JOURNAL OF REGIONAL SCIENCE, VOL. 26, NO. 3, 1986

# MAXIMUM LIKELIHOOD ESTIMATION OF GRAVITY MODEL PARAMETERS

Ashish Sen*

**ABSTRACT.** It is shown that, under some very mild conditions, maximum likelihood estimates of gravity model parameters exist and are unique (up to a scale transformation for some parameters). An algorithm for finding such estimates is also proposed.

## 1. INTRODUCTION

One of the most general forms of the gravity model is the following:

$$T_{ij} = A_i B_j F_{ij} \tag{1}$$

where $T_{ij}$ is the expected flow (e.g., of people, vehicles, money, or goods) from zone $i = 1, \ldots, I$ to zone $j = 1, \ldots, J$. The quantities $A_i$ and $B_j$ may be called origin and destination factors, respectively, and for the purposes of this paper, as indeed in most recent papers on the subject, $A_i$'s and $B_j$'s will be considered to be unknown parameters, the values of which may be estimated from observations of flows. (In the somewhat distant past the $A_i$'s and $B_j$'s have been, a priori, set equal to functions of factors such as population, and this has led to unpleasant properties of the model.) We shall call $F_{ij}$ a separation factor and assume that it has the form

$$F_{ij} = F(\mathbf{c}_{ij}, \theta) = \exp\left[\sum_{k=1}^{K} c_{ij}^{(k)} \theta_k\right] \tag{2}$$

where $c_{ij}^{(k)}$ are $K$ measures of separation (e.g., travel time, distance, generalized cost, psychological distance, etc.) and $\theta_k$ are unknown parameters. It may readily be verified that both the frequently used forms

$$F_{ij} = \exp[-c_{ij}\beta] \text{ and } F_{ij} = c_{ij}^{\gamma} \exp[-\beta c_{ij}]$$

and the step function are particular cases of (2). Hence, since $\Sigma_{k=1}^{K} c_{ij}^{(k)} \theta_k$ includes polynomials, (2) is as general a function as we are likely to need.

Let $X_{ij}$ be the observed flow from $i$ to $j$. Then $E(X_{ij}) = T_{ij}$. The aim of this paper is to discuss the maximum likelihood (ML) estimation of the parameter vector

$$\zeta = (A_i, \ldots, A_I, B_1, \ldots, B_J, \theta_1, \ldots, \theta_k)'$$

from the observations $X_{ij}$ where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. These estimates will be denoted by a hat or circumflex; e.g., $\hat{A}_i$, $\hat{B}_j$, $\hat{\theta}_k$, and $\hat{\zeta}$ are, respectively, ML estimates of $A_i$, $B_j$, $\theta_k$, and $\zeta$; and $\hat{T}_{ij} = \hat{A}_i \hat{B}_j \hat{F}_{ij}$ is an estimate of $T_{ij} = A_i B_j F_{ij}$. Since,

*Professor of Urban Planning and Policy, University of Illinois at Chicago.

Received July, 1984; in revised form May, 1985; accepted June, 1985.

in most applications of the gravity model, little attention is paid to the $A_i$ and $B_j$ terms, we also will concentrate mainly on estimating $\theta_k$'s.

We shall assume that the $X_{ij}$'s have independent Poisson distributions. It can readily be shown that the results of this paper would not be altered if we had assumed that $X_{ij}$'s had the multinomial distribution. In most of the gravity model literature in which distributions are explicitly discussed, either a Poisson or a multinomial distribution is assumed. Intuitively, one should also expect one of these distributions to hold at least approximately, since the $X_{ij}$'s are usually counts of units whose behavior is roughly independent of that of other units. [See also Smith (1984a, 1984b).]

Since each $X_{ij}$ has a Poisson distribution with $E(X_{ij}) = T_{ij}$, the probability function for $X_{ij}$ is

$$P(X_{ij}|T_{ij}) = \exp[-T_{ij}]\ T_{ij}^{X_{ij}}/X_{ij}!$$

and since the $X_{ij}$'s have been assumed independent, their joint distribution is given by the probability function

$$\begin{aligned} &\prod_{ij} \exp[-T_{ij}]\ T_{ij}^{X_{ij}}/X_{ij}! \\ &= \prod_{ij} \{\exp[-A_iB_jF_{ij}]\}\ \{[A_iB_jF_{ij}]^{X_{ij}}\}/X_{ij}! \end{aligned} \tag{3}$$

If we treat the $T_{ij}$'s as constants and the expression (3) mainly as a function of the $X_{ij}$'s, (3) gives the probabilities of occurrence of each set of values $X_{11}, X_{12}, \ldots$, etc. On the other hand, if we treat the $X_{ij}$'s as constants and (3) as a function of the $A_i$'s, $B_j$'s, and $\theta_k$'s, (3) is a likelihood function and values of $A_i$, $B_j$, and $\theta_k$ which maximize (3) are the ML estimates. This paper is concerned with such estimates which have a number of very nice properties [Kendall and Stuart (1967), Rao (1973)].

Since $\ln \lambda$ is a monotonically increasing function of $\lambda$, maximizing (3) is equivalent to maximizing its logarithm, which is

$$\mathcal{L} = \sum_{ij} [-A_iB_jF_{ij} + X_{ij} \ln A_i + X_{ij} \ln B_j + X_{ij} \ln F_{ij} - \ln (X_{ij}!)] \tag{4}$$

In order to find necessary conditions for a maximum of (4) we need to set the partial derivatives of (4) with respect to each of $A_i$, $B_j$, and $\theta_k$ equal to zero. Thus, starting with $A_i$ we have

$$\frac{\partial \mathcal{L}}{\partial A_i} = \sum_i [-B_jF_{ij} + X_{ij}/A_i] = 0 \tag{5}$$

