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.

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 θ_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

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

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

$$\prod_{ij} \exp[-T_{ij}]\, T_{ij}^{X_{ij}}/X_{ij}! \tag{3}$$

$$= \prod_{ij} \{\exp[-A_i B_j F_{ij}]\}\{[A_i B_j F_{ij}]^{X_{ij}}\}/X_{ij}!$$

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

[redacted]

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

[redacted]

Using this notation, (6) becomes

(7) $$T_{+j} = X_{+j}$$

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

(8) [redacted] $$T_{i+} = X_{i+}$$

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

(9) [redacted]

In the next section of this paper we shall present the principal theorem of the

[redacted]

		Column Corresponding to							
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	[illegible]	[illegible]
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 which is the MLE of θ 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) $\quad y_{i+} = X_{i+} \qquad y_{+j} = X_{+j} \qquad y_{ij} > 0 \qquad$ (for all i and j)

K=1

$((c_{ij}))$

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

$((X_{ij}))$

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

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

[redacted]

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 $\mathcal{R}^{(0)}$.

[redacted]

or

$$\sum c_{ij}^{(n)} y_{ij} \geq \sum c_{ij}^{(n)} Y_{ij} \quad \text{for all } y \in \mathcal{P}^{(K_1)}$$

ij ij

Proof: Similar to the discussion above.

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 $\hat{T}$ of the form given by (1) and (2) is that there exists

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

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

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

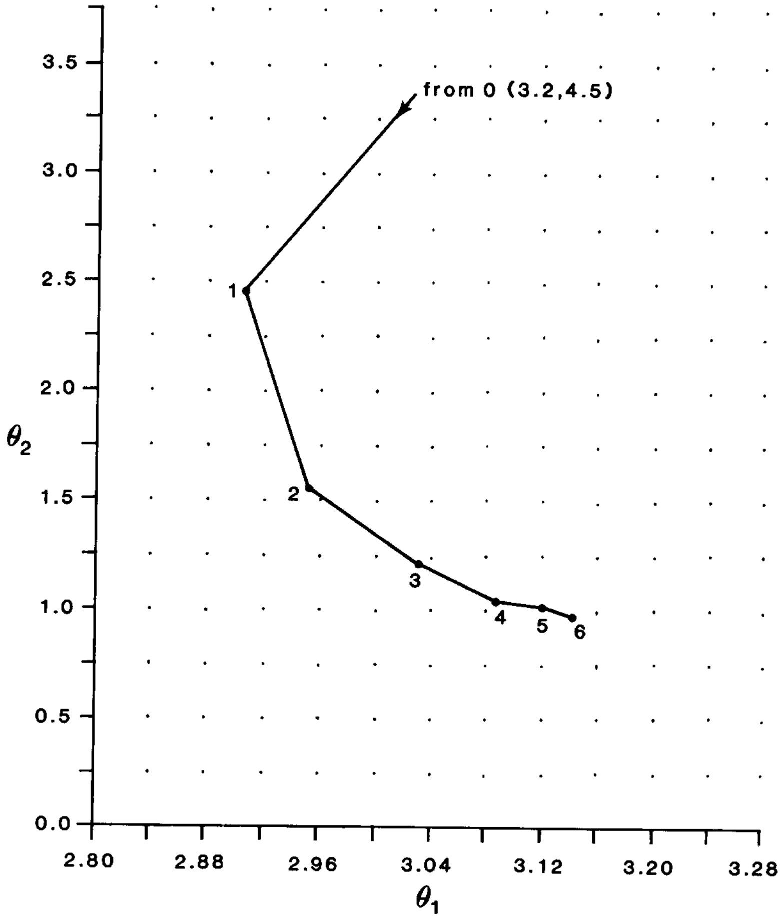

FIGURE 3: Trace of Gradient Search.

