Second, incorporation of limited OD data into the estimation method produced a substantial reduction in the average trip error. This suggests that even with the fundamental weaknesses that characterize simplistic algorithms like SYNODM, they could be useful in expanding limited survey data.

In any case, additional research on OD estimation procedures is needed. In this regard, Nihan's (7) incorporation of a trip impedance factor into the estimation method has produced some encouraging results.

ACKNOWLEDGMENT

The research reported in this paper was conducted as part of an ongoing research project entitled West Loop Authorized Vehicle Lane Conceptual Planning and Design sponsored by the Texas State Department of Highways and Public Transportation.

REFERENCES

1. E. Hauer and B.T. Shin. Estimation of Origin-Destination Matrices from Observed Flows: Simple Systems. Research Report 65. Department of Civil Engineering, University of Toronto, Toronto, Canada, March 1980.
2. P.P. Jovanis, W.-K. Yip, and A.D. May. FREQ6PE: A Freeway Priority Entry Control Simulation Model. Research Report UCB-ITS-RR-78-9. Institute of Transportation Studies, University of California, Berkeley, 1978.
3. G.H. English, H. Joiner, and W. Neyland. Traffic Study of Southwest Freeway (US 59) Between Shepherd Drive and Bissonnet Street in Houston, Texas Using the FREQ3CP Simulation Model. Texas State Department of Highways and Public Transportation, Houston, Oct. 1978.
4. R.W. Stokes, J.M. Mounce, D.E. Morris, and R.L. Peterson. Simulation Analyses of Proposed Improvements for the Southwest Freeway (US 59), Houston; Part 1: Long-Term, Capital-Intensive Improvements. Final Report. Texas Transportation Institute, Texas A&M University System, College Station, Aug. 1982.
5. K. Banning, R. Stokes, J. Mounce, and D. Morris. Simulation Analyses of Proposed Improvements for the Eastex Freeway (US 59N), Houston. Texas Transportation Institute, Texas A&M University System, College Station, Feb. 1983.
6. L.G. Nungesser, D.L. Christiansen, and A.J. Ballard. Priority Treatment for High-Occupancy Vehicles on I-10 and I-35 in San Antonio, Texas: A Needs Assessment. Texas Transportation Institute, Texas A&M University System, College Station, March 1983.
7. N.L. Nihan. Procedure for Estimating Freeway Trip Tables. In Transportation Research Record 895, TRB, National Research Council, Washington, D.C., 1982, pp. 1-5.
8. M. Eldor. Integrated Real-Time Freeway On-Ramp Control Strategies. Ph.D. dissertation. Institute of Transportation Studies, University of California, Berkeley, 1976.
9. R.W. Stokes, D. Morris, and J.M. Mounce. Katy Freeway (I-10W) Origin-Destination Study. Texas Transportation Institute, Texas A&M University System, College Station, Jan. 1982.

Estimation of Origin-Destination Matrices with Constrained Regression

CHRIS HENDRICKSON and SUE McNEIL

ABSTRACT

The use of constrained generalized least-squares (CGLS) regression to estimate origin-destination travel matrices from aggregate data is described. The CGLS method does not require general surveys but allows any available data to be included. Variances of matrix entry estimates can be estimated and used as measures of uncertainty or to suggest additional sampling strategy. Two case studies are described from applications to data from Portland, Oregon. The first involves expanding a matrix of transit work trips to all transit trips. Second, a gravity-type model of trip distribution for all work trips is estimated. Comparisons are made with other estimation methods with respect to accuracy, computational effort, and the use of uncertainty measures.

Origin-destination (OD) matrices representing the number of trips between zones or locations in a particular time period are widely used in transportation systems analysis. In particular, they are an important intermediate stage in the Urban Transportation Planning System (UTPS) (1). They are also used for modal operations planning, as in transit route planning [see, for example, discussions by Sheffi and Sugiyama (2) and Turnquist et al. (3)].

Surveys are widely used to estimate such matrices. For example, entries may be obtained from estimates of household travel between specified zones based on home interview surveys. Unfortunately, such surveys are time consuming and expensive. Moreover, updating and expanding survey-based matrices are common problems.

To overcome the costs and delays of general surveys, several alternative methods are used to obtain estimates of matrix entries. Nonsurvey matrix estimation methods range from applying a constant factor to update an existing matrix to more complicated ab initio estimates using, for example, gravity models. More recently, matrix entry estimation methods that include small-sample or traffic-count data or both have been developed. These include entropy maximization (4) and maximum-likelihood methods (5,6). For a survey of these methods see the paper by Chan et al. (7).

Methods used to estimate OD matrices should meet the following requirements. They should be able to

1. Include any available survey-derived or aggregate data,
2. Include socioeconomic characteristics and attitude such as travel time between zones,
3. Provide some measure of the reliability of the estimates, and
4. Account for errors in the available data.

Based on these criteria, a regression formulation is developed of the matrix entry estimation problem that is equivalent to a quadratic programming formulation (3,8). Although this formulation has been applied to the estimation of input-output tables (9) and Markov transition probability matrices (10), in the applications to transportation planning the statistical properties of the estimates have not been explored and these applications have been less general. The method is able to include any linear constraint on the matrix entries in the formulation as well as a linear-in-the-parameters distribution function with known or unknown parameters. The resulting estimates are best linear unbiased estimates and an estimate of their variance-covariance matrix may be obtained.

