In Survey of Current Business, U.S. Government Printing Office, Vols. 56-60, 1976-1980.
9. U.S. Department of Commerce. General Business Indicators: Monthly Series. In Survey of Current Business, U.S. Government Printing Office, Vols. 56-60, 1976-1980.
10. D.J. Pack. Computer Programs for the Analysis of Time Series Models Using the Box-Jenkins Philosophy. Automatic Forecasting Systems, Hatboro, PA, May 1977.
11. D.L. Greene. ORNL Highway Gasoline Demand Model: Volume V--The Aggregate Demand for Gasoline and Highway Passenger Vehicles in the United States: A Review of the Literature: 1938-1978. Energy Information Administration, U.S. Department of Energy, Draft Rept., 1980.

Impact of Travel Survey Sampling Error on Travel Demand Forecasting

CARMEN DIFIGLIO AND JAMES A. DISBROW

Alternative models of urban travel demand and the data used to estimate them are reviewed. The study focuses on the sampling error in origin-destination trip data and the impact that sampling error has on the estimation of a direct-demand travel model. High sampling errors in origin-destination trip data are found to significantly inhibit the performance of the direct-demand travel model.

The home-interview origin-destination (O-D) travel survey has been developed for most major metropolitan areas as a major data resource for the urban transportation planning process (UTPP). The Bureau of Public Roads, and later the U.S. Department of Transportation, provided funding for the UTPP and the home-interview O-D surveys that the UTPP required. The sample rates used in the home-interview surveys were typically less than 10 percent. The sample rate recommended for each urban area was based on total urban-area population, as given below (1):

Population	Sample of Households (%)
<50 000	20.0
50 000 to 150 000	12.5
150 000 to 300 000	10.0
300 000 to 500 000	6.7
500 000 to 1 000 000	5.0
>1 000 000	4.0

Larger cities (>500 000 population) were generally sampled at 4 or 5 percent.

A substantial amount of research was performed to guarantee that the chosen sampling strategy would be adequate for the UTPP models that the O-D data would be used to estimate. A major study by Sosslau and Brokke (2) showed that the chosen sampling strategy produced travel estimates that corresponded to screenline crossing data applicable to the corridor level. This level of aggregation corresponded to the UTPP model system typically used by local institutions that administered the UTPP [metropolitan planning organizations (MPOs)]. This model system consists of a series of sequential modeling steps: (a) trip generation, (b) trip distribution, (c) modal split, and (d) traffic assignment. Each of these models uses the available travel data in different ways. The trip-generation model uses data on the number of trip ends produced or attracted to an areal zone or district. The trip-distribution model is calibrated by using data aggregated to the corridor level (3). In modal split, the ratio of trips by highway or transit is the estimated variable. At no point in the conventional UTPP modeling process is the accuracy of the zone-to-zone or district-to-district O-D trip matrix ever a factor in model calibration or application. Consequently, the sampling error of the O-D trip matrix has never been examined.

The sequential modeling system used in the UTPP has a serious flaw. The trip-generation model has not typically responded to variables that characterize the transportation system (3). Since the endogenous variable of the trip-generation model is the number of trips produced by or attracted to a particular zone or district, changes in the transportation system have to be characterized in terms of how they affect the accessibility of the zone or district to all other zones or districts. Unfortunately, these accessibility measures have not been statistically significant variables in trip-generation models. When trip generation is insensitive to changes in transportation supply, the entire UTPP model process assumes that total travel demand is perfectly inelastic with respect to the quantity, quality, or cost of transportation services. This is not a novel observation but one that has been made before, as illustrated by the following quote from Wohl and Martin (4):

> [In] virtually every study this (calculation of trip ends by zone or district) has been accomplished independently from the travel conditions or the price of travel and with empirical observation of existing trip generation rates being used. Implicitly it has been assumed either that the price of travel will not change in the future compared with the present or that the demand for travel is entirely insensitive to the price of travel, i.e., that demand is perfectly inelastic.

In direct response to this deficiency in UTPP models, the direct-demand model was developed to make travel characteristics between zones or districts an important exogenous variable in determining not only the ratio of travel demand by automobile and transit but also the total number of trips. Direct-demand travel models accomplish this by integrating the three submodels (trip generation, trip distribution, and modal split) into a one-step model. This model has as its endogenous variable the demand for travel by a particular mode between origin district (or zone) i and destination district (or zone) j. Since in this direct-demand model the modal travel demand between any ij pair can be readily related to the travel characteristics of the modes connecting i and j, total travel becomes dependent on transportation supply. Consequently, total travel is not assumed to be inelastic with respect to the level and quality of available transportation.

Despite this considerable advantage, the one-step aggregate direct-demand model has not been used. A major application of this approach was made by Charles River Associates using data for Boston, Massachusetts (5). However, the level of explanatory power achieved by this model was low, and it did not inspire other urban-area applications of the method. It is unfortunate for the UTPP that the direct-demand model was not more successful, since it offered a practical approach to estimating the equilibrium level of travel that will occur after transportation investments have been made. The importance of developing an iterative equilibrium between travel demand models and networks was made clear by Wohl (6). Without such a process, there is no assurance that the travel characteristics assumed in estimating travel demand are consistent with those implied by the estimated level of travel demand and the capacity of the planned transportation system. Wohl calls for an iterative estimation of travel demand and network characteristics until the input travel time and cost entered into the last demand iteration are approximately the same as the output travel time and cost provided by the network (5). However, such a procedure is not feasible unless the total travel estimated by the travel demand model is sensitive to transportation characteristics and the model can be used expeditiously enough to engage in an iteration of travel forecasts.

The purpose of this paper is to reexamine the direct-demand model and, in particular, to evaluate whether the available data on which it must be estimated are adequate. Recall that the O-D matrix has not been used in any step of the sequential UTPP model process and, in addition, is not used by disaggregate models at all. Therefore, in comparison with other travel demand methods, the direct-demand model imposes a unique requirement on the accuracy of the O-D data.

SAMPLING-ERROR ANALYSIS OF TRAVEL DEMAND DATA

The sampling error inherent in home-interview O-D travel surveys can be analyzed in two distinctly different ways. The first is to estimate the confidence interval about a sample estimate of the number of trips produced or attracted to a district or zone. The second approach is to estimate the confidence interval about a sample estimate of the number of trips from a district (zone) to a district (zone). We will begin with the first approach for work trips.

Presume that work trips per household are, within each district, an independent random variable so that the sampled variables S_{hwhh1} (home-to-work trips made by household number 1 in district i), ..., S_{hwhhn} (household number n in district i) have the same distribution with mean S_{slis}/HH_{is}, variance σ^2, and moment generating function $M_x(t)$. If $n \to \infty$, the limiting distribution of

$$Z = [\overline{S}_{hwhh} - (S_{slis}/HH_{is})]/(\sigma/n^{1/2}) \tag{1}$$

is the standard normal distribution. If the population from which S_{hwhh} is sampled is normally distributed, Equation 1 is exact (the assumption that $n \to \infty$ and the limiting distribution are no longer required). The sampling distribution of $\overline{S}_{hwhh}$ is normal as long as the population is normal.

A confidence interval about the mean S_{slis}/HH_{is} is provided by Equation (inequality) 2:

$$\overline{S}_{hwhh} - z_a(\sigma/n^{1/2}) < (S_{slis}/HH_{is}) < \overline{S}_{hwhh} + z_a(\sigma/n^{1/2}) \tag{2}$$

where $1 - z_a$ is the probability that $(\overline{S}_{hwhh} - S_{slis}/HH_{is})n^{1/2}/\sigma$ will assume a value between $-z_a$ and z_a and that the integral of the standard normal density from z_a to ∞ is equal to a. Since σ in inequality 2 is generally unknown, the sample variance S^2 must be used to estimate a confidence interval for S_{slis}/HH_{is} =

$$\overline{S}_{hwhh} - t_{a,n-1}(S/n^{1/2}) < (S_{slis}/HH_{is}) < \overline{S}_{hwhh} + t_{a,n-1}(S/n^{1/2}) \tag{3}$$

where S^2 is the sample variance and $(\overline{S}_{hwhh} - S_{slis}/HH_{is})n^{1/2}/S$ has a t-distribution with (n - 1) degrees of freedom.