Note that in (5) terms not involving $A_i$ are zero, which is why we were able to go from $\Sigma_{ij}$ in (4) to $\Sigma_i$ in (5). From (5), we have

$$\sum_i A_iB_jF_{ij} = \sum_i T_{ij} = \sum_i X_{ij} \tag{6}$$

A notation that is very convenient and one we shall use in this paper is to replace a

subscript by a "+" when we add over all values of it. For example,

$$T_{i+} = \sum_{j=1}^{J} T_{ij} \qquad T_{+j} = \sum_{i=1}^{I} T_{ij} \qquad T_{++} = \sum_{i=1}^{I} \sum_{j=1}^{J} T_{ij}$$

Using this notation, (6) becomes

$$T_{+j} = X_{+j} \tag{7}$$

There are obviously $I$ equations of this form—one for each $A_i$. Similarly, by considering the $\partial \mathcal{L}/\partial B_j$'s we have $J$ equations

$$T_{i+} = X_{i+} \tag{8}$$

and considering $\partial \mathcal{L}/\partial \theta$'s we obtain, using (2), the $K$ equations

$$\sum_{ij} c_{ij}^{(k)} T_{ij} = \sum_{ij} c_{ij}^{(k)} X_{ij} \tag{9}$$

In the next section of this paper we shall present the principal theorem of the paper. This theorem provides conditions under which Equations (7), (8), and (9) have solutions. Moreover, the theorem states that these solutions are ML estimates (MLE's), and we also make statements about uniqueness. In particular, we provide conditions under which the MLE's $\theta_k$ of $\theta_k$ are unique. (Notice that $\hat{A}_i$'s and $\hat{B}_j$'s can never be unique, since if $\hat{A}_i$ and $\hat{B}_j$ are MLE's, then $\gamma \hat{A}_i$ and $\gamma^{-1} \hat{B}_j$ must be MLE's for arbitrary $\gamma$'s.)

The theorem will be proved in Section 3, and in Section 4 we propose an algorithm for obtaining the MLE's. A concluding thought is presented briefly in Section 5.

## 2. THE PRINCIPAL THEOREM

Before we can state the theorem we need to introduce some further notation. If we use lower-case letters to denote the logarithms of variables denoted by corresponding capital letters (e.g., $\ln A_i = a_i$), we can write the gravity model (1) as

$$t_{ij} = a_i + b_j + f_{ij} = a_i + b_j + \sum_{k=1}^{K} c_{ij}^{(k)} \theta_k \tag{10}$$

We shall assume that $t_{ij}$'s exist—i.e., all $T_{ij}$'s are positive. The only way this could fail to occur is if some $A_i$ or $B_j$ is zero. Then all $T_{ij}$'s for that origin or destination would be zero and we could delete them from consideration. Then it follows from (7) and (8) that $X_{i+} > 0$ and $X_{+j} > 0$. Since $\ln(X)$ is a one-to-one function, knowing $a_i$ is equivalent to knowing $A_i$. Thus, with the gravity model expressed as (10), the vector of unknown parameters is

$$\xi = (a_1, \ldots, a_I, b_1, \ldots, b_J, \theta_1, \ldots, \theta_k)'$$

When we wish to refer to the $a_i$'s, $b_j$'s, and $\theta_k$'s without concern for which ones we are referring to, we shall refer to them as components $\xi_n$ of $\xi$.

There are $IJ$ equations of the form (10). Consider the coefficient matrix of the right-hand side of this system (i.e., the matrix of coefficients of $a_i$'s, $b_j$'s, and $\theta_k$'s).

| | | Column Corresponding to | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Row | (i,j) | $A_1$ | $A_2$ | $A_3$ | $B_1$ | $B_2$ | $B_3$ | $\theta_1$ | $\theta_2$ |
| 1 | (1,1) | 1 | 0 | 0 | 1 | 0 | 0 | $c_{11}^{(1)}$ | $c_{11}^{(2)}$ |
| 2 | (1,2) | 1 | 0 | 0 | 0 | 1 | 0 | $c_{12}^{(1)}$ | $c_{12}^{(2)}$ |
| 3 | (1,3) | 1 | 0 | 0 | 0 | 0 | 1 | $c_{13}^{(1)}$ | $c_{13}^{(2)}$ |
| 4 | (2,1) | 0 | 1 | 0 | 1 | 0 | 0 | $c_{21}^{(1)}$ | $c_{21}^{(2)}$ |
| 5 | (2,2) | 0 | 1 | 0 | 0 | 1 | 0 | $c_{22}^{(1)}$ | $c_{22}^{(2)}$ |
| 6 | (2,3) | 0 | 1 | 0 | 0 | 0 | 1 | $c_{23}^{(1)}$ | $c_{23}^{(2)}$ |
| 7 | (3,1) | 0 | 0 | 1 | 1 | 0 | 0 | $c_{31}^{(1)}$ | $c_{31}^{(2)}$ |
| 8 | (3,2) | 0 | 0 | 1 | 0 | 1 | 0 | $c_{32}^{(1)}$ | $c_{32}^{(2)}$ |
| 9 | (3,3) | 0 | 0 | 1 | 0 | 0 | 1 | $c_{33}^{(1)}$ | $c_{33}^{(2)}$ |

FIGURE 1. Example of Matrix $M$.

Call this matrix $M$. Such a matrix for $I = J = 3$ and $K = 2$ is illustrated in Figure 1. The transpose $M'$ of $M$ is the matrix of coefficients of $T_{ij}$'s in the system of $I + J + K$ equations in (7), (8), and (9).

It may be seen that the sum of the first $I$ columns of $M$ is a vector consisting only of 1's as is the sum of columns $I + 1$ to $I + J$. Thus the first $I + J$ columns of $M$ are linearly dependent, and the rank of $M$ is $I + J + K - \upsilon$ where $\upsilon$ must be at least 1. It is not difficult to show [e.g., following Evans (1970, 1971)] that the rank of the matrix consisting of the first $I + J$ columns of $M$ is exactly $I + J - 1$. Thus, whether $\upsilon = 1$ or $\upsilon > 1$ depends principally on the last $K$ columns of $M$—i.e., on the $c_{ij}^{(k)}$'s. If these columns are either (i) mutually linearly dependent or (ii) linearly dependent on the remaining columns, then $\upsilon$ will be greater than 1; otherwise $\upsilon = 1$. An example of (ii) is where $c_{ij}^{(k)}$ is the parking cost which is destination specific. Then it is too much for any estimating procedure to separate the effect of such a $c_{ij}^{(k)}$ from those of $b_j$'s. Usually, for a well-specified problem we should have $\upsilon = 1$.

