area	Percentage
Hong Kong, 1979-1980	
All	5.68
Hong Kong Island	5.32
Kowloon	4.61
New Kowloon	6.66
New Territories	5.85
Sri Lanka, 1980-1981	
All	3.2
Urban	4.4
Rural	3.1
Estates	1.7
Tunisia, 1979-1980	
All	3.06
Cities	4.22
Towns	2.23
Rural	2.96
Thailand, 1975-1976	
All	4.07
Cities	4.91
Towns	4.17
Rural	3.50
Delhi suburbs, 1979-1982[a]	
Nand nagri	11
Shakarpur	12
W. Patel Nagar	11
Dakshin puri	8
Janakpuri	9
Saket	12

[a] As shares of income, not total outlay.

TABLE 2 Proportions of Households Reporting No Expenditures on Travel

Area	Percentage
Hong Kong	
All	3.02
Hong Kong Island	1.93
Kowloon	6.22
New Kowloon	1.69
New Territories	2.97
Sri Lanka	
All	25.0
Urban	26.1
Rural	24.1
Estates	33.7
Tunisia	
All	35.2
Cities	18.6
Towns	47.9
Rural	36.0
Thailand	
All	21.8
Cities[a]	21.0
Towns	22.5
Rural	22.1

[a] Bangkok, 14.4.

TABLE 3 Proportions of the Budget Devoted to Travel by Expenditure Groups: Four Countries

Group	Percentage of Population in Group	Percentage Spending Nothing on Travel	Mean Share
Thailand			
1	3.7	64.3	1.90
2	11.1	45.7	2.51
3	15.2	29.7	2.87
4	22.6	22.3	3.37
5	15.9	16.4	4.07
6	15.4	39.0	4.93
7	5.8	7.5	6.64
8	6.5	2.7	6.19
9	3.7	7.0	8.70
Hong Kong			
1	1.0	40.0	1.75
2	3.8	17.8	4.10
3	6.4	10.1	5.27
4	10.9	3.3	5.25
5	12.0	1.7	5.12
6	11.8	1.6	5.40
7	10.5	0.5	5.70
8	7.9	1.7	5.79
9	6.8	1.6	6.02
10	5.6	0.0	6.30
11	7.8	0.3	6.62
12	5.2	0.4	6.73
13	6.0	0.3	6.72
14	1.8	0.0	5.92
15	2.4	0.0	8.06
Tunisia			
1	2.7	72.2	2.14
2	9.0	57.0	2.20
3	13.3	44.0	2.55
4	13.9	39.7	2.68
5	11.4	34.6	2.74
6	9.8	28.8	3.36
7	14.8	25.1	3.17
8	9.2	23.4	3.65
9	5.3	16.9	3.85
10	5.6	16.7	4.35
11	2.2	7.5	6.00
12	2.8	14.5	5.14
Sri Lanka			
1	5.0	51.5	1.4
2	5.0	39.0	1.9
3	10.0	31.3	1.8
4	10.0	32.4	2.0
5	10.0	28.6	2.2
6	10.0	27.2	2.2
7	10.0	24.1	2.7
8	10.0	22.4	2.9
9	10.0	16.5	3.8
10	10.0	14.2	4.5
11	5.0	9.1	6.3
12	5.0	5.9	11.0

Table 3 shows expenditure groups for each country from poorest to richest together with the estimated proportions of the population in each group. The third column gives the proportion of households in the expenditure group spending nothing on travel and the fourth the estimated mean of shares within the group. In all cases, the proportion spending nothing on travel declines steadily with increasing total expenditure (except possibly among the very rich, whose expenditure on private transport is more irregular and therefore less likely to occur in the survey period). Among the very poor, more than half spend nothing on travel, so subsidizing travel can do little for them. Note, however, the possibility that for some consumers, travel expenditures may be necessary for them to earn anything at all. Low or zero travel expenditures among the poor could therefore reflect high unemployment and poor work opportunities in those groups. As total outlays increase, the share devoted to travel also increases in all four surveys. The overall expenditure elasticity of travel is therefore typically greater than unity; overall subsidization will therefore tend to proportionally favor the rich over the poor. Depending on the price elasticity, efficiency considerations may offset the undesirability of this situation. Assessing the balance requires more detailed analysis than is possible here.

In Maunder's work (1) on the Delhi suburbs, travel shares are rather different from those in Table 3. For all six suburbs, the fraction of income devoted to travel is remarkably constant over all income groups except the poorest, where the shares are extraordinarily high, with averages as high as 30 to 40 percent in three of the districts. However, it should be noted that there are very small numbers of households in the poorest groups in the Delhi surveys, and that these groups, almost by definition, contain an abnormal representation of individuals whose incomes are temporarily low. In consequence, if travel expenditures remain at their normal level while incomes fluctuate, there will always be a few individuals or households with extremely large shares of income devoted to travel. This is no more a matter of concern than is the fact that a daily commuter who gets paid only weekly spends an infinitesimal fraction of his Monday's income on Monday's transport. Because total expenditures are much more stable than are incomes, this phenomenon is much less pronounced if shares of the budget are used in place of shares of income. Nevertheless, and in spite of the averages in Table 3, it is still the case that some poor households spend remarkably high proportions of their budgets on travel and that a higher proportion of poor households do so than is the case among groups that are better off, though not among the richest. For example, in Hong Kong, among the poorest 15 percent of the population, 14 percent spend more than 10 percent of their outlay on travel; in the next richest 25 percent, only 8 percent spend more than 10 percent. In Tunisia, of the poorest 15 percent, 6 percent spend more than 10 percent; the same is true of the next 15 percent in spite of their higher total budgets.

As is to be expected, modal choice shows much greater diversity across space than do the aggregate travel expenditure patterns. Tables 4-6 give the travel expenditure shares disaggregated by mode for Tunisia, Thailand, Hong Kong, and Sri Lanka. At this level of aggregation, these figures hold few surprises. Buses are the major item of expenditure throughout; in Tunisia, the hire car element that dominates outside the cities is essentially a form of bus service. Vehicle ownership is most important in the cities, and operating expenses are closely linked to automobile and motorcycle ownership. More interesting is to disaggregate these figures further, both by income level and by region. However, cross-tabulation is too clumsy a tool for this, and it is necessary to summarize the patterns more succinctly. Regression analysis is one way of doing this, and the next section contains some illustrative results from the Tunisian survey.

REGRESSION ANALYSIS OF TUNISIAN TRAVEL PATTERNS

One immediate problem with the application of regression analysis to travel expenditures is that a large fraction of households report zero expenditures. To some extent, these zeroes reflect low

TABLE 4 Travel Budget Shares by Mode: Tunisia and Thailand

Mode	All	Cities	Towns	Rural
Tunisia				
Total	3.06	4.22	2.23	2.96
Private	0.92	1.88	0.78	0.48
Public	2.14	2.34	1.45	2.48
Bus	0.52	1.10	0.30	0.35
Hire automobile	1.19	0.42	0.87	1.83
Taxi	0.20	0.35	0.10	0.18
Other (including season tickets)	0.23	0.47	0.19	0.12
Automobile ownership	15.8	25.4	14.5	11.4
Thailand				
Total	4.07	4.91	4.17	3.50
Local	1.97	2.52	1.93	1.64
Bus	1.51	1.71	1.68	1.31
Taxi	0.07	0.20	0.01	0.02
Other	0.39	0.61	0.24	0.31
Nonlocal	1.06	1.01	0.92	1.16
Bus	0.76	0.56	0.69	0.92
Other	0.30	0.45	0.23	0.24
Private	1.04	1.38	1.32	0.71
Automobile ownership	4.3	9.1	4.3	1.4
Motorcycle ownership	11.3	18.1	11.3	7.1
Bicycle ownership	26.8	25.2	31.6	25.8

TABLE 5 Travel Budget Shares by Mode: Hong Kong

Mode	All	Hong Kong Island	Kowloon	New Kowloon	New Territories
Total	5.68	5.32	4.61	6.66	5.85
Public	5.19	4.77	4.22	6.12	5.40
Bus	3.13	2.70	2.27	3.82	3.90
Taxi	0.86	0.74	1.00	0.86	0.76
Air	0.07	0.11	0.07	0.05	0.02
Other	1.13	1.22	0.88	1.39	0.72
Private	0.49	0.55	0.39	0.54	0.45

TABLE 6 Travel Budget Shares by Mode: Sri Lanka

Mode	All	Urban	Rural	Estates
Total	3.2	4.4	3.1	1.7
Public	2.6	3.0	2.6	1.6
Bus	2.3	2.5	2.3	1.4
Taxi	0.1	0.2	0.1	0.1
Train	0.2	0.3	0.1	0.1
Other	0.0	0.0	0.0	0.0
Private	0.6	1.4	0.5	0.1

frequencies of purchase; if trips on a particular mode are taken once a month and the survey period is a week, then a quarter of the households will register four times their normal weekly expenditures and three-quarters will register nothing. It is clear in this case that the expected value of expenditures is correct, and for that reason, a regression with expenditure as the dependent variable will yield unbiased and consistent parameter estimates. More difficult to deal with are those zeroes that occur because the household never makes that type of expenditure, and presumably this is frequently the case for certain travel modes and occasionally even for all travel expenditures. Unfortunately, there is no way of telling these "genuine" zeroes from the "infrequent purchase" zeroes, and even if this problem could be solved, there is as yet no agreed-on technique for estimating such models that is feasible on anything other than small data sets (8-10). In this paper, for lack of anything better, ordinary least-squares regressions are reported, and these include all the observations, zero or nonzero. This has the advantage of preserving the same sample for all regressions so that regressions for subcategories of expenditures relate to the same households as does the regression for the sum. Coefficients then add up across regressions, allowing decomposition of totals. The parameters of such regressions can also be straightforwardly interpreted, at least under certain assumptions. The coefficient on an explanatory variable estimates the corresponding coefficient for households that purchase that particular category multiplied by the proportion of such households. Dividing by this proportion yields the conditional coefficient, so that the regression context corresponds exactly to that in the cross-tabulations where the mean shares are unconditional means (including zeroes) and the means conditional on traveling can be obtained by dividing by the proportion of travelers.

The final issue is how to allow for ownership of motor vehicles. Clearly, the short-run travel decisions of household members depend crucially on whether the household owns a motor vehicle. In the long run, of course, vehicle ownership is determined along with other consumption decisions. Formally, the following expressions can be written:

$$y_i = \underline{\beta}_1' \underline{x}_{1i} + \theta d_i + u_{1i} \tag{1}$$

$$d_i^* = \underline{\beta}_2' \underline{x}_{2i} + u_{2i} \tag{2}$$

$$d_i = 1 \quad \text{if } d_i^* \geq 0$$

$$= 0 \quad \text{if } d_i^* < 0 \tag{3}$$

where

- u_1, u_2 = endowed with a joint distribution, usually bivariate normal;
- y_i = expenditure by household i on some mode, say buses;
- d_i = dummy variable that is 1 if a vehicle is owned and zero otherwise; and
- $\underline{x}_1$ and $\underline{x}_2$ = vectors of variables influencing bus travel and vehicle ownership, respectively.

The estimation of this model is discussed, for example, by Heckman (11) and ideally is handled as follows. Equations 2 and 3, the vehicle ownership equations, are estimated by a standard probit. Equation 1 is then estimated by "instrumental variables" by replacing d_i by its estimated probability of being unity from the probit. For this to work properly, there must be variables in $\underline{x}_2$ that do not appear in $\underline{x}_1$; otherwise the model is essentially underidentified (except for functional form, which is a poor crutch on which to lean). For the current Tunisian data, the most plausible variables appear to be the employment status of heads of households. Presumably certain types of workers will need private means of transport, for example, those for whom there is a high penalty for persistently being late for work. Otherwise there appears little reason to expect household transport budget shares to depend directly on employment status.

One possibility that is followed here is to estimate Equation 1 as it stands; this produces consistent estimates only if u_1 and u_2 are independent, that is, only when there are no common omitted variables. If one could believe this, the estimates from Equation 1 could be regarded as those of short-run demands. The second line is to estimate Equation 1 excluding d_i; this can be thought of as a linearized reduced form or long-run demand. For example, conditional on vehicle ownership, the income elasticity of the demand for bus tickets may be positive, whereas the long-run elasticity, taking into account higher vehicle ownership, may be negative.

Therefore the determinants of the probability of vehicle ownership are presented first. For convenience, the formulation used was logit rather than probit. Hence, the parameters shown in Table 7 represent the derivative with respect to each explanatory variable of the log odds in favor of owning an automobile.

TABLE 7 Vehicle-Ownership Logistic Regression: Tunisia

Variable	Abbreviation	Coefficient	Chi-Square
Regression constant	CONST	−2.71	281.8
Total household expenditure	THE	0.0004	173.6
No. of female workers	NFW	−0.96	1.9
No. of male workers	NMW	0.07	2.0
No. of children in primary school	NCP	0.04	1.3
No. of children in high school and tertiary education	NCH	−0.08	1.9
Cities			
Northeast	CNE	−0.68	18.5
Center	CC	−0.87	6.2
South	CS	1.45	62.5
Towns			
Northeast	TNE	−0.85	18.5
Northwest	TNW	−0.73	11.3
Center	TC	−0.78	19.9
South	TS	−0.36	4.0
Rural areas			
Northeast	RNE	−1.04	31.1
Northwest	RNW	−0.20	2.0
Center	RC	−1.08	42.4
Employer	PATRON	0.88	14.9
Self-employed	INDEP	0.48	9.2
Laborer	OUVRIER	0.57	15.4
Wage earner	EMPLOYEE	1.26	60.5
Salaried employee	SALARIE	0.75	1.3
Family worker	AIDEFAM	1.85	6.9

Note: The model was estimated without weighting the sample data. Model χ^2 = 675.73; degrees of freedom = 21.

As is to be expected, the level of total household expenditure is the dominant explanatory variable. The coefficient suggests that an increase in total housing expenditure of 1,200 dinars, say for just below the first quartile to just above the third quartile, would raise the log odds by 0.53, and the probability of ownership from, say, 0.25 to 0.38. Additional male workers and primary school children have a positive effect on vehicle ownership; female workers and school children have negative coefficients. All of these effects, however, appear to be rather weak. The regional and urbanization dummies indicate that the probability of owning a vehicle is greatest in the south, especially in the cities. Adopting rural south as the base, the probability of ownership is slightly lower in the towns and significantly lower in the rural northeast and center. The concentration of private modes in the south presumably reflects a relative shortage of public transport in that region. The base for employment status of the head of the household incorporates persons not working and a small group of apprentices and persons for whom occupations are unknown, which accounts, in all, for about 16.5 percent of the sample. The probability of owning a vehicle is greater for all other groups and is surprisingly large for wage earners and family workers.

Tables 8 and 9 show what, with some presumption, are labeled short-run and long-run travel regressions; Table 8 contains the ownership dummy and Table 9 does not. The responses are not inconsistent with this basic interpretation. For example, Table 8 shows the total expenditure elasticities of the transport, private, and public categories to be (at the mean) 1.1, 1.3, and 1.0, respectively. In Table 9, when the long-run effects operating through vehicle ownership are also included, these become 1.3, 2.0, and 0.98, respectively. Because vehicle ownership itself responds to changes in per-capita household expenditures (PCEs), long-run elasticities are higher for those categories that are positively affected by vehicle ownership and lower for those that are negatively affected. Similar patterns of short-run versus long-run responses can be seen for the coefficient on the number of male workers in the household; once again, it is the strong effect on vehicle ownership that accounts for the differences in parameter estimates between Tables 8 and 9.

In reading these tables it is helpful to note that because total transport is the sum of the public and private categories, column 1 is the sum of columns 2 and 3. Similarly, public transport is the sum of five modes shown plus an unimportant "other" category so that column 3 is the sum of columns 4 through 8 approximately. Hence, looking along rows reveals how the structure (as well as total) of travel demand responds to changes in the variable concerned. Taking PCE first, it may be seen that better-off households spend a larger share of their outlay on travel, an increase that is almost totally in the short run and more than totally in the long run accounted for by the luxury nature of private travel expenditures. Among the public modes, taxis and car hiring tend to replace buses among better-off households, other things held constant. The next group of variables shows the impact of work and education patterns on travel expenditures. From Table 8, extra workers, male or female, have a similar effect on the public travel share, as do extra high school children. Primary school children have little impact on the budget, presumably because primary schools are relatively close to residences and therefore do not involve paid trips. Extra male workers, conditional on automobile ownership status, cause a switch from private to public transport; in the long run such workers tend to lead to higher probabilities of automobile ownership. These results are clearly consistent with fixed trip patterns in relation to work and higher education. The public modes associated with these trips are of some interest. The additional public share associated with male workers goes to buses and to hires, presumably the former in the towns and the latter in the countryside. Hires are also associated with extra female workers and high school children, but there is no effect on bus fares, only on season tickets. Presumably there is some explanation for this anomaly.

The regional dummies are of interest in assessing how much of the regional variations in patterns remain once the other variables, particularly PCE, have been controlled for. Notably, most of the variations in the share of private transport over regions and levels of urbanization are explained by the other variables, although there is still a significant positive dummy for southern cities. Otherwise, public transport tends to be low in the towns; the cities are heavy on buses and the rural areas on hiring, and neither is very important in the towns, hence the difference.

SUMMARY AND CONCLUSIONS

In this paper, it is proposed that household expenditure surveys be regarded as a useful supplementary source of data on household travel patterns and the point is illustrated with travel data from a number of household surveys from developing countries around

TABLE 8 Short-Run Travel Regression

Variable	Transport	Private	Public	Buses	Hires	Taxis	Seasons	Trains
CONST								
Coefficient	.54	−1.5	2.1	.48	1.4	.06	.09	−.004
t-Statistic	1.2	−5.5	5.3	2.5	4.7	0.7	1.0	−.04
LN PCE								
Coefficient	.33	.28	.05	−.06	.06	.04	−.01	.02
t-Statistic	4.1	5.8	0.8	1.9	1.1	2.8	−.9	1.3
NFW								
Coefficient	.20	.04	.16	−.01	.08	.01	.06	.01
t-Statistic	2.2	0.8	2.1	−0.1	1.3	.4	3.9	0.7
NMW								
Coefficient	.04	−.13	.17	.11	.09	−.003	−.001	−.03
t-Statistic	0.7	−3.5	3.2	4.5	2.2	−.3	−.13	−2.2
NCP								
Coefficient	−.06	.02	−.08	−.04	−.03	.01	−.01	−.01
t-Statistic	−1.2	0.8	−1.9	−2.3	−0.9	1.3	1.1	−0.9
NCH								
Coefficient	.20	−.01	.21	.03	.08	−.01	.13	−.01
t-Statistic	2.5	−0.2	3.1	0.9	1.4	−1.0	9.1	−0.6
MVD								
Coefficient	4.6	5.4	−.84	−.33	−.34	−.05	−.05	−.09
t-Statistic	28.1	56.0	−6.2	−5.0	−3.2	−1.5	−1.6	−2.3
CNE								
Coefficient	.41	.74	−.33	.72	−1.4	.07	.25	.12
t-Statistic	1.8	5.4	−1.7	7.8	−9.7	1.6	6.0	2.2
CC								
Coefficient	−.40	.11	−.51	1.1	−1.7	.07	−.07	.12
t-Statistic	−0.9	0.4	−1.3	6.1	−5.8	0.9	−0.9	1.2
CS								
Coefficient	.76	−.10	.87	1.8	−1.2	−.04	.05	.34
t-Statistic	2.4	−0.6	3.3	13.9	−5.9	−0.6	0.8	4.7
TNE								
Coefficient	−.89	.11	−.99	.22	−1.1	−.19	.15	−.02
t-Statistic	−3.2	0.7	−4.3	2.0	−6.1	−3.6	3.0	−0.4
TNW								
Coefficient	−1.3	.30	−1.6	−.20	−1.3	−.20	−.07	.25
t-Statistic	−4.0	1.6	−5.9	−1.6	−6.4	−3.3	−1.2	3.4
TC								
Coefficient	−.50	.21	−.71	.25	−.70	−.24	.03	−.01
t-Statistic	−2.0	1.4	−3.4	2.5	−4.2	−5.2	0.6	−0.2
TS								
Coefficient	−1.4	−.19	−1.2	−.002	−1.0	−.14	.003	−.01
t-Statistic	−5.3	−1.2	−5.4	−.02	−5.9	−2.9	.06	−0.1
RNE								
Coefficient	.61	.33	.28	.79	−.55	−.09	.10	.07
t-Statistic	2.4	2.1	1.3	7.6	−3.2	−1.8	2.0	1.2
RNW								
Coefficient	−1.0	−.13	−.91	−.26	−.50	−.12	−.03	.02
t-Statistic	−4.4	−0.9	−4.5	−2.7	−3.2	−2.6	−.6	0.3
RC								
Coefficient	.96	.42	.54	.02	.77	−.15	−.05	−.02
t-Statistic	4.1	3.0	2.8	0.2	5.1	−3.4	−1.1	−0.3

Note: All coefficients (times 100) express the shares in total expenditures of each mode. LNPCE = log of per-capita household expenditure; MVD = dummy (1) if vehicles owned. All other abbreviations are given in Table 7.

the world. The share of the budget devoted to travel appears to increase slightly with income within countries, and the limited evidence here reveals that the share also increases as does the level of development. Travel expenditures in relation to the total budget are greater in urban than in rural areas, though travel expenditures are still substantial in the latter. The vast majority of the households sampled, whether in Sri Lanka, Hong Kong, Thailand, or Tunisia, show some expenditure on travel. As with the share devoted to travel, the fraction of households spending anything on travel tends to increase with income in all the surveys. The pattern of mode choice is less uniform across surveys than is the broad characterization of total travel expenditures; as is to be expected, local availability exerts a strong influence on the details of transport modes. For the Tunisian data, the pattern of vehicle ownership was studied together with its relation to patterns of household expenditures on travel. Total household resources exercise the dominant influence on both, though other factors, such as the presence of additional male members in the household, are important for determining the probability of vehicle ownership. Travel expenditures themselves are significantly influenced by vehicle ownership, so that factors such as income exert quite different long- and short-run effects. In particular, conditional on vehicle ownership, both public and private transport are income elastic, but once the effects of income on promoting automobile ownership are allowed for, private transport becomes more elastic and the elasticity of demand for public transport falls below unity. As is to be expected, regional effects are strong, as are the influences of the demographic composition of the household. The latter are consistent with findings in other developing countries that the majority of

TABLE 9 Long-Run Travel Regression

Variable	Transport	Private	Public	Buses	Hires	Taxis	Seasons	Trains
CONST								
Coefficient	−1.6	−4.1	2.5	6.3	1.6	.08	.11	.04
t-Statistic	−3.3	−12.1	6.4	3.4	5.3	0.9	1.3	0.4
RNPCE								
Coefficient	0.86	.91	−.04	−.10	.02	.04	−.02	.01
t-Statistic	10.2	15.5	−.7	−3.1	0.3	2.6	−1.3	0.7
NFW								
Coefficient	.20	.04	.16	−.01	.08	.01	−.06	.01
t-Statistic	2.1	0.7	2.1	−0.2	1.3	0.4	3.9	0.6
NMW								
Coefficient	.18	.04	.14	.10	.08	−.005	−.003	−.03
t-Statistic	2.8	0.5	2.8	4.1	2.0	−0.4	−0.3	−2.4
NCP								
Coefficient	.05	.15	−.10	−.05	−.04	.01	−.01	−.01
t-Statistic	0.9	4.2	−2.4	−2.6	−1.1	1.1	−1.3	−1.1
NCH								
Coefficient	.21	0	.21	.03	.07	−.01	.13	−.01
t-Statistic	2.5	0.1	3.0	0.9	1.4	−1.0	9.1	−0.6
CNE								
Coefficient	.06	.32	−.27	.75	−1.4	.07	.25	.12
t-Statistic	0.2	1.9	−1.4	8.1	−9.5	1.7	6.1	2.4
CC								
Coefficient	−.94	−.53	−.41	1.2	−1.7	.08	−0.7	.13
t-Statistic	−2.0	−1.6	−1.1	6.3	−5.6	1.0	−0.8	1.3
CS								
Coefficient	2.2	1.6	.61	1.7	−1.3	−.05	.03	.31
t-Statistic	6.6	6.9	2.4	13.3	−6.5	−0.9	0.5	4.4
TNE								
Coefficient	−1.4	− .47	−.90	.25	−1.1	−.18	.16	−.01
t-Statistic	4.6	−2.3	−3.9	2.3	−5.9	−3.5	3.1	−0.2
TNW								
Coefficient	−1.6	−.14	−1.5	−.18	−1.3	−.19	−.06	.26
t-Statistic	−4.9	−0.6	−5.6	−1.4	−6.3	−3.2	−1.1	3.5
TC								
Coefficient	−.93	−.29	−.64	.28	−.67	−.24	.03	−.01
t-Statistic	−3.5	−1.6	−3.0	2.8	−4.1	−5.2	0.7	−0.1
TS								
Coefficient	−1.5	−.41	−1.2	.01	−1.0	−.14	.005	−.002
t-Statistic	−5.6	−2.1	−5.3	0.1	−5.8	−2.9	0.1	−0.03
RNE								
Coefficient	.16	−.21	.37	.83	−.52	−.08	.10	.08
t-Statistic	0.6	−1.1	1.7	8.0	−3.1	−1.7	2.1	1.4
RNW								
Coefficient	−1.0	−.09	−.92	−.26	−.50	−.12	−.03	.02
t-Statistic	−4.0	−0.5	−4.6	−2.7	−3.2	−2.6	−0.6	0.3
RC								
Coefficient	.56	−.06	.62	.05	.80	−.14	−.04	−.01
t-Statistic	2.3	−0.3	3.2	0.5	5.3	−3.3	−1.0	−0.2

Note: Coefficients (times 100) express the shares in total expenditures of each mode. Abbreviations are as given in Table 7.

household trips are associated with either work or school. Although not all such trips give rise to expenditures, the pattern of paid-for trips appears to be similar to that for all trips. These results appear to be of considerable interest in their own right and, in the author's view, they demonstrate the usefulness of household expenditure survey data for the analysis of travel behavior.

ACKNOWLEDGMENT

The research reported here is part of a World Bank study into the demand for personal travel in developing countries. The author is grateful to the editor of that study and three referees for helpful comments and suggestions.

REFERENCES

1. D.A.C. Maunder. The Effects of Household Income and Location on Trip Rates and Travel Patterns of Different Socio-Economic Groups in Delhi, India. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1983.
2. D.A.C. Maunder. Household and Travel Characteristics in Jaipur, India. Working Paper 159. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
3. D.A.C. Maunder. Household and Travel Characteristics in Vadodara, India. Working paper 171. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
4. C.R. Eastman and D. Pickering. Transport Problems of the Urban Poor in Kuala Lumpur. TRRL Supplementary Report 683. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
5. M.J. Heraty. Public Transport in Kingston, Jamaica and Its Relation to Low Income Households. TRRL Supplementary Report 546. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1980.
6. M. Thobani. A Nested Logit Model of Travel Mode to Work and Auto Ownership. Journal of Urban Economics, Vol. 15, 1982, pp. 287-301.

7. A.B. Atkinson and J.E. Stiglitz. Lectures on Public Economics. McGraw-Hill, New York, 1980.
8. T.J. Wales and A.D. Woodland. Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints. Journal of Econometrics, Vol. 21, 1983, pp. 263-85.
9. L.F. Lee and M.M. Pitt. Specification and Estimation of Demand Systems with Limited Dependent Variables. Department of Economics, University of Minnesota, Minneapolis, 1983.
10. A.S. Deaton. Demand Analysis. In The Handbook of Econometrics (Z. Griliches and M.D. Intriligator, eds.), North-Holland, Amsterdam, in preparation.
11. J.J. Heckman. Dummy Endogenous Variables in a Simultaneous Equations System. Econometrica, Vol. 46, 1978, pp. 931-61.