As will be shown below, the level of sampling error incurred to estimate a variable such as S_{slis}/HH_{is} or a similar characteristic variable for a district (e.g., automobiles per household in district i) is acceptably small when the sample rates recommended in the table given earlier are used in conducting home-interview travel surveys. Generally, sampling errors for such variables are not larger than ±5 percent for a 67 percent level of confidence or ±10 percent for a 95 percent level of confidence. Trip matrices, however, are an entirely different matter. If there are n districts, there are n^2 district-to-district interchanges. This increased disaggregation of trip record data will cause trip interchange data to have the largest sampling errors of all data collected in home-interview O-D surveys.

To analyze the sampling error of trip matrices derived from home-interview surveys, it is necessary to associate all trips with the district from which the trip maker's household was sampled. Consider trips that originate in district i and have the "home" trip purpose at the origin. The trip destinations may be any of the n districts in the study area. Let P_{ij} be the actual proportion of trips that begin in the home district i and end in the nonhome district j. Let $\hat{P}_{ij}$ be the proportion of trips sampled that began in the home district i and ended in the nonhome district j. If the total number of trips beginning in home district i is designated S_i and its sample estimate is designated $\hat{S}_i$, then

$$S_{ij} = S_i P_{ij} \tag{4}$$

and

$$\hat{S}_{ij} = \hat{S}_i \hat{P}_{ij} \tag{5}$$

Equation 4 is definitionally true. Equation 5 separates the statistical problem of estimating the sampling error of S_{ij} into two parts:

1. Estimating the total number of home-to-nonhome trips from district i (S_i) and
2. Estimating the distribution of destinations of S_i (P_{ij}).

Since the confidence interval about S_i is acceptably small, we can assume, with relatively little error, that $\hat{S}_i$ is known with certainty. Thus,

$$\hat{S}_{ij} = S_i \hat{P}_{ij} \tag{6}$$

The sampling statistic $\hat{P}_{ij}$ is based on N_i trip observations drawn from district i (home district), of which X_{ij} trips have their destination in district j:

$$\hat{P}_{ij} = X_{ij}/N_i \quad (7)$$

The probability of selecting X_{ij} trips out of the N_i trips sampled is constant and equal to P_{ij}. Therefore, $\hat{P}_{ij}$ is drawn from a binomial distribution with the single parameter P_{ij}. If N_i is large, the statistic $(X_{ij} - N_iP_{ij})/\sqrt{N_iP_{ij}(1 - P_{ij})}$ can be treated as if it were a random variable that has the standard normal distribution. A (1 - 2a) confidence interval for P_{ij} is

$$\left(\{X_{ij} + \tfrac{1}{2}z_a^2 - z_a\sqrt{[X_{ij}(N_i - X_{ij})/N_i] + \tfrac{1}{4}z_a^2}\}/(N_i + z_a^2)\right) < P_{ij} < \{X_{ij} + \tfrac{1}{2}z_a^2 + z_a\sqrt{[X_{ij}(N_i - X_{ij})/N_i] + \tfrac{1}{4}z_a^2}\}/(N_i + z_a^2) \quad (8)$$

From Equations 4 and 5,

$$\hat{S}_i\left(\{X_{ij} + \tfrac{1}{2}z_a^2 - z_a\sqrt{[X_{ij}(N_i - X_{ij})/N_i] + \tfrac{1}{4}z_a^2}\}/(N_i + z_a^2)\right) < S_{ij} < \hat{S}_i\left(\{X_{ij} + \tfrac{1}{2}z_a^2 + z_a\sqrt{[X_{ij}(N_i - X_{ij})/N_i] + \tfrac{1}{4}z_a^2}\}/(N_i + z_a^2)\right) \quad (9)$$

Inequality 9 represents a (1 - 2a) confidence interval for S_{ij} and can be used to evaluate the sampling error of trip matrices derived from home-interview O-D surveys.

SAMPLING-ERROR ANALYSIS OF MWCOG TRAVEL-DEMAND DATA

The Metropolitan Washington Council of Governments (MWCOG) undertook an inventory of transportation supply and demand for the Washington, D.C., metropolitan area in 1968. The MWCOG home-interview O-D survey was designed to survey a 5 percent sample of households living inside the Capital Beltway (I-495, the circumferential highway surrounding Washington, D.C., and a major portion of the Maryland and Virginia suburbs) and a 3 percent sample of households living outside the beltway but within the study area (7). Approximately 30 000 households were interviewed to achieve these sample rates. The interviews provided information about the trip-making activity of approximately 100 000 individuals. This sample of trip information is adequate to reliably determine important trip-making characteristics of each district in the study area (there are 134 districts within the study area). For example, the 67 percent confidence interval about the number of work trips per household (Equation 3) varies between 6 percent in the district with the smallest number of sample observations (115 households) to 3.5 percent in the district with the largest number of sample observations (220 households) (8).

To determine the reliability of the MWCOG home-interview survey for constructing matrices of trips from origins to destinations, inequality 9 was calculated for 12 combinations of trip mode and trip purpose (see Table 1). In using the notation given in Table 1, the confidence-interval inequality is

$$S_{mpis}\left(\{X_{mpij} + \tfrac{1}{2}z_a^2 - z_a\sqrt{[X_{mpij}(N_{mpis} - X_{mpis})/N_{mpis}] + \tfrac{1}{4}z_a^2}\} \div (N_{mpis} + z_a^2)\right) < S_{mpij} < S_{mpis}\left(\{X_{mpij} + \tfrac{1}{2}z_a^2 + z_a\sqrt{[X_{mpij}(N_{mpis} - X_{mpis})/N_{mpis}] + \tfrac{1}{4}z_a^2}\}/(N_{mpis} + z_a^2)\right) \quad (10)$$

where m is travel mode (m = 1, transit passenger trip; m = 2, automobile driver trip; m = 3, automobile passenger trip; and m = 4, automobile person trip) and p is travel purpose (p = 1, home to work; p = 2, home to shop; and p = 3, home to other).

Since there are 134 MWCOG internal districts, there are 17 956 confidence intervals for each mode-purpose combination. Therefore, inequality 10 involves 215 472 confidence intervals for the 12 sets of mode-purpose combinations (m = 1, ..., 4 and p = 1, ..., 3). Obviously, this much information cannot be presented or, if presented, cannot be comprehended. In addition, there are other useful ways of using the confidence-interval values (a lower and upper bound on S_{mpij}) that increase the amount of information that should be reported. Therefore, the following aggregation system was used to evaluate the data derived from inequality 10. The lower-bound (= LS_{mpij}) and upper-bound (= US_{mpij}) statistics derived from inequality 10 were aggregated over all trip interchanges from each district of origin and were also aggregated over all trip interchanges to each district of destination. This reduces the number of lower- and upper-bound statistics to 134 origin aggregations per mode-purpose combination and 134 destination aggregations, or 3216 aggregations for all 12 mode-purpose combinations.