The plan of the remainder of this paper is as follows. The estimation problem is formulated as a regression problem and the assumptions that are required are described. The estimation method is applied to the problem of expanding a matrix of transit work trips to all transit trips for Portland, Oregon. A work trip OD matrix for Portland is estimated using a distribution function with unknown parameters. Last, the advantages and disadvantages of the regression formulation are discussed and some conclusions are presented.

ESTIMATION PROBLEM FORMULATION

The objective of matrix entry estimation is to form a consistent estimate ($\tilde{Q}$) of an actual but unknown matrix Q of size nm. It is required that the estimate $\tilde{Q}$ be consistent with all relevant information, in the sense that the estimates do not conflict with any available data known a priori.

To formulate the estimation problem, let the matrices Q and $\tilde{Q}$ be rearranged to form vectors $\underline{q}$ and $\tilde{\underline{q}}$ (of size nm) with the columns of the matrices arranged end to end. (Throughout this paper capital letters will be used to represent matrices and underlined lower-case letters to represent vectors.) It is assumed that a linear-in-the-parameters functional relationship exists between some set of attributes X and the values of matrix entries $X\underline{a}$, where $\underline{a}$ is an h vector of either known or unknown parameters. For example, $X\underline{a}$ might be a linear form of a gravity model. With these assumptions a regression equation is formulated as follows:

$$\underline{q} = X\underline{a} + \underline{\varepsilon} \tag{1}$$

where $\underline{\varepsilon}$ is an nm vector of errors that are described in the following.

It is often the case that observations or constraints exist on the sums of some subsets of the elements of q. For example, if Q is an OD matrix, the row totals are the number of trips generated at each origin and the column totals are the number of trips attracted to each destination. Alternatively, the number of trips crossing a corridor may be known. Accordingly, it is assumed that some linear constraints exist on values of $\underline{q}$:

$$\underline{r} = A\underline{q} \tag{2}$$

where $\underline{r}$ is a p vector of constraint totals and A is a p x (mn) incidence matrix of zeros and weights (which will usually be 1's). Known values of $\underline{q}$ can be constrained to equal their values by a simple constraint $r_i = q_i$. The constraint totals $\underline{r}$ may contain measurement errors that may be accounted for in the estimation procedure (11). The existence of inequality constraints is also of interest in many cases, but such constraints introduce some complications. For simplicity it is also assumed that X is a matrix of fixed numbers of full column rank and A is of full row rank. Thus, constraints that are linear combinations of other constraints should be eliminated. For example, if constraints exist on each row and column of Q, one of these n + m constraints is a linear combination of the others and should be discarded because it adds no information for estimation purposes.

The special case in which all $\underline{a}$ are known is of particular interest, and it will be considered first. This case would occur if hypothesized or base-period values of $\underline{q}$ were available rather than the functional form $X\underline{a}$. Such base values, denoted $\underline{y}$, may be obtained from a gravity model or be outdated estimates. In any case, the estimates are generally not consistent with the constraints given in Equation 2. Consistent estimates are obtained by assuming that the base matrix vector $\underline{y}$ is equal to the true vector $\underline{q}$ plus an error term:

$$\underline{y} = I\underline{q} + \underline{\varepsilon} \tag{3}$$

As before, a set of constraints (Equation 2) on $\underline{q}$ is imposed.

The standard assumptions of the generalized least-squares regression model concerning the errors in regression Equation 1 are introduced:

$$E[\underline{\varepsilon}] = \underline{0} \tag{4}$$

$$E[\underline{\varepsilon}\ \underline{\varepsilon}'] = \sigma^2 V \tag{5}$$

so that the expected value of entries in $\underline{\varepsilon}$ is zero and the covariance matrix of $\underline{\varepsilon}$ is a constant (σ^2) times a known nonsingular matrix V, which is discussed in the following. In this case, the model given by regression Equation 3, the linear constraints (Equation 2), and the assumptions about the error terms (Equations 4 and 5) are identical to that of constrained generalized least-squares regression (CGLS) (12). The CGLS estimate of $\underline{q}$ is

$$\tilde{\underline{q}} = \underline{y} + VA'(AVA')^{-1}(\underline{r} - A\underline{y}) \tag{6}$$

The estimator $\tilde{\underline{q}}$ produces best linear unbiased estimates of $\underline{q}$ and has a covariance matrix:

$$COV[\tilde{\underline{q}}] = \sigma^2[V - VA'(AVA')^{-1}AV] \tag{7}$$

where an unbiased estimator of σ^2 is

$$s^2 = [\underline{y} - \tilde{\underline{q}}]'V^{-1}[\underline{y} - \tilde{\underline{q}}]/p$$

where p is the number of constraints. Assuming normality of observations, confidence intervals and hypothesis testing may be performed using Equations 6 and 7. The square roots of the diagonal elements of $COV[\tilde{\underline{q}}]$ are the standard errors of estimated parameters that are normally reported in regression.