The views expressed are those of the author and should not be attributed to the World Bank, to its affiliated organizations, or to any individual acting on their behalf.

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

Analysis of Automobile Ownership by Using a Divisive Hierarchical Technique

JOKE M. JAGER and WIM H. SCHELTES

ABSTRACT

A method of analysis of personal automobile ownership is presented that differs from the well-known aggregate and disaggregate methods. The analysis consists of two steps. First, a cluster-segmentation method is applied to data from the Dutch National Travel Survey. The results show that personal automobile ownership is mainly determined by personal net income, age, and sex. Second, a model has been specified that includes these factors. When income and age are accounted for, a structural difference in automobile ownership is shown between men and women. Furthermore, for the period studied (1979-1982) the results indicate that for the age group 65 and older, automobile ownership increased significantly, whereas for those 25 years and younger, it decreased. Advantages of the method are (a) the relative stability of the homogeneous population groups independent of accidental changes in the survey population and (b) insight into the relationship of automobile ownership with the most essential determining factors. Because of these advantages, the method presented can be used to improve both analysis and forecast of automobile ownership.

It has been widely recognized that travel behavior is strongly influenced by automobile ownership. This applies to mode choice as well as to trip frequency and daily mileage (1-3). Therefore numerous models of automobile ownership have been developed, both at aggregate and disaggregate levels (4). Some models are based on time-series data under the assumption that a certain saturation level exists, whereas others are disaggregate at the household level and are based on cross-sectional data (5,6). Because both aggregate and disaggregate methods suffer from a number of disadvantages (7-9), another method is applied in this paper.

Phase 1 of this work aims at finding those demographic and socioeconomic factors that influence personal automobile ownership most. The survey population is split into homogeneous population groups according to these most important factors. Personal automobile ownership is preferred here to household automobile ownership because models of travel behavior are usually specified at a personal level and because this analysis is part of a comprehensive transportation study. Phase 2 of this work focuses on the level of influence of these factors on automobile ownership and investigates trends in the development of automobile ownership in the homogeneous population groups.

DATA BASE

The data base used was Onderzoek Verplaatsingsgedrag (OVG), the Dutch National Travel Survey (10). It contains extensive information, both demographic and socioeconomic data on the interviewees and data on the trips they made. Since the survey started in 1978, about 25,000 persons have been interviewed annually. In contrast to most other surveys, the interviews were held throughout the year. Because of the large number of questions, the home-interview technique was used. A trained interviewer registered the information concerning the household. The data on trips were recorded by the respondents (12 years old or older) themselves, usually during a 2- or 3-day period.

VARIABLES OF POTENTIAL IMPORTANCE

A whole range of demographic and socioeconomic variables correlate with automobile ownership. A portion of this range formed by the variables that were collected in the Dutch National Travel Survey has been analyzed for the relation with automobile ownership.

These characteristics of interviewees and their environment are listed alphabetically (the number of distinct classes is given in parentheses after each variable):

- Age (5)
- City size (3)
- Educational level (7)
- Employment status (5)
- Household income (6)
- Household size (6)
- Marital status (4)
- Personal income (6)
- Position in household (7)
- Province (13)
- Pattern of visit to office or school (6)
- Sex (2)

Note that a number of other possibly relevant variables (such as population density and public transport quality) were not included.

METHOD OF ANALYSIS

It is obvious that the method of analysis should belong to the group of multivariate techniques. From a mathematical point of view, the interviewees are observations in a multidimensional space formed by the selected demographic and socioeconomic variables and their distinct classes.

Clustered observations in this hyperspace indicate dominating variables or dependencies or both. A well-known technique to find these data concentrations is the cluster-segmentation analysis. Its mathematical robustness makes it easy to use. Several algorithms exist [see, e.g., studies by Everitt (11) and Späth (12)], most of them combining nearby observations to form clusters. However, dividing the data into separate groups is (in the case of many observations) preferable, as it often leads to better group distinction. These algorithms are called segmentation or divisive hierarchical techniques.

The likelihood-segmentation method is one of these. Its algorithm was developed by Hamerslag (1) by using the likelihood-estimation theory. When the number of clusters is equal to the number of observations (so each cluster contains only one element), the likelihood value is maximal. Each data combination causes the likelihood value to decrease, so the lowest value will be found when all observations are merged into one single cluster. Thus the value of the likelihood can be used to measure loss of information during the clustering process. All variables have discrete classes; for each variable, classes are combined and the information loss is calculated. The variable that shows the largest drop is (normally) the most significant.

The opposite is also valid: Dividing a data cluster will increase the likelihood value and the information level. Thus the data base will be split into several segments (subspaces), every segment being a class from the most discriminating variable (dimension). The obtained data groups can be analyzed in the same manner. Theoretically the process ends when all clusters contain only one observation. In practice, however, segmentation is stopped in an earlier phase.

The formula for calculating the decrease or increase of information (using an amount k of simultaneous criteria factors) is

$$d(X,Y) = \sum_k [(N_x\bar{X}_k)\ln(\bar{X}_k + \text{eps}) + (N_y\bar{Y}_k)\ln(\bar{Y}_k + \text{eps}) - (N_x\bar{X}_k + N_y\bar{Y}_k)\ln(\bar{U}_k)]$$

$$\bar{U}_k = (N_x\bar{X}_k + N_y\bar{Y}_k)/(N_x + N_y) \qquad (1)$$

where

$d(X,Y)$ = information loss when group X and group Y are combined;
N_x, N_y = number of elements in groups X and Y, respectively;
$\bar{X}_k, \bar{Y}_k$ = average value of the observations in groups X and Y, respectively;
eps = small number to prevent ln(0) from occurring; and
k = number of criteria factors, used to measure the data distance (normally k = 1).

For more details, see work by Hamerslag (1).

SEGMENTATION RESULTS

In the Netherlands the minimum age for drivers is 18. To make certain that statistics are useful and understandable, all younger interviewees (and those with unstated answers) are eliminated from the 1979 data base, leaving 16,777 persons to analyze.

The first computer run showed that the variables personal income, position in household, and sex had a significant influence on automobile ownership. Although the information loss among these variables did not differ much, personal income was pointed out as most important, for the following reasons:

- It is an important indicator for potential expenditure, in contrast to both other variables mentioned. Fairly static characteristics such as sex and household position cannot easily be used in research applications like forecasting. By using income, continuous annual research is also possible.
- Research on travel behavior and in particular travel performance also showed personal income to be of great significance (1). Linking up to this might bear advantages in interpreting the results.

To maximize the information level, the data base was split up into its distinct income classes. As seen in Figure 1, low income and high income form quite homogeneous groups (small data dispersion). This indicates that there are no other important influences, so further analysis of these groups is not necessary.

In the second computer run, all income classes were analyzed separately. Age, position in household, and sex turned out to be the most discriminating factors. After all information loss was totaled for each class (this is allowed because of interdependencies), age was found to be the most important

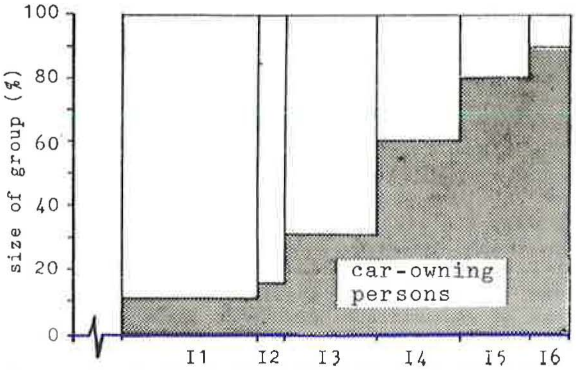

FIGURE 1 Relationship between automobile ownership and personal net income (source: Dutch National Travel Survey).

variable. Thus each income group was split into age classes. The effect of personal income and age on automobile ownership is shown in Figure 2. In most cases the peak level is located around the ages of 36 to 45.

In the third computer run the population groups formed were analyzed further. Depending on the group

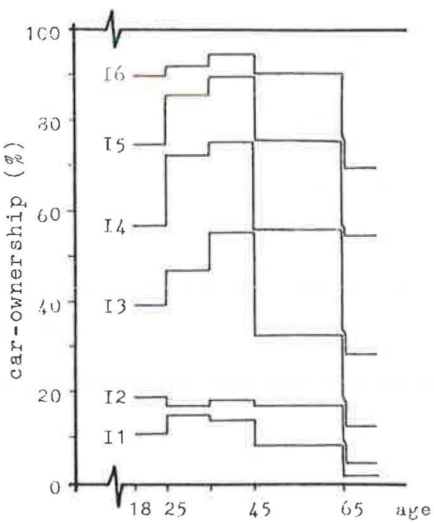

FIGURE 2 Automobile ownership in relation to personal net income and age (source: Dutch National Travel Survey).

characteristics, sex and position in household turned out to be the most important factors. New groups were created based on this knowledge. At this stage some groups did not contain enough interviewees to be analyzed in further detail, whereas the remaining ones had small data dispersion (homogeneous data). Therefore the analysis stopped.

In Figure 3 the results of each computer run are shown. Note that the population groups given are quite homogeneous, or too small to analyze.

CLASSIFICATION OF POPULATION GROUPS

As was demonstrated in Figure 3, the two most relevant factors to personal automobile ownership are personal net income and age. Another important factor is sex, which sometimes appears in a different form, such as position in the household or employment status.

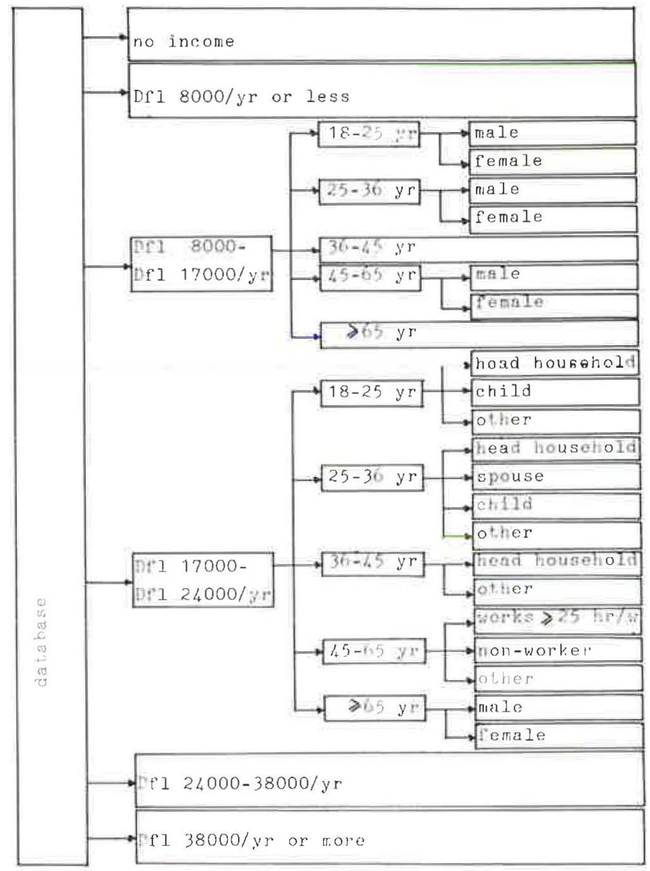

FIGURE 3 Segmentation results visually summarized.

It was therefore decided to classify the survey population according to personal net income, age, and sex. For each group the percentage of personal automobile ownership was calculated. This was done for the survey years 1979-1982, which span the turbulent period of the second oil crisis and the subsequent income stagnation in the Netherlands. The classification is given in Table 1.

TABLE 1 Classification of the Relevant Factors for Automobile Ownership

Personal Income (Dfl x 1,000)	Age (yr)	Sex	Year
No income (I1)	18-25 (A1)	Male (S1)	1979 (Y1)
0-8 (I2)	25-36 (A2)	Female (S2)	1980 (Y2)
8-17 (I3)	36-45 (A3)		1981 (Y3)
17-24 (I4)	45-65 (A4)		1982 (Y4)
24-38 (I5)	≥65 (A5)		
≥38 (I6)			

Source: Dutch National Travel Survey.

COMPOSITION OF SURVEY POPULATION

The composition of the survey population in 1982 according to age and income is given in Figures 4 and 5. As expected, there is a somewhat higher percentage of women than men in the group over 65 years.

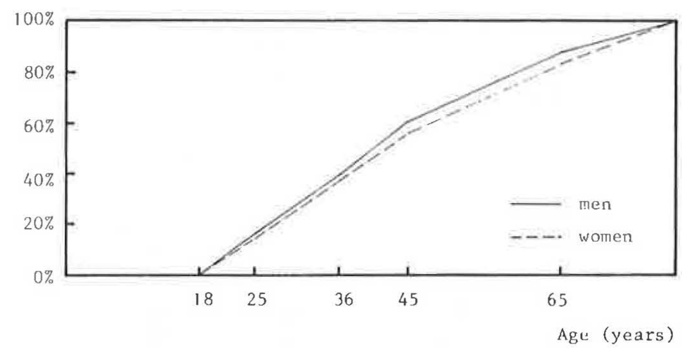

FIGURE 4 Cumulative age distribution for 1982 (source: Dutch National Travel Survey).

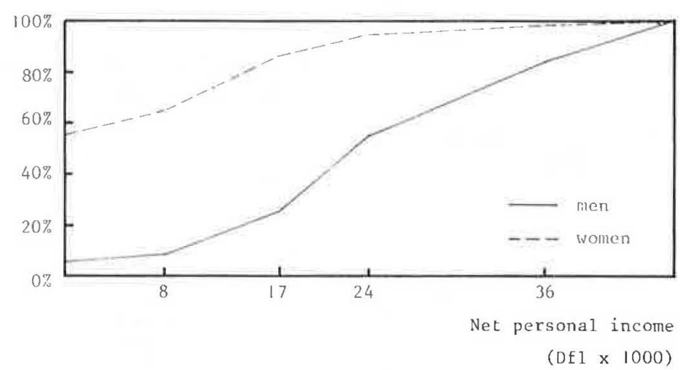

FIGURE 5 Cumulative income distribution for 1982 (source: Dutch National Travel Survey).

The distribution of income by sex shows great differences between men and women: about 65 percent of the women have a personal net income of less than Dfl 8,000 per year versus 7 percent of the men. The percentage of men and women in the annual survey population is found to be stable, ranging from 50.16 to 49.25 percent men. The composition of the survey sample can have important consequences for aggregate data such as personal automobile ownership.

Table 2 gives the percentage of personal automobile ownership for 1979-1982 for both the total population and men and women separately. It can be seen that a 1 percent increase in the proportion of women in the survey may cause a drop of more than 0.5 percent in automobile ownership for the total group.

TABLE 2 Percentage of Automobile Ownership

Year	Total Group	Men	Women
1979	41.9	68.3	15.4
1980	43.2	69.9	16.6
1981	43.4	70.7	16.5
1982	42.2	69.5	15.8

Source: Dutch National Travel Survey.

MODEL FORMULATION

To analyze the influence of the most relevant factors (income, age, and sex) on automobile ownership and also the development of automobile ownership over the years 1979-1982 for the distinct groups, a multiplicative model was used:

$$\mathrm{NUMCO}(i,j,k) = C * \mathrm{NPERS}(i,j,k) * I(i) * A(j) * S(k) \tag{2}$$

where

NUMCO(i,j,k) = number of automobile-owning persons in groups i, j, and k;
C = constant;
NPERS(i,j,k) = total number of persons in groups i, j, and k;
I(i) = coefficient of income class i;
A(j) = coefficient of age class j; and
S(k) = coefficient of sex class k.

The estimation was carried out by using a weighted multiproportional Poisson estimation method (13,14). This was done both for marginal factors and for the simultaneous inclusion of all factors.

The results of the general estimations are shown in Tables 3 and 4. Table 3 indicates a strong increase in automobile ownership with increasing income as well as a relation to age (lowest in the group of 65 and older, highest in the group 36 to 45). As shown, women reach only about 22 percent of the automobile-ownership level of men.

Table 4 indicates, however, that the simultaneous inclusion of all relevant factors leads to somewhat less marked differences per factor, which is caused by correlation between the factors. The difference between men and women is sharply reduced with the inclusion of the income factor. This is because of the previously mentioned wide difference in personal income distribution between men and women.

TABLE 3 Results of the General Estimation of Automobile Ownership with Marginal Factors

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Income	Constant	-	-	C(I) = 0.11
(Dfl)	No income	16,513	10.9	I1 = 1.00
	0-8,000	4,471	16.9	I2 = 1.55
	8,000-17,000	10,844	30.8	I3 = 2.82
	17,000-24,000	10,320	62.1	I4 = 5.69
	24,000-38,000	8,563	80.2	I5 = 7.35
	≥38,000	4,402	89.0	I6 = 8.20
Age	Constant	-	-	C(A) = 0.20
(yr)	18-25	8,219	32.4	A1 = 1.60
	25-36	13,959	50.9	A2 = 2.51
	36-45	8,930	54.9	A3 = 2.70
	45-65	16,290	42.2	A4 = 2.08
	≥65	7,715	20.3	A5 = 1.00
Sex	Constant	-	-	C(S) = 0.16
	Men	26,561	70.1	S1 = 4.44
	Women	28,552	15.8	S2 = 1.00

Source: Dutch National Travel Survey.

TABLE 4 Results of the General Estimation of Automobile Ownership with Simultaneous Inclusion of All Factors

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Constant	-	-	-	C = 0.052
Income	No income	16,513	10.9	I1 = 1.00
(Dfl)	0-8,000	4,471	16.9	I2 = 1.42
	8,000-17,000	10,844	30.8	I3 = 2.26
	17,000-24,000	10,320	62.1	I4 = 3.25
	24,000-38,000	8,563	80.2	I5 = 3.75
	≥38,000	4,402	89.0	I6 = 4.17
Age	18-25	8,219	32.4	A1 = 1.89
(yr)	25-36	13,959	50.9	A2 = 2.17
	36-45	8,930	54.9	A3 = 2.21
	45-65	16,290	42.2	A4 = 1.88
	≥65	7,715	20.3	A5 = 1.00
Sex	Men	26,561	70.1	S1 = 2.18
	Women	28,552	15.8	S2 = 1.00

Source: Dutch National Travel Survey.

It should be noted, however, that when the effects of income and age have been taken into account, the percentage of automobile ownership by women is less than half that of men. The difference in automobile ownership between men and women is least for middle-aged persons and the highest income group.

Similar estimations were performed with the model including a factor for the distinct survey years. No overall trends for automobile ownership were found.

SEPARATE ESTIMATION BY SEX

In Tables 5 and 6 separate estimations are presented for men and women. These results are shown in Figure 6. It is found that the relative effect of the distinct factor classes is much greater for women than for men (although absolute automobile ownership for women is much lower in any case).

The age effect for women is especially strong; there is a sharp drop in automobile ownership for the group over 65 years. Similar estimations have been made with the inclusion of a year factor for the total group and for separate income and age groups.

It was found that there are significant changes in the automobile ownership of the oldest age group; automobile ownership increased in 1982 compared with previous years. This is sometimes referred to as a "generation effect." In the youngest age groups, automobile ownership decreased. There are no significant changes in automobile ownership by income group.

TABLE 5 Results of the Simultaneous Estimation of Automobile Ownership for Men

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Constant	-	-	-	C = 0.134
Income	No income	1,254	22.1	I1 = 1.00
(Dfl)	0-8,000	677	32.6	I2 = 1.50
	8,000-17,000	4,950	47.6	I3 = 2.37
	17,000-24,000	8,047	69.9	I4 = 3.16
	24,000-38,000	7,532	84.5	I5 = 3.65
	≥38,000	4,101	91.9	I6 = 4.06
Age	18-25	4,122	47.1	A1 = 1.62
(yr)	25-36	6,959	81.9	A2 = 1.83
	36-45	4,569	86.8	A3 = 1.83
	45-65	7,455	75.6	A4 = 1.64
	≥65	3,456	39.7	A5 = 1.00

Source: Dutch National Travel Survey.

TABLE 6 Results of the Simultaneous Estimation of Automobile Ownership for Women

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Constant	-	-	-	C = 0.024
Income	No income	15,259	9.9	I1 = 1.00
(Dfl)	0-8,000	3,794	14.1	I2 = 1.37
	8,000-17,000	5,894	16.7	I3 = 2.05
	17,000-24,000	2,273	34.5	I4 = 3.62
	24,000-38,000	1,031	48.8	I5 = 4.85
	≥38,000	301	57.5	I6 = 5.98
Age	18-25	4,097	17.7	A1 = 3.75
(yr)	25-36	7,000	20.0	A2 = 4.85
	36-45	4,361	21.5	A3 = 5.63
	45-65	8,835	14.1	A4 = 3.71
	≥65	4,259	4.5	A5 = 1.00

Source: Dutch National Travel Survey.

CONCLUSIONS

The method applied in this paper is a useful alternative to the generally used aggregate or disaggregate methods. It leads to distinguishable population groups, the behavior of which with respect to automobile ownership is different. The groups themselves are homogeneous because they are differentiated on the basis of the most influential factors that were documented in the data base.

With data from the Dutch National Travel Survey, it was found that automobile ownership is mainly determined by personal net income, age, and sex. It is essential to distinguish between the right influential factors, because changes in the composition of the (survey) population in relation to these factors will cause significant changes in the automobile ownership of the aggregate group, whereas other factors will have much less effect.

Analysis of a multiplicative model of automobile ownership versus income, age, and sex for four consecutive years has clarified these relationships and their interdependencies. Because there is a strong relation between automobile ownership and income, it can be expected that future income development will influence automobile ownership as well.

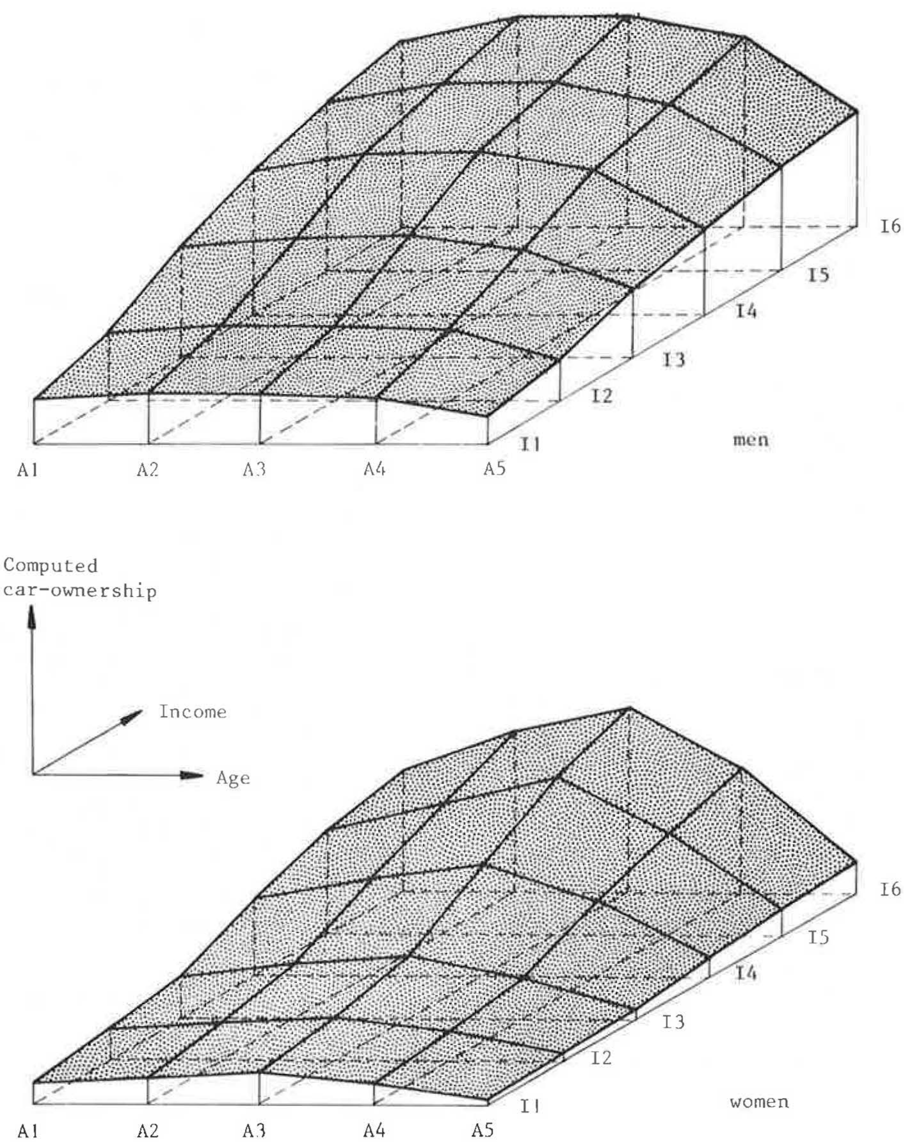

FIGURE 6 Results of estimated automobile ownership coefficients for income and age by sex [CAROW% = C x I(i) x A(j)].

Furthermore, continuing generation effects are expected, which means that the automobile ownership in the group older than 65 will increase. Although the relative effect is especially strong for women, this will only have a limited effect on total automobile ownership, because of the small absolute values. The same applies to developments such as increasing participation of women in the labor force, which would influence their personal income.

Automobile ownership in the age group of 18-25 has decreased slightly. This may have been caused by unemployment, but further analysis will be necessary to confirm this.

It is remarkable that there is a structural difference in automobile ownership between men and women, even in the middle-aged and high-income categories. This can most probably be explained by the position in the household and employment status, which are not very likely to change rapidly in the near future.

The consequences of the established relationships for forecasting purposes are interesting and a number of exercises for possible developments can be performed, assuming that the established relationships remain valid. Because the population group that will be 18-25 years old in the year 2000 has already been born, the group size is known. The same applies to the other age groups, depending on their life expectations. By using the relative group sizes of the survey sample, the automobile ownership in, for instance, the year 2000 (ignoring any income effects) can be calculated. For the total population that is 18 years or older the result of such a calculation shows 48 percent automobile ownership (76 percent for men and 20 percent for women).

A similar exercise is to calculate automobile ownership for several scenarios of future economy (without age effects). An increase of 1 percent per year in personal net income for all income groups may lead to 46 percent automobile ownership (75 percent for men and 18 percent for women) in the year 2000.

If women obtained exactly the same income distribution as men have at present, total automobile ownership would rise to about 50 percent (women would reach 32 percent automobile ownership). This, however, assumes that there still is a structural difference in automobile ownership between men and women. (This could be caused by factors that are assumed to remain unchanged, e.g., position in the household.) Other exercises such as calculation of combined effects could of course be performed as well.

A time-trend calculation based on historic data or a backward calculation in time using the survey data could be applied to confirm the validity of these coefficients for forecasting purposes.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Rudy Hamerslag of Delft University of Technology for his assistance with the application of the segmentation method and to Velibor Vidakovic of the Municipality of Amsterdam, who provided many useful comments on earlier versions of the paper.

