Estimating Optimal Flow Values From Aggregate Data in Undetermined Spatial Systems

Richard W. Janson, Lalita B. Krishna, and Daniel W. Janson, Department of Geography, Kent State University, Kent, OH 44242, Department of Mathematical Sciences, University of Akron, Akron, OH 44325-4002, and 3-COM Inc., Santa Clara, CA 95052

ABSTRACT. A very rapid method to estimate optimal flow values of commodities within and among all regions of a spatial system is shown to be functionally dependent on the number of regions comprising the system. Aggregate data for the regions, such as total shipments from producing establishments are used as rim totals of the unknown spatial matrix. In general, the spatial matrix is an undetermined system, and the solution for the unknown cell values requires an inverse. A formula for the required inverse is derived and is incorporated into a distributor matrix, for which a formula is presented. To compute the unknown cells of the spatial matrix, as a vector, the distributor matrix is post multiplied by the vector of rim totals. The values estimated by this approach are based on least squares method.

INTRODUCTION

A spatial system is defined as an exhaustive set of mutually exclusive non-overlapping spatial units (k regions). Interindustry commodity flows (n commodities) link the regions into the system-wide trading pattern that enables each region to produce the array of commodities (both goods and services) that represents the outputs of the region. The variation in the lists of outputs among the regions corresponds to the variation of comparative advantage among the regions (Janson et al. 1987).

A spatial matrix, X, records the flows of a specified commodity (the ith commodity) among and within the k regions of the spatial systems. The row headings are the regions as exporting units of the system, and the column headings are the regions as importing units of the system. The cell values within the matrix are usually unknown, but frequently the row totals and column totals (together called rim totals) are known or can be estimated (Schaffer and Chu 1963). Customarily, the flow values are recorded in value terms.

The flow among the spatial categories may be material substance such as commodity or immaterial such as technology or other information. The data required for solution is the aggregate value of the substance produced (shipped) in each spatial unit (category) of the system and the aggregate value of the substance consumed (received) in each spatial unit (category) of the system. Estimating the interregional and intraregional optimal flows among the spatial units of an all-region trading system is the objective of the estimating method elucidated in the next section. Note that the method can be extended to non-spatial categories. For example, the flow of energy production and consumption among industrial sectors may be estimated by the procedure that will be described.

MATERIALS AND METHODS

Flows within an undetermined spatial system can be modeled using a stochastic assumption in a least squares sense (Janson et al. 1989, Lawson 1974). A formula for the exact inverse of \( B^T(BB^T)^{-1} \) (the distributor matrix) will be derived. All of the information required to stochastically distribute the rim totals of the spatial matrix, X, to the empty cells of the spatial matrix, is contained in the distributor matrix, which will be discussed later.

The quantity of spatial units alone determines each entire row in \( B^T(BB^T)^{-1} \); and every row corresponds to a cell value in the vectorized spatial matrix. Only rows of the distributor that correspond to exports from the region and imports into the region may be of immediate interest. Our method produces an abstraction that is totally devoid of distance decay effect and commodity-specific transport cost. The values computed by the procedure represent the underlying pattern that is implied solely as a result of even-handed (equally likely) flows within and among spatial units, given the actual rim totals, which are the aggregated flows from and to all producing units within a corresponding spatial unit (region).

The spatial matrix, X, will be defined mathematically next. After the spatial discussion, a binary matrix B will be defined in order to derive the distributor matrix, \( B^T(BB^T)^{-1} \). To solve for the commodity flows among and within the system of regions, the distributor matrix is simply post-multiplied by the vectorized rim totals of the spatial matrix. The computed solution is \( x \), which is the vectorized spatial matrix.

FORMULATION OF THE MODEL PROBLEM

The least squares solution depends on k, the number of interacting regions that comprise the interregional trade system. The generalized inverse formula of the matrix associated with the problem is given in terms of k. We will define the spatial matrix X which will give rise to a so-called binary matrix B. The distributor matrix will be a function of matrix B. Since the formula of the generalized inverse of the least squares matrix is explicitly given in terms of k, the round off error will be negligible.