Since the importance of LS_{mpij} and US_{mpij} is the range of confidence they provide about the sample estimate of S_{mpij}, the following statistics involving LS_{mpij} and S_{mpij} or US_{mpij} and S_{mpij} were calculated based on aggregations over all trip interchanges from each district of origin and aggregations over all trip interchanges to each district of destination:

Low $(S_{mpij} - LS_{mpij})$ for all $S_{mpij} \neq 0$
Mean $(S_{mpij} - LS_{mpij})$ for all $S_{mpij} \neq 0$
Mean $(S_{mpij} - LS_{mpij})$ for all S_{mpij}
Low $(S_{mpij} - LS_{mpij})/S_{mpij}$ for all $S_{mpij} \neq 0$
Mean $(S_{mpij} - LS_{mpij})/S_{mpij}$ for all $S_{mpij} \neq 0$
High $(US_{mpij} - S_{mpij})$ for all $S_{mpij} \neq 0$
High $(US_{mpij} - S_{mpij})$ for all $S_{mpij} = 0$
Mean $(US_{mpij} - S_{mpij})$ for all $S_{mpij} \neq 0$
Mean $(US_{mpij} - S_{mpij})$ for all S_{mpij}
Mean $(US_{mpij} - S_{mpij})$ for all $S_{mpij} = 0$
High $(US_{mpij} - S_{mpij})/S_{mpij}$ for all $S_{mpij} \neq 0$
Mean $(US_{mpij} - S_{mpij})/S_{mpij}$ for all $S_{mpij} \neq 0$

Table 1. Trip-record sets selected for sampling-error analysis.

Record Set	Trip Mode	Trip Purpose	Sampled Home Trip Ends	Sampled Trips	Home-Trip-End Variable	Trip Variable
1	Transit passenger	Home to work	N_{11is}	X_{11ij}	S_{11is}	S_{11ij}
2	Automobile driver	Home to work	N_{21is}	X_{21ij}	S_{21is}	S_{21ij}
3	Automobile passenger	Home to work	N_{31is}	X_{31ij}	S_{31is}	S_{31ij}
4	Automobile person	Home to work	N_{41is}	X_{41ij}	S_{41is}	S_{41ij}
5	Transit passenger	Home to shop	N_{12is}	X_{12ij}	S_{12is}	S_{12ij}
6	Automobile driver	Home to shop	N_{22is}	X_{22ij}	S_{22is}	S_{22ij}
7	Automobile passenger	Home to shop	N_{32is}	X_{32ij}	S_{32is}	S_{32ij}
8	Automobile person	Home to shop	N_{42is}	X_{42ij}	S_{42is}	S_{42ij}
9	Transit passenger	Home to other	N_{13is}	X_{13ij}	S_{13is}	S_{13ij}
10	Automobile driver	Home to other	N_{23is}	X_{23ij}	S_{23is}	S_{23ij}
11	Automobile passenger	Home to other	N_{33is}	X_{33ij}	S_{33is}	S_{33ij}
12	Automobile person	Home to other	N_{43is}	X_{43ij}	S_{43is}	S_{43ij}

Although these statistics cannot be easily listed for all combinations of trip mode and trip purpose for which they were calculated and for all of the 134 origin districts and 134 destination districts, it is instructive to list these statistics for the most important and numerous types of trips between the most-traveled interchanges in the Washington area. For this purpose, home-to-work trips by transit passengers (S_{11ij}) and automobile drivers and passengers (S_{41ij}) were selected. In order to select a few districts that contain the bulk of these trips, the destination-based aggregations to the 14 districts that constitute the Washington central business district (CBD) are listed in Tables 2-7. All confidence intervals are based on a z_a value of 1.96 providing a 95 percent level of confidence. The average upper-bound and high upper-bound statistics are given in Table 2 (transit trips) and Table 5 (highway trips). The average lower-bound statistics are given in Table 3 (transit trips) and Table 6 (highway trips). These statistics are expressed in absolute terms (number of trips) and as a percentage of the sample estimate of S_{mpij}. A high observation represents either the largest observation of ($US_{mpij} - S_{mpij}$) or ($US_{mpij} - S_{mpij})/S_{mpij}$ observed from all origin districts (i) to each of the 14 destination districts (j) included in Tables 2 and 5. Likewise, the mean of ($US_{mpij} - S_{mpij}$) and of ($US_{mpij} - S_{mpij})/S_{mpij}$ is included in Tables 2 and 5. These averages represent the mean observation from all origin districts (i) to each of the 14 CBD destination districts (j). In order to discuss all of the 14 CBD destination districts as a group, several averages over all districts for each statistic will be used. Thus, "average mean value of ($US_{11ij} - S_{11ij})/S_{11ij}$" will refer to the unweighted average of the last column of numbers in Table 2.

The 14 CBD districts account for a substantial portion of all home-to-work travel in the Washington, D.C., study area. The 14 CBD destination districts receive 55 percent of all transit home-to-work trips and 24 percent of all automobile home-to-work trips. On average, the CBD districts receive transit home-to-work trips from 51.7 origin districts (maximum = 81 for destination district 5 and minimum = 18 for destination district 6) and receive automobile home-to-work trips from 80.1 origin districts (maximum = 100 for destination districts 1 and 13 and minimum = 25 for district 10). Thus, the CBD districts are connected to far more origin districts than the remaining non-CBD destination districts.

Despite the high population of trips represented in these CBD destination district summaries, the estimated error statistics are very high. For example, the average mean upper bound ($US_{11ij} - S_{11ij}$) for transit trips is 246 percent, and the average high upper bound for transit trips is 439 percent. Furthermore, these percentage errors represent only the observations of S_{11ij} that are nonzero, since a percentage error cannot be defined for estimates of S_{11ij} that equal zero. The average upper bound for the zero estimates of S_{11ij} is equal to 56.6 trips (95 percent confidence interval). The average maximum upper bound for zero transit trip estimates within each CBD destination district is 182.6 trips. Consequently, the number of cells in the trip matrix that were estimated to contain no trips is likely to be overestimated by a significant factor.

The lower-bound statistics for transit trips are no more reassuring (Table 3). Since negative trip observations are not possible (i.e., the probability of X_{ij} observations cannot be below zero), the maximum lower-bound percentage is 100 percent. The mean lower-bound error averaged over the 14 CBD destination districts for transit passenger home-to-work trips is 71 percent (95 percent confidence interval). Thus, the mean lower-bound error is typically only 29 percent better than having no lower-bound confidence at all.

The error statistics for highway trips show a similar pattern. The average mean percentage upper bound is 241 percent, and the average high observation is 463 percent. The average mean upper bound is 93.6 trips for zero estimates of S_{41ij} and 191.3 trips for nonzero estimates of S_{41ij}. Lower-bound statistics indicate an average mean lower bound of 67 percent and a corresponding average of 74.4 trips.

The error indicated by these sampling-error statistics significantly reduces the accuracy that can be achieved by using a direct-demand model. Any model is attempting to explain the variation of one or more endogenous variables. If the endogenous variable itself possesses sampling error of the magnitude documented for the S_{mpij} variables, there will be an upper limit on the amount of variation that even a perfect demand model can explain. To the extent that noise in the data increases unexplained variance, it also increases the standard error about the parameter estimates.

IMPACT OF TRAVEL SURVEY SAMPLING ERROR ON ESTIMATION OF DIRECT-DEMAND TRAVEL MODEL

The direct estimation of urban travel between specific interchanges is an appealing approach to fulfilling urban transportation planning requirements. It can provide estimates of induced travel brought about by transportation improvements, be sensitive to O-D transportation characteristics, and be relatively easy to use. All of these characteristics would make possible an equilibration of transportation demand and transportation supply in the transportation planning process (6). Direct-demand models also avoid the problem of aggregation bias, since travel forecasts are made at the same level of aggregation at which behavioral parameters are estimated.