REFERENCES

1. R. Hamerslag. Population Groups with Homogeneous Travel Performance by Car, Public Transport and Bicycle. Department of Civil Engineering, Delft University of Technology, Netherlands, 1983.
2. C.G.B. Mitchell. The Influence of the Car on Personal Travel. TRRL Report SR 681. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
3. J. Supernak. Car Availability Versus Car Ownership Modelling: Theoretical Discussion and Empirical Findings. Presented at World Conference on Transport Research, Hamburg, West Germany, 1983.
4. Car Ownership and Use. Organization for Economic Cooperation and Development, Paris, France, 1982.
5. Models for Car Ownership and Licence Holding. Zuidvleugel Report 9. Rijkswaterstaat and Cambridge Systematics Europe, The Hague, Netherlands, 1981.
6. J.C. Tanner. Saturation Levels in Car Ownership Models: Some Recent Data. TRRL Report SR 669. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
7. C.J. Ruijgrok. Disaggregate Choice Models: An Evaluation of New Developments in Modelling Travel-Demand and Urban Systems. Presented at Transportation Planning Colloquium, The Hague, Netherlands, 1979.
8. J. Supernak. Toward a Better Understanding of Transportation Related Choices. Presented at Transportation Planning Colloquium, Zandvoort, Netherlands, 1983.
9. J. Supernak. Transportation Modeling: Lessons from the Past and Tasks for the Future. Transportation, Vol. 12, No. 1, 1983.
10. F.W.M. Hendrikx. Framework and Execution of the Dutch National Travel Survey (Opzet en Uitvoering van het Onderzoek Verplaatsingsgedrag). Stichting Postacademiale Vorming Verkeerskunde, Delft, Netherlands, 1982 (in Dutch).
11. B. Everitt. Cluster Analysis. Heinemann Educational Books, Inc., Portsmouth, N.H., 1980.
12. H. Späth. Cluster Analysis Algorithms (English translation). Queen Elizabeth College, University of London, United Kingdom, 1980.
13. H. Derriks. Computer Documentation of the Weighted Multiproportional Poisson Estimation Program. Department of Civil Engineering, Delft University of Technology, Netherlands, 1984.
14. R. Hamerslag, J.P. Roos, and M. Kwakernaak. Analysis of Accidents in Traffic Situations by Means of the Weighted Multiproportional Poisson Model. In Transportation Research Record 847, TRB, National Research Council, Washington, D.C., 1982, pp. 29-36.

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

Automobile Availability and Its Application in Transportation Studies

JANUSZ SUPERNAK and DAVID SCHOENDORFER

ABSTRACT

The usefulness of personal automobile availability in travel behavior analyses as an alternative to the more familiar automobile-ownership approach is discussed. A measure of automobile availability for an individual based over a longer period of time (an average situation) rather than a person's actual access to the automobile at a specific time is offered. Four versions of the proposed definition of automobile availability are formulated and their performance is studied by examining the relationship between automobile availability and modal split. This is done by using data sets from Baltimore and the KONTIV data sets from Germany. Results of this analysis give preference to a simple three-level stratification [automobile (never/sometimes/always) available] in defining automobile availability for its apparent application in travel demand and policy analyses.

Until recently, the term "automobile availability" appeared in the literature just as a synonym for family automobile ownership. A forecast level of the family automobile ownership (0, 1, or 2+ automobiles) was a simple and natural input into the household-based trip generation models and an important explanatory variable for modal-split models.

However, in recent years interest in the variable "automobile availability," which describes an individual's access to a private automobile, has increased noticeably. In order to better understand transportation-related behavior it became necessary to focus on analyses oriented toward individuals within the household rather than on the household as a whole, because individual needs, options, and constraints are ultimately responsible for the travel choices made. Both individually oriented modal-split models and recent person-based trip generation models (1,2) required a compatible term: "individual automobile availability" rather than "household automobile ownership." Indeed, automobile availability, or its equivalent variable, automobile competition, appeared crucial for both modal choices (1, 3-7) and mobility analysis (1,2).

The usefulness of the automobile availability concept in transportation studies is examined with a primary focus on modal-split analysis. The following issues are addressed: (a) comparison between automobile availability and automobile ownership concepts and (b) comparison of alternative definitions and measurements of automobile availability with primary reference to the relationship between automobile availability and modal choices in different geographic contexts.

Two data sets are utilized in this study: the Baltimore Disaggregate Data Set from 1977 and the German data set called KONTIV, gathered in 1976. Only data records from German cities with more than 500,000 inhabitants (code 7) were considered for the KONTIV set, to make it comparable with the Baltimore set.

COMPARISON BETWEEN AUTOMOBILE AVAILABILITY AND AUTOMOBILE OWNERSHIP APPROACHES

For all household-oriented modeling approaches, an automobile ownership description (0, 1, or 2+ automobiles in the family) is a natural and simple one. For years, the forecast level of family automobile ownership has been commonly used as an input to the household-based trip generation models and as an important explanatory variable for modal-split models. However, a closer look at this problem from the point of view of an individual--the true decision maker and traveler--can raise some doubts about the adequacy of the term "automobile ownership" for travel behavior analyses and disaggregate travelchoice models.

First, it is apparent that any given level of family automobile ownership seldom means equal access to automobiles for all family members. For example, some will be primary users, whereas others will have to wait for the automobile until it is not needed for a more important activity. Also, not all family members may have a driver's license. A seemingly easy automobile-sharing arrangement among family members may often be significantly restricted if their activities outside the home are, for different reasons, temporarily or spatially inflexible.

Therefore, the total number of automobiles owned by a family may not be an absolutely objective description of high or low ownership level because it does not refer to the real need for an automobile by each family member. For example, family ownership of two automobiles (seemingly high) may not fully satisfy the needs of a family with four drivers if three of them are employed at different, widely dispersed locations. On the other hand, a low ownership level of one automobile will warrant unrestricted access to the automobile if there is only one driver in the household.

In addition, modal choices made by different family members depend primarily on the availability of private transportation to each individual family member rather than on the overall automobile ownership of the family. "Family modal choice" is virtually an undefinable term because the individual mode choices are often dramatically different among family members. Thus, the automobile availability description may be more suitable than the automobile ownership approach in describing the behavioral background of the modal-split choices, which are always closely related to automobile ownership and automobile availability issues.

Finally, the household-based automobile ownership concept encounters several problems when dynamic changes in the family "life cycle" (family size or number of employed members) are accounted for. Such changes contribute to the changing attitudes toward possessing a given number of automobiles. These changes are crucial considerations for long-range travel forecasts. Also there is difficulty in capturing such commonly observed trends relevant to the family automobile ownership issue as (a) decrease in the household size (fewer children and lower percentage of three-generation families), (b) increase in families with two or more breadwinners (increase in percentage of female employment), (c) increase in percentage of single-parent families, (d) increase in percentage of single persons, (e) increase in the average age of the population, and (f) increase in the percentage of women possessing a driver's license.

Because the automobile availability concept addresses the issue of access to an automobile at the individual level, it provides the potential for a more precise and better behaviorally based description of the complex relationship among a person's need for travel, travel opportunities, and actual travel itself. It should be stressed that any description of automobile availability should not ignore obvious family links and constraints, which may affect both access to the automobile and its actual use.

DEFINITIONS AND MEASUREMENTS OF AUTOMOBILE AVAILABILITY

Previous Studies

The majority of studies on automobile availability have based their measures on direct questions asked in household surveys, for example, "Was there an automobile available that you could have used for this trip?" Bailey (4) argues that this simple question can be difficult to interpret and is likely to be difficult for a respondent to answer quickly.

Therefore, in a number of studies, attempts have been made to overcome some of the problems and ambiguities by not asking the question about actual automobile availability at all (4). Instead, a number of assumptions had to be made regarding the priority of use of the automobile in potentially conflicting and nonconflicting situations. From these assumptions a judgment about actual automobile availability was made.

Gwilliam and Banister (8), for example, made the following assumptions: (a) an automobile was considered available for a particular trip when it was not in use and located at the point from which that trip was to begin, (b) availability of the automobile for passenger travel was excluded from the analysis, and (c) all trips measured as "automobile available" would be made by automobile.

Bailey and Layzell (9) postulated a clarification of the automobile availability concept. They compared the number of license holders and number of automobiles in the household, and defined the automobile as being available only if it remained unused for the duration of the period for which a particular traveler would be away from home.

It has to be emphasized that all the aforementioned definitions of automobile availability consistently attempted to represent actual access to an automobile by an individual at the analyzed specific period of time.

Proposed Definition of Automobile Availability

Stopher and Wilmot (7) developed individual-choice models of modal split by using a variable defined as "automobile competition." This variable was defined as the ratio of automobiles to licensed drivers and was a continuous measure of automobile availability. When the automobile competition variable was added to the multinomial logit model as a mode-specific variable for automobile driver, it added significant explanatory power to the work-trip model of mode choice.

Similar to Stopher and Wilmot's definition of automobile competition, the definition of automobile availability proposed here refers to the average situation over a longer period of time rather than just a survey day, which could be atypical (e.g., an automobile normally available for a given individual could be in for repair during the survey day; a person without an automobile can use a friend's automobile, etc.). Instead of delving into a complex system of dependencies to understand why an automobile is available for a given family member at certain times or directly asking the question about automobile availability in the household survey (which will never bring any clear-cut answers), it appears to be more beneficial and practical to investigate the general travel choices made by a person who has no access to the automobile as a driver, limited access, or unlimited access. The difference between the analysis presented in this paper and the one made by Stopher and Wilmot (7) is that the purpose of the automobile availability variable used here is to stratify the population into homogeneous groups that can be examined individually rather than using the automobile competition variable as an explanatory variable for multinomial logit (MNL) models.

The concept of automobile availability presented in this paper was originally proposed by Supernak et al. (5,6,10). The three-level choice described in this approach [automobile (never/sometimes/always) available] replaces a two-level choice [automobile (available/ not available) for a particular trip]. Automobile availability levels are described as follows (N_c = number of automobiles in the household and N_d = number of persons with a driver's license in the household):

Criterion	Automobile Availability: Drivers	Nondrivers
$N_c = 0$	Never	Never
$N_c > 0$, $N_d > N_c$	Sometimes	Never
$N_c > 0$, $N_d \leq N_c$	Always	Never

It should be noted that similar to previous work [e.g., that of Bailey (4)], the level of automobile availability refers to the availability to drive an automobile at any given time (not to being a passenger). Theoretically, any person can be a passenger in an automobile at any time (by hiring a taxi, for example). Also, any ridesharing arrangements are not limited to the same household.

Automobile Availability as an Element of a Person Category Travel-Demand Analysis

The concept of automobile availability was developed as part of an integrated modeling system based on homogeneous person categories. In particular, the result of the automobile availability model--a forecast share of the population with an automobile available (never/sometimes/always)--is the direct input into the person category trip generation model (1).

In developing this model, a multistage, multivariate analysis of factors influencing a person's travel behavior proposed that the most significant variables describing differences in travel behavior were age, employment status, and automobile availability. This analysis resulted in the formulation of eight homogeneous person categories as seen in Figure 1.

Age reflects obvious differences in demand for travel among (a) preemployment, (b) employment, and (c) postemployment stages in everyone's life. Employment status reflects a basic distinction between employed and nonemployed adults with respect to their demand for activities and travel. The former

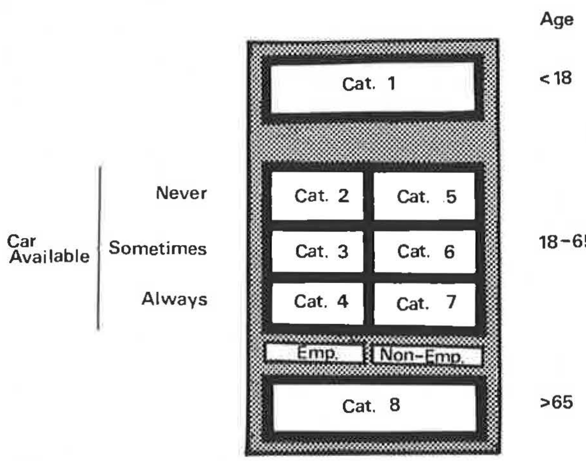

FIGURE 1 Description of the eight-person homogeneous categories.

group participates regularly in both obligatory and discretionary activities, whereas the latter participates primarily in discretionary outside-home activities. The third variable describes a person's ability to fulfill his or her travel needs through "purchasing" the services offered by the most convenient transportation mode: an automobile.

The aim of the person category automobile availability model is to describe the proportions $\alpha_2 : \alpha_3 : \alpha_4$ and $\alpha_5 : \alpha_6 : \alpha_7$, where α_i is the share of the population in category i. This is the only remaining element needed to forecast category percentages $\alpha_1, \alpha_2, \ldots, \alpha_8$ for the trip generation model. Shares of α_1 and α_8 are known from the demographic forecasts, whereas the split between the employed and nonemployed adults $(\alpha_2 + \alpha_3 + \alpha_4)/(\alpha_5 + \alpha_6 + \alpha_7)$ is known from the labor force and employment projections (which have to be made anyway for trip generation and trip distribution forecasting).

It should be noted that the level of automobile availability has to be described separately for employed $(\alpha_2 : \alpha_3 : \alpha_4)$ and nonemployed $(\alpha_5 : \alpha_6 : \alpha_7)$ adults because it can be reasonably expected that the need for an automobile is significantly differentiated between these two groups. Supernak has discussed the description of these shares briefly (6).

Modifications of Proposed Definition of Automobile Availability

The description of individual automobile availability proposed earlier is only one of many possible formulations. Although the situations "automobile never available" and "automobile always available" are clearly specified, there could be several alternative definitions of the situation "automobile sometimes available" to capture the difference between, say, one automobile shared by three drivers or three automobiles shared by four drivers. Therefore, three other descriptions of the situation "automobile sometimes available" are presented as modifications to the proposed definition of automobile availability.

Modification 1

Obligatory trips, which include work and school trips, usually must occur at a specific time of the day. Discretionary trips, which include personal business, shopping, and social-recreational trips, are more flexible and can be scheduled for more convenient times during the day. Because obligatory trips are on this rigid time schedule, they would usually be considered to have priority over discretionary trips. Therefore, in a household with less automobiles than drivers, the employed individuals may often have priority for the automobile to go to work.

Schoendorfer (11) proposed a modification of the automobile availability description that could account for the priority given to obligatory trips. The definition of the situation "automobile sometimes available" was revised by introducing a variable N_e, which is the number of employed persons in the household with a valid driver's license. The N_c variable (number of automobiles in the household) maintained the same definition that was used in the previous version.

If $N_e > N_c (N_c/N_e < 1)$, then some employed persons in that household may not always have unrestricted access to the automobile for their obligatory trips. If $N_e \le N_c (N_c/N_e \ge 1)$, then the employed persons in this household may not need to compete for the automobile for obligatory trips or "always" have access to the automobile for these trips. Taking this into account, the original Category 3 (employed, automobile sometimes available), was divided into two groups, Categories 3A and 3B. Category 3A includes those individuals who may "sometimes" have access to the automobile for obligatory trips and Category 3B is those individuals who may "always" have access to the automobile for these trips.

Modification 2

Modification 2 examines in more detail Category 3A according to the actual ratio N_c/N_e. Although theoretically continuous, in reality this ratio is reduced to a relatively few discrete values resulting from possible combinations of number of automobiles, drivers, and employees in a household.

It was found that this ratio was often 1/3, 1/2, and 2/3. Based on this finding, the stratification of the segment of the population that is employed with an automobile sometimes available resulted in four ranges; $0 < x \le 1/3$; $1/3 < x < 2/3$; $2/3 \le x < 1$; and $x \ge 1$. The categories were named 3A.1, 3A.2, 3A.3, and 3B, respectively.

Modification 3

Modification 3 is an extension of the original description of automobile availability based on the ratio N_c/N_d to introduce more segments in the "automobile sometimes available" category $(0 < N_c/N_d < 1)$. The stratification of the segment of the population that is employed with an automobile sometimes available resulted in three groups, 3.1, 3.2, and 3.3, corresponding to N_c/N_d ranges $0 < x \le 1/3$, $1/3 < x < 2/3$, and $2/3 \le x < 1$, respectively.

COMPARISON OF ALTERNATIVE DEFINITIONS OF AUTOMOBILE AVAILABILITY

The aim in this section is to recommend the preferred version of automobile availability descriptions by comparing the proposed version and Modifications 1, 2, and 3. Data sets from Baltimore, Maryland, and German cities are utilized in order to (a) examine population representations within each automobile availability segment, (b) examine the consistency of category-specific modal-split characteristics for alternative descriptions of automobile availability, (c) compare modal-split characteristics of person categories between Baltimore and the German cities (this is done for the purpose of comparing automobile availability definitions only; the modal-split relationship will be investigated in greater detail in a separate paper), and (d) examine prospects for transferring category-specific modal shares from Germany to predict the use of the automobile and public transit modes in Baltimore.

Population Representations for Each Category

Table 1 shows the percentage of the population represented by each category. Most of the emphasis in this section is directed at Category 3 (age 18 to 65, employed, automobile sometimes available), which represents only about 20 percent of the population in both Baltimore and the German cities. Although this may be considered a small percentage of the entire population, it represents those who sometimes have an automobile available, a category that is not as clearly defined as situations in which an automobile is never or always available. However, when the individuals in Category 3 were reclassified according to Modifications 1, 2, and 3, it was recognized that this group was not as ambiguous as originally thought. In large German cities, for example, 74.6 percent of Category 3 fell into Category 3A.2 (N_c/N_e > 1/3 and < 2/3) for Modification 2 and 86.3 percent of Category 3 fell into Category 3.2 (N_c/N_d > 1/3 and < 2/3) for Modification 3. This suggests that because such a small segment of the population is represented by some of the other categories (Categories 3A.1 and 3A.2 in Modification 2 and Categories 3.1 and 3.3 in Modification 3), little accuracy is to be gained by the further stratification of Category 3, as was done in all modifications.

For Baltimore it is interesting to note that a larger percentage of the employed population has an automobile always available (23.6 percent of the population, or 46.8 percent of the work force). Almost half of the population represented by Category 3 always has an automobile available when availability is defined by the ratio of N_c/N_e as in Modification 1. This means that nearly 62 percent of the work force always has access to the automobile for obligatory trips. This suggests that the number of employees in a household directly affects the desired level of automobile availability within a household.

Automobile Availability and Modal Split in Baltimore and German Cities

The results summarizing modal-split characteristics are presented in Tables 2 and 3 for Baltimore and in Tables 4 and 5 for the German cities. The findings may be summarized as follows:

1. Independent of version, an increase in automobile availability results in consistent increases in shares for the automobile-driver mode and decreases in the automobile-passenger and public transit share. The walk share also decreases, although less consistently. This applies to both obligatory and discretionary trips for both data sets, although the overall differences in modal-split characteris-

TABLE 1 Category Representations for Alternative Descriptions of Automobile Availability

				Category Representations (%)			
				Entire Population		Category 3	
Age	Employment Status	Automobile Availability	Category[a]	Baltimore	German Cities	Baltimore	German Cities
Proposed Definition: N_c/N_d							
<18	N/A	N/A	1	20.5	15.0		
18-65	Employed	$N_c = 0$	2	10.2	20.9		
18-65	Employed	<1	3	16.6	20.5	100.0	100.0
18-65	Employed	≥1	4	23.6	18.2		
18-65	Nonemployed	$N_c = 0$	5	9.5	10.7		
18-65	Nonemployed	<1	6	6.9	4.0		
18-65	Nonemployed	≥1	7	6.8	2.5		
>65	N/A	N/A	8	5.9	8.2		
Modification 1[b]: N_c/N_e							
18-65	Employed	<1	3A	8.9	16.2	53.6	79.0
18-65	Employed	≥1	3B	7.7	4.3	46.4	21.0
Modification 2: N_c/N_e							
18-65	Employed	>0; ≤1/3	3A.1	0.7	0.3	4.2	1.5
18-65	Employed	>1/3; <2/3	3A.2	4.7	15.3	28.3	74.6
18-65	Employed	≥2/3; <1	3A.3	3.5	0.6	21.1	2.9
18-65	Employed	≥1	3B	7.7	4.3	46.4	21.0
Modification 3: N_c/N_d							
18-65	Employed	>0; ≤1/3	3.1	1.7	1.2	10.2	5.9
18-65	Employed	>1/3; <2/3	3.2	10.3	17.7	62.0	86.3
18-65	Employed	≥2/3; <1	3.3	4.6	1.6	27.7	7.8

Note: N_c = number of automobiles in the household; N_d = number of persons in household with a driver's license; N_e = number of employed persons in household with a driver's license; N/A = not applicable.

[a]Categories 1, 2, and 4 through 8 remain the same for all versions.
[b]For Modifications 1, 2, and 3, N_c is always greater than zero.

TABLE 2 Modal-Split Shares for Obligatory Trips in Baltimore for All Category Descriptions

Age	Employment Status	Automobile Availability	Category[a]	α_{trav} (%)	Trip Rate (N)	Mode-Split Shares (%) Automobile Driver	Automobile Passenger	Public Transit	Walk	Other Modes
Proposed Definition: N_c/N_d										
<18	N/A	N/A	1	20.5	1.79	4.6	14.5	24.2	43.7	13.0
18-65	Employed	$N_c = 0$	2	10.2	1.75	2.4	30.7	49.2	14.6	3.1
18-65	Employed	<1	3	16.6	2.05	65.1	18.2	9.3	7.0	0.4
18-65	Employed	⩾1	4	23.6	2.00	88.1	6.2	1.8	2.7	1.2
18-65	Nonemployed	$N_c = 0$	5	9.5	0.34	0.0	15.2	30.4	54.3	0.0
18-65	Nonemployed	<1	6	6.9	0.50	46.9	28.6	8.2	16.3	0.0
18-65	Nonemployed	⩾1	7	6.8	0.37	66.7	8.3	5.6	11.1	8.3
>65	N/A	N/A	8	5.9	0.29	83.3	12.5	4.2	0.0	0.0
Modification 1[b]: N_c/N_e										
18-65	Employed	<1	3A	8.9	1.96	54.6	22.5	11.6	10.8	0.4
18-65	Employed	⩾1	3B	7.7	2.16	76.2	13.6	6.8	3.0	0.4
Modification 2: N_c/N_e										
18-65	Employed	>0; ⩽1/3	3A.1	0.7	1.60	37.5	25.0	12.5	25.0	0.0
18-65	Employed	>1/3; <2/3	3A.2	4.7	1.83	50.4	26.4	14.1	8.3	0.8
18-65	Employed	⩾2/3; <1	3A.3	3.5	1.98	58.6	16.2	12.1	13.1	0.0
18-65	Employed	⩾1	3B	7.7	2.16	76.2	13.6	6.8	3.0	0.4
Modification 3: N_c/N_d										
18-65	Employed	>0; ⩽1/3	3.1	1.7	1.83	50.0	25.0	11.4	13.6	0.0
18-65	Employed	>1/3; <2/3	3.2	10.3	1.99	58.6	21.0	11.7	8.3	0.3
18-65	Employed	⩾2/3; <1	3.3	4.6	2.27	82.0	10.7	4.0	2.7	0.7

Note: N_c = number of automobiles in the household; N_d = number of persons in household with a driver's license; N_e = number of employed persons in household with a driver's license; α_{trav} = percentage of travelers in each category of the population; N/A = not applicable.
[a]Categories 1, 2, and 4 through 8 remain the same for all versions.
[b]For Modifications 1, 2, and 3, N_c is always greater than zero.

TABLE 3 Modal-Split Shares for Discretionary Trips in Baltimore for All Category Descriptions

Age	Employment Status	Automobile Availability	Category[a]	α_{trav} (%)	Trip Rate (N)	Mode-Split Shares (%) Automobile Driver	Automobile Passenger	Public Transit	Walk	Other Modes
Proposed: N_c/N_d										
<18	N/A	N/A	1	20.5	1.72	5.8	28.3	4.6	52.4	8.9
18-65	Employed	$N_c = 0$	2	10.2	1.14	0.0	24.8	21.8	46.1	7.3
18-65	Employed	<1	3	16.6	1.33	70.5	17.5	0.6	10.8	0.6
18-65	Employed	⩾1	4	23.6	1.64	83.0	7.5	1.3	7.3	0.9
18-65	Nonemployed	$N_c = 0$	5	9.5	2.70	1.1	33.5	11.0	50.8	3.6
18-65	Nonemployed	<1	6	6.9	2.91	50.5	27.0	2.8	16.5	3.2
18-65	Nonemployed	⩾1	7	6.8	3.33	73.4	22.0	0.3	4.0	0.3
>65	N/A	N/A	8	5.9	2.64	39.6	11.7	12.6	30.2	5.9
Modification 1[b]: N_c/N_e										
18-65	Employed	<1	3A	8.9	1.47	61.5	23.0	1.1	13.4	1.1
18-65	Employed	⩾1	3B	7.7	1.17	83.6	9.4	0.0	7.0	0.0
Modification 2: N_c/N_e										
18-65	Employed	>0; ⩽1/3	3A.1	0.7	1.60	25.0	43.8	6.2	25.0	0.0
18-65	Employed	>1/3; <2/3	3A.2	4.7	1.50	70.7	13.1	1.0	13.1	2.0
18-65	Employed	⩾2/3; <1	3A.3	3.5	1.22	57.4	29.5	0.0	13.1	0.0
18-65	Employed	⩾1	3B	7.7	1.17	83.6	9.4	0.0	7.0	0.0
Modification 3: N_c/N_d										
18-65	Employed	>0; ⩽1/3	3.1	1.7	1.25	36.7	33.3	3.3	26.7	0.0
18-65	Employed	>1/3; <2/3	3.2	10.3	1.35	73.6	14.7	0.5	10.2	1.0
18-65	Employed	⩾2/3; <1	3.3	4.6	1.33	75.0	18.2	0.0	6.8	0.0

Note: N_c = number of automobiles in the household; N_d = number of persons in household with a driver's license; N_e = number of employed persons in household with a driver's license; α_{trav} = percentage of travelers in each category of the population; N/A = not applicable.
[a]Categories 1, 2, and 4 through 8 remain the same for all versions.
[b]For Modifications 1, 2, and 3, N_c is always greater than zero.

TABLE 4 Modal-Split Shares for Obligatory Trips in German Cities (KONTIV Code 7) for All Category Descriptions

Age	Employment Status	Automobile Availability	Category[a]	α_{trav} (%)	Trip Rate (N)	Mode-Split Shares (%): Automobile Driver	Automobile Passenger	Public Transit	Walk	Other Modes
Proposed: N_c/N_d										
<18	N/A	N/A	1	15.0	1.85	0.1	4.1	37.0	38.4	20.4
18-65	Employed	$N_c = 0$	2	20.9	1.52	4.7	9.1	46.6	32.6	7.0
18-65	Employed	<1	3	20.5	1.94	57.3	6.4	16.6	16.3	3.4
18-65	Employed	≥1	4	18.2	2.06	81.1	2.1	4.5	11.2	1.1
18-65	Nonemployed	$N_c = 0$	5	10.7	0.92	2.9	6.9	44.9	36.9	8.4
18-65	Nonemployed	<1	6	4.0	1.47	41.7	6.7	27.3	16.1	8.2
18-65	Nonemployed	≥1	7	2.5	1.49	71.4	1.6	8.3	15.7	3.0
>65	N/A	N/A	8	8.2	0.59	17.5	3.8	25.4	50.3	3.0
Modification 1[b]: N_c/N_e										
18-65	Employed	<1	3A	16.2	1.97	54.4	7.1	17.3	17.8	3.5
18-65	Employed	≥1	3B	4.3	1.80	69.5	3.8	13.2	10.2	3.3
Modification 2: N_c/N_e										
18-65	Employed	>0; ≤1/3	3A.1	0.3	2.22	44.9	8.2	20.4	22.4	4.1
18-65	Employed	>1/3; <2/3	3A.2	15.3	2.01	54.2	7.1	17.0	18.1	3.5
18-65	Employed	≥2/3; <1	3A.3	0.6	1.63	62.5	4.5	28.4	4.5	0.0
18-65	Employed	≥1	3B	4.3	1.80	69.5	3.8	13.2	10.2	3.3
Modification 3: N_c/N_d										
18-65	Employed	>0; ≤1/3	3.1	1.2	1.93	50.4	7.2	23.9	14.4	4.1
18-65	Employed	>1/3; <2/3	3.2	17.7	1.91	55.8	6.7	16.4	17.6	3.5
18-65	Employed	≥2/3; <1	3.3	1.6	1.88	78.0	2.8	13.6	4.9	0.7

Note: N_c = number of automobiles in the household; N_d = number of persons in household with a driver's license; N_e = number of employed persons in household with a driver's license; α_{trav} = percentage of travelers in each category of the population; N/A = not applicable.
[a]Categories 1, 2, and 4 through 8 remain the same for all versions.
[b]For Modifications 1, 2, and 3, N_c is always greater than zero.