Let \( R_1, R_2, \ldots, R_k \) represent k categories called regions and \( x_{ij} \) is the flow of substance from region i to region j. The ith row sum and jth column sum are denoted by \( y_i \) and \( y_j \), respectively. The matrix representation of flows of the substance among the k regions of a multi-region system are given (Fig. 1). This can be written as a system of 2k linear equations in the following form in k² unknowns \( x_{ij} \).
Thus we must drop one of the equations in (2). A convention will be adopted to drop the last equation.

If we let

\[ x_i = y_{i+} , \quad 1 \leq i \leq k, \]

\[ y_j = y_{j+} , \quad 1 \leq j \leq k \]

Note that \( \sum_{i=1}^{k} y_{i+} = \sum_{j=1}^{k} y_{j+} \).

Thus we must drop one of the equations in (2). A convention will be adopted to drop the last equation.

If we let

\[ x = [x_{11}, x_{12}, \ldots, x_{1k}, x_{21}, x_{22}, \ldots, x_{2k}, \ldots, x_{kk}]^T, \]

\[ y = [y_{1+}, y_{2+}, \ldots, y_{k+}, y_{+1}, y_{+2}, \ldots, y_{+(k-1)}]^T. \]

Equations (2) can be written in the matrix form as

\[ Bx = y, \]

where \( B \) is a binary matrix given below for \( k=4 \).

\[
B = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Note \( Bx = y \) is an undetermined linear system which has infinitely many solutions. The least squares solution is obtained by finding

\[
\text{Minimum} \quad ||x||_2; \quad Bx = y, \quad x \in \mathbb{R}^{k^2},
\]

where \( \mathbb{R}^{k^2} \) represents vectors with \( k^2 \) real components. And it is known that the minimum is given by (Hager, 1988)

\[ x = B^T (BB^T)^{-1} y. \]

Note, here \( B^T (BB^T)^{-1} \) is generalized inverse of \( B \).

Next we will derive \( B^T (BB^T)^{-1} \) explicitly in terms of \( k \). In block form \( B^T \) is given by

\[
\begin{bmatrix}
[k]_k & A \\
A^T & [k]_{k-1}
\end{bmatrix},
\]

where \( I_k \) and \( I_{k-1} \) are \( kxk \) and \( (k-1)x(k-1) \) identity matrices and \( A \) is \( kx(k-1) \) matrix of the form

\[
\begin{bmatrix}
1 & 1 & \ldots & 1 \\
1 & 1 & \ldots & 1 \\
\vdots & \ddots & \ddots & \vdots \\
1 & 1 & \ldots & 1
\end{bmatrix}
\]

It can be easily seen that \( B^T \) is nonsingular and the inverse can be computed as follows. Let the inverse of \( B^T \) in the block form be given by

\[
\begin{bmatrix}
E & F \\
F^T & G
\end{bmatrix}
\]

These result in the following matrix equations.

\[ kE + AF^T = I_k; \]

\[ kF + AG = 0; \]

\[ A^T E + kF^T = 0; \]

\[ A^T F + kG = I_{k-1}; \]

or,

\[ E = [k]_k - \begin{bmatrix}
1 \\
(k-1)_k
\end{bmatrix} A A^T - 1; \]

\[ F^T = \begin{bmatrix}
(k-1)_k
\end{bmatrix} A^T E; \quad G = \begin{bmatrix}
1 \\
(k-1)_k
\end{bmatrix} A^T F + kG \]

The \( (k-1)x(k-1) \) matrix \( A A^T = (a_{ij}) \) is given by

\[ a_{i,j} = (k-1), \quad i,j = 1,2,\ldots,(k-1) \]

and it can be shown that the entries of \( E = (e_{ij}) \) is given by

\[
e_{i,j} = \begin{cases}
\frac{(2k-1)}{k^2}, & i = j \\
\frac{(k-1)}{k^2}, & i \neq j, \quad i,j = 1,2,\ldots,k.
\end{cases}
\]
Moreover, on simplification we have
\[ f^T = - \frac{1}{k} A^T e = - \frac{1}{k} A^T \]
and
\[ G = (g_{i,j}) \text{ is given by } \]
\[ g_{i,j} = \begin{cases} \frac{2}{k}, & i = j \\ \frac{1}{k}, & i \neq j, \quad i,j = 1,2, \ldots, (k-1) \end{cases} \]  
(12)

Finally, premultiplying \((BB^T)^{-1}\) by \(B^T\), \(B^T(BB^T)^{-1}\) matrix of order \(k^2 \times (2k-1)\) is shown (Fig. 2).