The direct-demand travel model was first developed by Quandt and Baumol (9) for interurban travel analysis and by Charles River Associates (5) for intraurban travel. The latter model is the single urban application of the direct-demand travel model. Regional planning agencies, consultants, and academic researchers have used either aggregate-sequential models as evolved in the UTPP or models developed from disaggregate travel demand data. The UTPP-type models are insensitive to transportation supply to the point that the elasticity of total travel demand to transportation system characteristics is structurally confined to be zero (4). Disaggregate models have varying levels of sensitivity to transportation system characteristics but are difficult to use in such a way as to achieve an equilibrium between assumed zone-to-zone travel characteristics and travel characteristics implied by the allocation of forecast travel on the highway and transit networks. The direct-demand travel model, unlike any other type of travel demand model, has its parameters estimated by fitting to a single endogenous variable: the number of trips of a specific mode and purpose that have origin district i and destination district j for all districts or a subset of districts in the study area. Consequently, sampling errors in the sampled matrix of trips have the most direct impact on a direct-demand travel model.

Two alternative functional forms of the direct-demand model were tested:

Table 2. Upper-bound-based statistics for 95 percent confidence interval estimates for transit passenger home-to-work trips with CBD destination.

CBD Destination District	High (US − S) for All S ≠ 0	High (US − S) for All S = 0	Mean (US − S) for All S	Mean (US − S) for All S = 0	Mean (US − S) for All S ≠ 0	High [(US − S)/S] for All S ≠ 0	Mean [(US − S)/S] for All S ≠ 0
1	441	189	118.0	39.8	181.4	4.294	1.879
2	437	189	109.4	50.1	165.3	4.333	2.188
3	462	181	119.1	35.8	179.0	4.121	1.574
4	363	152	110.4	47.2	179.5	4.275	2.047
5	453	189	111.9	44.7	168.0	4.379	2.021
6	447	152	119.3	36.5	173.5	4.019	1.646
7	250	189	89.4	75.5	178.5	4.596	3.935
8	250	189	91.5	72.6	182.6	4.459	3.337
9	322	189	96.6	70.9	198.2	4.439	2.650
10	217	189	87.9	78.8	172.1	4.579	3.661
11	474	189	105.4	63.8	199.7	4.439	2.332
12	269	189	94.0	68.0	159.6	4.585	3.079
13	415	181	10.5	52.2	180.3	4.348	2.177
14	336	189	105.5	56.2	159.3	4.426	1.972

Table 3. Lower-bound-based statistics for 95 percent confidence interval estimates for transit passenger home-to-work trips with CBD destination.

CBD Destination District	Low (S − LS) for All S ≠ 0	Mean (S − LS) for All S ≠ 0	Mean (S − LS) for All S	Low [(S − LS)/S] for All S ≠ 0	Mean [(S − LS)/S] for All S ≠ 0
1	14	85.6	47.3	0.311	0.653
2	14	68.6	35.3	0.337	0.701
3	14	90.4	52.6	0.312	0.633
4	19	74.1	35.4	0.377	0.681
5	14	73.2	39.9	0.301	0.682
6	17	83.0	50.2	0.306	0.653
7	26	39.8	5.3	0.610	0.795
8	26	47.9	8.2	0.610	0.770
9	14	71.3	14.4	0.423	0.706
10	20	40.5	3.9	0.608	0.790
11	17	85.1	26.0	0.284	0.679
12	17	45.3	12.9	0.539	0.770
13	17	76.4	34.8	0.335	0.681
14	19	64.5	30.8	0.425	0.692

Table 4. Miscellaneous confidence interval statistics for transit passenger home-to-work trips with CBD destination.

CBD Destination District	Mean S for All S ≠ 0	Number of S ≠ 0	High S	(US − S) for High S	(S − LS) for High S
1	167.1	74	1010	414	314
2	124.4	69	912	437	307
3	179.4	78	1085	462	338
4	129.0	64	586	340	221
5	132.5	73	1088	453	328
6	159.7	81	1049	447	321
7	52.6	18	164	250	100
8	65.5	23	164	250	100
9	120.4	27	456	322	193
10	53.6	13	148	217	90
11	173.7	41	1244	474	353
12	63.8	38	233	269	126
13	139.1	61	855	415	286
14	104.4	64	466	336	198

$$S_{mpij} = aPC_{is}^{b} AC_{sj}^{c} T_{1ij}^{d} T_{2ij}^{e} C_{1ij}^{f} C_{2ij}^{g} \quad (11)$$

and

$$S_{mpij} = aPC_{is}^{b} AC_{sj}^{c} \exp(dT_{1ij}) \exp(eT_{2ij}) \exp(fC_{1ij}) \exp(gC_{2ij}) \quad (12)$$

where

S_{mpij} = trips by mode m for purpose p between origin district i and destination district j,

PC_{is} = production characteristic variable for origin district i,

AC_{sj} = attraction characteristic variable for destination district j,

T_{1ij}, T_{2ij} = travel time by mode m = 1 or m = 2 between origin district i and destination j, and

C_{1ij}, C_{2ij} = travel cost by mode m = 1 or m = 2 between origin district i and destination j.

However, in all subsequent regressions tested, the ln-linear form (Equation 11) performed as well as or better than the negative exponential cost function (Equation 12). Therefore, results below are only reported for the full ln-linear form of the direct-demand model.

The production and attraction characteristic variables identified in Equations 11 and 12 depend on the purpose of the travel that is estimated. For the home-to-work trip purpose, the MWCOG data base provides the labor force in the trip origin district (L_{is}) and total employment in the trip destination district (TE_{sj}). For the home-to-shop trip purpose, the following variables are available: number of households (HH_{is}) in the trip origin district and shop and service employment (SSE_{sj}) in the trip destination district. In addition, the level of automobile ownership and mean household income in the origin district can potentially affect home origin trip demand and are available in the MWCOG data base (AOR_{is} and HHI_{is}, respectively). All trip production and attraction characteristic variables are defined below in terms of whether they apply to the home-trip origin district, work-trip destination district, or shopping-trip destination district:

L_{is} = number of employed persons who reside in trip origin district i (home-to-work trips);

Table 5. Upper-bound-based statistics for 95 percent confidence interval estimates for automobile person home-to-work trips with CBD destination.

CBD Destination District	High (US − S) for All S ≠ 0	High (US − S) for All S = 0	Mean (US − S) for All S	Mean (US − S) for All S = 0	Mean (US − S) for All S ≠ 0	High [(US − S)/S] for All S ≠ 0	Mean [(US − S)/S] for All S ≠ 0
1	419	154	180.0	81.3	213.6	4.438	1.682
2	362	177	167.3	88.4	197.3	4.682	2.060
3	403	172	183.9	82.9	218.3	4.591	1.708
4	334	177	153.2	92.9	190.3	4.639	2.423
5	380	182	160.6	90.2	190.6	4.607	2.343
6	378	182	183.7	91.5	195.6	4.682	2.077
7	274	182	128.7	103.0	166.8	4.682	3.530
8	260	172	131.6	102.4	168.7	4.682	3.079
9	310	182	139.7	101.3	182.9	4.657	2.749
10	264	182	121.3	107.6	180.8	4.649	3.252
11	439	182	172.4	80.8	206.1	4.682	2.131
12	337	182	146.2	98.2	178.5	4.682	2.581
13	437	154	172.1	88.0	200.7	4.586	1.991
14	307	182	150.4	101.4	187.7	4.537	2.096

Table 6. Lower-bound-based statistics for 95 percent confidence interval estimates for automobile person home-to-work trips with CBD destination.

CBD Destination District	Low (S − LS) for All S ≠ 0	Mean (S − LS) for All S ≠ 0	Mean (S − LS) for All S	Low [(S − LS)/S] for All S ≠ 0	Mean [(S − LS)/S] for All S ≠ 0
1	17	100.9	75.3	0.354	0.590
2	17	83.9	60.7	0.373	0.633
3	18	106.6	79.6	0.304	0.581
4	17	71.7	44.4	0.426	0.672
5	20	75.0	52.6	0.317	0.661
6	17	82.9	57.6	0.350	0.638
7	17	43.2	17.4	0.537	0.763
8	17	48.0	21.1	0.567	0.740
9	17	62.3	29.3	0.438	0.701
10	17	50.6	9.4	0.538	0.747
11	17	92.6	67.7	0.309	0.631
12	17	63.8	38.1	0.401	0.694
13	17	88.1	65.8	0.313	0.622
14	18	71.5	40.6	0.437	0.651

Table 7. Miscellaneous confidence interval statistics for automobile person home-to-work trips with CBD destination.