TABLE 5 Modal-Split Shares for Discretionary Trips in German Cities (KONTIV Code 7) for All Category Descriptions

Age	Employment Status	Automobile Availability	Category[a]	α_{trav} (%)	Trip Rate (N)	Mode-Split Shares (%): Automobile Driver	Automobile Passenger	Public Transit	Walk	Other Modes
Proposed: N_c/N_d										
<18	N/A	N/A	1	15.0	1.49	0.3	10.5	22.3	43.4	23.5
18-65	Employed	$N_c = 0$	2	20.9	1.51	3.7	11.5	22.6	52.5	9.7
18-65	Employed	<1	3	20.5	1.69	49.6	10.0	7.3	29.5	3.6
18-65	Employed	≥1	4	18.2	1.71	77.2	3.6	2.7	15.1	1.4
18-65	Nonemployed	$N_c = 0$	5	10.7	2.12	1.9	11.8	20.8	57.7	7.7
18-65	Nonemployed	<1	6	4.0	2.14	43.4	9.8	10.6	30.2	6.0
18-65	Nonemployed	≥1	7	2.5	2.25	65.7	0.7	4.4	24.1	5.1
>65	N/A	N/A	8	8.2	2.42	9.0	3.8	24.6	57.3	5.4
Modification 1[b]: N_c/N_e										
18-65	Employed	<1	3A	16.2	1.65	45.1	10.6	8.3	32.1	3.9
18-65	Employed	≥1	3B	4.3	1.81	65.4	7.6	3.7	20.5	2.8
Modification 2: N_c/N_e										
18-65	Employed	>0; ≤1/3	3A.1	0.3	0.68	53.3	13.3	0.0	26.7	6.7
18-65	Employed	>1/3; <2/3	3A.2	15.3	1.70	44.8	10.7	8.2	32.4	3.9
18-65	Employed	≥2/3; <1	3A.3	0.6	1.32	50.7	9.9	14.1	22.5	2.8
18-65	Employed	≥1	3B	4.3	1.81	65.4	7.6	3.7	20.5	2.8
Modification 3: N_c/N_d										
18-65	Employed	>0; ≤1/3	3.1	1.2	1.36	53.8	7.7	10.3	26.3	1.9
18-65	Employed	>1/3; <2/3	3.2	17.7	1.67	47.1	10.3	7.6	31.2	3.9
18-65	Employed	≥2/3; <1	3.3	1.6	1.87	69.8	8.4	2.8	16.1	2.8

Note: N_c = number of automobiles in the household; N_d = number of persons in household with a driver's license; N_e = number of employed persons in household with a driver's license; α_{trav} = percentage of travelers in each category of the population; N/A = not applicable.
[a]Categories 1, 2, and 4 through 8 remain the same for all versions.
[b]For Modifications 1, 2, and 3, N_c is always greater than zero.

tics between Baltimore and the German cities should be kept in mind.

2. The ratios N_c/N_d and N_c/N_e are higher in Baltimore than in the German cities. Both of these ratios were most often 1/2 for the German cities, whereas for Baltimore the N_c/N_d ratio shifts from 1/2 toward 2/3 and the N_c/N_e ratio is frequently 1. This offers much better ridesharing opportunities in Baltimore as compared with the German cities.

3. For Modifications 1 and 2 it is clear that preference is given to employed persons for their obligatory trips. In Modification 1, for example, Category 3B modal-split shares shift from Category 3 toward Category 4, indicating a tendency to make the automobile always available for obligatory trips for employed members. The modal shares are not exactly like Category 4, because those in Category 3B still may compete with other household members for the automobile when they make discretionary trips.

Transferability Tests

Simple transferability tests were performed to determine (a) how consistent the category-specific modal-split travel behavior is in Baltimore and the German cities and (b) which version of automobile availability description performs best for different population segments. The category-specific modal shares from Germany (β_{ij}^{Germ}) were "borrowed" to explain modal share j in Baltimore (β_j^{Balt}):

$$\beta_j^{Balt,pred} = \left(\sum_i \beta_{ij}^{Germ} \cdot N_i^{Balt} \cdot \alpha_i^{Balt}\right) \div \left(\sum_i N_i^{Balt} \cdot \alpha_i^{Balt}\right) \quad (1)$$

where

$\beta_j^{Balt,pred}$ = predicted share of mode j in Baltimore,

β_{ij}^{Germ} = actual share of mode j for person category i in the German cities,

N_i^{Balt} = actual trip rate per traveler belonging to Category i in Baltimore, and

α_i^{Balt} = actual share of the traveler Category i in Baltimore.

The errors were calculated as follows:

$$\text{Error} = (\beta_j^{Balt,pred} - \beta_j^{Balt,act})/\beta_j^{Balt,act} \quad (2)$$

where $\beta_j^{Balt,act}$ is the actual share of mode j in Baltimore.

The results of such an analysis are shown in Table 6. It may be seen that (a) transferability errors are small for the automobile-driver mode and much higher for the public transit mode; (b) errors are much smaller for "organized" obligatory trips than for more area-specific discretionary trips; (c) errors are much smaller for the employed segment than for the entire population; (d) Categories 1 and 8, not defined around the automobile availability variable, are least transferable; and (e) level-of-service variables are needed to explain the public transit share of the modal split, because the errors without these variables are too large, particularly for discretionary trips. [Note, however, that transfers of MNL models of mode choice may result in much higher errors if the model is transferred into a different urban environment; for example, the transfer of the Baltimore model to Twin Cities resulted in 500 percent error to the public transit share (12). Recent work by Supernak (13,pp. 533-559) shows a method for updating alternative specific constants of MNL models of mode choice by utilizing category-specific modal shares.]

It appears from Table 6 that no one version of automobile availability performs much better than any of the others. In most cases, the original proposed definition of automobile availability will be considered superior because of its simplicity as compared with the modified versions. Although further stratification of Category 3 has provided some interesting observations (e.g., that the higher the ratio of number of drivers to number of automobiles, or the number of employed drivers to the number of automobiles, the higher is the likelihood of driving the automobile), the small representations of each of the modified groups make these descriptions of automobile availability more cumbersome than they are worth.

Table 7 confirms the usefulness of categorization of the population according to such variables as

TABLE 6 Transferability Errors for Automobile-Driver and Public Transit Modes in Application of Category-Specific Modal Shares from German Cities to Baltimore

		Errors (%)					
		Obligatory Trips		Discretionary Trips		All Trips	
Version	Category	Automobile Driver	Public Transit	Automobile Driver	Public Transit	Automobile Driver	Public Transit
Proposed version	2-4	−8.65	+25.75	−13.48	+66.68	−11.03	+27.23
	2-7	−8.17	+31.61	−12.57	+107.55	−10.47	+63.17
	1-8	−11.75	+41.28	−19.41	+149.24	−15.79	+80.24
Modification							
1	2-4	−6.30	+22.44	−11.74	+60.03	−11.11	+12.82
	2-7	−5.94	+28.64	−11.46	+104.29	−10.53	+49.55
	1-8	−9.61	+39.46	−18.42	+147.08	−15.85	+70.80
2	2-4	−5.24	+25.80	−11.08	+63.20	−10.10	+11.35
	2-7	−4.93	+31.60	−11.04	+105.84	−13.84	+46.89
	1-8	−8.69	+41.21	−18.06	+148.10	−15.15	+68.51
3	2-4	−6.04	+24.72	−13.82	+61.12	−11.08	+7.45
	2-7	−5.69	+27.62	−12.79	+104.84	−10.51	+43.44
	1-8	−9.44	+38.89	−19.60	+147.45	−15.83	+66.12

TABLE 7 Analysis of Variance Results for All Trip Purposes

Factor	F^{calc} Automobile Driver	Automobile Passenger	Public Transit	Walk	Other Modes	$F^{0.01}$
2 Cities and 8 Categories						
Categories	33.65	3.27	10.41	8.80	7.77	6.99
Cities	4.30	41.22	11.38	7.25	9.02	12.25
2 Cities and 10 Categories[a]						
Categories	32.33	2.88	13.71	9.64	8.36	5.35
Cities	3.35	51.10	16.59	7.87	11.45	10.56

Note: Calculated F-values are shown and compared with $F^{0.01}$

[a]As defined in Modification 3.

automobile availability, employment status, and age (the factor "categories" is significant at the 1 percent level), for all modes except automobile passenger. When the analysis of variance is applied to the modified category description (Modification 3), the significance of area characteristics increases and the significance of category characteristics remains unchanged (see Table 7). Most significantly, this table confirms the usefulness of the automobile availability concept to analyze travel behavior, particularly modal split.

CONCLUSIONS AND RECOMMENDATIONS

The following conclusions and recommendations may be stated:

1. The concept of automobile availability appears to be a valid alternative to the automobile ownership concept. An individual's access to the most convenient transportation mode, an automobile, is a primary factor in determining mode choice, particularly the share of the automobile-driven trips. The relationship between automobile availability and automobile use is strong, consistent, and very similar in Baltimore and certain large German cities.
2. This paper tested with success the automobile availability description based on average potential access of an individual to the automobile (never/sometimes/always) rather than actual access to the automobile for a given trip (available/not available). The recommended description is simple and easily applicable.
3. The proposed version of automobile availability was extended by a more detailed description of the situation "automobile sometimes available." Three modifications were considered. They consistently show that the higher the ratio of number of drivers to number of automobiles, or number of employed drivers to number of automobiles, the higher is the likelihood of driving the automobile. In families, priority for automobile use is commonly given to employed persons and their obligatory activities. The proposed version in its original form was preferred because of its simplicity.
4. The automobile availability issue deserves more studies, such as (a) an international comparison of the relationship between automobile availability and modal split, (b) explanatory variables for automobile availability levels, and (c) the potential for practical applications.

ACKNOWLEDGMENTS

The research reported on in this paper was sponsored in part by a grant from the San Diego State University Foundation. The authors would like to thank Werner Brög from SOCIALDATA, Munich, and Raimund Herz from the University of Karlsruhe, West Germany, for providing them with the KONTIV data set used in the analysis reported in this paper.

REFERENCES

1. J. Supernak, A. Talvitie, and A. DeJohn. A Person-Category Trip Generation Model. In Transportation Research Record 944, TRB, National Research Council, Washington, D.C., 1984, pp. 74-83.
2. J. Supernak and A. DeJohn. Comparison Between the Person and Household Category Trip Generation Models. Presented at 1983 World Conference on Transport Research, Hamburg, West Germany, 1983.
3. R.G. Donald. Modal Split Based on Car Availability: The Application of Such Models in Studies of Medium Size Towns. Transportation Planning and Technology, Vol. 6, 1980, pp. 149-158.
4. J.M. Bailey. Car Availability and Mode Choice--A Discussion Paper. Presented at the Summer Annual Meeting of PTRC, University of Sussex, Brighton, England, 1983.
5. J. Supernak and A. Talvitie. A Person Category Trip Generation Model: A Concept. Working Paper 782-09. State University of New York at Buffalo, 1980.
6. J. Supernak. Car Availability Versus Car Ownership Modeling: Theoretical Discussion and Empirical Findings. Proceedings of the World Conference on Transport Research, Hamburg, Germany, 1983, pp. 304-320.
7. P.R. Stopher and C.G. Wilmot. Work-Trip Mode-Choice Models for South Africa. Transportation Engineering Journal, ASCE, Nov. 1979, pp. 595-608.
8. K.M. Gwilliam and D.J. Banister. Patterns of Car Usage and Restraint Modeling. Transportation, Vol. 6, No. 4, 1977, pp. 345-363.
9. J.M. Bailey and A.D. Layzell. Travel Behavior and Traffic Restraint: A Study of Households in Oxford. TRRL Supplementary Report 734. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1982.
10. J. Supernak, A. Talvitie, and B. Bogan. A Person Category Car Availability/Ownership Model. Working Paper 782-17. State University of New York at Buffalo, 1981.
11. D.L. Schoendorfer. Car Availability--Accounting for the Priority Given to Obligatory Trips. Working Paper 280-84. Drexel University, Philadelphia, Pa., 1984.
12. A. Talvitie, Y. Dehghani, J. Supernak, M. Morris, M. Anderson, A. DeJohn, B. Kousheshi, and S. Feuerstein. Refinement and Application of Individual Choice Models in Travel Demand Forecasting. State University of New York at Buffalo, 1981.
13. J. Supernak. Disaggregate Models of Mode Choice: An Assessment of Performance and Suggestions for Improvement. In Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht, Netherlands, 1984.

14. M. Ben-Akiva and S. Lerman; Cambridge Systematics, Inc. A Behavioral Analysis of Automobile Ownership and Modes of Travel. U.S. Department of Transportation, 1976.
15. J. Supernak. Toward a Better Understanding of Transportation-Related Choices. Proceedings of the 10th International Colloquium "Transportation and Stagnation: Challenges for Planning and Research," Delft, The Netherlands, 1983, pp. 269-288.

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

Attribute Thresholds and Logit Mode-Choice Models

W. YOUNG and D. BERTRAM

ABSTRACT

The concept of thresholds has been mentioned in the transport choice literature from time to time. Few studies of mode choice have attempted to incorporate them into a modeling context, however. In this paper the concept of minimally perceived attribute differences is introduced into a logit choice model. For estimating the parameters of the model, maximum likelihood is employed and an experimental test is carried out on a sample of trip makers going to the Melbourne central business district. It was found that the average respondent required a 12-min (22 percent) difference in travel time or a 12-cent (32 percent) difference in travel cost before he would react to the variation in attribute ratings. The model is compared with a more traditional logit model with a linear additive measure of utility.

Transport planners have developed a variety of statistical techniques for analyzing mode choice (1-4). The common feature of all these models is that choice is seen as a function of the utility gained from each alternative. To calculate utility it was assumed that an alternative was characterized by a set of attributes that contribute to an index of total utility. A linear additive function was used to combine the attribute utilities into the index. In turn attribute utilities were assumed to be a continuous function of the satisfaction gained from each attribute. That is, every change in satisfaction, no matter how small, will influence the utility gained from an alternative and hence an individual's choice.

Evidence in the psychology (5,6), economics (7), and biology (8) literature suggests that people may be indifferent to changes in a stimulus unless it crosses a threshold of indifference. In the transport literature this suggestion has found support in several studies of the application of transport-choice models. Kovak and Demetsky (9) and Burns et al. (10) found that models that did not incorporate indifference thresholds tended to overestimate mode shift for small changes in attribute satisfaction. It was suggested that the inclusion of thresholds of indifference may overcome this problem because they would tend to dampen the effect of small changes in attribute satisfaction. In this paper the incorporation of such thresholds into logit choice models is investigated.

The paper is divided into six sections. The next section describes the incorporation of thresholds as used in a number of disciplines. The third section describes the incorporation of thresholds into logit choice models. The fourth section describes the data used in the study, and the fifth to seventh sections discuss the model estimation and compare model performance.

BACKGROUND

The existence of thresholds of acceptance has been discussed in many disciplines.

In psychology, sensory thresholds were suggested by Weber in 1830 (5). He introduced the concept of just noticeable differences and related their size to the magnitude of the stimulus. Fechner (6) extended Weber's law by relating the strength of the sensory process to the logarithm of the stimulus. Experimental studies that followed appeared to support Fechner's logarithm law and the existence of thresholds was accepted.

Similarly, economists analyze consumer choice of commodities by the application of indifference curves (5). In this approach it is considered that, in a choice between two commodities, the decision maker will choose one or the other or be indifferent. If the decision maker is indifferent, he will tend to randomize his decision. Indifference curves define all situations where the consumer is indifferent.

Biological experiments indicate that thresholds vary between subjects. The distribution of these thresholds was hypothesized to be normal. The resulting relationship between response and stimuli was therefore described by a probit model. Finney (8) analyzed a number of situations and concluded that the probit model predicted response relatively accurately. He also considered other relationships for the form of the threshold distribution; one of these was the logit model.

In transport planning, thresholds have been suggested in a number of contexts. Choice inertia, perception, and constraints (11) may to some extent exhibit threshold effects. Empirical studies of these thresholds have been directed along two lines.

The first related to thresholds in the comparison of the utility gained from each alternative (12,13). Krishnan (12) contended that the difference in the utility gained from a number of alternatives must be large enough for the individual to recognize the difference; otherwise he will be indifferent. Krishnan introduced a threshold (δ) into the choice situation such that in a choice between A_1 with utility U_1 and A_2 with utility U_2

$$\text{Prob } (A_1 > A_2) = \text{prob } (U_1 > U_2 + \delta)$$

$$\text{Prob } (A_2 > A_1) = \text{prob } (U_2 > U_1 + \delta)$$

$$\text{Prob } (A_1 \sim A_2) = \text{prob } (|U_1 - U_2| \leq \delta)$$

This model was found to fit the travel data better than the traditional logit model [i.e., the model with the threshold (δ) equal to zero]. Kawakami and Hirobata (13) argued that the utility of an alternative must change by an amount greater than a threshold of inertia before people will change mode. Their study of mode choice on the Nagaya-Tokyo railway line, using before-and-after data, confirmed this hypothesis.

The second approach was developed in the area of noncompensatory lexicographic or elimination-by-aspects (EBA) models. These models hypothesize that the decision maker considers the attributes describing a set of alternatives in order of importance. An alternative is eliminated if its attribute satisfaction level falls below an acceptance threshold. The most common method for calculating the acceptable threshold (14,15) has been to use a criterion whereby attribute satisfaction levels are considered to be acceptable if they lie within a specific fractional tolerance of the best satisfaction level for the attribute over all alternatives for each individual. Thus

$$\text{Acceptable } S_{kjq} \geq (1 - T_k) \max_j (S_{kjq}) \tag{1}$$

where

S_{kjq} = satisfaction with the kth attribute of the jth mode for the qth individual,
T_k = tolerance for the kth attribute, and
$\max_j(S_{kjq})$ = maximum satisfaction for the kth attribute for the qth individual over all j modes.

This approach enables the concept of just noticeable differences to be incorporated into the model as well as the size of the stimulus. It is therefore in line with the psychological research of Weber.

This review illustrates that there is considerable evidence for the existence of thresholds. Particular emphasis in the literature appears to be directed at the decision maker's inability to discern small changes in stimulus levels.

INCORPORATION OF THRESHOLD TYPES

The literature review in the previous section alluded to the existence of two apparently different approaches to incorporating thresholds in mode-choice models. The first concentrates on thresholds in total utility and the second on attribute thresholds. Before these differences are discussed further, it is necessary to develop the modeling framework for the incorporation of the thresholds.

The most commonly used choice model in transport is the logit model. The most popular derivations of this model are the constant-utility approach (16) and the random-utility approach (17). With the latter approach, in the choice between two alternatives, it is assumed that the choice alternative will be the one that maximizes the decision maker's utility. That is, if x is chosen,

$$U_x > U_y \tag{2}$$

where U_x is the utility of alternative x.

If it is assumed that the total utility is an additive function of the utility gained from each attribute of the alternative, then x will be chosen if

$$\sum_{k=1}^{k} u_{kx} > \sum_{k=1}^{k} u_{ky} \tag{3}$$

where u_{kx} is the utility of attribute k for alternative x.

Further the attribute utility is assumed to be composed of two independent elements. These are the degree of importance associated with and the satisfaction gained from an attribute. Hence the utility function takes the form

$$u_{kx} = I_k \cdot S_{kxq} \tag{4}$$

where I_k is the importance of attribute k to the decision maker and S_{kxq} is the satisfaction gained from attribute k for mode x for individual q.

It can be shown that if there is an error function associated with the decision maker's perception of utility and that the error function is described by a Weibull distribution, then the multinomial logit model will describe the choice process (18). The form of this model is

$$p(x|K) = \exp(U_x) / \sum_{k=1}^{k} \exp(U_k) \tag{5}$$

For the binary case the logit model takes the form

$$p(x/xy) = \exp(U_x)/[\exp(U_x + \exp(U_y)] \tag{6}$$

As stated earlier, there are two methods for incorporating thresholds into this model. The first concentrates on attributes and the second on total utility. An approach that combines both methods into the binary logit model can be illustrated by reference to Figure 1. If the satisfaction levels are equal for the two alternatives, then there is no difference in the two alternatives. As the difference in the satisfaction levels increases, there is still no perceived difference until the difference

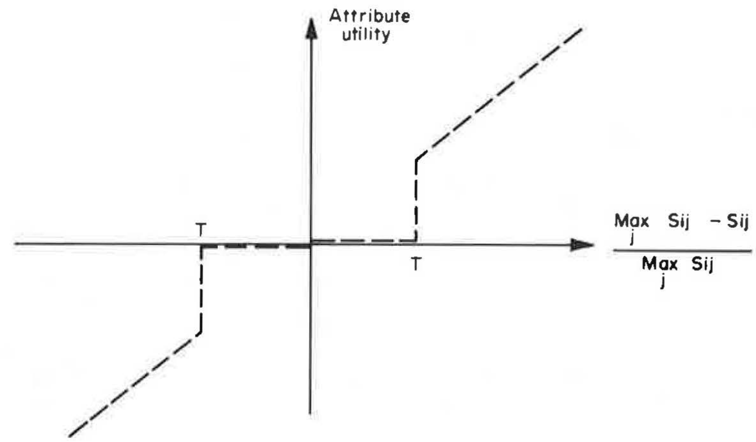

FIGURE 1 Threshold Type I.

crosses the acceptance tolerance. Once this occurs, the utility obtained from the attributes for each alternative is equal to the product of the importance and satisfaction ratings. This will be referred to as the Type I threshold in the ensuing discussion.

In mathematical terms this can be written as follows. If

$$|(S_{kxq} - S_{kyq})/S_{kxq}| < T \quad (7)$$

then

$$u_{kx} = u_{ky} = 0 \quad (8)$$

whereas if

$$|(S_{kxq} - S_{kyq})/S_{kxq}| \geq T \quad (9)$$

then

$$u_{kx} = I_k \cdot S_{kxq} \quad (10)$$

where $S_{kxq} > S_{kyq}$.

The utility for alternative x is then given by

$$U_x = \sum_{k=1}^{k} u_{kx} \quad (11)$$

A similar procedure can be used to obtain the utility associated with alternative y. The utility for alternatives x and y can then be substituted into Equation 6 and the choice probability calculated.

Another approach to thresholds is illustrated in Figure 2. Here the total utility gained from each attribute for each alternative is obtained once the threshold tolerance is crossed. This approach is consistent with studies by Recker and Golob (14) and Young and Brown (15). This approach will be referred to as the Type II threshold in the ensuing discussion. In mathematical terms this can be written using Equations 7-9. Then

$$u_{kx} = I_k \quad (12)$$

The utility for alternative x is then given by

$$U_x = \sum_{k=1}^{k} u_{kx} \quad (13)$$

A similar procedure can be used to obtain the utility associated with alternative y. The utility for alternatives x and y can then be substituted into Equation 6 and the choice probability calculated.

DESCRIPTION OF THE DATA

The data used in this study came from a survey of commuters going to the Melbourne central business district (CBD) in 1974 (19). A questionnaire survey was distributed to the employees of 35 CBD firms selected on a representative geographical and classification basis. A total of 3,737 correctly completed responses was received from a total of 7,400 issued questionnaires; of these 1,205 respondents reported a choice between automobile and train travel. It is these respondents who have been considered throughout the study.

The survey provided detailed information regard-

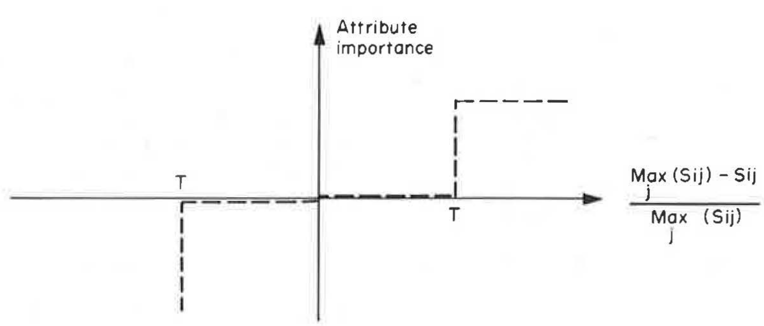

FIGURE 2 Threshold Type II.

ing the usual and next-best alternative mode available for the work trip. As well as actual time and cost data, perceptual data were solicited. The perceptual data related to the level of satisfaction experienced with the overall descriptors--comfort, convenience, and reliability. More specifically, satisfaction scores for the three factors relating to the overall trip, for both usual and alternative modes, were registered on a semantic scale of 1 to 7.

PARAMETER ESTIMATION AND MEASURES OF MODEL PERFORMANCE

The review of the literature indicated that decision makers may not be sensitive to small differences in attribute satisfaction. These differences were also thought to be related to the magnitude of the attribute satisfaction. The form of threshold most commonly used in transport research therefore takes the form presented in Equation 1. This expression incorporates both the maximum available attribute satisfaction ($\mathrm{Max}_j\ S_{kjq}$) and the tolerable difference (T_k). It was therefore used in this study.

The aim of the model estimation procedures is to determine the most appropriate value of the tolerance (T_k). It was also necessary to determine the importance (I_k) placed on each attribute. Because the logit model was probabilistic, maximum likelihood was used to estimate these two parameters. The likelihood function took the form

$$L(\hat{I},\hat{T}) = \prod_q \prod_j P_q(j)^{g_{jq}} \tag{14}$$

where

$L(\hat{I},\hat{T})$ = likelihood at tolerance level $\hat{T}$ and importance level $\hat{I}$,
$P_q(j)$ = probability from the model that individual q chooses alternative j, and
g_{jq} = 1 if alternative j was selected by individual q, 0 otherwise.

Because both the threshold and logit models used maximum likelihood to estimate the parameters, it was possible to compare the overall fit of the model. Two tests were used. The first was the generalized likelihood ratio test (18), which was used to test the hypothesis that the probability that an individual would choose an alternative was independent of the value of the parameters in the choice model. If this hypothesis cannot be rejected, the tolerance and importance estimates used in the model may be assumed to have no effect on choice (i.e., the choice was a random one). The likelihood ratio test for the models took the form

$$-2\ln\lambda_T = -2[L^*(0,\infty) - L^*(\hat{I},\hat{T})] \tag{15}$$

where $L^*(0,\infty)$ is the log of the likelihood when the importance estimates are constrained to 0 and the tolerance estimates are constrained to a very high value, ∞, and $L^*(\hat{I},\hat{T})$ is the log of the likelihood for the best estimates of the importance and tolerance parameters. This $-2\ln\lambda_T$ value is distributed like a chi-squared distribution, with degrees of freedom equal to the number of parameters in the model.

The second test was the pseudo-r^2 value (18) where

$$\rho^2 = 1 - [L^*(\hat{I},\hat{T})./L^*(0,\infty)] \tag{16}$$

Although the statistic has the range $0 \le \rho^2 \le 1$, a value between 0.2 and 0.4 was considered to represent a good fit (18).

MODEL ESTIMATION

Model development consisted of a number of stages. First the model containing all the attributes was estimated. The attribute that was associated with the parameters that had the least influence on the model fit (lowest level of significance) was removed and the parameters for the new set of attributes were estimated. This refining procedure was continued until all remaining attributes had significant parameter estimates at the 5 percent level. The models presented in Tables 1 and 2 are the final product of this refining process.