As with other regression models, this special case can be formulated as a quadratic programming problem:

$$P1: \mathrm{Min}_{\underline{q}}[(\underline{q} - \underline{y})'V^{-1}(q - \underline{y}) \mid \underline{r} = A_q] \tag{8}$$

in which the weighted sum of squared errors is minimized subject to a set of constraints.

For the general case, parameters $\underline{a}$ of regression Equation 1 are to be estimated in addition to the matrix entries $\tilde{\underline{q}}$. Estimates of the model parameters $\tilde{\underline{q}}$ and of the matrix entries $\tilde{\underline{q}}$ can be obtained (13) as follows. Substituting Equation 1 into Equation 2 yields

$$\begin{aligned}\underline{r} &= A(X\underline{a} + \underline{\varepsilon})\\ &= AX\underline{a} + A\underline{\varepsilon}\\ &= X^*\underline{a} + \underline{\varepsilon}^*\end{aligned} \tag{9}$$

This is a linear-regression model with observations $\underline{r}$, with $X^* = AX$, and a random vector $\varepsilon^* = A\varepsilon$. If it is assumed that the assumptions about the error term in Equation 1 are given by Equations 4 and 5, the expectation of the error term is zero, because $E[\varepsilon^*] = E[A\underline{\varepsilon}] = \underline{0}$, from Equation 4. Also, the covariance matrix of error terms is $COV[\underline{\varepsilon}^*] = E[A\underline{\varepsilon}(A\varepsilon')] = AE[\underline{\varepsilon\varepsilon}']A' = \sigma^2 AVA'$. Assuming that AVA' is nonsingular, this regression model fulfills the assumptions of Equations 3 and 4 with $V^* = AVA'$, and is a generalized least-squares model.

Following Theil (12), the GLS estimator of $\underline{a}$ is

$$\begin{aligned}\tilde{\underline{a}} &= [(AX)'(AVA')^{-1}(AX)]^{-1}(AX)'(AVA')^{-1}\underline{r}\\ &= [X'A'(AVA')^{-1}AX]^{-1}X'A'(AVA)^{-1}\underline{r}\end{aligned} \tag{10}$$

which has the normal regression properties of being a best linear unbiased estimate of $\underline{a}$. Moreover, the covariance matrix of $\tilde{\underline{a}}$ is

$$COV[\tilde{\underline{a}}] = \sigma^2\ [X'A'(AVA')^{-1}AX]^{-1} \tag{11}$$

where an unbiased estimator of σ^2 is

$$s^2 = (\underline{r} - AX\tilde{\underline{a}})'(\underline{r} - AX\tilde{\underline{a}})/(p - h)$$

The predictor $X\tilde{\underline{a}}$ provides estimates of $\underline{q}$ that can then be used in the CGLS formulation to obtain best linear unbiased estimates of the matrix entries that are consistent with the constraints.

In passing, it might also be noted that the GLS model could be equivalently formulated as a quadratic programming problem. In this case, it is desired to minimize the weighted sum of squared errors, subject to the known constraints, by choosing values of $\underline{q} + \underline{a}$:

$$P1: \mathrm{Min}_{\underline{q},\underline{a}}[(\underline{q} - X\underline{a}')V^{-1}(\underline{q} - X\underline{a}) \mid \underline{r} = A\underline{q}] \tag{12}$$

This programming problem has the same solution for $\underline{a}$ as the problem in Equation 8 [assuming, of course that $(AVA')^{-1}$ exists].

Thus, the estimation problem can be formulated as either a generalized least-squares regression problem or a quadratic programming problem. A third interpretation as a Bayesian estimation problem with a quadratic loss function is also possible (11).

Solution of the estimation equations (Equation 6 or 10) can be performed by any general matrix inversion and multiplication package. Note that estimation of the covariance matrix (Equation 7 or 11) can use the matrix inversion required for calculation of the actual estimates. Alternatively, specialized techniques such as conjugate gradient algorithms can be employed to speed calculations for large problems. Sparse matrix calculation methods can also be used (14).

Before turning to examples, the specification of the error term $\underline{\varepsilon}$ deserves mention. It is assumed that the covariance matrix of error terms is nonsingular and specified in advance. Here some possible specifications for the error term covariance matrix (V) are considered. This is an important subject because specification of the covariance matrix will affect estimates and should reflect the analyst's beliefs concerning errors. In many applications of quadratic estimation techniques, the assumption of independent and identically distributed error terms seems to be made implicitly and without due consideration, solely because of analytical convenience.

The classical assumption in least-squares regression is that error terms are independently and identically distributed. In this case, the matrix V is the identity matrix (V = I) and $V[\underline{\varepsilon}] = \sigma^2 I$. For this common case, the generalized covariance matrix AVA' = AIA' = AA' must be nonsingular because A is of full row rank. Consequently, estimates of $\tilde{\underline{a}}$ can almost always be calculated from Equation 10; the only exception occurs in the unlikely case that the solution matrix $[X'A'(AA')^{-1}AX]$ is singular.