CBD Destination District	Mean S for All S ≠ 0	Number of S ≠ 0	High S	(US − S) for High S	(S − LS) for High S
1	205.0	100	794	419	281
2	157.0	97	597	362	228
3	228.3	100	738	383	257
4	125.9	83	458	334	195
5	140.0	94	845	380	268
6	156.5	93	672	378	245
7	61.5	54	244	274	131
8	69.5	59	204	260	117
9	103.0	63	406	310	178
10	73.6	25	221	256	119
11	196.4	98	1008	439	311
12	106.5	80	489	337	206
13	177.0	100	970	433	304
14	122.1	76	399	301	175

HH_{is} = number of households in trip origin district i (home-to-shop trips);

AOR_{is} = automobile ownership rate in home-trip origin district i = A_{is}/L_{is} for home-to-work trips and A_{is}/HH_{is} for home-to-shop trips, where A_{is} is the number of automobiles available to households in district i;

HHI_{is} = mean household income in trip origin district i;

TE_{sj} = total employment in work-trip destination district j; and

SSE_{sj} = shop and service employment in shopping-trip destination district j.

All variables tested in the direct-demand model are summarized below:

S_{11ij} = transit passenger home-to-work trips from origin district i to destination district j,

S_{41ij} = automobile person home-to-work trips from origin district i to destination district j,

S_{12ij} = transit passenger home-to-shop trips from origin district i to destination district j,

S_{42ij} = automobile person home-to-shop trips from origin district i to destination district j,

HH_{is} = total number of households in the district of origin i

L_{is} = labor force in the district of origin i,

A_{is} = total number of automobiles available to all households in the district of origin i,

HHI_{is} = mean household income in the district of origin i,

TE_{sj} = total employment in the district of destination j,

SSE_{sj} = shop and service employment in the district of destination j,

T_{1ij} = total transit travel time between districts i and j,

T_{2ij} = total highway travel time between districts i and j,

C_{1ij} = transit fare between districts i and j, and

C_{2ij} = highway cost between districts i and j (average parking cost in j).

Equations 11 and 12 were tested for the endogenous variables S_{11ij}, S_{41ij}, S_{12ij}, and S_{42ij} by using the stepwise regression package from the Biomedical Computer Program package, version BMDP

(10). The data set containing all of the variables included in the list above was combined with the BMDP stepwise regression model in an interactively controlled program. This program included the ability to screen available observations of all variables in the data set used in each equation on the basis of the value of each observation of the endogenous variable. This feature allows the elimination of interchanges that contain no observed trips or a small number of estimated trips, which reduces the number of observations included in the regression from the maximum of 17 822 observations. In addition, a second screening procedure was included that allowed observations that have large confidence intervals associated with the sample estimate of the endogenous variable S_{mpij} to be eliminated from the regression. Two sampling-error statistics--the percentage upper bound and the percentage lower bound--were selected to screen observations of S_{mpij} before the entering of the regressions used to estimate direct-demand model parameters. Recall that US_{mpij} was defined to be the upper limit (95 percent confidence) of the estimate of S_{mpij} and that LS_{mpij} was defined to be the lower limit of the estimate of S_{mpij}. The percentage upper and lower bounds for any observation of S_{mpij} are $(US_{mpij} - S_{mpij})/S_{mpij}$ for all $S_{mpij} \neq 0$ and $(S_{mpij} - LS_{mpij})/S_{mpij}$ for all $S_{mpij} \neq 0$.

The calculation of US_{mpij} and LS_{mpij} is based on Equation 10, where the lower bound of the inequality is defined to be LS_{mpij} and the upper bound is defined to be US_{mpij}. The range of values that $(S - LS)/S$ can take is 0 to 1 (or 100 percent), and the range of values that $(US - S)/S$ can take is 0 to ∞. For this reason, the screening process was designed to accept a criterion for both the lower and the upper bound independently of each other. Since the statistics used to screen observations are not defined for observations estimated to be zero ($S_{mpij} = 0$), all zero observations were also excluded when sampling-error criteria were applied to regression observations.

The basic strategy was to estimate the direct-demand model with three sets of data. The first set of data includes all of the trip interchanges regardless of whether trips were actually sampled in particular cells of the trip matrix for the travel mode and purpose being estimated. There are 17 822 observations in this data set (134 origin districts multiplied by 133 destination districts). The second set of data included observations for which trips were sampled. All trip interchanges (ij pairs) that had zero trips were eliminated from the regression. This reduced the number of observations from 17 822 to 1524 for transit work trips, 5168 for automobile work trips, 161 for transit shopping trips, and 1349 for automobile shopping trips.

A third set of data included trip interchanges for which the estimated sampling error was above a prescribed minimum. For the minimum sampling error allowed for work trips, the upper bound was between 75 and 90 percent and the lower bound was between 50 and 90 percent. For the minimum sampling error allowed for shopping trips, the upper bound was between 100 and 300 percent and the lower bound was 90 percent. The criterion used to establish the allowed sampling error was to allow the minimum possible sampling error yet still have enough observations left with which to estimate the regression parameters without serious colinearity problems. That is why the allowed sampling error for shopping trips was higher than that allowed for work trips.

The three data sets tested have substantially different average sampling errors for the estimated (endogenous) variable. When all trip interchanges are included, the zero trip observations dominate the data. They represent 91 percent of all transit work-trip interchanges, 71 percent of all automobile work-trip interchanges, 99 percent of all transit shopping-trip interchanges, and 92 percent of all automobile shopping-trip interchanges. The average upper bound for all transit passenger home-to-work trips (S_{11ij}) is 80.3 trips. Since the average number of trips observed for nonzero observations of S_{11ij} is 99.8 trips, it is clear that the zero observations of transit work-trip interchanges are not reliably estimated. In the case of automobile work-trip interchanges, the average upper bound is 109.0 trips and the average number of trips for nonzero trips is 123.3. Sampling-error data calculated for shopping trips show similar patterns. Therefore, elimination of the zero observations of the estimated variable is likely to improve the performance of the direct-demand model, and elimination of the nonzero observations that have high sampling errors can be expected to improve it further.

A detailed account of results for all of the equations estimated need not be provided. Instead, a brief summary will identify the major conclusions that can be derived from the results displayed in Tables 8-10. When the direct-demand model was estimated by using all observations, the equations provided very low levels of explanatory power. This is evident in Table 8, which provides the model parameters and R^2 for each estimation (transit and automobile, work trips, and shopping trips). The level of explained variation (R^2) ranged from 6 to 31 percent. In addition, several irrational parameter values were estimated. For example, in work-trip estimation (Equation 13, Table 8) a high correlation of 0.79 between highway travel time (T_{tij}) and transit fare (C_{1ij}) caused highway travel time to take an irrational coefficient of -0.686. The cross elasticity of transit travel with highway travel time should be positive, not negative. This correlation also affected the highway travel-time parameter in the transit shopping-trip estimator (Equation 15, Table 8). Other irrational parameter values in Table 8 can also be traced to high correlations among the exogenous variables.