In order to describe the procedure, we shall first outline gradient search procedures in general. For some value $\hat{\zeta}^{(r)}$ of ζ, 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 ζ is as in Section 1, and the partial derivatives are evaluated at $\hat{\zeta}^{(r)}$. It is

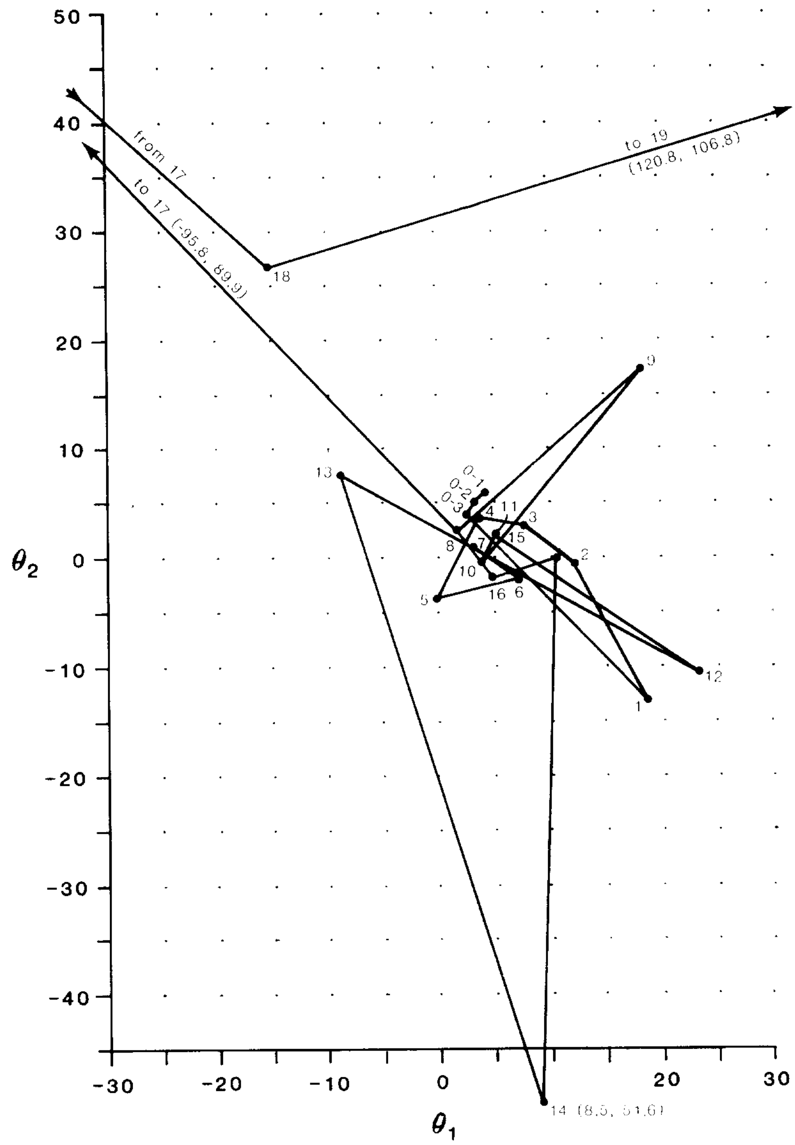

FIGURE 4: Trace of Scoring Method.

is the only variable in (15).] This value of ρ can be obtained sometimes by setting

[redacted]

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

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

[redacted]

and

$$\frac{\partial \mathcal{L}}{\text{[redacted]}} = \frac{\partial \mathcal{L}}{\text{[redacted]}} = \sum X_{ij} c_{ij}^{(k)} - \sum T_{ij} c_{ij}^{(k)} = \nu_k^{(r)} \text{ (say)} \qquad \text{(when } 1 \le k \le K\text{)}$$

[redacted]

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

[redacted]

Thus solving the equation $\delta(\rho) = 0$ for ρ 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 $d\delta(\rho)/d\rho$

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

[redacted]

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

[redacted]

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),

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

where

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

Let us now summarize the gradient search procedure:

[redacted]

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

[redacted]

Transportation Research, 5 (1971), 15–38.

[redacted]