We can now state our principal theorem:

*Theorem* 1: The two conditions

(C1) There exist $IJ$ positive numbers $y_{ij}^{(0)}$ such that

$$(11) \qquad y_{i+}^{(0)} = X_{i+} \qquad y_{+j}^{(0)} = X_{+j} \qquad \sum_{ij} c_{ij}^{(k)} y_{ij}^{(0)} = \sum_{ij} c_{ij}^{(k)} X_{ij}$$

for all $i, j, k$ (i.e., there exists a solution to (7), (8), and (9), but not necessarily of the form (1)], and

(C2) The rank of $M$ is $I + J + K - 1$

are necessary and sufficient for the existence of a unique vector $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)'$

which is the MLE of $\theta$ and for which numbers $\hat{A}_1, \ldots, \hat{A}_I$ and $\hat{B}_1, \ldots, \hat{B}_J$ can be found (not uniquely) so as to satisfy (1), (2), (7), (8), and (9).

Condition (C1) deserves some attention. Figure 2 illustrates a situation where $y_{ij}^{(0)}$'s satisfying (11) do not exist. But such situations are rare. In order to investigate them further, first set $K = 1$. Then let

$$\mathbf{y} = \mathcal{V}(y_{ij}) = (y_{11}, y_{12}, \ldots, y_{1J}, y_{21}, \ldots, y_{2j}, \ldots, y_{I1}, \ldots, y_{IJ})'$$

[The symbol $\mathcal{V}$ simply writes the matrix $(y_{ij})$ of $y_{ij}$'s as a vector.] Consider the region $\mathcal{R}^{(0)}$ of $\mathbf{y}$'s such that

$$(12) \qquad y_{i+} = X_{i+} \qquad y_{+j} = X_{+j} \qquad y_{ij} > 0 \qquad \text{(for all } i \text{ and } j\text{)}$$

K=1

$((c_{ij}))$

$$\begin{pmatrix} 1 & 50 \\ 50 & 1 \end{pmatrix}$$

$((X_{ij}))$

$$\begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}$$

For any $\chi^{(0)}$ such that $y_{i+}^{(0)} = X_{i+}$, $y_{+j}^{(0)} = X_{+j}$ for all i and j, we have

$$\begin{aligned} \sum_{ij} c_{ij} y_{ij}^{(0)} - \sum_{ij} c_{ij} X_{ij} &= 50 (y_{12}^{(0)} + y_{21}^{(0)}) \\ &\quad - (X_{11} + X_{22} - y_{11}^{(0)} - y_{22}^{(0)}) \\ &= 50 (y_{12}^{(0)} + y_{21}^{(0)}) - (15 - y_{11}^{(0)} - y_{22}^{(0)}) \\ &= 50 (y_{12}^{(0)} + y_{21}^{(0)}) - (y_{12}^{(0)} + y_{21}^{(0)}) \\ &\quad (\text{since } \sum_{ij} y_{ij}^{(0)} = 15) \\ &= 49 (y_{12}^{(0)} + y_{21}^{(0)}) > 0 \end{aligned}$$

whenever the $y_{ij}^{(0)}$'s are positive. Thus

$$\sum_{ij} c_{ij} y_{ij}^{(0)} \neq \sum_{ij} c_{ij} X_{ij}$$

This example also illustrates that the $X_{ij}$'s are in a lowest cost configuration.

FIGURE 2. Example in Which $y_{ij}^{(0)}$'s Satisfying (11) Do Not Exist.

To see that $\mathcal{R}^{(0)}$ is nonempty, note that if $X_{ij}$'s are positive we can set $y_{ij} = X_{ij}$. If some $X_{st} = 0$, then there must be a nonzero $X_{sv}$ and a nonzero $X_{ut}$ (since $X_{i+} > 0$ and $X_{ij} > 0$). Then, for a small enough $\delta > 0$, if we set $z_{st} = \delta$, $z_{sv} = X_{sv} - \delta$, $z_{ut} = X_{ut} - \delta$, $z_{uv} = X_{uv} + \delta$ and $z_{ij} = X_{ij}$ otherwise, then $z_{st} > 0$, $z_{ij} > 0$ whenever $X_{ij} > 0$ and $z_{i+} = X_{i+}$, $Z_{+j} = X_{+j}$. If we repeat a similar procedure for all nonzero $X_{ij}$'s we would ultimately reach a point in $\mathcal{R}^{(0)}$.

Now it can be shown that if $y_{ij}^{(0)}$'s obeying (12) fail to exist [i.e., they obey (12), but not (11)], $X_{ij}$'s are in an extremal configuration with respect to costs $c_{ij}^{(1)}$—i.e., we have either

$$\sum_{ij} c_{ij}^{(1)} y_{ij} > \sum_{ij} c_{ij}^{(1)} X_{ij} \text{ for all } \mathbf{y} \in \mathcal{R}^{(0)}$$

or

$$\sum_{ij} c_{ij}^{(1)} y_{ij} < \sum_{ij} c_{ij}^{(1)} X_{ij} \text{ for all } \mathbf{y} \in \mathcal{R}^{(0)}$$

To prove this, assume that we have $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ in $\mathcal{R}^{(0)}$ such that

$$\sum_{ij} c_{ij}^{(1)} y_{ij}^{(1)} < \sum_{ij} c_{ij}^{(1)} X_{ij} < \sum_{ij} c_{ij}^{(1)} y_{ij}^{(2)}$$

Then for all $\alpha$, $0 < \alpha < 1$, $\mathbf{y}^{(\alpha)} = \alpha \mathbf{y}^{(1)} + (1 - \alpha)\, \mathbf{y}^{(2)}$ is in $\mathcal{R}^{(0)}$ and for some $\alpha$

$$\sum_{ij} c_{ij}^{(1)} y_{ij}^{(\alpha)} = \sum_{ij} c_{ij}^{(1)} X_{ij}$$

violating the nonexistence of $y_{ij}^{(0)}$'s.