When zero observations of the endogenous variable were excluded from the regressions, the level of explanatory power increased substantially but still remained below 50 percent for all equations, as Table 9 indicates. An unexpected benefit of eliminating zero observations was to reduce colinearity among several exogenous variables. Partly because of this, no irrational parameter values were produced in this set of estimators. In addition, the parameter estimates indicate elasticities of travel demand with respect to modal characteristics that are more consistent with other studies (11). For example, the elasticity of transit work travel demand with respect to transit travel time increases from -0.086 (Equation 13, Table 8) to -0.368 (Equation 17, Table 9) and the demand elasticity with respect to transit fare drops from -1.043 to -0.412. Another improvement is the approximate equality of the parameters that measure the impact of labor force in the trip origin (C_{is}) and total employment in the trip destination (TE_{sj}).

All of the estimates presented in Table 9 do not include the variable mean household income (HHI_{is}), which failed to pass the partial F-criterion. Unscreened regressions generally included this variable with an irrational sign and/or with a disturbing influence on the estimate of the elasticity of travel demand with respect to automobile ownership. This result occurs because of the reduced colinearity among HHI_{is}, AOR_{is}, and L_{is}.

The level of explanatory power of each trip estimator has increased substantially by screening zero observations of S_{mpij}, as given below:

Endogenous Variable	R^2 Unscreened	S_{mpij} = 0 Screened	Increase (%)
S_{11ij}	0.252	0.447	77
S_{41ij}	0.314	0.450	43
S_{12ij}	0.059	0.149	153
S_{42ij}	0.180	0.316	76

The improved performance of the direct-demand estimators can be attributed largely to the reduction of sampling error associated with the survey sample estimation of low-level interchanges. But the entire data set of trip observations has high sampling errors, and there is substantial sampling-error variation. The average percentage upper confidence interval for all nonzero observations of S_{11ij} (home-to-work transit passenger trips) is 278 percent, and the high sampling error is 461

Table 8. Parameters and t-statistics for direct-demand model: all trip interchanges regressed as observations.

Equation No.	Endogenous Variable	L_{is}	HH_{is}	HHI_{is}	AOR_{is}	TE_{si}	SSE_{si}	T_{1ij}	C_{1ij}	T_{2ij}	C_{2ij}	R^2	No. of Observations
Parameter													
13	S_{11ij}	0.566	NA	-0.396	-0.204	0.152	NA	-0.086	-1.043	-0.686	0.244	0.052	17 822
14	S_{41ij}	1.107	NA	-1.225	0.890	0.427	NA	NS	NS	-2.923	0.197	0.314	17 822
15	S_{12ij}	NA	0.046	NS	-0.077	NA	0.009	-0.016	NS	-0.294	0.043	0.059	17 822
16	S_{42ij}	NA	0.080	-0.203	0.405	NA	0.090	NS	0.450	-2.584	-0.055	0.180	17 822
t-Statistic													
13	S_{11ij}	35.4	NA	-18.9	-6.8	16.9	NA	-14.3	-15.4	-10.1	37.4		
14	S_{41ij}	44.3	NA	-37.1	18.9	30.5	NA	NS	NS	-47.9	19.7		
15	S_{12ij}	NA	9.2	NS	-7.7	NA	3.0	-8.0	NS	-19.6	21.5		
16	S_{42ij}	NA	5.7	-11.9	12.3	NA	12.9	NS	6.3	-40.4	-9.2		

Notes: $AOR_{is} = A_{is}/L_{is}$ for estimators of S_{11ij} and S_{41ij}; $AOR_{is} = A_{is}/HH_{is}$ for estimators of S_{12ij} and S_{42ij}.
NS = estimated parameter value either not significant or variable did not pass partial-F criterion to enter equation.

Table 9. Parameters and t-statistics for direct-demand model: nonzero trip interchanges regressed as observations.

Equation No.	Endogenous Variable	L_{is}	HH_{is}	AOR_{is}	TE_{si}	SSE_{si}	T_{1ij}	C_{1ij}	T_{2ij}	C_{2ij}	R^2	No. of Observations
Parameter												
17	S_{11ij}	0.413	NA	-0.270	0.360	NA	-0.368	-0.412	NS	0.011	0.447	1524
18	S_{41ij}	0.448	NA	0.476	0.382	NA	NS	NS	-0.783	NS	0.450	5168
19	S_{12ij}	NA	0.188	-0.244	NA	0.201	NS	NS	NS	NS	0.149	161
20	S_{42ij}	NA	0.319	0.890	NA	0.232	NS	NS	-0.949	-0.055	0.316	1349
t-Statistic												
17	S_{11ij}	18.8	NA	-7.1	14.4	NA	-9.4	-7.8	NS	2.8		
18	S_{41ij}	40.7	NA	17.6	38.2	NA	NS	NS	-43.5	NS		
19	S_{12ij}	NA	3.0	-2.5	NA	3.4	NS	NS	NS	NS		
20	S_{42ij}	NA	11.8	12.2	NA	10.1	NS	NS	-20.6	-9.2		

Notes: $AOR_{is} = A_{is}/L_{is}$ for estimators of S_{11ij} and S_{41ij}; $AOR_{is} = A_{is}/HH_{is}$ for estimators of S_{12ij} and S_{42ij}.
NS = estimated parameter value either not significant or variable did not pass partial-F criterion to enter equation.

Table 10. Parameters and t-statistics for direct-demand model: sampling-error screening of endogenous variable.

Equation No.	Endogenous Variable	L_{is}	HH_{is}	AOR_{is}	TE_{sj}	SSE_{sj}	T_{1ij}	C_{1ij}	T_{2ij}	C_{2ij}	R^2	No. of Observations	Maximum Sampling Error (US - S)/S	(S - LS)/S
Parameter														
21	S_{11ij}	0.464	NA	-0.126	0.448	NA	-0.263	-0.539	NS	NS	0.555	474	0.75	0.75
22	S_{11ij}	0.532	NA	NS	0.378	NA	-0.442	-0.650	NS	NS	0.842	105	0.75	0.50
23	S_{41ij}	0.396	NA	0.230	0.294	NA	0.134	NS	-0.585	-0.018	0.500	481	0.90	0.90
24	S_{12ij}	NA	0.384	-0.386	NA	0.174	NS	-0.529	NS	NS	0.347	135	3.00	0.90
25	S_{42ij}	NA	0.444	1.044	NA	0.222	NS	NS	-0.727	-0.047	0.397	705	2.50	0.90
26	S_{42ij}	NA	0.485	0.994	NA	0.151	NS	NS	-0.541	NS	0.499	267	1.00	0.90
t-Statistic														
21	S_{11ij}	15.5	NA	-2.1	13.2	NA	-5.5	-7.0	NS	NS				
22	S_{11ij}	11.8	NA	NS	4.6	NA	-4.9	-4.2	NS	NS				
23	S_{41ij}	17.2	NA	5.8	9.2	NA	3.2	NS	-11.3	-4.5				
24	S_{12ij}	NA	5.5	-3.2	NA	2.8	NS	-3.3	NS	NS				
25	S_{42ij}	NA	15.3	12.6	NA	6.9	NS	NS	-13.0	-5.9				
26	S_{42ij}	NA	13.9	7.6	NA	3.2	NS	NS	-6.4	NS				

Notes: $AOR_{is} = A_{is}/L_{is}$ for estimators of S_{11ij} and S_{41ij}; $AOR_{is} = A_{is}/HH_{is}$ for estimators of S_{12ij} and S_{42ij}.
NS = estimated parameter value either not significant or variable did not pass partial-F criterion to enter equation.

percent. The average percentage upper confidence interval for all nonzero observations of S_{41ij} (home-to-work automobile person trips) is 281 percent, and the high sampling error is 468 percent. It is likely that a screening of observations based on the sampling error estimated for each observation of the endogenous variable would result in further improvements in direct-demand estimation. We used the sampling-error statistics (US - S)/S and (S - LS)/S defined above for this purpose, using the interactive software developed for this project.