For physical matrix entries, an appealing alternative to this classical assumption is that error variances are proportional to the corresponding entry in the estimate $\underline{y}$. Thus, entries with small expected value would have small variance, as might be expected in many situations. This assumption results in a weighting function in P1 similar to that of a chi-square goodness-of-fit test:

$$P2: \mathrm{Min}_{\underline{q}}[(\underline{q} - \underline{y})'V^{-1}(q - \underline{y}) \mid \underline{r} = A\underline{q}]$$

$$= \left[\sum_{k=1}^{nm} (q_k - y_k)^2/y_k \mid \underline{r} = A\underline{q}\right] \tag{13}$$

where q_k and y_k are the kth entries in $\underline{y}$ and $\underline{q}$, respectively.

One practical advantage of a chi-square weighting scheme of this sort is that the likelihood of obtaining negative entry estimates is much reduced from the case of the classical assumption of constant variance. This result is due to an assumption of small variance for entries of $\underline{q}$ having small initial estimates $\underline{y}$. This result may often remove the necessity to introduce inequality constraints to ensure that entry estimates are nonnegative.

These results are now applied to the problem of expanding an OD matrix.

EXAMPLE 1: EXPANSION OF A TRANSIT WORK TRIP OD TRAVEL MATRIX TO ALL TRANSIT TRIPS

The purpose of the application problems in this and the next section is to illustrate quadratic matrix entry estimation methods. The problems are intended to demonstrate the feasibility and flexibility of the methods developed in the preceding section. The applications also provide some experience with computational problems and illustrate the formulations that might be appropriate for different applications and how these formulations might vary with the quality and quantity of data available.

In the examples the quadratic matrix entry estimation method is applied and evaluated. For validation and evaluation, matrices obtained from surveys are used for comparison with the quadratic estimates. For estimation, a base matrix or data to estimate a matrix entry function as well as some constraints are used. The more constraints, the larger the degrees of freedom and the more reliable are the estimates. The first example expands a transit OD matrix from Portland, Oregon. The second estimates ab initio an OD matrix for work trips in Portland.

The first application is an expansion of the transit work trip matrix to an OD matrix for all transit trips on the Portland Transit Authority's system (Tri-Met). In practice, the transit work trip matrix might have been obtained from data used for another purpose, inferred from census data, or estimated from journey-to-work survey data. The constraints required to estimate the complete matrix might be the result of data routinely collected for each route and some simple surveys. The expansion problem might be a problem facing a transit authority that does not wish to administer a general survey but would like to estimate an OD matrix from such available data.

The OD matrix used as a base matrix and the matrix to be estimated in this application differ slightly from the usual OD matrices. Instead of using OD zones, an OD matrix with transit routes as origins and four types of destinations is estimated. The destinations are defined in two categories--transfer or nontransfer trips. Then within these categories the place where the person alights from the bus is defined as in the central business district (CBD) or outside the CBD, giving a total of four destinations. The transfer trips are particularly interesting to transit planners because although route volumes are generally known, the origin and number of transfer trips can rarely be estimated unless systemwide on-board surveys have been undertaken.

Therefore, the OD matrices in this example are estimated by origin route and four defined destinations. There are 71 routes, so the matrix to be estimated is 71 by 4 (with 284 entries). The ijth element in the matrix represents the number of people beginning their trip on route i with destination j. The Tri-Met route numbers associated with each row (origin) are described elsewhere (11). The destinations are defined to be

j = 1 if the destination is outside the CBD for a nontransfer trip,

j = 2 if the destination is in the CBD for a nontransfer trip,

j = 3 if the destination is outside the CBD for a transfer trip, and

j = 4 if the destination is in the CBD for a transfer trip.

Five different formulations were used to expand the Portland transit OD matrix for work trips to all trips. The five formulations permit a comparison of the estimates using the biproportional method (described in the following) and an ad hoc procedure with estimates obtained using the quadratic method as well as consideration of the effect of additional constraints. For comparison and basic data, a matrix derived from 12 percent on-board survey data is used.

The first formulation is an ad hoc procedure for expanding the matrix and was devised so that the entry estimate equals the matrix entries for work trips multiplied by a factor equal to the total number of all transit trips divided by the number of work transit trips. Thus, this is simply a constant expansion. The expansion factor in this case was 1.683.

The second formulation is a quadratic method using a chi-square objective function, implying that changes in the base matrix are likely to be more variable (larger) if the relative magnitude of the entry is larger. This is believed to be a reasonable assumption. As base values in the formulation, the ad hoc estimates obtained with the first formulation are used. The constraints are row and column totals representing the total number that use a route and the total number that have each of the four destinations, respectively. These data would normally be available from aggregate route counts and some simple surveys.

The third formulation is a biproportional estimation problem. The biproportional estimation is also known as the Fratar method, Bregman's balancing method, the Furness iterative procedure, or the RAS method. The base matrix used in the biproportional formulation is the work trip matrix. Again, row and column totals are assumed known. The biproportional method factors up each entry by a row and column factor so that the constraints are met. This method is a common alternative to least-squares methods.