TABLE 1 Comparison of Statistical Performance of Threshold and Logit Models: Part 1

	Threshold Type I		Threshold Type II		
Attribute	Tolerance	Importance	Tolerance	Importance	Logit
Time	−0.20	−0.40	−0.22	1.2	−0.041
Cost	−0.29	−0.28	−0.32	1.3	−0.030
Convenience	0.31	−0.40	0.33	1.6	−0.414
Constant					
Train		−1.32		1.54	−1.486
$-2\ln\lambda_T$		481		471	464
ρ^2		0.35		0.34	0.34

TABLE 2 Comparison of Statistical Performance of Threshold and Logit Models: Part 2

	Threshold Type I	Threshold Type II	Logit
Correct prediction			
Train	753	752	751
Market share	895	895	895
Percent	84	84	84
Automobile	168	167	166
Market share	310	310	310
Percent	54	54	54

Threshold Type I Model

The physical measures of travel time and travel cost and the perceptual measure of convenience remained in the threshold Type I model. Convenience was found to have the highest importance (−0.40) and largest tolerance (0.31), whereas travel time had the lowest tolerance (−0.20) and travel cost had the lowest importance (−0.28).

In terms of the overall fit, the model was encouraging. The $-2\ln\lambda_T$ value was significant at the 5 percent level ($-2\ln\lambda_T = 481 > 12.6 = \chi^2_{6,0.05}$) and the ρ^2-value was in the generally accepted range of 0.20 and 0.40. Further, the train mode was correctly predicted for 84 percent of the train users and the automobile mode was correctly predicted for 54 percent of the automobile users.

Threshold Type II Model

The threshold Type I and II models showed a number of similarities. Both models contained the same attribute set after refinement. There was also a marked similarity in the tolerance estimates. The major difference in the two models was the magnitude of the importance estimates. The threshold Type I model had importance ratings an order of magnitude lower than those of the Type II model.

The overall fit of the threshold Type II model was acceptable but was slightly poorer than that of the threshold Type I model.

Interpretation of Parameters

Three aspects of the estimated threshold model require further discussion: the interpretation of the estimated tolerance levels, the significance of the constant terms, and the relative magnitude of the estimated importance parameters. The threshold Type II model will be used to illustrate these points.

To facilitate this discussion, and to obtain a clearer picture of how the threshold model works, it is useful to consider the choice process for an average respondent in the sample, where such a respondent experiences the average satisfaction ratings of the sample. These average ratings (Table 3) are 54 min, 38 cents, and 4 units of convenience for the train user and 56 min, 64 cents, and 4 units of convenience for the automobile user.

TABLE 3 Model Allocation Process for Average Respondent

	Attribute				
	Travel Time (min)	Travel Cost (cents)	Convenience Units	Constant	Total
Satisfaction					
Train	54	38	4		
Automobile	56	64	4		
Tolerance	−0.22	−0.32	0.33		
Acceptable satisfaction	<66	<50	>3		
Estimated importances	1.2	1.3	1.6		
Allocation to sets					
Automobile	–	–	–	0.0	0
Train	–	1.3	0	1.5	2.8
Both	1.2	–	1.6	–	2.8

The first step in the threshold process is to determine the acceptance levels. These are obtained by using Equation 1. For example, consider the attribute travel time. The best satisfaction level for this attribute is the minimum travel time for each mode--54 min for the train mode. The tolerance level for this attribute is -0.22. Hence for the average respondent to react to any difference in the two modes there must be a 12-min [$T_k \times \mathrm{Max}(S_{kjq}) = 0.22 \times 54$] difference in travel time. Given that the best travel time is 54 min, all travel times under 66 min are acceptable. That is, both the automobile and train modes are acceptable to the average respondent.

In the case of the travel-cost attribute, the average respondent will react to a difference in the two modes if there is a 12-cent [$T_k \times \mathrm{Max}(S_{kjq}) = 0.320 \times 38$] variation in the cost of travel between the two modes. This is in fact the case; the train is more than 12 cents cheaper than the automobile for this trip.

Given the composition of each attribute set, its magnitude can be determined by summing the importances of each attribute as shown in Table 3. It is evident that time and convenience allocate their importance to the set that is satisfactory for both train and automobile. That is to say, these attributes have no influence on the final choice. The allocation of the cost importance level is to the train set.

The size of the alternative specific constants is large when compared with the importances of the other attribute sets and hence it may be concluded that in this model unspecified attributes have a large effect on the final choice.

Furthermore, it can be seen from Table 3 that travel cost, travel time, and convenience have equal importance ratings.

COMPARISON OF THRESHOLD AND TRADITIONAL LOGIT MODELS

The empirical comparison between the threshold models and the logit model will be carried out on two levels. First, the statistical performance of the models and then the predictions resulting from the models will be compared.

Statistical Performance

The parameter estimates for the refined logit model were presented in Table 1. It can be seen that the refined logit model contained the same three attributes as the two threshold models. Further comparison of the parameter estimates is unlikely to be of value because of the difference in interpretation of the attribute satisfaction.

In terms of the overall fit there appeared to be little difference among the three models. All appeared to perform equally well. Hence on statistical grounds there appears to be little difference in these models.

Predictive Sensitivity

A full comparison of the predictive sensitivity of the three models would require the models to predict changes in the transport system. Those predictions could be compared with what actually takes place and the accuracy of the model determined. However, the data used in this study could not be used for such a test. It is worthwhile, nonetheless, to have the models predict what might occur if a system change were made. These predictions could then be used to determine whether the three models would in fact indicate different changes to the transport system for the same changes in attribute satisfaction.

The models were required to predict the magnitude of mode shift resulting from changes in attribute ratings between -95 and +100 percent in 5 percent increments for both time and cost. A similar prediction for changes in convenience ratings was not carried out because it was based on a semantic scale and would inevitably be of a discontinuous nature.

The changes in use of the train mode consequent on changes in the cost of travel by car and train are shown in Figure 3. It can be seen that the three models give a very similar prediction of changes in train use due to changes in train travel cost.

The changes in the use of the train mode consequent on changes in the travel time by car and train are given in Figure 4. Unlike Figure 3, there are a number of differences between the predictions provided by each model. In fact the only similarity in prediction is found when the three models predict changes in mode choice consequent on changes in automobile travel time between -40 and +100 percent. It is also of note that the traditional logit model tends to provide predictions that are greater than the threshold models.

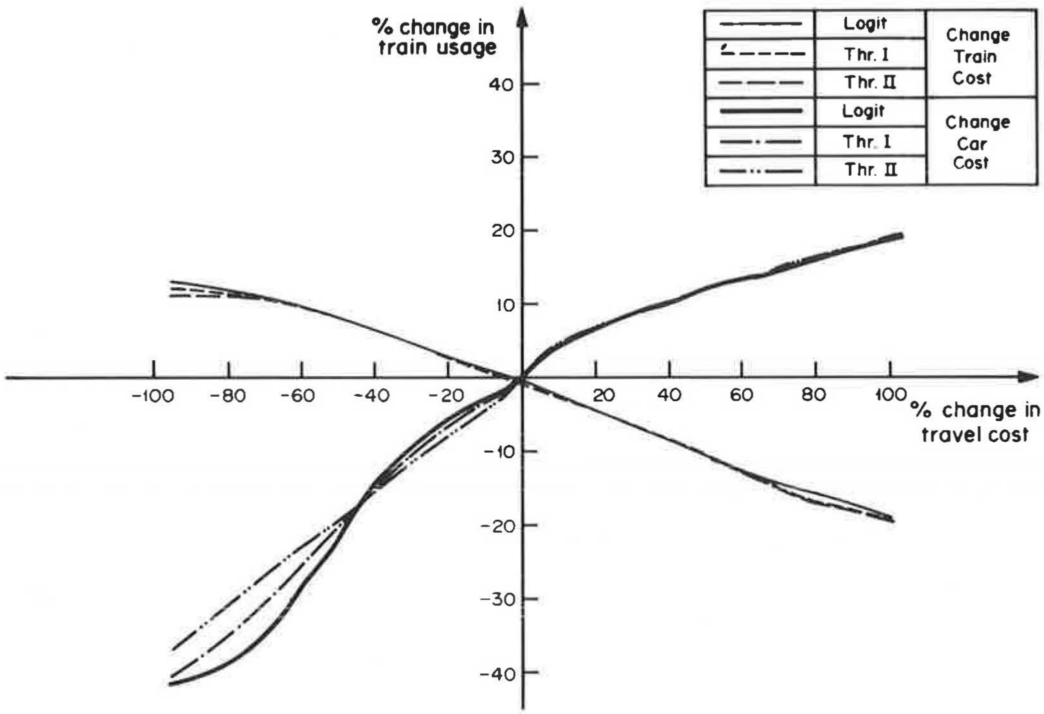

FIGURE 3 Response sensitivity to changes in travel cost.

The differences among the predictions of the three models imply that each model would give a different valuation of travel time. The logit model provides a value of travel time of 16 percent of the wage rate. The importance parameters from the threshold model cannot be interpreted in this way, but inspection of the sensitivity curves indicates that the value of travel time implicit in the threshold models is lower than that given by the logit model.

CONCLUSION

The inclusion of attribute thresholds into logit mode-choice models has been investigated. The sta-

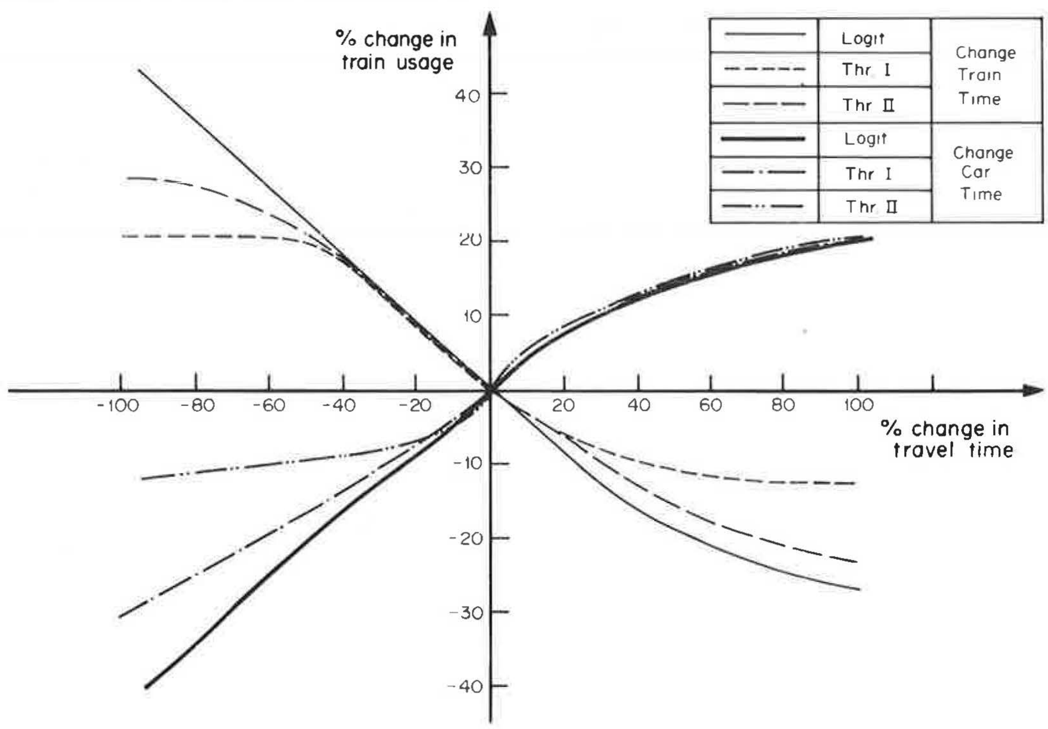

FIGURE 4 Response sensitivity to changes in travel time.

tistical performance and the predictions of these models were compared with a traditional logit model that did not contain thresholds. It was found that on the basis of statistical fit there was little difference in the performance of the models. However, each model responded differently to changes in the attribute satisfaction level and would thus predict different outcomes for certain system changes. The traditional logit model was found to be more sensitive to attribute level changes than the models that incorporated thresholds.

All three models indicated that the main attributes influencing choice were travel time, cost and convenience. The threshold models indicated that the average traveler would not react to differences in travel time of less than 12 min and travel costs of less than 12 cents from the best available alternatives satisfaction rating.

Unfortunately, this study provided no clear indication of the need to include thresholds. Rather, it was found that if thresholds are included, the model will perform well statistically and provide a different prediction to the model that does not include thresholds. The unfortunate conclusion is therefore a call for further research to reconcile this dilemma.

REFERENCES

1. D.A. Quarmby. Choice of Travel Mode for the Journey to Work: Some Findings. Journal of Transport Economic Policy, Vol. 1, 1967, pp. 273-314.
2. C.A. Lave. A Behavioural Approach to Mode Split Forecasting. Transportation Research, Vol. 3, 1969, pp. 463-480.
3. P.R. Stopher. A Probability of Travel Mode Choice for the Work Journey. In Highway Research Record 283, HRB, National Research Council, Washington, D.C., 1969, pp. 57-65.
4. C.F. Daganzo. Calibration and Prediction with Random Utility Models: Some Recent Advances and Unresolved Questions. Presented at Fourth International Conference on Behavioural Travel Modelling, Eibsee, West Germany, 1980.
5. J.P. Guilford. Psychometric Methods. McGraw-Hill, New York, 1954.
6. S.S. Stevens. The Surprising Simplicity of Sensory Metrics. American Psychologist, Vol. 17, 1962, pp. 26-29.
7. E.E. Slutsky. On the Budget of the Consumer. In Readings in Price Theory, Irwin, New York, 1952.
8. D.J. Finney. Probit Analysis. Cambridge University Press, Cambridge, England, 1964.
9. C.F. Kovak and M.H. Demetsky. Behavioural Modelling of Express Bus-Fringe Parking Decisions. In Transportation Research Record 534, TRB, National Research Council, Washington, D.C., 1975, pp. 10-23.
10. L.D. Burns, T.F. Golob, and G.C. Nicolaidis. A Theory of Urban Households' Automobile Ownership Decisions. In Transportation Research Record 569, TRB, National Research Council, Washington, D.C., 1975, pp. 56-72.
11. A.J. Richardson. Advanced Demand Modelling for Advanced Transport Systems. Journal of Advanced Transportation, Vol. 14, No. 2, 1980, pp. 113-131.
12. K.S. Krishnan. Incorporating Thresholds of Indifference in Probabilistic Choice Models. Management Science 23, 1977, pp. 1224-1233.
13. S. Kawakami and Y. Hirobata. Predicting Changes of Modal Choice after Improvement of Transport Services. Presented at World Conference on Transport, Hamburg, West Germany, 1983.
14. W.W. Recker and T.F. Golob. A Non-Compensatory Model of Transportation Behaviour Based on Sequential Consideration of Attributes. Transportation Research, Vol. 13B, 1979, pp. 269-280.
15. W. Young and H.P. Brown. A Revealed Importance Elimination-By-Aspects Model of Mode Choice. Proc., PTRC Summer Annual Meeting, Brighton, England, PTRC Education and Research Services, Ltd., London, 1983.
16. R.D. Luce. Individual Choice Behaviour: A Theoretical Analysis. John Wiley and Sons Inc., New York, 1959.
17. J. Marschak. Binary Choice Constraints and Random Utility Indicators. In Mathematical Methods in Social Sciences (K.J. Arrow et al., eds.), Stanford University Press, Stanford, Calif., 1959.
18. D.A. Hensher and L.W. Johnson. Applied Discrete Choice Modelling. Croom Helm, London, 1981.
19. H.P. Brown. Attitudinal Measures in Models of Mode Choice. Presented at Third Annual Meeting, Australian Transport Research Forum, 1977.

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

Dynamic Aspects of Departure-Time Choice Behavior in a Commuting System: Theoretical Framework and Experimental Analysis

HANI S. MAHMASSANI and GANG-LEN CHANG

ABSTRACT

The day-to-day dynamics of departure-time decisions of urban commuters and the underlying behavioral mechanisms determining user responses to dynamically varying time-dependent congestion patterns are addressed. A conceptual model is presented incorporating the boundedly-rational notion of an indifference band of tolerable schedule delay. The results of an experiment involving real commuters interacting daily within a simulated traffic corridor are examined, with particular emphasis on the dynamics of user behavior.

The departure-time decision of urban commuters is of fundamental importance to the study of peak-period traffic congestion and to the analysis of traffic control as well as to broader, demandside congestion relief measures, such as pricing, ridesharing incentives, flex time, and others (1). Previous work on the departure-time problem has followed one of two principal lines: (a) econometric models of individuals' departure-time choices under fixed and known transportation level-of-service attributes (2-5) and (b) dynamic user equilibrium formulations in idealized traffic systems consisting of a single origin-destination pair connected by either a single route (6-13) or multiple routes (14,15), with congestion modeled either by using deterministic queues (6,14) or traffic-flow relationships (15). More elaborate reviews of these studies may be found elsewhere (15-17).

There is, however, an important dimension of the dynamics of this problem that has received very little attention, namely, the processes governing commuters' day-to-day responses to the system's performance, including the effect of prior experience and perceptions on current decisions. These processes are undoubtedly complex because they involve behavioral aspects of individual decision making, learning, and judgment in the context of a complex interactive system. However, the understanding of these processes and the ability to represent them analytically are of considerable importance to the design and evaluation of congestion relief measures, particularly with regard to time lags that may be associated with users' responses to these measures and information dissemination programs that could influence these responses. Furthermore, these dynamic aspects have significant implications for the stability of the system, as demonstrated by Horowitz in the context of route choice in a simplified transportation network (18). It is these dynamic processes underlying users' departure-time decisions in an urban commuting corridor that form the focus of this paper.

An effort in this direction was recently presented by Mahmassani and Chang (16), who addressed the day-to-day evolution of the time-dependent demand pattern resulting from the interaction between system congestion and user decisions. In addition, that study differed from the previous lines of research in its use of a process model of individual behavior consisting of a combination of relatively simple decision rules and heuristics, including explicit mechanisms for learning over time, and incorporating the notion of an "indifference band" of tolerable schedule delay. The latter reflects boundedly-rational, or satisficing (19), behavior of users in their daily commuting choices in an effort to explore behaviorally realistic decision rules as an alternative to the more restrictive but convenient utility maximization rule adopted in all previous studies. Another different feature in that study was the use of a special-purpose traffic simulation model for the performance side, thus allowing for greater flexibility and realism in system representation.

Understanding of these processes can of course best be furthered when coupled with observations of actual behavior. However, the acquisition of the necessary data at the desired level of richness in the real world presents formidable difficulties, including (a) the need to monitor in great detail both user decisions and the facility's time-varying congestion levels over a period of at least a few weeks and (b) the high degree of experimental control required. An alternative approach has recently been used by Mahmassani et al. (20) whereby the behavior of actual commuters is observed under controlled conditions. Participants, facing a hypothetical though realistic commuting situation, supply daily departure-time choices in response to congestion conditions, which are in turn obtained by using a special-purpose traffic simulation model, given the time-varying demand pattern (resulting from the aggregation of the individual participants' decisions).

In this paper the results of the first such experiment, involving 100 participants over 24 days, are examined from the perspective of the processes governing the dynamics of the users' behavior. Other aspects, such as traffic conditions or convergence properties of the system, are discussed elsewhere (20). The related conceptual background is presented in the next section, followed by a brief description of the experiment. The principal results are then examined and compared with the simulation results obtained earlier (16), and concluding comments are presented.

CONCEPTUAL BACKGROUND

Because the principal concern here is with the departure-time decision for home-to-work trips, it will be assumed that it is the only short-term decision available to trip makers. This would be the case in a commuting corridor consisting of a single highway facility with residences and workplaces distributed along this facility. As such, other choice dimensions normally available to trip makers, such as the choice of mode or route, do not unduly divert the discussion from its central focus. Extension to the more general case would be possible, though it would require considerably more complexity in the presentation and notation.

Given a work starting time WS_i, a trip maker i will select, on day t, a departure time $DT_{i,t}$. The outcome of this decision will be an arrival time $AT_{i,t}$, which follows the identity

$$AT_{i,t} \equiv DT_{i,t} + TT_{i,t} \tag{1}$$

where $TT_{i,t}$ is the trip time experienced on day t (including travel time on the facility and all other components). The trip time naturally depends on the user's departure-time decision as well as that of all other users of the facility, that is,

$$TT_{i,t} = f(DT_{i,t}, \text{all } i) \tag{2}$$

As mentioned in the previous section, prior studies have assumed that users select their respective $DT_{i,t}$ so as to maximize their utility, which is usually formulated as a weighted sum of the attributes of the departure-time opportunities. Although theoretically appealing, the maximization paradigm has a number of limitations from a behavioral standpoint, especially in the context of a descriptive model of day-to-day choice dynamics. For instance, it requires users to possess information on all the decision alternatives, that is, that they know a priori or can predict the time-dependent congestion pattern on any given day. This is clearly a difficult task in view of the often substantial and well-documented variability of travel time during the peak period (9,21). In addition, it is not clear that the parameters of one's utility function would remain constant from day to day, but rather that users may update their relative trade-offs as they learn about the system's performance. Another assumption that is difficult to support in this context is that of the individual ability to evaluate the optimal solutions of rather complicated objective functions (22).

An alternative behavioral notion that suggests itself here is that of satisficing, proposed by Simon (19) as a model of so-called boundedly-rational decision makers in search of an acceptable solution as opposed to a necessarily optimal one. Acceptability is usually defined relative to some aspiration level. In addition, it is well established in behavioral science that decision rules employed by individuals are greatly influenced by the nature of the task and the decision environment (22,23). In everyday decisions, the predominance of mental heuristics in individual judgment, learning, and decision making is generally well accepted (23, 24). Preference is usually for simpler, less demanding (in terms of cognitive strain on the decision maker) rules, which subsequently may become more demanding in response to a more complex decision environment.

A useful analogy here is between commuting behavior and consumers' repurchase decisions, to the extent that the latter are repeated daily or frequently and involve nonmajor items. As such, the marketing research literature can provide some useful insights and possible guidance. Satisficing models have received increased attention and acceptance in marketing research because of their ability to capture consumer choice behavior (25-27). The role of satisfaction in the consumer decision process has been documented in a number of studies (28,29), including that of Oliver (30), who identified the adaption-level theory (31) as an appropriate one for explaining how past experience and current satisfaction interact in affecting repeat purchase behavior [see also the paper by Labarbera (32)]. Worthy of note is the recognition that the very basis for satisfaction and acceptability of a given outcome itself dynamically varies in response to prior outcomes as well as recent experience (28).

The application of some of these concepts to the dynamics of the departure-time choice problem is presented next. Trip makers can be viewed as searching for a departure time that yields a satisfactory outcome or arrival time. When user i is satisfied with $AT_{i,t}$, he is expected to maintain the same departure time on the following day; thus $DT_{i,t+1} = DT_{i,t}$. The acceptability of a particular outcome is evaluated with respect to the trip maker's own desired arrival time. Note here that the latter quantity is generally different from the work start time WS_i, as shown empirically by Hendrickson and Plank (5). Users typically possess a preferred arrival time PAT_i, which would prevail in the absence of congestion (yet still be within the constraints of the workplace). It generally incorporates a safety margin to protect against lateness at work and allow some time for preparation at the onset of the working day. One can thus expect a distribution of preferred arrival times across the population, reflecting both workplace conditions as well as inherent individual preferences and risk attitudes. Further support for this notion is presented later in this paper based on analysis of the experiment.

A plausible satisficing mechanism used in earlier simulations (16) is based on the notion of an "indifference interval" or band of acceptable schedule delay. On a given day t, user i's schedule delay is $SD_{i,t} \equiv PAT_i - AT_{i,t}$. Letting $\delta_{i,t}$ be a binary variable that takes the value 1 if the actual arrival time on day t is acceptable to user i, and 0 otherwise, the decision rule can be stated as

$$\delta_{i,t} = \begin{cases} 1 \text{ if } 0 \le SD_{i,t} < IB^e_{i,t} \text{ or } IB^\ell_{i,t} < SD_{i,t} \le 0 \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

where $IB^e_{i,t}$ and $IB^\ell_{i,t}$ are two nonnegative threshold values reflecting what user i considers tolerable earliness and tolerable lateness, respectively. The time interval within which an arrival time $AT_{i,t}$ will be considered acceptable then becomes $(PAT_i - IB^e_{i,t}, PAT_i + IB^\ell_{i,t}]$.

The threshold values $IB^e_{i,t}$ and $IB^\ell_{i,t}$ can be expected to vary across individuals, reflecting differing preferential attitudes as well as workplace conditions. To the extent that the preferred arrival time PAT_i reflects some of these same sources of variation, it can be expected to be correlated with those threshold values. For this reason, in the empirical analysis section, various user groups will be considered on the basis of their preferred arrival times.

An individual's indifference band, reflecting his aspiration level on a given day, is not necessarily constant over time, particularly if the system is not in a steady state, such as after the implementation of a major new control or policy. This band is instead dynamically changing in response to the user's personal experience with the facility as well as information that he may have actively or passively acquired from other sources. Insight into this phenomenon was obtained in earlier simulations (16). In particular, more distant users (relative to a common work destination) would tend to adjust their aspirations more frequently than closer users in order to accommodate greater day-to-day variability and fluctuation in their longer commutes. Similarly, more distant users appear to require wider indifference bands. These aspects of user behavior are explored in the section on analysis of experimental results.

Information acquired through repeated usage of the facility, as well as from other possible sources, influences trip makers' short-term departure-time choice behavior in two major ways: (a) the previously mentioned effort on the aspiration level, defining the acceptability of particular outcomes, and (b) learning about the facility's performance, which provides the basis for the user's travel time estimate and the subsequent departure-time adjustment in the event that the latest outcome was not acceptable. This adjustment is determined by both the current indifference band and the user's perception of the system's travel time characteristics; it can thus be viewed as the following function:

$$DT_{i,t+1} - DT_{i,t} = g(TT_{i,s}, SD_{i,s};\ s = 1,\ldots,t) \qquad (4)$$

In this expression, the relative importance of terms corresponding to different values of s (days) is not expected to be uniform. Clearly, recent experience is likely to contribute more heavily than that of more distant days. At one end of the spectrum, user behavior could be purely myopic and affected by the latest day only. At the other extreme, all days from 1 to t could contribute with equal intensity to the user's decision on day (t+1). However, because of memory capacity limitations, the retrieval of prior information is not likely to go beyond a relatively small number of recent days.

In summary, user behavior in this commuting system can be viewed as a boundedly-rational search for a satisfactory departure time. Conceptually, it consists of two principal components: (a) the acceptance or rejection of a given day's decision outcome, which determines, respectively, whether the user will or will not maintain the same departure time on the following day and (b) the amount by which departure time should be adjusted, if that is needed. The first component can be viewed as the stopping criterion in the user's search process, whereas the second is analogous to the "step size." The former is based on the key notion of an indifference band of tolerable schedule delay. Prior experience with the facility, as the principal mechanism of information acquisition, enters the first component through its effect on the indifference band, and the second component through its contribution to the user's learning about the facility's performance.