Application of the screening procedures described above to the direct-demand model estimate for transit passenger home-to-work trips resulted in Equation 21 (Table 10) after various levels of (US - S)/S and (S - LS)/S used to screen observations of S_{11ij} were tested. Explanatory power is increased to 56 percent, or 24 percent higher than a regression that only eliminates zero observations of S_{11ij} (Equation 17, Table 9). The parameter values of these two equations are similar. The most significant difference is the reduced elasticity of transit trip demand with respect to the automobile-ownership rate (A_{is}/L_{is}). Automobile-ownership elasticity drops from -0.2700 to -0.126 and the t-value drops to a low but marginally significant value of -2.1. However, the correlation of the automobile-ownership rate with transit travel time and cost has increased to 0.49 (with T_{1ij}) and 0.63 (with C_{1ij}). Lower t values for all three variables (A_{is}/L_{is}, T_{1ij}, and C_{1ij}) partly reflect the increasing difficulty of isolating the specific impact of each variable as correlations between them increase. The increased correlations result from the reduced data set (474 versus 1524 observations) remaining after the screening procedure.

Further tightening of the lower confidence interval to 0.50 reduces the number of accepted observations to 105 for Equation 22 (an additional 57 percent reduction from the number of observations accepted in Equation 21). The automobile-ownership rate does not pass the partial F-test to enter the regression and is also more correlated (negatively) with labor force at the trip origin (L_{is}) than in previous regressions (correlation = 0.53). Correlations of A_{is}/L_{is} and transit travel characteristic variables remain high. Parameter values for the four accepted variables change somewhat with an increased elasticity of transit travel with respect to transit travel time (T_{1ij}: -0.263 to -0.442) and transit fare (C_{1ij}: -0.412 to -0.650). The overall explanatory power of the equation increases to a very high 84 percent, or 88.4 percent higher than Equation 17 (Table 9), which only screened observations of $S_{1ij} = 0$ from the regression. However, it is not entirely clear that Equation 22 is a better estimator than Equation 21, which has a lower R^2 (0.56). Colinearity among variables is probably responsible for the absence of the automobile-ownership rate in Equation 22. Colinearity is a more serious problem, since Equation 22 is estimated with 105 observations compared with 474 observations accepted for Equation 21. In addition, as mentioned above, correlations between the automobile-ownership rate and other exogenous variables happen to be larger among the 105 observations accepted for Equation 22 than among the 474 observations accepted for Equation 21. The smaller number of observations combined with higher colinearity involving A_{is}/L_{is} would lead one to believe that the automobile-ownership rate should still appear in Equation 22 and that its absence may bias the parameters for the variables L_{is}, T_{1ij}, and C_{1ij}. However, any bias is likely to be slight, since the parameters for these variables do not differ substantially from those estimated in Equation 21, which includes the automobile-ownership rate as an exogenous variable.

Application of sampling-error confidence-interval screening to the estimator of automobile passenger home-to-work trips does not affect the overall level of explanatory power as significantly as when observation screening was applied to transit passenger home-to-work trips.

Exactly the same variables are included in Equation 23 (Table 10) as were included in Equation 18 (Table 9), and the parameter values calculated for them are quite similar. The elasticity of home-to-work automobile person trips with respect to highway travel time (T_{2ij}) decreases slightly, from -0.783 to -0.613, whereas other parameters remain more similar. The increase in explanatory power from 0.450 to 0.499 (an 11 percent increase) is modest but represents a significant improvement. The data set used to estimate Equation 23 has a reduced correlation between C_{1ij} and T_{2ij} of 0.644 and a consequent reduced propensity for C_{1ij} to enter the regression with a negative and irrational coefficient. The upper and lower bounds that produce this data set are both equal to 0.9. In the regression used to estimate Equation 23, transit fare (C_{1ij}) becomes the only mode characteristic variable that does not enter because it does not meet the partial-F test. The elasticity of automobile demand with respect to highway travel time (T_{2ij}) remains at approximately the same value estimated in earlier regressions. A slight but significant elasticity of automobile travel with respect to parking cost is estimated at -0.018. More important, a significant and reasonably large cross elasticity of automobile travel with respect to transit travel time (T_{1ij}) is estimated at 0.134.

When zero observations of transit passenger home-to-shop trips were screened from the regression analysis, no travel characteristic variables were accepted into the regression. The percentage of explained variation of this estimator was 14.9 percent. The only variables accepted were households in the trip origin district (HH_{is}), the automobile ownership rate (A_{is}/HH_{is}), and shop and service employment in the destination district (SSE_{sj}). When observations are screened on the basis of sampling error, the results are improved substantially. Explanatory power increases from 15 to 35 percent, a 133 percent increase (Equation 24, Table 10). Even though observations are reduced by 16 percent, the F-ratio increases from 9 to 17, and t-statistics improve overall for the three variables common to both equations (HH_{is}, A_{is}/HH_{is}, and SSE_{sj}). Furthermore, a significant elasticity of transit travel with regard to transit fare (-0.529) is estimated.

Application of sampling-error criteria to the estimation of home-to-shop automobile person trips yields estimators with similar parameter values as before but with a substantial improvement in explained variation of the endogenous variable (Equations 25 and 26, Table 10).

Use of the estimated sampling-error confidence interval to screen observations of S_{mpij} has resulted in the improvements in overall explanatory power given in Table 11.

CONCLUSIONS

The analysis and data presented above provide a comprehensive basis for evaluating direct-demand urban travel models and the data required to estimate them. Because of the uniform survey sampling procedures used in urban transportation data-collection projects, the MWCOG data base used in this project

Table 11. Improvements in overall explanatory power resulting from use of estimated sampling-error confidence interval to screen observations of S_{mpij}.

Endogenous Variable	R^2 Unscreened (A)	R^2 S_{mpij} = 0 Screened (B)	Confidence Interval Screened R^2: R^2	Increase (%) Over A	Increase (%) Over B
S_{11ij}	0.252	0.447	0.842	234	88
S_{41ij}	0.314	0.450	0.500	59	11
S_{12ij}	0.059	0.149	0.347	488	133
S_{42ij}	0.180	0.316	0.499	177	58

is typical of transportation data available for any urban area in the United States. Therefore, many conclusions that derive specifically from the tested data can be expected to hold for direct-demand models when they are applied to any urban area in the United States.

One objective of this study was to determine whether the potential benefits of direct-demand urban travel models could be realized and, if not, what factors prevent direct-demand models from fulfilling their potential. Direct-demand models, being a simultaneous or one-step determination of trip generation, mode split, and trip distribution, allow O-D travel characteristics to affect total travel estimation in addition to estimation of modal split and trip distribution. In contrast, the UTPP model system is structurally incapable of estimating induced travel. Although simultaneous disaggregate models can account for induced travel, they are substantially more difficult to use in the forecasting mode. They require the application of aggregation bias estimation in order to use disaggregate models with forecasts of socioeconomic and travel characteristic data. This is required because disaggregate-model parameters are estimated by using household-level data whereas forecasts of the exogenous variables are generally available only at more aggregated levels (districts). Furthermore, aggregation bias techniques are difficult to use and complicate the use of travel demand models to assist in the UTPP. In contrast, direct-demand-model parameters are estimated at the same level of aggregation as that available for future forecasts of exogenous variables. But it is exactly this difference in the level of data aggregation between aggregate and disaggregate models that causes the direct-demand model to have problems. If sampling errors are present in aggregate O-D trip data, the accuracy of the direct-demand-model parameters and the achieved goodness of fit are adversely affected whereas the disaggregate model would not experience similar problems.

This study has documented very large sampling errors in the O-D trip data collected for the Washington, D.C., metropolitan area. Screening observations to eliminate high sampling errors has significantly increased the level of explained variance that the direct-demand model can provide. However, unless practically all observations are eliminated, the sampling error of the accepted observations is still high. In addition, as pointed out above, the reduced data sets that remain after sampling-error criteria have been tightened often have more colinearity among exogenous variables. With fewer observations, the increased colinearity results in an inability to identify important variables in the estimator. Direct-demand models are therefore limited in their ability to provide reliable estimates of the response of travelers to transportation system characteristic variables by the high sampling errors contained in the travel demand data available for estimation of direct-demand models.