This result, although worthwhile, is not entirely startling. The maximum value of $\mathcal{L}$, if it does not lie in $\mathcal{R}^{(0)}$, must lie on the boundary of $\mathcal{R}^{(0)}$ where $\mathcal{V}(X_{ij})$ is. The result also reminds one of the fact that minimum and maximum cost configurations occur when $\theta_1 \to \infty$ or $\theta_1 \to -\infty$ [Evans (1973)]. It can be extended to $K \geq 1$ in the following way:

*Lemma* 1: Let $R^{(K_1)}$ denote a nonempty region

$$y_{i+} = X_{i+} \qquad y_{+j} = X_{+j} \qquad \sum_{ij} c_{ij}^{(k)} y_{ij} = \sum_{ij} c_{ij}^{(k)} X_{ij}$$

$$y_{ij} > 0 \qquad \text{for all } i \text{ and } j \text{ and } k \leq K_1$$

Then if

$$\sum_{ij} c_{ij}^{(n)} y_{ij} \neq \sum_{ij} c_{ij}^{(n)} X_{ij}$$

for all $\mathbf{y} \in \mathcal{R}^{(k_1)}$ and $K_1 < n \leq K$ we must have either

$$\sum_{ij} c_{ij}^{(n)} y_{ij} < \sum_{ij} c_{ij}^{(n)} X_{ij} \qquad \text{for all } \mathbf{y} \in \mathcal{R}^{(K_1)}$$

or

$$\sum_{ij} c_{ij}^{(n)} y_{ij} > \sum_{ij} c_{ij}^{(n)} X_{ij} \qquad \text{for all } \mathbf{y} \in \mathcal{R}^{(K_1)}$$

*Proof:* Similar to the discussion above.

## 3. PROOF OF THEOREM 1

Results similar to Theorem 1 have been proved by Haberman (1974) and others for contexts other than the gravity model. Hence, we shall mainly adapt Haberman's work to the gravity model.

*Lemma* 2 (Haberman): A necessary and sufficient condition for the existence of MLE's $\hat{T}_{ij}$ of the form given by (1) and (2) is that there exists a vector $\eta = \mathcal{V}(\eta_{ij})$ such that $M'\eta = 0$ and $X_{ij} + \eta_{ij} > 0$ for all $i$ and $j$. Moreover, such $\hat{T}_{ij}$'s are unique, and solve Equations (7), (8), and (9).

This is essentially a restatement of Theorem 2.1 and 2.2 in Haberman (1974, pp. 35–37).

*Lemma* 3: A necessary and sufficient condition for the existence of MLE's of the form given by (1) and (2) is condition (C1). Moreover, such $\hat{T}_{ij}$'s are unique and solve Equations (7), (8), and (9).

*Proof:* Let $\mathbf{X} = \mathcal{V}(X_{ij})$ and $\mathbf{y}^{(0)} = \mathcal{V}(y_{ij}^{(0)})$. Then from (11), $M'X = M'y^{(0)}$. Thus setting $\eta = \mathbf{y}^{(0)} - \mathbf{X}$ the lemma follows from Lemma 2.

From the above we have the uniqueness of the $\hat{T}_{ij}$'s but we still need to show the uniqueness of $\theta$.

*Lemma* 4: Let $T_{ij}$'s be positive numbers of the form given by (1) and (2). Then (C2) is necessary and sufficient for the uniqueness of $\theta$.

*Proof:* Call the columns of $M$, $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{I+J+K}$. Recall that under conditions of the lemma, $t_{ij} = \ln T_{ij}$ can be written in the form (10).

Suppose there were two distinct values $\theta = (\theta_1, \ldots, \theta_k)'$ and $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_K)'$ such that $(\xi_1, \ldots, \xi_{I+J}, \theta_1, \ldots, \theta_K)'$ and $(\tilde{\xi}_1, \ldots, \tilde{\xi}_{I+J}, \tilde{\theta}_1, \ldots, \tilde{\theta}_K)'$ were both solutions to (10), that is

$$\mathbf{t} = \sum_{s=1}^{I+J} \xi_s \mathbf{m}_s + \sum_{k=1}^{K} \theta_k \mathbf{m}_{I+J+K}$$

and

$$\mathbf{t} = \sum_{s=1}^{I+J} \tilde{\xi}_s \mathbf{m}_s + \sum_{k=1}^{K} \tilde{\theta}_k \mathbf{m}_{I+J+K}$$

On subtracting the second of these equations from the first we have

$$0 = \sum_{s=1}^{I+J} (\tilde{\xi}_s - \xi_s)\, \mathbf{m}_s + \sum_{k=1}^{K} (\tilde{\theta}_k - \theta_k)\, \mathbf{m}_{I+J+K} \tag{13}$$

Since $\tilde{\theta} \neq \theta$, some $\theta_k \neq \tilde{\theta}_k$. Let $\theta_1 \neq \tilde{\theta}_1$. Then from (13)

$$\mathbf{m}_{I+J+1} = \sum_{s=1}^{I+J} (\tilde{\xi}_s - \xi_s)/(\theta_1 - \tilde{\theta}_1)\, \mathbf{m}_s + \sum_{k=2}^{K} (\tilde{\theta}_k - \theta_k)/(\theta_1 - \tilde{\theta}_1)\, \mathbf{m}_{I+J+k} \tag{14}$$

One of the $\mathbf{m}_s$, say $\mathbf{m}_1$, can be eliminated from the right-hand side of (14) since $\Sigma_{s=1}^{I} \mathbf{m}_s = \Sigma_{s=I+1}^{I+J} \mathbf{m}_s$. Then we have two columns $\mathbf{m}_{I+J+1}$ and $\mathbf{m}_1$, both of which can be expressed in terms of the remaining $I + J + K - 2$ columns of $M$. Therefore $M$ can have at most $I + J + K - 2$ independent columns. This contradicts (C2) and proves that when $M$ has rank $I + J + K - 1$, $\theta$ is unique.