It should be noted here that the use of schedule delay as the principal criterion for acceptability of a given decision outcome should not be taken to imply that other attributes, particularly travel time, will under no circumstances be explicitly evaluated by trip makers. Implicit is the assumption that the range of travel times encountered by individuals in this urban commuting system is such that users are effectively indifferent among the travel time outcomes of their departure decisions. Naturally, for excessively long travel times, this assumption is not likely to hold. In an intercity context, where travel times are much more substantial, explicit trade-offs between schedule delay and travel time should be expected, as in airline flight selection. However, in an urban commuting context, particularly for short-range, day-to-day decisions, schedule delay is clearly significantly more highly valued (negatively) than travel time, as evidenced by the findings of Hendrickson and Plank (5). Users are likely to control for their travel times through longer-run choices such as that of residence or workplace location. In a dynamically changing context, where users possess only limited information on the system's performance, boundedly-rational behavior predicated on the most-important attribute appears to be plausible descriptivity.

In the remainder of this paper, the results of an experiment for additional insight into the foregoing aspects of user behavior and the extent to which they appear consistent with the conceptual model presented in this section are analyzed. However, no formal functional specification and estimation will be conducted herein, because the analysis is exploratory in nature and is intended at this stage primarily as an indication of the usefulness of this general approach to studying the complex day-to-day dynamics of commuter behavior. The experiment itself is described in the next section.

DESCRIPTION OF EXPERIMENT

Given the previously mentioned difficulties of obtaining adequate data for the study of the day-to-day dynamics of commuter behavior, the approach recently described by Mahmassani et al. (20) consists of observing the decisions of real commuters placed in controlled and carefully designed hypothetical commuting situations. A number of important features characterize this type of experiment, including the following:

1. All the departure-time decisions that collectively determine the system's service levels can be observed,
2. The analyst has a high degree of control over the information available to participants, and
3. The interactions in the traffic system, which determine the user's decisions, are realistically captured by a special-purpose traffic simulation model.

The commuting context considered in this experiment consists of an urban corridor composed of a four-lane highway (two lanes in each direction) used by residents who live adjacent to it for their daily home-to-work trips to a single work destination, such as a central business district (CBD) or a major industrial park. Concern here is with the inbound, or home-to-work, direction. The corridor is subdivided into nine identical 1-mi sectors, with the common destination located at the end of the last sector. Sectors are numbered from 1 to 9 in order of decreasing distance from the destination; Sector 1 is the farthest outbound. Commuter residences are located in Sectors 1 through 5 only, each of which is treated as distinct trip origin, whereas Sectors 6 through 9 are treated as a nonresidential fringe area in which no trips are generated.

The time-dependent departure pattern from each residential sector on any given day results from departure-time decisions made by the participants.

Each participant is assigned to only one sector and is assumed to represent a group of 20 trip makers who make identical decisions. A total of 400 trip makers was assumed in each of the residential sectors (or 200 trip makers per lane per sector), resulting in 20 participants for each of the five sectors in this experiment.

The following information was initially provided to each participant: (a) a general description of the foregoing commuting context, (b) the participant's residential sector, (c) the highway facility's characteristics (number of lanes, free-flow speed), and (d) the work start time.

With regard to the third item, note that similar facilities in the Austin area were indicated to the participants for anchoring purposes. With regard to the fourth item, all participants were placed in the familiar situation of having to start work at 8:00 a.m. Although it would have been more representative of the real world to have had a distribution of work start times, still strongly peaked at 8:00 a.m. (5), it would have required considerably more participants to attain a meaningful level of interaction in the system. The specification of a single work start time in this first experiment captures all the key phenomena of interest and avoids undue complexity.

At the onset of the experiment, each participant was asked to state his or her preferred arrival time at work (PAT_i for participant i), in the absence of traffic congestion, given the official work start time WS. Naturally, $PAT_i \leq WS$ for all i.

Every simulation day, each participant supplied a departure time and an anticipated arrival time, denoted hereafter by $DT_{i,t}$ and $AAT_{i,t}$, respectively, for user i on day t. The departure-time decisions of all individuals in a given sector were aggregated into a time-dependent departure pattern for that sector. These patterns formed the input to the highway traffic flow simulation model, briefly described later in this section. The outcome of each participant's decision (the actual arrival time $AT_{i,t}$) and the corresponding travel time $TT_{i,t}$ were determined by the simulation and supplied to each participant individually on the following day before that day's choice. This iterative interactive process covered 24 simulation days, by the end of which the system had evolved to a stable state, with all participants maintaining the same choices from one day to the next. In order to relate the experiment to the participants' daily commute, it was administered daily, 5 days per week, during the entire period.

The importance of information acquired through one's own commuting experience and from other possible sources was discussed in the previous section. In this experiment, the informational scenario under which users have only their own actual experience to rely on is considered. Furthermore, to the extent that commuters do not usually maintain a written log of their departure and arrival times over a number of days, only the latest day's decision outcome was displayed to each participant. Other informational scenarios involving additional sources, such as mass media reports or word of mouth, were outside the scope of this particular experiment and may be addressed in future work.

In order to achieve the desired quality of the results, participants were selected very carefully, especially because their involvement was required for a period of several weeks. All 100 participants were affiliated with the University of Texas at Austin, and most were staff members or graduate students with formal work experience. In addition, these participants were scattered over various parts of the campus, thus controlling for information exchange among participants during the survey period.

Before this experiment was conducted, a pretest was administered to a smaller and different group of individuals. Responses and suggestions from this pretest group led to helpful improvements in the procedure as well as initial insights into the behavior of the system.

A special-purpose, fixed-step macroscopic highway traffic simulation model was developed in conjunction with this experiment. The highway facility is segmented into a number of sections, in which traffic flow is modeled by using well-established fundamental traffic flow relationships; of particular interest is the speed-density model, which has the following form:

$$v = (V_f - V_0)(1 - K/K_0)^c + V_0 \qquad (5)$$

where

- V and K = speed and density prevailing on a given highway section, respectively;
- V_f and V_0 = free-flow speed and the minimum allowable speed on the facility, respectively;
- K_0 = maximum or "jam" density; and
- c = parameter reflecting the sensitivity of travel speed to density variations.

In this experiment, the following parameter values were used: V_f = 40 mph, V_0 = 6 mph, K_0 = 180 vehicles/lane-mile, and c = 1.0. Further details of the simulation model are outside the scope of the present paper and can be found elsewhere (20,33).

ANALYSIS OF EXPERIMENTAL RESULTS

The presentation of the principal results of interest to the behavioral processes underlying user decision dynamics is organized around four types of quantities:

1. Actions, meaning the actual departure-time decisions of users over the survey period (i.e., $DT_{i,t}$) for i = 1,...,100 and t = 1,...,24;
2. Outcomes, which result from the foregoing actions, namely, the actual arrival time $AT_{i,t}$ and associated travel time and schedule delay ($TT_{i,t}$ and $SD_{i,t}$, respectively);
3. Perceptions, by users, of the foregoing outcomes, translating into anticipated travel times and schedule delays ($ATT_{i,t}$ and $ASD_{i,t}$, respectively); and
4. Intentions, or preferences, which, when combined with the foregoing anticipated quantities, result in actual decisions; of concern here are the preferred arrival times PAT_i for all i and the anticipated arrival times $AAT_{i,t}$ stated by all users along with their departure decisions on any given day.

In addition to the description of the evolution of the foregoing quantities and their variation by geographic sector (as a function of distance from the destination) and other factors, their interrelation, as discussed in the previous section, is explored. However, first the overall evolution of the system's behavior is summarized.

Summary of System Evolution

The system equilibrates when all users are essentially satisfied with the outcome of their departure-time choices, thus maintaining the same daily departure pattern. In this experiment, no user changed his or her departure time as of day 21; however, the steady-state values were first attained on day 18 but were perturbed by a few participants who tried, unsuccessfully, to improve their outcome and subsequently returned to their steady-state choices. A clear geographic pattern in the evolution to the steady-state choices was apparent, with sectors closer to the destination generally reaching their steady-state earlier than more distant sectors. For instance, the steady-state departure patterns were reached (and maintained) in Sectors 1 through 5 as of days 20, 17, 16, 16, and 5, respectively. Furthermore, only a small fraction of the users in Sector 4 kept searching for a satisfactory outcome beyond day 7, as revealed by Figure 1, which shows the day-to-day evolution of the departure-time distribution (i.e., the fraction of users departing before time t on a given day) in Sector 4. The "ease" with which users in different sectors are able to attain a satisfactory outcome is further documented later in this section by looking at the frequency of departure time as well as anticipated arrival-time changes.

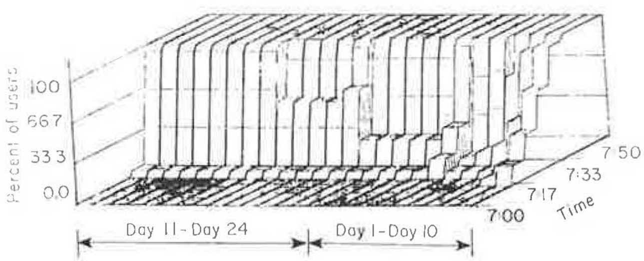

FIGURE 1 Cumulative departure pattern evolution for Sector 4.

Of course, the overall system cannot be considered in equilibrium so long as some sector has not yet reached its steady state, because changes in any sector will affect the outcomes of user decisions in other sectors through the traffic interactions. Actually, the fact that many users maintained their departure-time choice despite the continued variation of travel times and schedule delays suggests the existence of the tolerable range associated with the boundedly-rational behavior described earlier. For instance, the day-to-day variation of the average of the absolute value of schedule delay, per sector, is shown in Figure 2, which reveals that

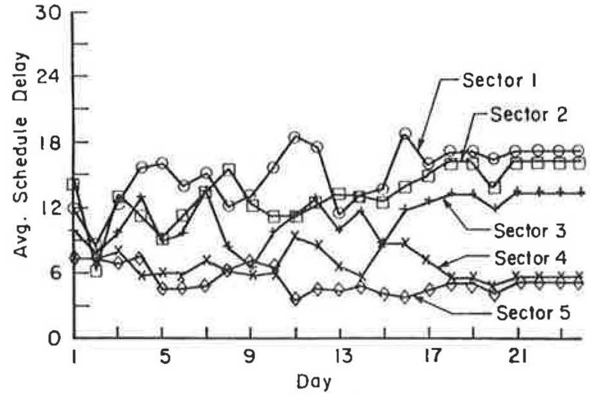

FIGURE 2 Evolution of average schedule delay in absolute values per sector.

this quantity still varied for Sector 5 for many days after users in that sector had stopped adjusting their departure times.

Although it is clear that convergence was attained, it is not possible to ascertain, on the basis of this single experiment, the uniqueness of this pattern nor to derive conditions for its existence. Further discussion of the convergence properties of this experimental system may be found elsewhere (20).

Preferred Arrival Time

As mentioned earlier, commuters have different preferred work arrival times. In this experiment, although the same common work start time (8:00 a.m.) was specified for all participants, the stated preferred arrival time followed the distribution shown in Figure 3, which reveals that over 40 percent prefer to reach their workplace at least 15 min before the official work start time. This distribution is primarily a reflection of inherent differences in individual preferences and does not exhibit any systematic variation across sectors. To the extent that

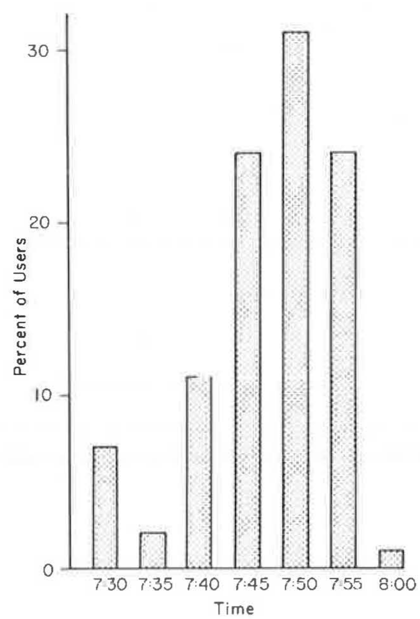

FIGURE 3 Preferred arrival time distribution.

the excess time preferred by users can be interpreted as a safety margin for avoiding lateness, the preferred arrival time provides a useful indication of a user's risk attitudes. Therefore, it has been used as a basis for segmenting the participants, and indeed significant differences in behavior across the three groups that were defined were found:

- Group 1, including all users i such that 7:30 $\leq$ PAT_i < 7:40 a.m.,
- Group 2, where 7:40 $\leq$ PAT_i < 7:50 a.m., and
- Group 3, where 7:50 $\leq$ PAT_i $\leq$ 8:00 a.m.

The relative frequency distribution of users into each of the foregoing three categories is given in Table 1 per sector, as well as overall.

The preferred arrival times represent the initial intentions of users before their experience with and

TABLE 1 Relative Frequency Distribution of Users into Preferred Arrival-Time Groups per Sector

	Percentage of Users by Sector					
Group	1	2	3	4	5	All
1 (7:30-7:39 a.m.)	10	15	5	5	10	9
2 (7:40-7:49 a.m.)	40	30	30	40	35	35
3 (7:50-8:00 a.m.)	50	55	65	55	55	56

subsequent learning about the system's performance. However, as learning develops through usage, these intentions evolve, as seen later in this analysis of the daily anticipated arrival times.

Actions: Departure-Time Decisions

Patterns exhibited by the frequency of departure-time changes and the time interval between successive changes across sectors and across user groups are examined first. Also the effect of the previous day's outcome on the decision to adjust one's departure time, particularly with regard to the existence of an indifference band of schedule delay, is highlighted. In addition, the magnitude of this adjustment is examined relative to the previous day's schedule delay.

Table 2 shows the respective proportion of participants in each sector who changed their daily departure time at least n times, where n = 1, ..., 15 (15 was the highest number of changes observed out of a maximum of 23 possible changes in 24 days). The overwhelming pattern is that the frequency of these changes increases with distance from the destination, thus confirming the observation that more distant sectors experience greater difficulty in converging to a steady state.

TABLE 2 Proportion of Users in Each Sector with at Least n Departure-Time Changes

	Percentage of Users by Sector				
No. of Changes[a]	1	2	3	4	5
1	100	100	100	100	75
2	100	100	95	65	25
3	100	100	90	60	
4	100	95	85	15	
5	90	90	70	5	
6	90	80	40		
7	90	75	30		
8	80	60	10		
9	65	50			
10	50	30			
11	35	15			
12	25				
13	20				
14	10				
15	5				

[a]Minimum number.

Table 3 presents the same information as Table 2, but for each of the previously defined user groups within each sector. As expected, users in Group 1, who were initially willing to accept a wide safety margin, were able to conclude their search for an acceptable departure time significantly sooner than the other groups. (It should be noted here that comparisons of Group 1 users across sectors is not meaningful given the small number of participants in this group in any one sector.) The same general trend is present for Groups 2 and 3, especially in Sector 1, in which residents encounter greater travel time fluctuation than in closer sectors, thus making it particularly difficult to successfully maintain a departure time that results in arrival within less than 10 min from the work start time. In addition, the preferential differences captured by the user groups may correspond to varying degrees of individual persistence, whereby users in Group 3 are less willing to adjust their indifference band to accommodate otherwise unacceptable outcomes. This particular aspect is more specifically explored in the context of the discussion of intentions.

Table 4 shows, per user group within each sector, the mean number of days since the previous change for the nth change (n = 1,...,15) as well as its standard deviation. Naturally, these numbers must be interpreted with caution because many of these averages, particularly for higher values of n, are taken over a small number of participants. Although no strong patterns are present in a uniform manner, the time until the first change appears to be of the same order of magnitude across the categories considered, with the notable exception of Sector 4, in which a large fraction of users did not have to change their initial selection for a long time, which resulted in the large means and standard deviations seen in Table 4. It is also apparent that the variability of the interval between changes is greater for the closer sectors (though not for Sector 5, where very few changes took place). More accurately, this variability is more evident for user groups in sectors where the decision to change was not clear-cut. For instance, users in Groups 2 and 3 in Sectors 1 and 2 experienced outcomes that were clearly unacceptable to most, resulting in the low observed standard deviations in Table 4. This was less the case in Sectors 3 and 4, where the time between consecutive changes varied considerably across users. These results will be contrasted later in this section with the time interval between changes in anticipated or intended arrival times reported by users.

To ascertain the effect of the previous day's outcome on the decision to change departure time, the response, in each sector, to different levels of schedule delay (in 5-min increments) has been examined. Thus for each sector, the fraction of those users experiencing a given schedule delay on day t - 1 that have changed their departure time on day t has been calculated. In order to detect the postulated evolution (see section on conceptual background) of the users' indifference bands as the search progresses and still have enough observations to yield meaningful fractions, the data were aggregated on a weekly basis (each including 5 days). Although not all schedule delay levels are sufficiently represented, two rather clear trends are suggested by these data.

First, as expected, the fraction of users who find a particular schedule delay unacceptable and thus change departure time on the next day increases with the magnitude of the delay. This is exemplified in Table 5, which shows these fractions for Sector 1 during the third week of the survey. Interestingly, no user experiencing lateness of up to 5 min or earliness of up to 10 min (relative to his or her respective preferred arrival time) decided to adjust departure time on the following day.

The second trend concerns the evolution of the indifference band, whereby the fraction of users rejecting a given outcome appears to decrease as the search progresses, shown as follows for selected

TABLE 3 Proportion of Users in Each User Group Within Sectors 1, 2, and 3 with at Least n Departure-Time Changes

No. of Changes[a]	Percentage of Users by Sector								
	1			2			3		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
1	100	100	100	100	100	100	100	100	100
2	100	100	100	100	100	100		100	100
3	100	100	100	100	100	100		83.4	100
4	100	100	100	66.7	100	100		83.4	92.8
5		100	100	33.3	100	100		83.4	69.7
6		100	100	33.3	66.7	100		66.7	31.2
7		100	100	33.3	66.7	90.9		50.0	23.4
8		71.5	100		50.0	81.8		16.7	7.8
9		28.6	100		50.0	63.6			
10		14.3	81.9		33.2	36.3			
11			63.7			27.3			
12			45.5						
13			36.4						
14			18.2						
15			9.1						

[a]Minimum number.

TABLE 4 Mean and Standard Deviation of Number of Days Between Consecutive Departure-Time Changes per User Group Within Each Sector

Change Sequence No.	Sector and Group														
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3	4,1	4,2	4,3	5,1	5,2	5,3
1															
Mean	1.50	1.71	1.50	1.0	1.25	1.50	1.0	2.83	1.84	1.0	5.3	5.58	–	1.50	2.50
SD	0.70	1.25	0.53	0.0	0.50	0.85	–[a]	2.31	0.89	0.0	6.0	4.28	–	1.00	0.68
2															
Mean	2.50	2.14	1.27	1.0	1.40	1.70		2.85	1.76		1.5	3.67		1.0	1.0
SD	1.00	0.78	0.47	0.0	0.55	0.48		3.10	1.64		1.0	4.30		0.0	0.0
3															
Mean	2.50	1.28	1.09	5.67	1.20	1.60		2.80	2.92		4.75	7.57			
SD	1.00	0.49	0.30	2.89	0.45	1.07		1.64	1.78		4.30	5.50			
4															
Mean	7.50	2.28	1.20	3.50	1.80	1.60		1.0	3.83		4.50	4.0			
SD	0.70	1.25	0.63	1.85	0.83	1.58		0.0	3.00		4.90	–[a]			
5															
Mean		1.29	2.00	4.0	1.60	1.40		1.75	4.0		4.0				
SD		0.76	1.61	–[a]	0.89	0.52		0.95	3.24		–[a]				
6															
Mean		1.71	1.36	2.0	1.40	1.90		2.25	3.0						
SD		1.25	0.92	–[a]	0.55	1.28		0.78	2.3						
7															
Mean		1.85	1.36	5.0	6.00	1.44		5.0	3.67						
SD		1.46	0.94	–[a]	4.12	1.33		4.20	1.15						
8															
Mean		1.50	1.40		1.0	1.87		4.0	1.0						
SD		1.00	0.96		0.0	1.64		–[a]	–[a]						
9															
Mean		2.00	1.33		1.0	2.83									
SD		0.85	1.00		0.0	1.94									
10															
Mean			1.50		1.0	1.0									
SD			1.24		–[a]	0.0									
11															
Mean			1.40			3.30									
SD			0.55			0.58									
12															
Mean			1.25												
SD			0.50												
13															
Mean			1.0												
SD			0.0												
14															
Mean			1.0												
SD			–[a]												
15															
Mean			2.0												
SD			–[a]												

Note: SD = standard deviation in days.

[a]Only one participant accounted for this change.

TABLE 5 User Response to Previous Day's Schedule Delay, Sector 1, Week 3

Schedule Delay on Day t − 1 (min)[a]	Proportion of Users Experiencing Delay Who Change Departure Time on Day t (%)
<−15	100.0
−15 to −11	84.6
−10 to −6	56.3
−5 to −1	0.0
0-5	0.0
6-10	0.0
11-15	11.1
16-20	33.3
21-25	37.5
>25	100.0

[a]Note that schedule delay on day t for a given user is defined as the difference between that user's preferred arrival time and his or her actual arrival time on day t. As such, negative values of schedule delay correspond to late arrivals, whereas positive values correspond to early arrivals (relative to the preferred arrival time).

schedule delays (the first column for lateness and the second for earliness) for Sector 1:

Week	Proportion of Users (%) by Schedule Delay (min) -10 to -6	16-20
1	81.8	100.0
2	45.5	78.6
3	56.3	33.3
4	40.0	0.0

A comparison of the responses between the first and the final weeks reveals this decrease in all cases where sufficient data exist. However, the path during the intervening weeks is not necessarily monotonic, particularly for the negative schedule delays as shown in the second column of the foregoing tabulation. Naturally, there are other factors affecting this response, such as the user preferential group, in addition to daily effects (aggregated in the weekly data) and random variation across users and days, which is of particular concern when the number of participants is relatively small.

Further support for the above two trends can be obtained by examining the magnitude of the departure time adjustment on day t (i.e., $DT_{i,t} - DT_{i,t-1}$) as a function of $SD_{i,t-1}$. Figures 4-8 show this adjustment versus $SD_{i,t-1}$, for all users in the system for t = 2, 6, 11, 16, and 24. In Figures 4-8, an asterisk corresponds to a single observation, a plotted number (2 to 9) refers to the number of participants with identical coordinates, whereas a plus sign represents at least 10 participants. When the focus is on the evolution of the points corresponding to a zero adjustment, these plots suggest that (a) as expected, there is a range of schedule delay that users are willing to tolerate and for which they do not adjust their departure time and (b) this range increases over time, indicating that users progressively accept greater schedule delay. In addition, examining the relative magnitudes of the two plotted variables reveals that (a) earliness on a given day implies a later (or same) departure on the next day, whereas lateness implies an earlier (or same) departure, and (b) the magnitude of the adjustment on day t is in most cases less than the corresponding magnitude of the earliness or lateness on day t - 1, which is consistent with a hypothesized rule in earlier simulations (16).

Further insight into the relation between this adjustment and schedule delay on the previous day is obtained by examining $(DT_{i,t} - DT_{i,t-1})/SD_{i,t-1}$. For each user group in each sector, the average of this ratio was calculated for the nth change given that the user was respectively late and early on day t - 1, with n = 1,...,8, and that the adjustment was nonzero. Table 6 shows these averages for Sector 1, which is representative of the other sectors. The variation of this ratio across user groups takes place in opposite directions depending on whether the adjustment is in response to an early or late arrival on the previous day. A plausible explanation

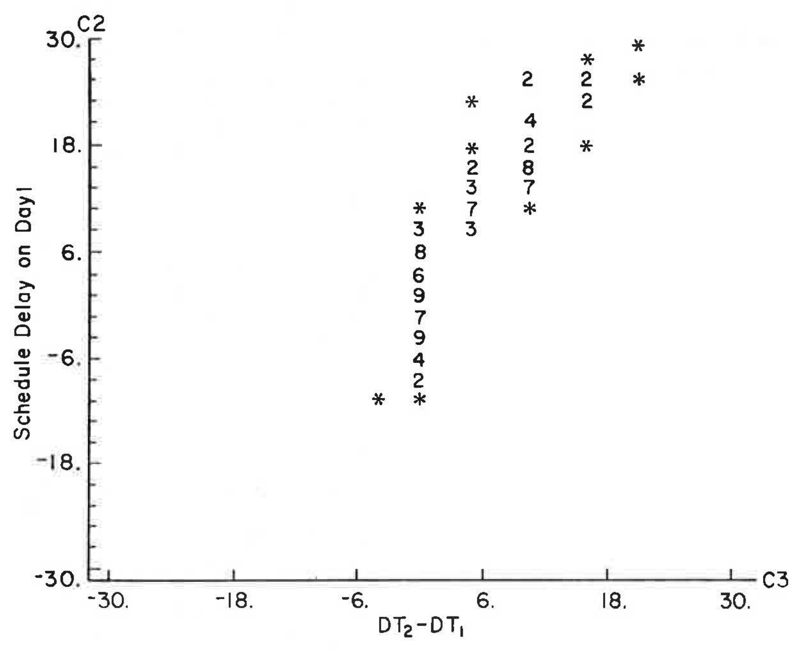

FIGURE 4 Departure time adjustment versus deviation from preferred arrival time on previous day: Day 2.

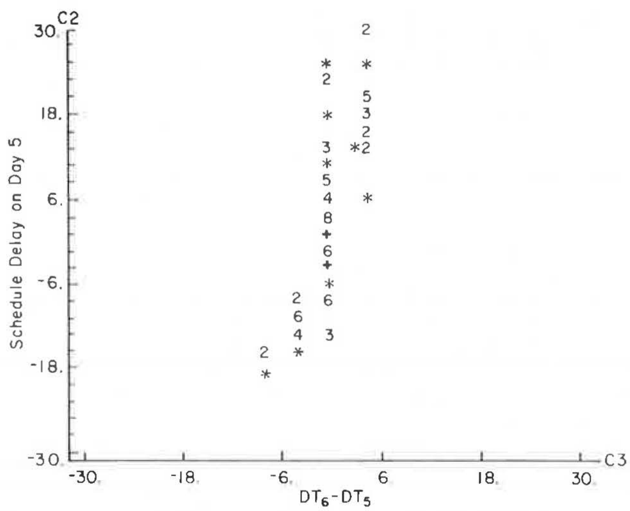

FIGURE 5 Departure time adjustment versus deviation from preferred arrival time on previous day: Day 6.

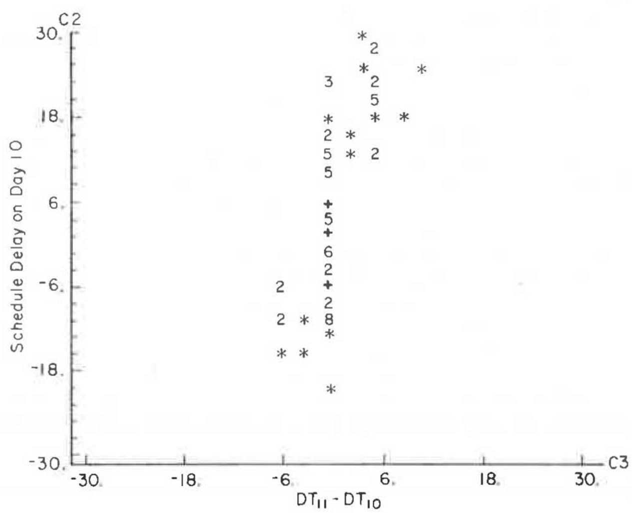

FIGURE 6 Departure time adjustment versus deviation from preferred arrival time on previous day: Day 11.