The fourth and fifth formulations use additional constraints in the form of the total number who transfer on each route. The fourth and fifth formulations are quadratic and biproportional formulations as in the second and third formulations but with the additional constraints. The quadratic formulation simply adds the additional constraints to the first formulation with the row and column totals. To formulate the biproportional problem with the additional constraints, the problem can be split into two independent biproportional problems. The first problem operates on columns 1 and 3 of the matrix and the second problem operates on columns 2 and 4.

The formulations are summarized as follows, where w_{ij} is the ijth entry in the matrix of work trips, t_i is the number of transfers from route i, $\underline{u}$ and $\underline{v}$ are row and column totals, and all other notation has been defined previously.

Ad hoc method, constant factor expansion:

AH: $q_{ij} = aw_{ij}$

where $a = \Sigma_{ij}q_{ij}/\Sigma_{ij}w_{ij}$ with $\Sigma_{ij}q_{ij}$ known from passenger counts.

Quadratic with row and column constraints:

Q1: Min $(\underline{q} - a\underline{w})'V^{-1}(\underline{q} - a\underline{w})$

subject to $\Sigma_i q_{ij} = v_j$ and $\Sigma_j q_{ij} = u_i$.

Biproportional with row and column constraints:

B1: Q = B W C

subject to $\Sigma_i q_{ij} = v_j$ and $\Sigma_j q_{ij} = u_i$. (Note that B and C are diagonal matrices of unknown factors.)

Quadratic with row and column totals and total number of transfers on each route (identical to Q1 but with added constraints):

Q2: $q_{ix2} + q_{i4} = t_i \; \forall i$

Biproportional with row and column totals and the total number of transfers on each route:

B2I: $q_{i1} = b_i w_{i1} c_1 \qquad q_{i3} = b_i w_{i3} c_3$

subject to $\sum_i q_{ij} = v_1$

$\sum_i q_{i3} = v_3$

$q_{i1} + q_{i3} = u_i - t_i$

B2II: $q_{i2} = b_i^* w_{i2} c_2 \qquad q_{i4} = b_i^* w_{i4} c_4$

subject to $\sum_i q_{i2} = v_2$

$\sum_i q_{i4} = v_4$

$q_{i2} + q_{i4} = t_i$

The chi-square formulation Q1 was estimated using a general-purpose quadratic programming package, and both formulations Q1 and Q2 were estimated using FORTRAN programs that used matrix manipulation methods. The specific quadratic programming package used for estimation uses Lemke's method and the matrix manipulation method used a matrix factorization routine from the International Mathematical and Statistical Library (IMSL) package. Similar results were obtained for formulation Q1 using both methods, and minor differences (less than 1 percent) can be attributed to rounding errors. The quadratic programming package has the advantage that the elements may be constrained to be nonnegative and inequality constraints may be included. for example, it is known that each entry in the estimated matrix of all trips has to be greater than or equal to the corresponding entry in the matrix for work trips. This constraint was not applied, because only 5 of the 284 entries (less than 2 percent) violated the constraint. The matrix inversion routines have the advantage that the variance of the estimates can easily be calculated.

Table 1 shows the central processing unit (CPU) computation time on a DEC-20 computer for the matrix manipulation routines and the quadratic programming package. The amounts of time include setting up the problem from the same basic data set. It is indicated in the table that the computation time using the quadratic programming package is significantly greater than that using the matrix manipulation routines. The matrix manipulation routines also calculate variance estimates.

TABLE 1 Computation Times for the Portland Expansion Example Estimates

Method	Formulation	Computation	CPU Time (sec)
Quadratic	Q1 matrix manipulation	Data preparation	14
		Estimation	41
	Q2 matrix manipulation	Data preparation	16
		Estimation	90
	Q1 quadratic programming package	Data preparation	3
		Estimation	93
Biproportional	B1	Data preparation and estimation	2
	B2	Data preparation and estimation	4

The biproportional estimates were obtained iteratively. The program terminated after 20 iterations or when all the row and column totals were within 1 percent of the constraints. Table 1 shows the computation time for obtaining the biproportional estimates on a DEC-20 to be about one-twentieth of the time required to obtain the quadratic estimates with the matrix inversion routines.

The evaluation of the estimation results presents a problem, because there is no unique method for comparing two or more methods. Several testing routines are possible (11,15,16). Table 2 gives three aggregate measures of estimation errors. Each is calculated by comparison with the on-board survey results. As can be seen, the quadratic and biproportional methods are comparable in accuracy for each case and generally superior to the ad hoc estimates. Adding additional information in the form of transfer totals improves the accuracy of estimates, as expected.

TABLE 2 Average Absolute and Relative Errors for Expanding the Portland Work Trip Transit Matrix

Method	Avg Absolute Error	Avg Relative Error[a]	Ratio of Avg Error to Avg Entry
Ad hoc	97.0	0.240	0.166
With row and column constraints			
Q1 quadratic	49.7	0.184	0.085
B1 biproportional	48.4	0.179	0.083
With row and column constraints and constraints on the number of transfers			
Q2 quadratic	35.0	0.155	0.060
B2 biproportional	34.4	0.151	0.059

[a]Calculated as the average ratio of errors to survey matrix entries excluding zero matrix entries.