The necessity of condition (C1) follows immediately on noting that if $\theta$ is unique and if we arbitrarily chose one other $\xi_i$ the remaining $\xi_i$'s would be uniquely determined by the $T_{ij}$'s.

## 4. A COMPUTATIONAL PROCEDURE

Several procudures have been given for obtaining ML estimates of $\theta_k$'s for some special cases of (2). For example, when $K = 1$, effective procedures have been supplied by Evans (1971), Hyman (1969), and others, and when $F_{ij}$ is a step function, a procedure given by Evans and Kirby (1974) works very well. Batty (1976) has presented and compared several procedures for obtaining ML estimates when $K = 2$, and these procedures can be easily generalized to when $K > 2$. In this section we propose a procedure which is possibly significantly faster than any of those described by Batty (1976).

The procedure we propose is a gradient search procedure; but unlike the gradient search procedure described by Batty (1976), a key step within each iteration is handled analytically (rather than through a search algorithm). We believe that this significantly cuts down on computer time. This conjecture has been borne out in trials using small ($5 \times 5$) matrices of $X_{ij}$'s. Figure 3 shows a trace of the iterations for our gradient search procedure. Figure 4 shows a trace using the same data and an alternative method—often called the method of scoring in statistics [Rao (1973)]. This latter procedure [also given in Batty (1976)] is essentially based on the multidimensional Newton-Raphson procedure and consists of taking the linear part of Taylor expansions of the nonlinear ML equations about the point $\xi^{(r)}$ obtained from the previous iteration and then solving these linear equations to obtain $\xi^{(r+1)}$ for the next iteration. For this problem $\theta = (\theta_1, \theta_2)$, and this procedure requires three initial points (0-1, 0-2, 0-3). If we notice the different scales used in the two figures we see that, at least in this case, the gradient search procedure does better than its competitor. While this was one of the less favorable performances of the scoring procedure among several hundred known to this author [some of these, including details of this one, are given in Sue (1985)], it does illustrate comparative merits of the two procedures. Our gradient search procedure always moved smoothly towards the estimate, while the scoring procedure zig-zagged, sometimes causing overflow or underflow. However, more and larger scale demonstrations are needed before its computational behavior is fully understood.

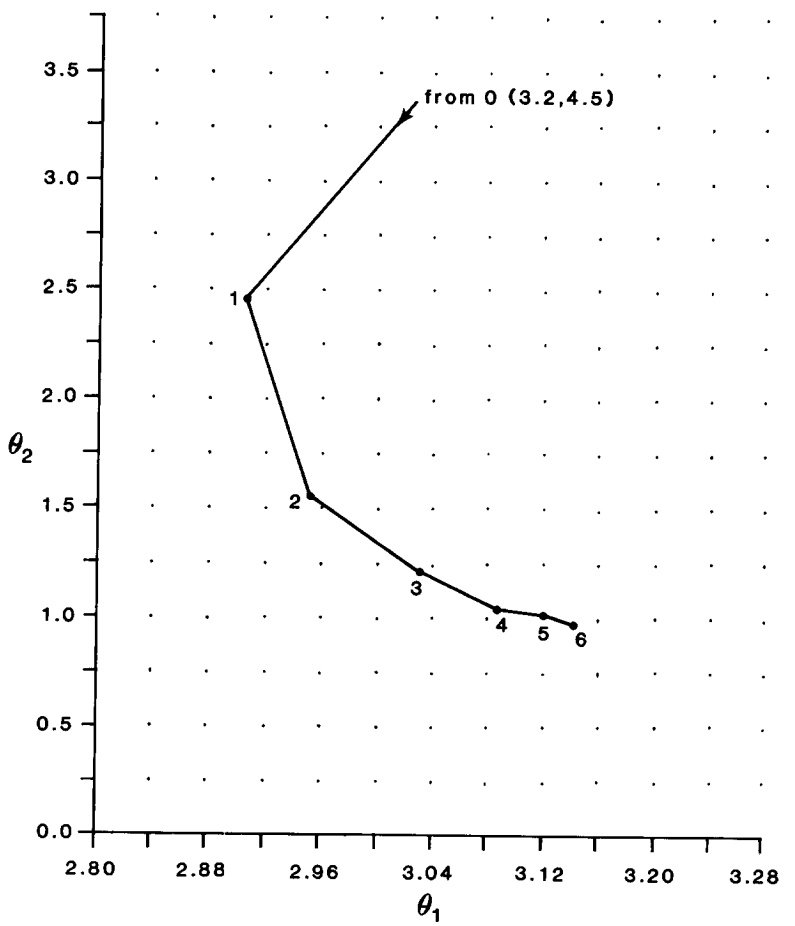


FIGURE 3: Trace of Gradient Search.