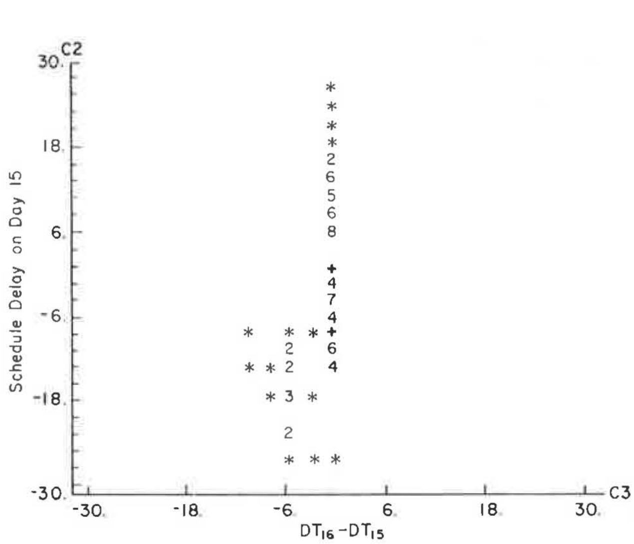

FIGURE 7 Departure time adjustment versus deviation from preferred arrival time on previous day: Day 16.

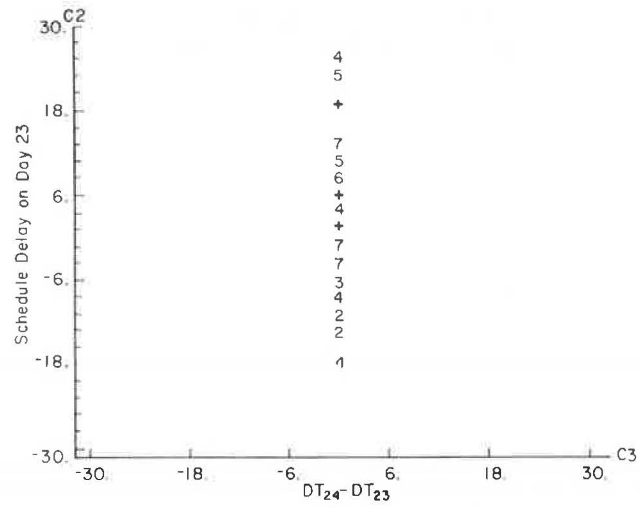

FIGURE 8 Departure time adjustment versus deviation from preferred arrival time on previous day: steady state.

TABLE 6 Average Ratio of Departure-Time Adjustment to Previous Day's Schedule Delay, Sector 1

	Average Ratio by User Group					
	1		2		3	
Change Sequence No.	E	L	E	L	E	L
1	0.67	0.29	0.56	0.38	0.50	0.52
2	0.55	0.22	0.24	0.40	0.26	0.46
3	0.38		0.23	0.24	0.19	0.46
4			0.20	0.21	0.17	0.46
5			0.11	0.25	0.12	0.51
6					0.11	0.63
7						0.90
8						0.36

Note: E and L refer to departure-time changes in response to earliness and lateness, respectively, on the previous day.

is that lateness relative to a preferred arrival time that is closer to the official work start time is more likely to result in actual lateness for work; the adjustment in this case is larger (relative to $SD_{i,t-1}$) than that when the lateness is entirely within the excess time between PAT_i and WS. On the other hand, adjustments in response to earliness are larger for users with earlier preferred arrival times, to avoid otherwise excessive earliness relative to the work start time. Table 6 also reveals the general trend of a decreasing adjustment ratio across successive changes, particularly in response to earliness. The trend is not as clear for responses to lateness. It should also be noted as one interprets Table 6 that the averages for the later changes are based on very few participants.

<u>Outcomes: Schedule Delay and Travel Time</u>

It was seen earlier that the average schedule delay ultimately accepted by users in each sector increases with distance from the destination (Figure 2), which suggests that more distant users ultimately accept larger schedule delays and as such possess wider indifference bands of tolerable schedule delay. This is supported by the more detailed analysis of the proportion of users accepting (at equilibrium) various levels of schedule delay, in 5-min increments, presented in Table 7.

TABLE 7 Relative Frequency per Sector of Difference Between Preferred Arrival Time and Actual Steady-State Arrival Time

Δ[a] (min)	Frequency by Sector 1	2	3	4	5	All
Early						
21–25	20	15	10	0	0	9
16–20	30	25	25	0	0	16
11–15	0	10	40	10	5	13
6–10	10	10	15	35	25	19
1–5	0	5	5	30	45	17
0	0	0	0	0	0	0
Late						
(-1)-(-5)	20	0	5	10	15	10
(-6)-(-10)	20	20	0	15	5	12
(-11)-(-15)	0	15	0	0	5	4

[a]Δ = preferred arrival time minus actual arrival time at steady state.

The evolution of average travel time per sector is shown in Figure 9, which reveals the greater day-to-day fluctuation encountered by commuters originating in more distant sectors and the ensuing difficulty in converging to a steady state. Further details on the facility's traffic flow performance and the travel time characteristics of the system may be found elsewhere (20,23).

Perceptions and Learning

Direct information on user perception of travel time and schedule delay was not available from this experiment. However, of related interest are the anticipated travel time and schedule delay derived from the anticipated arrival time reported daily by users along with their departure-time choice.

In order to examine how actual experience on a given day influences perception on the following day, the ratio of the actual travel time on day t - 1 to the anticipated travel time on day t is considered (i.e., $TT_{i,t-1}/ATT_{i,t}$). The average of this ratio is taken separately over users experiencing lateness and earliness (relative to PAT_i), respectively, on day t - 1 for each sector. Table 8 shows these averages for days 1 through 6 along with the corresponding standard deviations. If the ratio is greater than 1, travel time is anticipated to be lower than on the previous day. It can thus be seen that users arriving late on day t - 1 appear, on average, to anticipate travel time on day t to be lower than on the previous day, whereas those arriving early on day t - 1 anticipate higher travel time on the next day. This somewhat counterintuitive finding can be attributed to the allowance by early users of a safety margin over their latest experienced travel time when they reset their departure time. On the other hand, late users are somehow hoping to compensate for the latest experienced travel time by an earlier departure that would face less congestion.

In order to compare the anticipated travel time on a given day with the actual travel time on that day, Figure 10(a-e) shows the day-to-day evolution of the average difference $TT_{i,t} - ATT_{i,t}$ for Sectors 1 through 5. Figure 10 reveals considerable daily fluctuation, with no clear decreasing pattern appearing until day 16. Overall (with the exception of Sector 5), there seems to be no particular tendency of overestimation as opposed to underestimation. The key conclusion suggested here is that users can be good travel time predictors only when the system has essentially stabilized. There is therefore no support for the contention that users are systematically learning about the facility's time-dependent performance, as is usually implied in a perfectly rational decision framework. Instead, local and somewhat myopic rules seem to be governing users' perception of the facility's performance.

The day-to-day evolution of the average absolute value of the difference between actual and anticipated schedule delay on a given day is shown in Figure 11. The same conclusions apply here as previously because it can be established algebraically that $SD_{i,t} - ASD_{i,t} \equiv ATT_{i,t} - TT_{i,t}$.

Intentions: Anticipated Arrival Time

This analysis parallels that of the departure-time choices, particularly because the concern is primar-

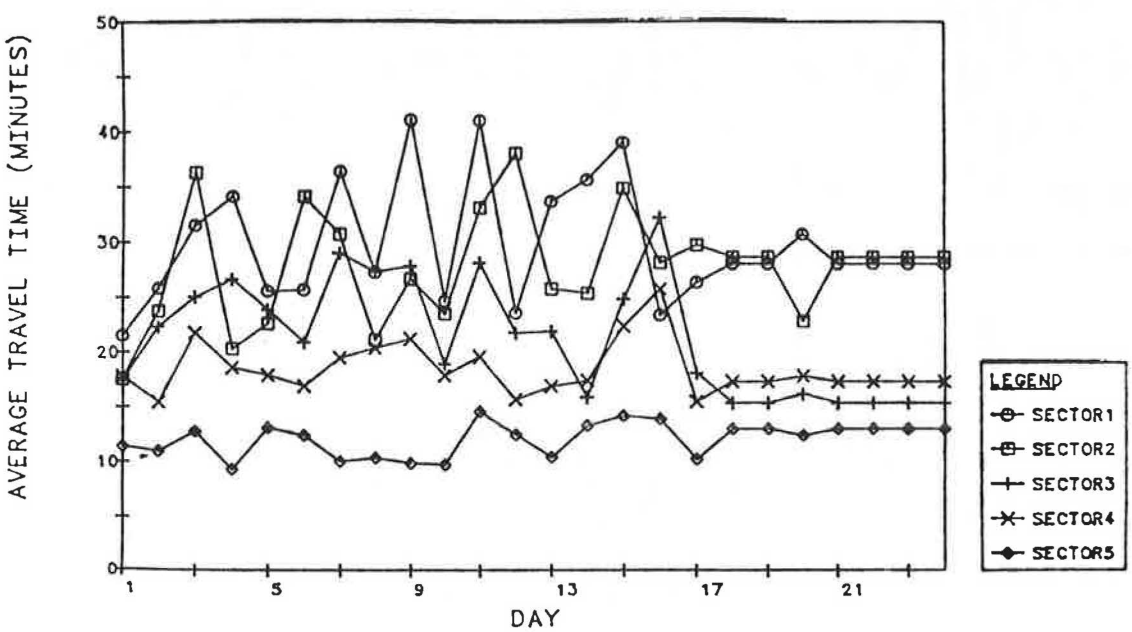

FIGURE 9 Evolution of average travel time for each sector.

TABLE 8 Average Ratio of Actual Travel Time on Day t – 1 to Anticipated Travel Time on Day t for Early Versus Late Users by Sector

	Day t									
	2		3		4		5		6	
Sector	Ratio	SD	Ratio	SD	Ratio	SD	Ratio	SD	Ratio	SD
1										
E	0.75	0.07	0.78	0.05	0.74	0.13	0.47	0.06	0.48	0.08
L	1.19	0.15	1.14	0.09	1.36	0.12	1.23	0.13	1.36	0.30
2										
E	0.65	0.22	0.76	0.09	0.64	0.21	0.66	0.11	0.69	0.16
L	1.04	0.03	1.15	0.30	1.26	0.24	1.35	0.27	1.15	0.09
3										
E	0.72	0.16	0.71	0.17	0.75	0.06	0.44	0.08	0.62	0.23
L	1.31	0.29	1.30	0.29	1.21	0.17	1.38	0.22	1.33	0.23
4										
E	0.76	0.14	0.76	0.13	0.75	0.13	0.85	0.11	0.88	0.08
L	1.26	0.11	1.29	0.14	1.33	0.31	1.17	0.13	1.29	0.19
5										
E	0.70	0.11	0.67	0.13	0.64	0.16	0.50	0.12	0.73	0.11
L	1.10	0.14	1.23	0.23	1.22	0.18	–	–	1.10	0.0

Note: E = group of users with early arrival on day t – 1. L = group of users with late arrival on day t – 1.

ily with the changes in intentions in response to experience with the facility. Table 9 shows the fraction of users in each sector who modified their anticipated arrival time at least n times, where n = 1,...,6. Comparing these data with Table 2 indicates that users are more prone to change actions before shifting intentions, as evidenced by the significantly fewer anticipated arrival-time changes. The same information is presented in Table 10 for each user group within Sectors 1, 2, and 3, respectively, thus confirming the general trend, discussed in conjunction with Table 3, that users with earlier preferred arrival times have to compromise less as the search progresses.

TABLE 9 Proportion of Users in Each Sector with at Least n Anticipated Arrival-Time Changes

No. of Changes[a]	Percentage of Users by Sector					
	1	2	3	4	5	All
1	75	75	80	30	35	59
2	50	60	25	5		28
3	20	30	10			12
4	15	10				5
5	15	5				4
6	10					2

[a]Minimum number.

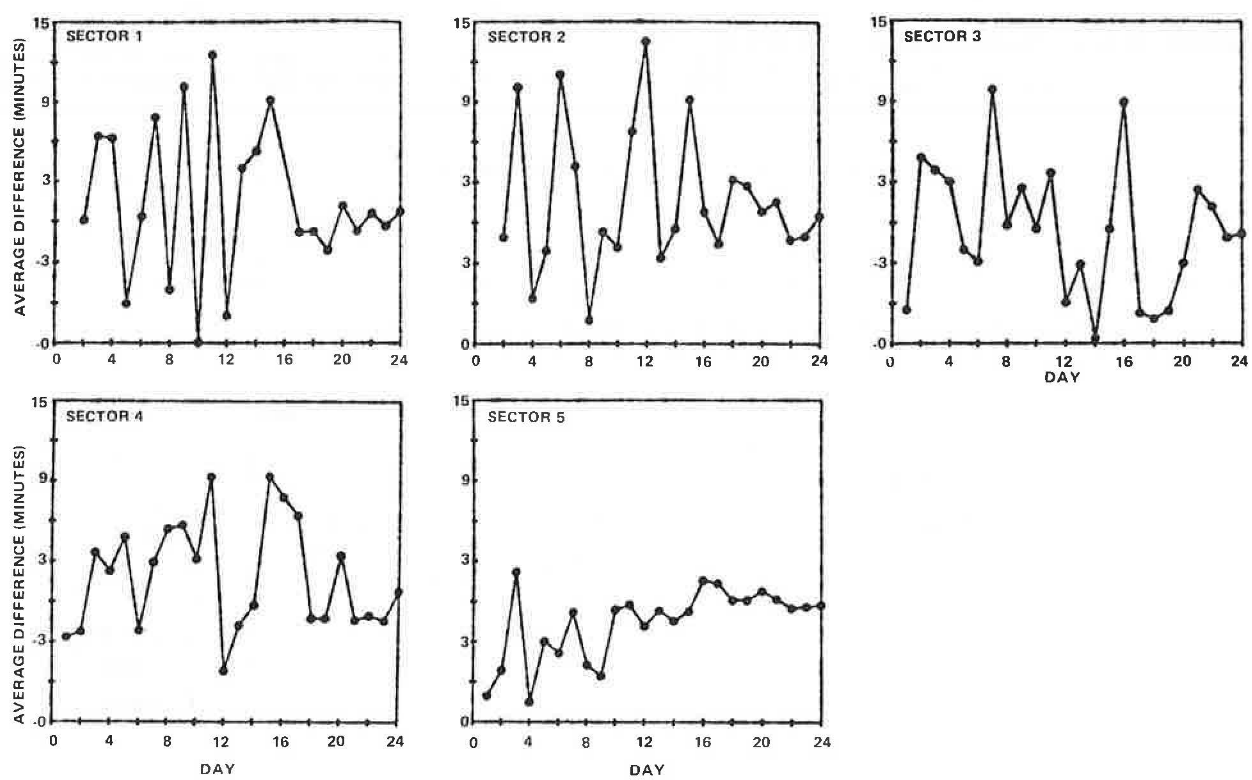

FIGURE 10 Day-to-day evolution of average difference between actual and anticipated travel time for each sector.

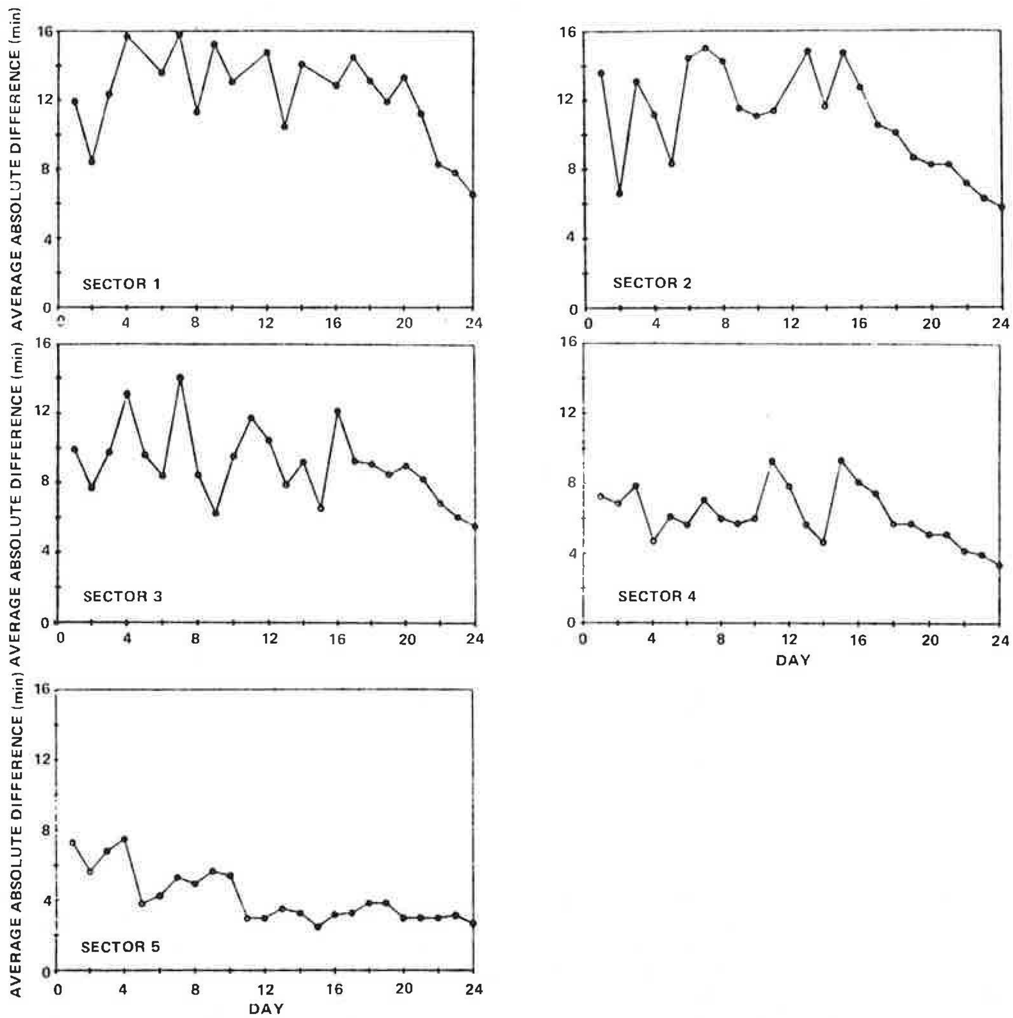

FIGURE 11 Day-to-day evolution of average absolute difference between actual and anticipated schedule delay for each sector.

Table 11 gives for each sector the mean number of days (and the standard deviation) since the previous revision for each of the n anticipated arrival-time changes, n = 1,...,6. Unlike Table 4, a clear decreasing trend is evident here, whereby users revise intentions at gradually smaller time intervals. As a matter of fact, the mean time until the first change is quite large, which indicates user persistence in initial intentions. However, as users progressively realize the inability to achieve their initial preference, and as they develop a better feel for the system's performance, they appear more willing to revise their anticipated arrival time. Increasingly, a number of participants updated their anticipated arrival time only after the system had reached steady state.

Table 12 presents the average number of departure-time changes (and the standard deviation) that took place since the previous revision for each of the six anticipated arrival-time revisions. The number of departure-time changes is an indication of the number of intervening unacceptable outcomes, and

TABLE 10 Proportion of Users in Each User Group Within Sectors 1, 2, and 3 with at Least n Anticipated Arrival-Time Changes

	Percentage of Users by Sector								
	1			2			3		
No. of Changes[a]	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
1	0	62.5	100	33.3	50.0	100	0	66.6	100
2		37.5	90.0		16.7	100		33.3	61.5
3		25.0	70.0		16.7	63.6		16.7	23.1
4		12.5	40.0			18.2		16.7	
5			30.0			9.1			
6			20.0			9.1			

[a]Minimum number.

TABLE 11 Mean and Standard Deviation of the Number of Days Between Consecutive Anticipated Arrival-Time Changes per Sector

Change Sequence No.	Sector 1 Mean	SD	2 Mean	SD	3 Mean	SD	4 Mean	SD	5 Mean	SD	All Users Mean	SD
1	7.62	4.12	6.53	4.42	9.63	5.32	14.52	6.98	9.14	3.92	8.77	5.21
2	4.54	2.99	5.08	3.68	6.89	5.13	6.00	–[a]			5.43	3.81
3	3.63	1.60	3.80	2.90	2.00	2.00					3.44	2.28
4	1.67	1.15	4.06	4.24							2.33	2.42
5	1.00	0.00	4.00	–[a]							1.75	1.55
6	1.50	0.72									1.50	0.72

[a]Only one participant was involved in this change.

TABLE 12 Mean and Standard Deviation of the Number of Departure-Time Changes Between the Consecutive Anticipated Arrival-Time Changes per Sector

Change Sequence No.	Sector 1 Mean	SD	2 Mean	SD	3 Mean	SD	4 Mean	SD	5 Mean	SD
No change	6.00	1.80	6.00	2.60	3.50	2.30	2.50	1.22	1.30	0.63
1	4.50	2.58	3.70	1.87	3.00	1.93	1.33	1.50	0.57	0.53
2	2.61	2.14	2.74	1.72	1.56	2.06	1.00	–[a]		
3	1.00	1.32	1.20	1.20	1.00	0.00				
4	0.80	0.83	1.50	0.70						
5	0.75	0.52	1.00	–[a]						
6	1.00	0.00								

[a]Only one participant was involved in this change.

is as such a measure of the number of failures until the next revision. As expected from the foregoing discussion, this number decreases as the search progresses, reflecting the initial resistance (to revising intentions), which appears to weaken progressively. Table 12 also reveals that users in closer sectors encounter fewer "failures," on average, than those in more distant sectors.

Finally, the direction of these readjustments is examined. Are users shifting their anticipated arrival to an earlier or a later time? And are they doing so consistently in one or the other direction? As suggested in the second section, users would tend to accept increasing earliness relative to their preferred arrival time in order to accommodate the fluctuations in system performance. This is indeed the case in this experiment, with 72 percent of all users who adjusted anticipated arrival time at least once consistently shifting to an earlier time. Only 7 percent consistently shifted to a later time (actually only two participants, both in Sector 4), with the remaining 21 percent moving at least once in each direction.

CONCLUSION

This paper has presented the principal elements of a theoretical framework to describe the processes governing commuters' daily departure-time decisions in response to experienced congestion patterns. Commuter behavior is viewed as a boundedly-rational search for an acceptable departure time. A key notion is that of an indifference band of tolerable schedule delay that determines the acceptability of a particular decision outcome on any given day. This indifference band, which varies across individuals, also shifts in response to users' experience with the facility.

Although not intended as a formal validation of the foregoing model, an experiment involving real commuters interacting daily with a hypothetical simulated traffic corridor was conducted over a period of 24 days, yielding valuable insights into the dynamics of the departure-time decision and its interaction with system performance. The results pertaining to the underlying behavioral processes were analyzed in this paper from the perspective of the key notions articulated in the conceptual framework.

Of course, this is only one such experiment, which involves obvious restrictions because of the hypothetical nature of the commuting corridor. Nevertheless, it has been quite insightful, particularly given the difficulty and the scale of corresponding real-world observations at the desired level of detail. As such, it offers a useful complementary approach to support the development of a comprehensive descriptive theory that would be subsequently validated, if only in part, in the field. Other experiments under different informational situations (e.g., where information about system congestion is available by word of mouth or through media reports) are also contemplated. In addition, formal mathematical model building and parameter estimation will be conducted.

ACKNOWLEDGMENTS

The contribution of Robert Herman to the overall research effort on which this paper is based is gratefully acknowledged, as are the many profitable discussions with him on various aspects of this problem.

Initial support for this work was provided by a grant from the Bureau of Engineering Research at the University of Texas at Austin. Computer support was provided, in part, by the Department of Civil Engineering. Further support for this work and its continuation is provided by a grant from the National Science Foundation.

REFERENCES

1. S. Rosenbloom. Peak-Period Traffic Congestion: A State-of-the-Art Analysis and Evaluation of Effective Solutions. Transportation, Vol. 7, 1978, pp. 167-191.
2. S. Cosslett. The Trip Timing Decision for Travel to Work by Automobile. In Demand Model Estimation and Validation, Urban Travel Demand Forecasting Project, Phase 1 Final Report, Vol. 5, Institute for Transportation Studies, University of California, Berkeley, 1977.
3. M. Abkowitz. Understanding the Effect of Transit Service Reliability on Work Travel Behavior. In Transportation Research Record 794, TRB, National Research Council, Washington, D.C., 1981, pp. 33-41.
4. K. Small. The Scheduling of Consumer Activities: Work Trips. American Economic Review, Vol. 72, 1982, pp. 467-479.
5. C. Hendrickson and E. Plank. The Flexibility of Departure Times for Work Trips. Transportation Research A, Vol. 18A, No. 1, 1984.
6. W.S. Vickrey. Congestion Theory and Transportation Investment. American Economic Review, Vol. 59, 1969, pp. 251-261.
7. J.V. Henderson. The Economics of Staggered Work Hours. Journal of Urban Economics, Vol. 9, 1981, pp. 349-364.
8. C. Hendrickson and G. Kocur. Schedule Delay and Departure Time Decisions in a Deterministic Model. Transportation Science, Vol. 15, No. 1, 1981, pp. 62-77.
9. C. Hendrickson, D. Nagin, and E. Plank. Characteristics of Travel Time and Dynamic User Equilibrium for Travel-to-Work. In Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, (V.F. Hurdle, E. Hauer, and G.N. Steuart, eds.), University of Toronto Press, 1983, pp. 321-347.
10. M.J. Smith. The Existence of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck. Presented at the International Symposium on Transportation Equilibrium and Supply Models, Montreal, Canada, 1981.
11. P.H. Fargier. Effects of the Choice of Departure Time on Road Traffic Congestion: Theoretical Approach. Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, (V.F. Hurdle, E. Hauer, and G.N. Steuart, eds.), University of Toronto Press, 1983, pp. 223-263.
12. A. De Palma, M. Ben-Akiva, C. Lefevre, and N. Litinas. Stochastic Equilibrium Model of Peak Period Traffic Congestion. Transportation Science, Vol. 17, No. 4, 1983.
13. M. Ben-Akiva, M. Cyna, and A. De Palma. Dynamic Model of Peak Period Congestion. Transportation Research B, Vol. 18B, No. 45, 1984, pp. 339-356.
14. M. Ben-Akiva, A. De Palma, and P. Kanaroglou. Analysis of Peak-Period Traffic Congestion with Elastic Demand. Presented at the Tenth Transportation Planning Research Colloquium, Netherlands, 1983.
15. H.S. Mahmassani and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. Transportation Science, Vol. 18, No. 4, 1984.
16. H.S. Mahmassani and G.-L. Chang. Experiments with Departure Time Choice Dynamics of Urban Commuters. Transportation Research Board (1986, in press).
17. A.S. Alfa. Modelling the Temporal Distribution of Peak Traffic Demand: State of the Art and Directions. Presented at the Spring Joint Annual Meeting of ORSA/TIMS, San Francisco, Calif., 1984.
18. J.L. Horowitz. The Stability of Stochastic Equilibrium in a Two-Link Transportation Network. Transportation Research B, Vol. 18B, No. 1, 1984.
19. H. Simon. A Behavioral Model of Rational Choice. Quarterly Journal of Economics, Vol. 69, 1955, pp. 99-118.
20. H.S. Mahmassani, G.-L. Chang, and R. Herman. Collective Effects and Individual Decisions in a Traffic System. Submitted to Transportation Science, 1984.
21. R. Herman and T. Lam. Trip Time Characteristics of Journeys to and from Work. In Proceedings of the Sixth International Symposium on Transportation and Traffic Flow, Sydney, Australia, 1974, pp. 57-85.
22. J.W. Payne. Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis. Organization Behavior and Human Performance, Vol. 16, 1967, pp. 366-387.
23. H.J. Einhorn and R.M. Hogarth. Behavioral Decision Theory: Processes of Judgment and Choice. Annual Review of Psychology, Vol. 32, 1981, pp. 53-88.
24. P. Slovic, B. Fischoff, and S. Lichtenstein. Behavioral Decision Theory. Annual Review of Psychology, Vol. 31, 1977, pp. 1-39.
25. R. Cardozo. An Experimental Study of Consumer Effort, Expectations and Satisfaction. Journal of Marketing Research, Vol. 2, 1965, pp. 240-249.
26. R.C. Oliver. Effect of Expectation and Disconfirmation on Post-Exposure Product Evaluations: An Alternative Interpretation. Journal of Applied Psychology, Vol. 62, 1977, pp. 480-486.
27. J.C. Olson and P.A. Dover. Disconfirmation of Consumer Expectation Through Product Trial. Journal of Applied Psychology, Vol. 64, 1979, pp. 179-189.
28. J.W. Thibout and H.H. Kelley. The Social Psychology of Groups. John Wiley and Sons, Inc., New York, 1959.
29. S. Latour and N. Peat. Conceptual and Methodological Issues in Satisfaction Research. Advances in Consumer Research, Vol. 6, 1979, pp. 431-437.
30. R.C. Oliver. A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. Journal of Marketing Research, Vol. 17, 1980, pp. 460-469.
31. H. Helson. Adaption Level Theory: An Experimental and Systematic Approach to Behavior. Harper and Row, New York, 1964.
32. P.A. Labarbera and D. Mazursky. A Longitudinal Assessment of Consumer Satisfaction/Dissatisfaction: The Dynamic Aspect of the Cognitive Process. Journal of Marketing Research, Vol. 20, 1983, pp. 393-404.
33. G.-L. Chang, H.S. Mahmassani, and R. Herman. A Macroparticle Traffic Simulation Model to Investigate Peak-Period Commuter Decision Dynamics. In Transportation Research Record 1005, 1985, pp. 107-121.