The similarity between the quadratic and biproportional results is not surprising. It has been shown (17) that the chi-square formulation is a first-order approximation to a biproportional problem.

When the estimation problem is formulated as a CGLS regression problem, the estimates can also be evaluated in terms of their uncertainty. This evaluation is based on the assumption that the errors are highly correlated with the standard deviations (variances) of the estimates. Table 3 shows some correlation coefficients between errors and measures of uncertainty for the quadratic estimates. For each set of estimates, there is a positive correlation between the standard deviation and the average absolute error and the coefficient of variation and the average relative error. If one is interested in reducing a particular type of error, entries can be selected for special surveying or data gathering on the basis of the estimated uncertainty measure that

TABLE 3 Correlation Coefficients Between Errors and Uncertainty Measures for Estimated Entries in Transit Matrix

Method	Uncertainty Measure	Avg Absolute Error	Avg Relative Error
Q1 quadratic with row and column constraints	Standard error	0.509	-0.542
	Coefficient of variation	-0.389	0.542
Q2 quadratic as in Q1 with constraints on transfers	Standard error	0.475	-0.187
	Coefficient of variation	-0.204	0.453

is highly correlated with that error. Using uncertainty measures in this way can be quite helpful. For example, examining the standard errors from the quadratic estimations, a planner might choose those with the largest uncertainty for special surveys.

Although the CGLS regression formulation produces similar results to the biproportional method and is computationally more expensive, there is no way to evaluate the uncertainty of biproportional estimates directly. The estimated variances of the entry estimates represent a measure of the uncertainty associated with each entry estimate and may be used to evaluate the estimates.

EXAMPLE 2: AB INITIO ESTIMATION OF A WORK TRIP OD MATRIX

This application illustrates the estimation of matrix entries ab initio using a distribution function with unknown parameters. The formulation follows that presented in the section on estimation problem formation, in which the matrix entries are assumed to be a linear function with unknown parameters (called a distribution function) of the attributes of the matrix entries. The estimation problem is then formulated as a least-squares estimation problem subject to the appropriate constraints.

One of the advantages of the quadratic estimation method in this application is that it uses any available aggregate data to estimate the matrix entries. Such data generally come from sources other than transportation surveys. For this application the required data can be obtained from sources that are likely to continue to provide the same kinds of information in the future and therefore the modeling approach is unlikely to become obsolete because the required data are not available. The methods used to illustrate ab initio estimation in this section also show how additional aggregate travel-specific data can be added to improve the estimates.

Four parameters must be specified to estimate the ab initio model: a definition of the origin and destination zones, the distribution function, the weighting matrix, and the constraints.

The origin and destination zones are defined in terms of census tracts. For computational ease, the 189 census tracts of the Portland metropolitan area are aggregated to 44 zones. The 57 census tracts that are in the Portland standard metropolitan statistical area (SMSA) but outside the metropolitan area are aggregated to four zones. The aggregation simply joins adjacent areas except where physical barriers such as rivers provide a natural division. Therefore, the OD matrix has 48 zones representing 48 x 48 = 2,304 trip elements to be estimated.

To estimate a model ab initio, a distribution function that is linear in the parameters is required to ensure convexity. A simple model based on one independent variable is used in this application. The dependent variable is a function of the number of workers resident in the origin zone, the employment in the destination zone, and the distance or travel time between the two zones. The number of residents serves as a measure of the trip-generating ability of a zone, whereas the employment serves a similar purpose with respect to the attractiveness of the zone in the sense that it measures the extent to which a zone provides employment. The distance or travel time are measures of the travel impedance, implying that large travel impedances discourage trips. The proposed model is

$$q_{ij} = a\, D_i E_j/(d_{ij}^2 + \varepsilon_{ij}) \quad (14)$$

where

q_{ij} = number of work trips from i to j,
a = parameter to be estimated,
D_i = number of workers resident in zone i,
E_j = employment in zone j,
d_{ij} = interzonal travel impedance from zone i to zone j, and
ε_{ij} = error term.

The model is a simple gravity model.

The formulation of the model also requires definition of the structure of the weighting matrix (V). The matrix V is the matrix of weights in the quadratic objective function or the variance-covariance matrix of the error terms in the regression equation. It is assumed that V is a diagonal matrix with elements on the diagonal proportional to $D_i E_j/d_{ij}^2$ and therefore the objective function is of the chi-square type.

There are many possible constraints $R\underline{q} = \underline{r}$ that can be used to estimate the unknown parameter (a) and the matrix entries. The number and type of constraints directly affect the accuracy of the estimation. Two different sets of constraints are used to estimate the problem. Each set represents feasible constraints that can be obtained without complete areawide surveys.

The first set of constraints is the most basic, consisting of the row and column totals. Each row represents an origin zone and the row total the number of trips from an origin; therefore the row total is the number of workers residing in that origin zone (D_i). Similarly, the column totals represent the number of trips made to a zone; therefore the column total is the employment (E_j) in that zone. With this set of constraints the formulation is for a doubly constrained gravity model.