In order to describe the procedure, we shall first outline gradient search procedures in general. For some value $\hat{\zeta}^{(r)}$ of $\zeta$, denote by grad $(\mathcal{L}, \hat{\zeta}^{(r)})$ the gradient vector of $\mathcal{L}$ at $\hat{\zeta}^{(r)}$

$$\text{grad } (\mathcal{L}, \hat{\zeta}^{(r)}) = \left(\frac{\partial \mathcal{L}}{\partial \zeta_1}, \frac{\partial \mathcal{L}}{\partial \zeta_2}, \ldots, \frac{\partial \mathcal{L}}{\partial \zeta_{I+J+K}}\right)'$$

where $\zeta$ is as in Section 1, and the partial derivatives are evaluated at $\hat{\zeta}^{(r)}$. It is known that of all directions around $\hat{\zeta}^{(r)}$, $\mathcal{L}$ increases the fastest if we choose the direction given by grad $(\mathcal{L}, \hat{\zeta}^{(r)})$. Therefore, we augment $\hat{\zeta}^{(r)}$ to

$$\hat{\zeta}^{(r)} + \rho \text{ grad } (\mathcal{L}, \hat{\zeta}^{(r)}) \tag{15}$$

and choose a value $\rho^{(r)}$ of $\rho$ as indicated below. If $\mathcal{L}$ were linear it would keep increasing as $\rho$ were increased. However, $\mathcal{L}$ is not linear and is bounded and for some value $\rho^{(r)}$ of $\rho$ it attains its maximum value over all points (15). [Notice that $\rho$

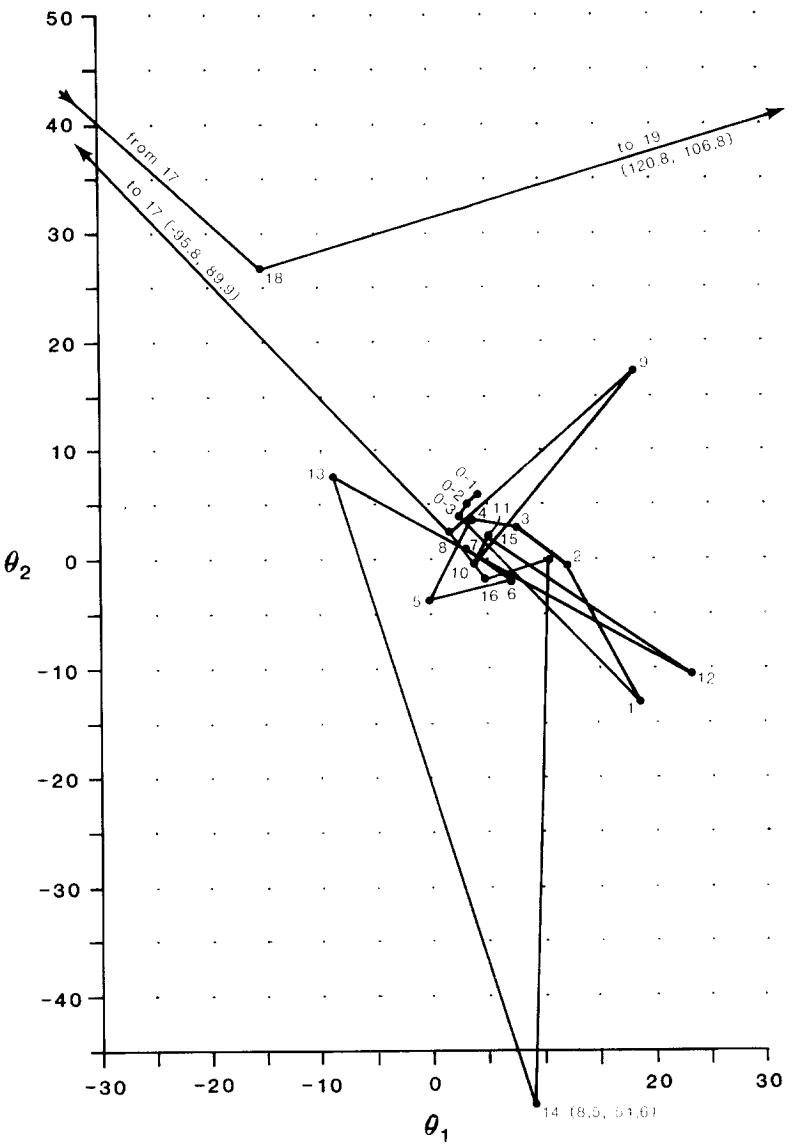


FIGURE 4: Trace of Scoring Method.

is the only variable in (15).] This value of $\rho$ can be obtained sometimes by setting the derivative of (15) with respect to $\rho$ equal to zero; when this is difficult, as it often is in practice, an approximate value of $\rho^{(r)}$ is found by evaluating (15) for several values of $\rho$ in a method of enumeration. Most gradient search algorithm packages provide the latter option. Once such a $\rho^{(r)}$ is found, we would set $\hat{\zeta}^{(r+1)}$ equal to (15) and continue our iterations, at each stage raising the value of $\mathcal{L}$ until we are satisfied.

Now let us return to the specific form of $\mathcal{L}$ that we have. As for (7), (8), and (9) we can readily compute

$$\frac{\partial \mathcal{L}}{\partial \zeta_i} = \frac{\partial \mathcal{L}}{\partial A_i} = A_i^{-1}(X_{i+} - T_{i+}) \qquad \text{(when } 1 \le i \le I\text{)}$$

$$\frac{\partial \mathcal{L}}{\partial \zeta_{I+j}} = \frac{\partial \mathcal{L}}{\partial B_j} = B_j^{-1}(X_{+j} - T_{+j}) \qquad \text{(when } 1 \le j \le J\text{)}$$

and

$$\frac{\partial \mathcal{L}}{\partial \zeta_{I+J+k}} = \frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_{ij} X_{ij} c_{ij}^{(k)} - \sum_{ij} T_{ij} c_{ij}^{(k)} = \nu_k^{(r)} \text{ (say)} \qquad \text{(when } 1 \le k \le K\text{)}$$

For any set of positive $F_{ij}$'s, we can easily solve $\partial \mathcal{L}/\partial A_i = 0$ and $\partial \mathcal{L}/\partial B_j = 0$ [which are, of course, the same as (7) and (8)] for all $i$ and $j$ using the well-known Furness iterations, which we shall call the DSF procedure, after Deming and Stephan (1940) and Furness (1965), who apparently independently discovered it. After choosing a set of arbitrary positive $A_i^{(0)}$'s, the procedure consists of using alternately and iteratively the two formulae

$$A_i^{(2p)} = X_{i+} \Big/ \sum_{j=1}^{J} B_j^{(2p-1)} F_{ij}$$

$$B_j^{(2p+1)} = X_{+j} \Big/ \sum_{i=1}^{I} A_i^{(2p)} F_{ij}$$

until an adequate level of convergence is attained.