The authors are solely responsible for the content of this paper.

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

Transfer Model Updating with Disaggregate Data

FRANK S. KOPPELMAN, GEOK-KOON KUAH, and CHESTER G. WILMOT

ABSTRACT

Model transfer provides an alternative to undertaking complete data collection and model development in every planning context. The effectiveness of the transferred model in the application context can be improved by updating selected model parameters by using limited data from the application context. The effect of model updating on the transferability of disaggregate travel choice models both within and between urban areas is examined. It is found that transfer effectiveness improves with updating alternative specific constants and improves further with updating the parameter scale for both intraurban and interurban transfers. Further, the sample size necessary to obtain a substantial improvement in model transferability is a small fraction of that needed to estimate a complete model in the application context. Thus, it appears that model transfer with updating may be preferable to either full model transfer or new model estimation in situations of constrained resources.

The transfer of a previously estimated model to a new application context can reduce or eliminate the need for a large data collection and model development effort in the application context. However, the usefulness of a transferred model depends on the degree to which it can provide useful information about the behavior or phenomenon of interest in the application context.

Models are not perfectly transferable between contexts. Thus, the general objective of model transfer is to obtain a model that reasonably approximates the behavior in the application context. The quality of this approximation can be improved by using available information about the application context to modify or update some or all of the model parameters. A wide range of updating procedures can be employed depending on the type of information available in the application context. Schultz and colleagues have employed updating by using transit corridor volumes and screen-line counts in Houston, Seattle, and New Orleans (1-3).

This study examines the effect of updating alternative specific constants and the scale of the model parameters on the transferability of disaggregate mode-choice models that use disaggregate data. This approach may be employed to facilitate the analysis process when a small sample of disaggregate data has been collected. Such an updating sample can be considerably smaller than the sample that would be necessary to calibrate a new model system. The approach is demonstrated and evaluated for both intraregional and interregional transfer of disaggregate models of mode choice to work.

This paper is organized as follows. The sources of differential transferability of model components are identified and the procedure is described that is used for the adjustment of alternative specific constants and the scale of the transferred parameters. The research approach, including the data used, model specifications, and the model estimation results, is described next. Then the effect of model updating on intraregional and interregional model transferability is evaluated, and the final section sets forth conclusions and implications.

PARTIAL TRANSFERABILITY AND MODEL ADJUSTMENT

Model transfer is expected to be effective when the underlying individual travel choice decision process is the same in both the estimation and application contexts and the model specification is appropriate (4). Perfect transferability of models cannot be achieved because of behavioral differences between contexts and limitations in model specification. The behavioral differences and specification limitations may result in differential transferability of different model components. Updating procedures can be used to modify selected parameters of transferred models by incorporating available information about the application context.

Sources of Differential Transferability of Model Components

McFadden (5) and Westin and Manski (6) identify three types of differences that may exist in models between estimation and application contexts. These are differences in the alternative specific constants, in the sensitivity or scale of the model parameters, and in the relative values of variable coefficients. These differences in expected transferability result from the differential effect of model specification errors on these classes of model parameters.

The impact of model specification error for the multinomial logit model can be seen from a review of the model derivation (similar results can be obtained for the multinomial probit model). Consider a decision maker faced with the problem of selecting one of a set of available alternatives. It is assumed that the decision maker will select that alternative which has the highest utility to him or her. The utility U_{it} of an alternative i to an individual t includes deterministic (V_{it}) and random (ε_{it}) components:

$$U_{it} = V_{it} + \varepsilon_{it} \tag{1}$$

The derivation of the multinomial logit model is based on the assumption that the random components (ε_{it}) are independently and identically Gumbel distributed over individuals and alternatives. Further, the systematic or deterministic portion of the utility function is generally assumed to be linear in parameters so that

$$V_{it} = X_{it}\beta \tag{2}$$

where X_{it} is a row vector of variables describing individual t and alternative i, and β is a column vector of parameters. Under these assumptions, the multinomial logit model has the form

$$P_{it} = \exp\,[(\eta_i + X_{it}\beta)/\omega]/\sum_j \exp\,[(\eta_j + X_{jt}\beta)/\omega] \qquad (3)$$

This model has location parameters (η_i), which represent the mode of the distribution of errors for each alternative; a scale parameter (ω), which represents the variance of the distribution of the error terms; and attribute importance parameters (β), which represent the attribute weighting that the individual employs in evaluating alternatives.

Tardiff (7) shows that the omission of explanatory variables will shift the mean of the error distribution represented in the model by η_i, increase the variance of the error distribution represented by ω, and bias the estimates of parameters associated with included variables. When different contexts are compared that have similar behavior but incompletely specified models, it is expected that the differences in the mean values of the error distribution will be relatively large, the differences in the error distribution variance will be smaller, and the differences in behavioral parameters will be the smallest. Thus, efforts to improve the transferability of a model to a specific application environment should emphasize adjustment of alternative specific constants first, parameter scale second, and relative parameter values last. Empirical results confirm the importance of adjusting alternative specific constants by using disaggregate data to improve the transferability of disaggregate choice models (8,9,4). However, there is no reported study of the effect of scale adjustment on model transferability.

Analytic Formulation of Updating Procedures

The parameters in Equation 3 are not uniquely identified and therefore cannot all be estimated. First, the η parameters can only be identified up to an additive constant. This limitation is dealt with by imposing an arbitrary constraint on one of these parameters (e.g., set $\eta_k = 0$). Second, it is not possible to estimate ω but only to estimate the ratios η/ω and β/ω. Defining ratios of these parameters by $\mu_j = \eta_j/\omega$, and $\theta = \beta/\omega$ and restating the multinomial logit model in terms of these new parameters obtains

$$P_{it} = \exp\,[\mu_i + X_{it}\theta]/\sum_j \exp\,[\mu_j + X_{jt}\theta] \qquad (4)$$

where one of the μ_j is constrained to zero.

Updating procedures can be used to modify or replace selected parameters in this model. In this study, the effectiveness is examined of updating the location parameters (μ) and the scale of the remaining parameters by using a sample of individual observations from the application context.

Parameter estimates for a choice model are obtained with disaggregate data by maximizing a log likelihood expression of the form

$$L = \sum_t \sum_i \delta_{it} \ln P_{it}\,(X_t, \mu, \theta) \qquad (5)$$

where

δ_{it} = indicator variable set to 1 if individual t chooses alternative i and to 0 otherwise,

$P_{it}(X_t,\mu,\theta)$ = probability that individual t chooses alternative i, and

μ = vector of alternative specific constants.

Embedded in the probability function in Equation 5 are expressions for the deterministic component of utility for each alternative formulated as

$$V_{it} = \mu_i + X_{it}\,\theta \qquad (6)$$

The transfer of the parameters describing the effect of time, cost, and other variables on travel choice is based on some expected generality of these factors across estimation and application contexts. There is no comparable basis for transferring the constant terms because average differences in the excluded factors between contexts are expected. Therefore, it is appropriate to consider transferring the θ parameters in Equation 6 to the application context while obtaining a local estimate of the alternative specific constants. In this case, the θ-parameters transferred to the application context are denoted with a subscript T (θ_T) and the transferred portion of the utility function is defined as

$$Z^A_{it} = X^A_{it}\,\theta_T \qquad (7)$$

where X^A_{it} is a vector of attributes of alternative i for individual t in the application context. The updating of the alternative specific constants is accomplished by modifying the utility function in Equation 6 for the application context to

$$V^A_{it} = \mu^A_i + Z^A_{it} \qquad (8)$$

where

V^A_{it} = deterministic component of utility for alternative i in the application context,

μ^A_i = updated alternative specific constant for alternative i in the application context, and

Z^A_{it} = transferred portion of the utility function defined in Equation 7.

The estimate of the updated alternative specific constants (μ^A_i) consists of those values that maximize the log likelihood function:

$$L = \sum_t \sum_i \delta_{it} \ln P_{it}\,(Z^A_t, \mu^A) \qquad (9)$$

where Z^A_t is a vector of variables defined in Equation 7 for individual t in the application context for all alternatives and μ^A is a vector of alternative specific constants. The final utility function employed for transfer prediction becomes

$$V^A_{it} = \mu^A_i + X^A_{it}\,\theta_T \qquad (10)$$

which includes all the transferred slope parameters (θ_T) and locally estimated alternative specific constants (μ^A_i).

The methodology just outlined can be extended to adjust the scale of the transferred parameters as well as the alternative specific constants. The coefficient of Z^A_{it} in Equation 8 was restricted to 1 in the preceding approach. When the parameter scale is updated, that restriction is relaxed and a coefficient is estimated for Z^A_{it}. The deterministic component of utility becomes

$$V^A_{it} = \mu^A_i + \lambda^A Z^A_{it} \qquad (11)$$

where λ^A is the scaling parameter for the application context relative to the estimation context.

Updating the alternative specific constants and the parameter scale amounts to selecting values of μ^A and λ^A that maximize the log likelihood function:

$$L = \sum_t \sum_i \delta_{it} \ln P_{it} (Z_t^A, \mu^A, \lambda^A) \quad (12)$$

The scaling parameter (λ^A) adjusts the scale of the explanatory variables but does not affect their relative importance. The adjusted expression for alternative utility becomes

$$V_{it} = \mu_i^A + \lambda^A X_{it}^A \theta_T \quad (13)$$

which differs from Equation 10 only by inclusion of the scaling constant (λ^A).

Practical Application of Updating Procedures

The updating procedures described can be readily implemented in standard packages for logit model estimation. The common application of such procedures includes the selection of variables to be included in the choice model. To use the same package for model updating, it is necessary to formulate the composite variable Z_{it}^A by means of Equation 7 and estimate the new model with a full set of alternative specific constants. For updating alternative specific constants only, the parameter of the composite variable (Z_{it}^A) must be restricted to 1. For updating the alternative specific constants and the parameter scale, the parameter is unrestricted. This procedure can be employed with a disaggregate data set of any size for the application context. The same data can be used both to estimate parameter scale adjustment and to update the alternative specific constants. It is expected, and this empirical study confirms, that a substantially smaller data set can be used to obtain satisfactory estimates of these parameters than would be necessary to estimate the complete model in the application context.

RESEARCH DESIGN

The analysis undertaken in the previous section suggests that transferability will be enhanced by adjustment of alternative specific constants and parameter scale and describes procedures for making such adjustments. However, the qualitative analysis does not provide information about the importance of these adjustments on transferability. An empirical exploration of these impacts is undertaken to increase the understanding of the effectiveness of these adjustments.

The research approach is to evaluate the transferability of models of mode choice to work within a single urban region and between urban regions. The intraregional transfers are among sectors in the Washington, D.C., metropolitan area. The interregional transfers are among the metropolitan areas of Minneapolis-St. Paul, Baltimore, and Washington, D.C. Model transfer effectiveness is evaluated for full model transfer, model transfer with updating of alternative specific constants, and model transfer with updating of alternative specific constants and parameter scale.

Data

The intraregional transferability analysis is undertaken by using Washington, D.C., data for those who reported traveling to work in the central business district (CBD) by driving alone, shared ride, or transit. These records are grouped into three geographic sectors between which model transferability is evaluated. The interregional transferability analysis is undertaken among the regions of Minneapolis-St. Paul, Baltimore, and Washington, D.C. Differences among sectors in Washington represent real differences in sociodemographic characteristics and transportation service attributes. Differences among the three regions include these real differences as well as apparent differences due to inconsistencies in data collection procedures. Thus, the analysis of transferability within Washington and between regions provides some insight into the additional limits on transferability that may be attributable to differences in data collection and other conventions between study areas.

Model Specification

The specification employed in the Washington, D.C., intraregional transferability analysis includes three level-of-service variables, a car-per-driver variable applied separately to the drive-alone and shared-ride alternatives, and alternative specific constants. This model is in the mid-range of specifications analyzed for transfer effectiveness by Koppelman and Wilmot (10).

The specification employed in the three-city interregional transferability analysis includes all the variables used in the intraregional transfer study with the addition of a variable that measures automobile access time to transit for zones in which automobile must be used to reach transit. The variables included in each transferability analysis are identified and defined in Table 1.

TABLE 1 Variables Included in Analysis of Intraregional and Interregional Transferability

Variable	Study Type: Intraregional	Study Type: Interregional
Dummy for drive-alone alternative (DAD)	X	X
Dummy for shared-ride alternative (SRD)	X	X
Cars per driver for drive-alone alternative (CPDDA)	X	X
Cars per driver for shared-ride alternative (CPDSR)	X	X
Out-of-pocket cost divided by income[a] (OPTCINC)	X	X
Total travel time[a] (TVTT)	X	X
Out-of-vehicle travel time divided by distance[a] (OVTTD)	X	X
Automobile access time to transit for zones not served by transit (AATR)	–	X

[a]Level-of-service variables (OPTCINC, TVTT, and OVTTD) are based on the simple home-work-home tour for the Washington, D.C., intraregional analysis and on the one-way home-work trip for the interregional analysis. This difference in variable definition will modify the scale of these parameters by a factor of 2 but will have no other impact on estimation and transferability results.

Disaggregate updating for both intraregional and interregional transfers is undertaken by using all available disaggregate data in the application environment. The earlier study by Atherton and Ben-Akiva (8) used a subsample of the available data for updating the alternative specific constants. Koppelman and Chu (11) show that the use of the full sample rather than a subsample will improve the precision of the obtained estimators but will not affect their consistency.

Evaluation of Transferability With and Without Parameter Updating

Transfer effectiveness is evaluated by the degree to which the transferred model with or without updating predicts the observed behavior in the application environment. Four measures, formulated by Koppelman and Wilmot (4), are used to evaluate transfer predictive accuracy. The transfer likelihood ratio index, analogous to the commonly used likelihood ratio index (12) for evaluating model goodness of fit, describes the extent to which the transferred model explains observed individual behavior in the application environment. The transfer index, a ratio of the transfer likelihood ratio index and the local likelihood ratio index, describes the degree to which the transferred model describes observed behavior relative to an identically specified local model. The root-mean-square-error measure is an index of the average proportional error in prediction of aggregate travel shares by any alternative. The relative root-mean-square error is the ratio between this measure and the corresponding measure for an identically specified local model.

Each of these measures describes the transfer effectiveness of a single estimated model applied in another context. These measures can be pooled across multiple transfers (13) to provide an overall indication of the effectiveness of a specific type of transfer over multiple applications.

Estimation Results

The estimation results for the Washington, D.C., sectors and for three urban regions are reported in Tables 2 and 3, respectively, and the supporting statistics for Tables 2 and 3 are given in Tables 4 and 5, respectively. The signs of all the estimated parameters are consistent with a priori expectations. The parameters for cars per driver and total travel time are significant in all cases. The other level-of-service parameters are significant in some, but not all, cases.

There are important differences in goodness of fit among sectors in the Washington region and among regions measured by the likelihood ratio index with either the equal-share or market-share reference. That is, models of identical specification are more able to explain the travel choices made in some contexts than in others.

EVALUATION OF MODEL TRANSFERABILITY WITH DISAGGREGATE UPDATING

The transferability of the estimated models within the Washington, D.C., region and among the selected regions is evaluated by using the four measures described previously. These measures are pooled over the full sets of intraregional and interregional transfers to provide an average index of the effect of differences in updating procedures.

Disaggregate Transferability Measures

The pooled transfer likelihood ratio index values for intraregional and interregional transfers are as follows:

	Pooled Transfer Index Values	
Adjustment	Intraregional Transfers	Interregional Transfers
None	.092	.089
Constants	.101	.128
Constants and scale	.106	.136
Local estimation	.113	.167

These values indicate that the adjustment of alternative specific constants and the additional adjustment of scale produce a substantial improvement in model transferability. These adjustments result in transfer model goodness-of-fit values that are much closer to the corresponding local goodness-of-fit values than those for full model transfer without adjustment. The magnitude of improvement due to scale adjustment is somewhat smaller than that at-

TABLE 2 Mode-Choice Model Estimates for Washington, D.C., Sectors

	Sector 1		Sector 2		Sector 3	
Variable	Parameter Value	t-Statistic	Parameter Value	t-Statistic	Parameter Value	t-Statistic
DAD	−3.455	9.4	−2.018	5.9	−2.875	7.2
SRD	−1.937	9.6	−1.401	8.2	−1.382	5.3
CPDDA	4.181	11.3	3.191	9.2	3.647	5.0
CPDSR	1.964	7.1	1.743	8.3	1.544	5.0
OPTCINC	−0.0055	0.4	−0.0168	1.5	−0.0196	1.2
TVTT	−0.0423	7.0	−0.0148	3.2	−0.0229	4.7
OVTTD	−0.0276	0.5	−0.1029	1.7	−0.0281	0.4

TABLE 3 Mode-Choice Model Estimates for Three Urban Regions

	Minneapolis-St. Paul		Baltimore		Washington, D.C.	
Variable	Parameter Value	t-Statistic	Parameter Value	t-Statistic	Parameter Value	t-Statistic
DAD	−2.387	8.0	−0.815	2.5	−2.799	11.5
SRD	−1.351	4.9	−1.776	6.5	−1.688	12.5
CPDDA	3.017	9.9	2.313	6.3	3.478	14.4
CPDSR	1.048	3.8	2.004	6.4	1.694	10.1
OPTCINC	−0.0967	6.7	−0.0313	1.1	−0.0345	2.1
TVTT	−0.0595	9.1	−0.0159	2.3	−0.0558	8.0
OVTTD	−0.0961	4.7	−0.102	4.0	−0.130	1.6
AATR	−0.0701	1.2	−2.24	2.2	−0.129	3.8

TABLE 4 Supporting Statistics for Table 2

	Sector 1	Sector 2	Sector 3
No. of cases	944	964	746
No. of observations	2,648	2,583	2,165
Log likelihood			
At zero	−962	−933	−790
At market shares	−904	−898	−771
At convergence	−766	−813	−705
Likelihood ratio index (ρ^2)			
Equal-shares base	0.204	0.129	0.108
Market-shares base	0.153	0.095	0.086

TABLE 5 Supporting Statistics for Table 3

	Sector 1	Sector 2	Sector 3
No. of cases	2,000	785	2,000
No. of observations	5,814	2,416	5,568
Log likelihood			
At zero	−1,976	−767.4	−2,022
At market shares	−1,772	−713.4	−1,957
At convergence	−1,416	−556.6	−1,731
Likelihood ratio index (ρ^2)			
Equal-shares base	0.285	0.275	0.135
Market-shares base	0.202	0.220	0.116

tributable to the adjustment of alternative specific constants.

The values for the pooled transfer index for full model transfer and for partial model transfer with adjustment of alternative specific constants without and with scaling factors for both intraregional and interregional transfers are as follows:

	Pooled Transfer Index Values	
Adjustment	Intraregional Transfers	Interregional Transfers
None	.805	.533
Constants	.890	.767
Constants and scale	.948	.814

Both the adjustments in constants and parameter scale substantially improve transferability. The differences in transfer effectiveness between intraregional and interregional transfers presumably reflect regional differences in context similarity and in measurement procedures.

Both pooled disaggregate measures of transferability give a strong indication of the effectiveness of model updating. The pooled values indicate strong improvement obtained by adjustment of alternative specific constants and strong but smaller improvements by the further adjustment of parameter scale in both intraregional and interregional transfer. Examination of the context pair transfer measures (not reported here) indicates some variability in transfer effectiveness. However, those results still support the overall interpretation obtained by analysis of the pooled values. Disaggregate transfer effectiveness can be substantially improved by adoption of these adjustment procedures.

Aggregate Transferability Measures

The pooled root-mean-square errors for local estimation and transfer prediction are as follows for both types of transfer:

	Pooled Root-Mean-Square Errors	
Adjustment	Intraregional Transfers	Interregional Transfers
None	.277	.460
Constants	.248	.369
Constants and scale	.242	.340
Local estimation	.231	.321

The adjustment of the alternative specific constants substantially reduces the pooled values of root-mean-square error for both intraregional and interregional transfers. The additional adjustment of parameter scale produces a small additional reduction in the root-mean-square error in both cases.

Differences in the magnitude of the root-mean-square error values for both types of transfer are largely attributable to the size of the prediction sample in the aggregate groups used in the two studies. Specifically, the use of large groupings in the intraregional analyses results in smaller errors in aggregate prediction (14). Thus, these measures are not directly comparable.

The pooled relative aggregate transfer errors for intraregional and interregional transfers are as follows:

	Pooled Aggregate Root-Mean-Square Errors	
Adjustment	Intraregional Transfers	Interregional Transfers
None	1.186	1.433
Constants	1.074	1.149
Constants and scale	1.048	1.059

The aggregate prediction errors using transferred models are not substantially larger than those using local models for intraregional transfers but are much larger for interregional transfers. In both cases, the relative error is substantially reduced by adjustment of alternative specific constants with or without parameter scale adjustment.

Thus, the pooled aggregate measures of transferability are consistent with the disaggregate measures. However, in this case, the individual context pair transfers (not reported here) show greater variability in the effectiveness of the updating procedure. Nevertheless, updating of alternative specific constants consistently reduced the aggregate transfer error. However, the additional updating of parameter scale does, on some occasions, produce a small increase in aggregate transfer error.

Sample Size for Transfer Model Updating

It is useful to obtain some estimate of the sample size required for model updating relative to that which would be required for full model estimation. An initial estimate can be obtained by comparison between the standard errors of estimate of the alternative specific constants for full model estimation and those for model updating. The average estimation variance for both alternative specific constants over the three Washington, D.C., sector estimations and the six Washington, D.C., sector transfers is as follows:

Alternative Specific Constant	Full Model Estimation	Transfer Model Updating
Drive alone	0.132	0.010
Shared ride	0.046	0.009

These values indicate that when the full available sample is used, estimation precision is increased by a factor of 13.2 for the drive-alone constant and 5.1 for the shared-ride constant. This suggests that sample sizes for model updating could be one-fifth or less than the corresponding sample size for full model estimation.

Interpretation of Transfer Updating Tests

The updating of alternative specific constants produces a substantial improvement in model transferability for both intraregional and interregional applications with respect to all four pooled measures and for each context pair transfer. Thus, adjustment of alternative specific constants appears to be a universally desirable procedure. The additional updating of model scale produces a smaller average improvement in all four pooled measures but results in a small increase in the aggregate, but not disaggregate, error measures in some cases. Thus, although parameter scale updating appears to be generally desirable, it may result in poorer model performance in some contexts. Further, the sample size required for model updating appears to be substantially smaller than that required for full estimation of a model in the application context.

CONCLUSIONS

This study examines the effectiveness of updating procedures to enhance model transferability. The study is undertaken in contexts where adequate data are available to estimate local models. The results obtained are used here to make inferences about the use of updating procedures in the application of these or other models to new contexts in which there are limitations on the availability of survey data.

The results of this study indicate that full model transfer provides a substantial improvement over using market-share information only but is quite deficient relative to the estimation of a local model (average transfer indices of 0.53 for interregional transfer). The use of updating procedures substantially improves the expected level of model effectiveness (adjustment of alternative specific constants explains almost half of the deficiency with respect to local models and adjustment of parameter scale provides a small incremental increase in model effectiveness). Although there is some variability in the improvement attributable to model updating, every case examined showed a substantial improvement in transferability due to model updating.

It is useful to think about the effectiveness of model updating relative to the extreme options of full model transfer without updating and estimation of new models in the application context. The advantage of full model transfer is the elimination of the need to collect any data on traveler behavior. The advantage of new data collection and model development is to obtain the best possible model estimation results. There is a clear trade-off between cost savings and model effectiveness. Adding the option of model updating offers the potential for obtaining a large portion of the potential improvement in model effectiveness for a small portion of the increased cost. These results indicate that almost one-half of the difference between full transfer and local estimation can be obtained by updating alternative specific constants and more can be obtained by updating constants and parameter scale. However, the amount of data needed for updating is less than one-fifth of that needed for full model development. Thus, it appears that model updating is a desirable alternative to either full model transfer or new model development.

REFERENCES

1. Development and Calibration of Travel Demand Models for the Houston-Galveston Area. Draft Report. Barton-Aschman Associates, Inc., Evanston, Ill., Aug. 1979.
2. Development and Calibration of Mode Choice Models--Seattle Urban Region. Draft Report. Barton-Aschman Associates, Inc., Evanston, Ill., 1982.
3. Development of Travel Demand Model Set for the New Orleans Region. Barton-Aschman Associates, Inc., Evanston, Ill., Jan. 1983.
4. F.S. Koppelman and C.G. Wilmot. Transferability Analysis of Disaggregate Choice Models. In Transportation Research Record 895, TRB, National Research Council, Washington, D.C., 1982, pp. 18-24.
5. D. McFadden. Properties of the Multinomial Model. Working Paper 7617. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, 1976.
6. R. Westin and C.F. Manski. Theoretical and Conceptual Developments in Demand Modelling. In Behavioral Travel Modelling (D.A. Hensher and P.R. Stopher, eds.), Croom-Helm, London, 1979.
7. T.J. Tardiff. Specification Analysis for Quantal Choice Models. Transportation Science, Vol. 13, No. 3, 1979.
8. T.J. Atherton and M.E. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. In Transportation Research Record 610, TRB, National Research Council, Washington, D.C., 1976, pp. 12-18.
9. K.E. Train. The Sensitivity of Parameter Estimates to Data Specification in Mode Choice Models. Transportation, Vol. 7, No. 3, Sept. 1978.
10. F.S. Koppelman and C.G. Wilmot. The Effect of Model Specification Improvement on Transferability. Transportation Research, Vol. 20, No. 3, June 1986.
11. F.S. Koppelman and C. Chu. Effect of Sample Size on Disaggregate Choice Model Estimation and Prediction. In Transportation Research Record 944, TRB, National Research Council, Washington, D.C., 1984, pp. 60-69.
12. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973.
13. F.S. Koppelman and G. Rose. Geographic Transfer of Travel Choice Models: Evaluation and Procedures. In Optimization and Discrete Choice in Urban Systems (B. Hutchinson, P. Nijkamp, and M. Batty, eds.), Springer-Verlag, New York, 1985.
14. F.S. Koppelman. Aggregate Travel Prediction with Models of Individual Choice Behavior. Ph.D. dissertation. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, 1975.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.