The quadratic programming problem is then formulated as follows:

$$\mathrm{Min}_{\underline{q},a} \sum_i \sum_j [(q_{ij} - aD_iE_j)/d_{ij}^2]^2/(D_iE_j/d_{ij}^2) \quad (15)$$

subject to $\Sigma_i q_{ij} = E_j$ and $\Sigma_j q_{ij} = D_i$, where the total employment (E_j) and the number of workers resident (D_i) in each zone are the column and row totals, respectively. It is assumed that those with no fixed place of employment work in the zone in which they reside.

The second set of constraints is added to the first set. The additional data are obtained by using the Willamette River as a cordon; this divides the Portland metropolitan area so that there are 15 zones on the east side of the river and 33 zones on the west side.

The additional constraints might be obtained by asking those crossing the cordon in either direction if they are going to work and either where they work or where they live, thereby adding 47 linearly independent constraints. For this estimation problem the 47 constraints obtained by surveying traffic going in both directions and asking travelers where they live are used. Algebraically, the constraints are as follows:

$$\Sigma_{j\varepsilon WS} q_{ij} = C^O_{iES} \qquad i \varepsilon\ ES$$

$$\Sigma_{j\varepsilon ES} q_{ij} = C^O_{iWS} \qquad i \varepsilon\ WS \quad (16)$$

where C^O_{iES} are the constraint totals for each of the origins on the east side, and C^O_{iWS} are the constraint totals for each of the origins on the west side.

The survey-derived matrix is from the 1976 travel-to-work supplement to the annual survey of housing for Portland. This matrix is used for comparison and to obtain much of the data for the estimation. The interzonal travel impedances were obtained from the Portland Metropolitan Service Center. Such data are commonly available within UTPS models. The peak-period travel time was used in this study because most trips to work are during the peak period and travel time is a more suitable measure of the travel impedance than distance.

Matrix manipulation programs were used to estimate the unknown parameter and matrix entries in each formulation. The results for estimation of the unknown parameter (a) were quite consistent. Using 95 or 145 constraints, an estimate for a of $\tilde{a} = 3.80 \times 10^{-4}$ was obtained, which was also the estimate obtained from regression on all 2,304 surveyed matrix entries. However, the t-statistic for the parameter ranged from 8.15 with 95 constraints to 8.76 with 142 constraints to 18.80 with all 2,304 observations. Clearly, more data reduced the level of uncertainty.

With the estimation of the unknown parameter ($\tilde{a}$), matrix entries could be calculated using Equation 14 as $q_{ij} = \tilde{a}\, D_i E_j / d_{ij}^2$. However, this would not ensure that the estimates $\tilde{q}_{ij}$ were consistent with the known totals D_i. Accordingly, a second-stage quadratic estimation was applied using the y_{ij} as initial estimates and the constraints defined previously.

Concerning the accuracy of the entry estimates themselves, the average absolute errors were 330 and 310 for the 95- and 142-constraint cases, respectively. Average relative errors were 1.2 in both cases.

In this example, the unknown parameter and matrix entries have been estimated using a linear form of a doubly constrained gravity model and the same model with additional constraints. Each estimation involved two stages. The first stage estimated the parameter in the gravity model, subject to the constraints, using generalized least-squares regression. The second stage uses the gravity model to predict the entries in the OD matrix and reconciles them with the constraints using CGLS regression. Thus, the estimation methods described here do not require iteration to obtain parameter estimates for a linear-in-the-parameters model. They also provide a measure of the reliability of the estimates.

RELATIVE ADVANTAGES AND DISADVANTAGES OF THE REGRESSION FORMULATION

In the preceding sections, CGLS formulations of the matrix entry estimation problem have been developed and applied to the problem of estimating two different OD matrices. The applications demonstrated that the methods are computationally feasible and produced estimates comparable in terms of accuracy with estimates obtained using other techniques. Some of the specific advantages and disadvantages of the regression formulation are as follows.

In particular, the regression formulation provides some unique advantages:

1. It is flexible in terms of the information that can be included as constraints. For example, individual elements can be constrained to particular values or a linear function of several entries; that is, the sum of two entries can be assumed known.

2. It is flexible in its formulation. Distribution functions with known or unknown parameters may be included in the formulation.

3. It is able to provide a measure of the reliability of the entry estimates in terms of the variances. In all the applications presented, the variances were relatively large because of the minimal amount of data used to estimate the matrix entries. The variances in turn can be used to evaluate the entries and derive strategies for obtaining additional information to improve estimates. In the expansion example the variance was found to be highly correlated with the absolute error in the estimates and the coefficient of variation with the percentage error, indicating that such reliability measures are highly desirable. The variance measures can also be used to derive sampling strategies for obtaining additional information.

4. It is of comparable accuracy with the biproportional methods when a chi-square formulation is used. The result is not surprising because the biproportional objective function may be shown to be a first-order approximation to the chi-square objective function. In cases in which the error distribution is not of a chi-square nature, a properly specified CGLS function can be expected to be more accurate than the biproportional method.

5. It is able to account for errors in the constraint totals. The problem can be reformulated to account for errors in the constraint totals (11).

6. It is able to produce estimates that are consistent with the available data. The quadratic methods also consistently improve the estimates and reduce the variance as more data are included in the formulation.