Suppose we start with some value $\theta^{(r)}$ of $\theta$ and use the DSF procedure to obtain the corresponding $A_i^{(r)}$'s, $B_j^{(r)}$'s, and $T_{ij}(\theta^{(r)})$'s. We see that $\partial \mathcal{L}/\partial A_i = \partial \mathcal{L}/\partial B_j = 0$ for all $i$ and $j$ (approximately so, since the DSF procedure is iterative). We use the symbol $T_{ij}(\theta^{(r)})$ to indicate that the $T_{ij}$'s are functions of $\theta^{(r)}$. Then, letting $\zeta^{(r)} = (A_i^{(r)}, \ldots, A_I^{(r)}, B_1^{(r)}, \ldots, B_j^{(r)}, \theta_1^{(r)}, \ldots, \theta_K^{(r)})'$, only the last $K$ components of grad $(\mathcal{L}, \zeta^{(r)})$ would be nonzero and these would be $\nu_k^{(r)}$. Thus in order to get (15) we would leave the $A_i^{(r)}$'s and $B_j^{(r)}$'s unchanged and augment each $\theta_k^{(r)}$ by $\rho^{(r)}\nu_k^{(r)}$.

In order to obtain $\rho^{(r)}$, consider

$$\frac{d\mathcal{L}}{d\rho} = \sum_{s=1}^{I+J+K} \frac{\partial \mathcal{L}}{\partial \zeta_s} \frac{d\zeta_s}{d\rho} \approx \sum_{k=1}^{K} \frac{d\mathcal{L}}{d\theta_k} \frac{d\theta_k}{d\rho} \tag{16}$$

(all other $\partial \mathcal{L}/\partial \zeta_s$'s being approximately zero)

$$= \sum_{k=1}^{K} \left( \sum_{ij} c_{ij}^{(k)} X_{ij} - \sum_{ij} c_{ij}^{(k)} A_i^{(r)} B_j^{(r)} \exp\left[ \sum_{l=1}^{K} c_{ij}^{(l)} (\theta_l^{(r)} + \rho \nu_l^{(r)}) \right] \nu_k^{(r)} \right)$$

$$= \sum_{k=1}^{K} \left( \sum_{ij} c_{ij}^{(k)} X_{ij} - \sum_{ij} c_{ij}^{(k)} T_{ij}(\theta^{(r)}) \exp\left[ \sum_{l=1}^{K} c_{ij}^{l} \rho \nu_l^{(r)} \right] \right) \nu_k^{(r)}$$

$$= \delta(\rho), \text{ say}$$

Thus solving the equation $\delta(\rho) = 0$ for $\rho$ will give us the desired value $\rho^{(r)}$. Since

$$\delta(\rho) = \sum_{k=1}^{K} (\nu_k^{(r)})^2 \geq 0$$

when $\rho = 0$ and $d\delta(\rho)/d\rho$

$$= -\sum_{k=1}^{K} \sum_{ij} \nu_k^{(r)} c_{ij}^{(k)} T_{ij}(\theta^{(r)}) \sum_{l=1}^{K} c_{ij}^{(l)} \nu_l^{(r)} \exp\left(\sum_{s=1}^{K} \rho c_{ij}^{(s)} \nu_s^{(r)}\right)$$

$$= -\sum_{ij} \left(T_{ij}(\theta^{(r)}) \exp\left(\sum_{s=1}^{K} c_{ij}^{(s)} \nu_s^{(r)}\right)\left(\sum_{k=1}^{K} c_{ij}^{(k)} \nu_k^{(r)}\right)^2\right) \leq 0$$

there is one and only one such solution.

However, in our efforts to work with (16) we found that a modification of it substantially reduced the number of interations. This modification consists of writing, instead of (16),

$$(17) \qquad \sum_{k=1}^{K}\left[\sum_{ij} c_{ij}^{(k)} X_{ij} - \tau \sum_{ij} c_{ij}^{(k)} T_{ij}(\theta^{(r)}) \exp\left(\sum_{l=1}^{K} c_{ij}^{(l)} \rho \nu_l^{(r)}\right)\right] \nu_k^{(r)} = \delta^*(\rho)$$

where

$$\tau^{-1} = \sum_{ij} T_{ij}(\theta^{(r)}) \exp\left(\sum_{l=1}^{K} c_{ij}^{(l)} \rho \nu_l^{(r)}\right) \Big/ T_{++}$$

Let us now summarize the gradient search procedure:

*Step* 1: Choose a value $\theta^{(0)}$ of $\theta$. For the procedure to work efficiently $\theta^{(0)}$ should be a reasonably good estimate of $\hat{\theta}$. Now compute $F(\mathbf{c}_{ij}, \theta)$ and (i) use the DSF procedure to obtain the $T_{ij}(\theta^{(0)})$'s corresponding to $\theta^0$. These $t_{ij}(\theta^{(0)})$'s approximately satisfy (7) and (8); (ii) then set

$$\nu_k^{(0)} = \sum_{ij} c_{ij}^{(k)} [X_{ij} - T_{ij}(\theta^{(0)}]$$

and compute the terms in $\delta^*(\rho)$, given in Equation (17). Solve the equation $\delta^*(\rho) = 0$ using the Newton-Raphson or other procedure for $\rho$ and call the unique solution $\rho^{(0)}$. Now obtain $\theta^{(1)}$ from (15).

This $\theta^{(1)}$ is used in Step 2 to obtain $\theta^{(2)}$, etc. At the $(r + 1)$th step we have:

*Step* $r + 1$: (i) From the value $\theta^{(r)}$ given by the previous step, compute $F(\mathbf{c}_{ij}, \theta^{(r)})$ and then use the DSF procedure to obtain the $T_{ij}(\theta^{(r)})$ which satisfy (7) and (8).