REFERENCES

1. Computing a Home Interview O-D Survey. Bureau of Public Roads, U.S. Department of Commerce, 1954.
2. A.B. Sosslau and G.E. Brokke. Appraisal of O-D Sample Size. Public Roads, Dec. 1960.
3. K.E. Heanue, L.B. Hamner, and R.M. Hall. Adequacy of Clustered Home Interview Sampling for Calibrating a Gravity Model Trip Distribution Formula. HRB, Highway Research Record 88, 1965, pp. 116-136.
4. M. Wohl and B.V. Martin. Traffic Systems Analysis for Engineers and Planners. McGraw-Hill, New York, 1967, pp. 128-129.
5. T.A. Domencich, G. Kraft, and J.-P. Valette. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. HRB, Highway Research Record 238, 1968, pp. 64-78.
6. M. Wohl. Demand, Cost, Price and Capacity Relationships Applied to Travel Forecasting. HRB, Highway Research Record 38, 1963, pp. 40-54.
7. Data Inventory: A General Description of Data Available for Planning Purposes. National Capital Region Transportation Planning Board, Washington, DC, Information Reprint 24, 1969.
8. R.T. Dunphy. Home Interview Survey Accuracy. Metropolitan Washington Council of Governments, Washington, DC, Tech. Notes 2, 1971.
9. R.W. Quandt and W.J. Baumol. The Demand for Abstract Transportation Modes. Journal of Regional Science, Vol. 6, No. 2, 1966.
10. BMDP: Biomedical Computer Programs. Univ. of California Press, Berkeley, 1975.
11. M.A. Kemp. Some Evidence on Transit Demand Elasticities. Transportation, Vol. 2, 1973.

Discussion

Kevin E. Heanue

Difiglio and Disbrow have done a very thorough job in examining the impact of survey sampling errors on travel demand forecasting. The paper is particularly useful in its coverage of statistical procedures. I am concerned, however, with the assumptions made regarding how survey data are actually used in travel demand forecasting. In general, the treatment of this aspect of forecasting does not track too well with actual practice.

The authors looked at the use of conventional major travel surveys and found them inadequate to the task of calibrating direct-demand models. They also found that conventional approaches to travel estimation are inadequate because induced demand is not separately estimated. I cannot support these conclusions.

In the recent past, competent analysts have tended to use special surveys rather than large surveys when attempting to calibrate direct-demand models, and they have been careful to eliminate coefficients based on small segments of their samples from the final models. They have also supplemented their direct-demand estimates with more conventional approaches to fill in the gaps where it was not possible to develop direct-demand coefficients.

The subject of induced demand is one that has been debated for some time, and I would argue that induced demand, although not separately estimated, is inherently accounted for in the combined generation-distribution effects. In using trip-generation procedures that consider increases in income and/or car ownership, induced travel is partially accounted for by increases in mobility (trips) that result from increases in vehicle supply. Similarly, as travel-time savings occur in the highway network, trip lengths increase and again induced travel is indirectly accounted for.

In conclusion, I recommend the paper to any student of travel demand forecasting but suggest that the problems of survey error can be and are overcome through the use of appropriate calibrating strategies. The results of well-calibrated forecasting models, when tested against measured travel volumes, are adequate for most planning applications.

Vanpool Energy Efficiency: A Reevaluation and Comparison with a Brokered-Carpooling Concept

AXEL B. ROSE

Since the first employer-operated vanpools began operating in 1973, much has been made of the considerable energy savings possible through vanpooling and it has been generally accepted that vanpools are the most efficient commuter transportation mode available. The analyses that formed the bases for these conclusions have seldom involved more than simple comparisons of the line-haul energies of vanpools and average commuter automobiles; rarely, if ever, have vanpools been compared with other innovative and efficient commuting modes. Based on data available through a recent survey of vanpool riders in Chattanooga, Tennessee, a more detailed calculation of vanpool energy intensities is presented that incorporates the line-haul, access-egress, and indirect energy uses of vanpools as well as a calculation of the energy uses arising from the use of pool vehicles for private purposes. The resultant energy intensity of vanpools is calculated at 1508 kJ/passenger-km (2300 Btu/passenger mile), which represents an increase of more than 100 percent over the line-haul energy intensity. Concurrently with the calculation of the vanpool energy intensities, values are calculated for an alternative commuting mode essentially identical to vanpools with the exception that efficient subcompact and compact automobiles are used instead of vans. In the final analysis it is shown that efficient brokered carpools could save up to 60 percent of the energy used by vanpools and also offer significant advantages over vanpools in ease of implementation and possible penetration of the commuting market.

In recent years, it has become a widely accepted conclusion that vanpools are the most efficient mode of commuter transportation and that consequently they should play a major role in any petroleum conservation program. Unfortunately, these conclusions have been largely based on incomplete or dated investigations of vanpool energy use; rarely, if ever, have vanpools been compared with other innovative commuter transportation modes. Within this context, this paper presents a more complete analysis of vanpool energy use and then compares the resultant energy uses with those of an alternative commuter transportation mode that has the potential of considerable energy savings over vanpool operations.

A vanpool can be described as a commuter ride-sharing transportation mode in which a group of people who live and work in proximity to each other commute together in an 8- to 15-passenger van. In return for a free ride and limited personal use of the van, one person in the group, typically the driver, assumes responsibility for the vehicle and its operation. The other pool members (and in some cases the employer and/or the government) share the costs of the whole operation. Three general types of vanpools are currently in operation: (a) employer-sponsored vanpools, in which the employer purchases the vans, furnishes them to the employees, and over time recovers the costs through the fares; (b) third-party-sponsored vanpools, in which a third party purchases the vans and acts as a broker between employees and employers; and (c) individually owned and operated vanpools.

The first employer-sponsored vanpool program became operational in April 1973 at the Minnesota Mining and Manufacturing (3M) Company in St. Paul, Minnesota. By April 1979, 4382 vanpools were known to be operating in addition to the 3000-5000 privately owned vanpools believed to be in existence (1, p. 6). From the first to the third quarter of 1980, the Tennessee Valley Authority (TVA) expanded its vanpool operations from 219 to 413 vans. By 1990, 1.15 million vanpools are forecast to be in operation in the United States (2, p. 10). Substantial government programs are under way to further ridesharing and vanpooling. Investment tax credits are being granted for the purchase of vans for pooling purposes, Highway Trust Fund money is available for the purchase of vans, and special lanes, to be reserved for high-occupancy vehicles, are being constructed in several areas. In summary, it can be stated that vanpooling has made substantial headway in the past few years toward penetrating the commuting market and that a variety of programs have been implemented that are aimed at increasing the growth of vanpooling in the future.

The impetus behind the movement can be found in a variety of perceived vanpool benefits frequently cited in the literature. Vanpool riders enjoy reduced commuting costs and the freedom of not having to drive, employers and/or localities need to provide and maintain fewer parking spaces, and everybody benefits from a reduction in congestion, vehicle emissions, and gasoline consumption. Of these benefits, lower commuting costs and energy savings are usually considered the most important. In comparison with traditional U.S. commuter transportation modes, a typical vanpool is generally credited with saving approximately 18 925 L (5000 gal) of gasoline per year, reducing emissions by 1.81 Mg (2 tons) per year, and removing six to nine vehicles from the road.

As stated, vanpool benefits have been calculated against a historical status quo. In view of the rapidly rising energy costs that tend to move people toward more efficient means of transportation and the significantly improved fuel economies for current and future automobiles, it is highly question-