7. It is able to estimate unknown parameters in a linear distribution function. The ab initio estimation example clearly demonstrates this and the quadratic method is the only way of doing this without survey-derived data for several individual entries.

The problems associated with the regression formulations are primarily related to the computational burden associated with calculating the estimates. The expansion example used approximately 20 times as much computer time to obtain estimates using CLGS regression compared with the biproportional method but also provided estimates of the reliability of the entry estimates. However, the reduced cost of computers and particularly the availability of personal computers reduces this problem considerably.

Another problem arises if the formulation does not explicitly include nonnegativity constraints. The applications indicated that regression formulation is unlikely to produce negative estimates when a reasonable formulation is used. In the expansion example, no entry was estimated to be negative. For the ab initio estimation problem with row and column totals, negative estimates were not a problem. For the other formulation, the negative values were constrained to zero and the problem was reestimated. Although this approach is not statistically rigorous, it is practically workable.

Although it is generally an advantage to have the flexibility obtained by allowing the analyst freedom to specify any linear function and the structure of the variance-covariance matrix of the errors, it does create some problems if the formulation is incorrectly specified. Fortunately, the impact of the specification error can be evaluated if the analyst is aware of its existence (11). For example, if the variance-covariance matrix is incorrectly specified, the estimates are unbiased but inefficient. Unfortunately, however, only linear function can be included in the formulation to ensure convexity and a global optimal solution (8).

CONCLUSIONS

In this paper methods for estimating matrix entries using generalized regression have been reviewed and developed. With these methods it is possible to include all available and relevant information, including uncertain information and judgment. As well as including all information, the form of the objective function in the formulation is flexible and the resulting estimates are best linear unbiased estimates. Associated with each entry estimate is an estimated measure of their reliability such as the variance. The variance can be used to evaluate the estimates and derive strategies for additional sampling. In applications where the quadratic method is the only possible formulation or the estimates of the entries' uncertainty are relevant, the quadratic method is a feasible estimation method and should be used. The applications provided examples of the use of the quadratic formulation and demonstrated that the techniques produce reasonable results as well as being computationally feasible for fairly large matrices.

REFERENCES

1. P.R. Stopher and A.H. Meyburg. Urban Transportation Modelling and Planning. Lexington Books, Lexington, Mass., 1975.
2. Y. Sheffi and M. Sugiyama. Optimal Bus Scheduling on a Single Route. In Transportation Research Record 895, TRB, National Research Council, Washington, D.C., 1982, pp. 46-52.
3. M.A. Turnquist, A.H. Meyburg, and S.G. Ritchie. TOP: A Transit Operations Planning Model. Technical Report DOT/RSPA/DMA-50/83-25. U.S. Department of Transportation, March 1983.
4. H.J. Van Zuylen and L.G. Willumsen. The Most Likely Trip Matrix Estimated from Traffic Counts. Transportation Research, Vol. 14B, No. 3, 1980, pp. 281-294.
5. H.R. Kirby and M.N. Lesse. Trip Distribution Calculations and Sampling Error: Some Theoretical Aspects. Environment and Planning A, Vol. 10, 1978, pp. 837-851.
6. 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., 1983, pp. 5-10.
7. V.K. Chan, R.C. Dowling, and A.E. Willis. Deriving Origin-Destination Information from Routinely Collected Traffic Counts. Working Paper UCB-ITS-WP-80-1. Institute of Transportation Studies, University of California, Berkeley, 1980.
8. M. Carey, C. Hendrickson, and K. Siddharthan. A Method for Estimation of Origin/Destination Trip Matrices. Transportation Science, Vol. 15, No. 1, Feb. 1981, pp. 32-49.
9. F. Van Der Ploeg. Reliability and Adjustment of Sequences of Large Economic Accounting Matrices. Journal of the Royal Statistical Society, Series A, Vol. 145, Part 2, 1982, pp. 169-194.
10. T.C. Lee, G.G. Judge, and A. Zellner. Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. North-Holland Publishing Co., Amsterdam, 1970.
11. S. McNeil. Quadratic Matrix Entry Estimation. Ph.D. thesis. Carnegie-Mellon University, Philadelphia, Pa., 1983.
12. H. Theil. Principles of Econometrics. Wiley, New York, 1971.
13. M. Carey. Constrained Estimation of Direct Demand Functions and Trip Matrices. Working Paper. Carnegie-Mellon University, Philadelphia, Pa., 1982.
14. R.P. Byron. The Estimation of Large Scale Social Accounts Matrices. Journal of the Royal Statistical Society, Series A, Vol. 141, 1978, pp. 359-367.
15. M. Butterfield and T. Mules. A Testing Routine for Evaluating Cell by Cell Accuracy in Short-Cut Regional Input-Output Tables. Journal of Regional Science, Vol. 20, No. 3, 1980, pp. 293-310.
16. G.J.D. Hewings and B.N. Janson. Exchanging Regional Input-Output Coefficients: A Reply and Further Comments. Environment and Planning A, Vol. 12, 1980, pp. 843-854.
17. M. Bacharach. Biproportional Matrices and Input-Output Change. Cambridge University Press, New York, 1970.