(ii) Compute $\nu_k^{(r)}$ and the terms in $\delta^*(\rho)$. Find the unique solution $\rho^{(r)}$ to $\delta^*(\rho) = 0$ and obtain $\theta^{(r+1)}$ from (15).

Continue this process until the $\nu_k^{(r)}$'s are small enough indicating that (7), (8), and (9) are approximately satisfied.

As already stated, in the above $\delta(\rho)$ could be used in lieu of $\delta^*(\rho)$ but our experience indicates that this would slow down the rate of convergence of the procedure. Notice that each step contains two substeps: in the first we hold $\theta^{(r)}$

constant and choose $A_i^{(r)}$'s and $B_j^{(r)}$'s to maximize $\mathcal{L}$, and in the second we hold $A_i^{(r)}$'s and $B_j^{(r)}$'s constant and choose $\theta^{(r+1)}$ such that $\mathcal{L}$ increases rapidly. From the form of the functions involved it would appear that convergence would be reasonably rapid, although gradient search procedures in general are often quite slow. Notice also that changes in $\theta^{(r)}$ from step to step will occur as long as all equations in (7), (8), and (9) are not satisfied, and the extent to which (9) is not satisfied determines the amount of this change.

Although we do not have enough experience to recommend them, some short-cuts may be possible in the procedure. One of the more time-consuming parts of the algorithm is the repeated computation of the "exp" function in $\delta^*(\rho)$. In the earlier part of the algorithm (i.e., for small $r$) perhaps we can be fairly rough; as such, rather than using the computer-supplied algorithm for "exp" a simpler but less accurate formula consisting of the first two or three terms of the Taylor Series for "exp" may suffice. During the later stages of the iteration the argument of "exp" will be small enough to render this Taylor Series formula quite accurate. Consequently, for all iterations such an approximation may be desirable. Furthermore, with this approximation the equation $\delta^*(\rho) = 0$ can be solved directly by formula; hence, the iterative Newton-Raphson algorithm would become unnecessary. On the other hand, it may increase the total number of iterations needed. Thus, while we conjecture that a linear approximation will lead to greater efficiencies in many cases, it is not clear whether this is true for all cases.

## 5. A CONCLUDING REMARK

Although we conjecture that the procedure given in the last section is simpler than other comparable procedures, it is still a fairly time-consuming procedure. Usually ML procedures are easy only when they are also least-squares (LS) procedures. If the $X_{ij}$'s are large enough so that their distributions, which we have assumed to be Poisson, are close enough to being normal, then we can approximate our ML procedures (since under the assumption of normality ML and LS procedures would coincide).

Two classes of particularly simple LS procedures have been given in Sen and Soot (1981), Gray and Sen (1983), and Sen and Pruthi (1983). For the data sets (on interstate flows of coal and food grains in India) in Sen and Pruthi (1983), the ML and LS estimates were nearly identical—which should have been anticipated, given that the observations were very large. How small do $X_{ij}$'s have to be for this to continue to hold? Do LS procedures continue to perform well when more than three-quarters of the $X_{ij}$'s are zeros, as frequently happens in intraurban O-D tables? These are questions that have not yet been answered. We know, however, that under realistic conditions ML estimates of $\theta_k$'s are unique as (of course) are LS estimates. Thus, it is at least valid to compare the two.

## REFERENCES

Batty, Michael. *Urban Modelling.* Cambridge, England: Cambridge University Press, 1976.

Deming, W. E. and F. S. Stephan. "On a Least Squares Adjustment of Sampled Frequency Tables When the Expected Marginal Totals are Known," *Annals of Mathematical Statistics,* 11 (1940), 427–446.

Evans, Andrew W. "Some Properties of Trip Distribution Models," *Transportation Research,* 4 (1970), 19–36.

———. "The Calibration of Trip Distribution Models with Exponential or Similar Cost Functions," *Transportation Research*, 5 (1971), 15–38.

Evans, Suzanne P. "A Relationship Between the Gravity Model for Trip Distribution and the Transportation Problem in Linear Programming," *Transportation Research*, 7 (1973), 39–61.

Evans, Suzanne P. and H. P. Kirby. "A Three-Dimensional Furness Procedure for Calibrating Gravity Models," *Transportation Research*, 8 (1974), 105–122.

Furness, K. P. "Time Function Iteration, Traffic Engineering and Control," 7 (1965), 458–460.

Gray, Robert H. and Ashish Sen. "Estimating Gravity Model Parameters: A Simplified Approach Based on the Odds Ratio," *Transportation Research*-B, 17B (1983), 117–131.

Haberman, Shelly J. *The Analysis of Frequency Data*. Chicago, Illinois: University of Chicago Press, 1974.

Hyman, G. M. "The Calibration of Trip Distribution Models," *Environment and Planning A*, 1 (1969), 105–112.

Kendall, Maurice G. and Alan Stuart. *The Advanced Theory of Statistics, Vol 2: Inference and Relationship*, 2nd Edition. New York: Hagner, 1967.

Rao, C. Radhakrishna. *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973.

Sen, Ashish and R. K. Pruthi. "Least Squares Calibration of the Gravity Model When Intrazonal Flows are Unknown," *Environment and Planning* A, 15 (1983), 1545–1550.

Sen, Ashish and Siim Soot. "Selected Procedures for Calibrating the Generalized Gravity Model," *Papers, Regional Science Association*, 48 (1981), 165–176.

Smith, Tony E. "A Foundational Framework for Poisson Frequency Analysis of Weakly Interacting Behavior," No. 55, Working Papers in Regional Science and Transportation, University of Pennsylvania, Philadelphia, PA, 1984a.

———. "Poisson Gravity Models of Spatial Flows," Lecture given at the University of Tsukuba, Japan, June, 1984b.

Sue, Lih-Ching. "Computational Aspects of Maximum Likelihood Methods for Estimating Gravity Model Parameters," Masters Project Report, School of Urban Planning, University of Illinois at Chicago, 1985.