The objective of the method is to derive a formula for the inverse in order to establish a methodology that permits estimation of optimal trade flows within, between, and among the spatial units (regions) of a free trade system. The method presented is for a single commodity, but the method can be easily generalized to a matrix formulation. The least squares assumption requires that the sum of squares for every row equals the sum of squares for every column. This implies that the sum of squares for the entire trade matrix is minimized given the constraints imposed by the known row totals and known column totals. This solution is optimal in a least cost sense, providing that the rim totals are recorded in units of delivered cost. The matrix generated by the method is thus a long run equilibrium solution, toward which a free trade system will trend, providing that ideal conditions of pure and perfect competition prevail in the system’s markets. To the extent that a measured value for a cell within the matrix differs from an observed value, the difference represents a target market for the under represented region and will indicate the existence of market imperfection, such as sanctions, or preferences, or communication costs.

The method can be profitably used for predicting the likely effect of plant investment policy. For example, Soviet planners in the past frequently located industries almost entirely within a single Soviet republic in order to increase the interdependence of the regions. Now the republics and individual enterprises need to evaluate likely results of plant investment in other republics. Using this methodology, the row total corresponding to the region being considered for the potential industry expansion is increased by the design capacity and the estimated flow values are recomputed in order to estimate the expected changes attributed to the investment strategy.

**EXAMPLE**

An example of a least squares matrix computed by the formula presented in this paper will be compared with the matrix of actual measured values (Table 1). The purpose of the discussion is to demonstrate the use of the method in order to gain insights in the analysis of data, and to formulate hypotheses for evaluation. By using a physical quantity (barrels of oil) a reasonable surrogate for delivered cost is obtained. This is because the physical quantity shipped implicitly includes the costs and difficulties of transfer and communication.

The optimal flows of international crude oil in 1988 differ substantially from the measured flows. It is clear that the observed pattern of commodity flow is far from optimal. Various hypotheses can be offered to explain the difference in important cell values. First, the flow of oil is the result of imperfect market forces subject to cartel manipulation and the legal and institutional constraints of national governments. For example, under optimal assumptions oil, from Alaska’s north slope should be sold to Japan, a nearby region. In fact, the United States Congress requires that Alaska oil be consigned to U.S. ports. As a result, observed flows from the United States to the Far East (Japan) are less than optimal, and observed
Table 1


(Thousands Barrels Per Day)

<table>
<thead>
<tr>
<th></th>
<th>North America</th>
<th>Central and South America</th>
<th>Western Europe</th>
<th>Eastern Europe and USSR</th>
<th>Middle East</th>
<th>Africa and Oceania</th>
<th>TOTAL EXPORTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>a. 1376.0</td>
<td>194.0</td>
<td>344.0</td>
<td>0.0</td>
<td>35.0</td>
<td>0.0</td>
<td>205.0</td>
</tr>
<tr>
<td></td>
<td>b. 576.1</td>
<td>14.0</td>
<td>1069.0</td>
<td>88.0</td>
<td>-140.2</td>
<td>-132.2</td>
<td>678.4</td>
</tr>
<tr>
<td>Central and South America</td>
<td>a. 694.0</td>
<td>484.0</td>
<td>189.0</td>
<td>0.0</td>
<td>0.0</td>
<td>16.0</td>
<td>80.0</td>
</tr>
<tr>
<td></td>
<td>b. 478.3</td>
<td>-84.6</td>
<td>970.4</td>
<td>-10.6</td>
<td>238.8</td>
<td>-230.8</td>
<td>579.5</td>
</tr>
<tr>
<td>Western Europe</td>
<td>a. 615.0</td>
<td>0.0</td>
<td>1861.0</td>
<td>16.0</td>
<td>9.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>b. 628.8</td>
<td>65.9</td>
<td>1120.0</td>
<td>139.9</td>
<td>-88.3</td>
<td>-80.3</td>
<td>730.2</td>
</tr>
<tr>
<td>Eastern Europe and USSR</td>
<td>a. 0.0</td>
<td>85.0</td>
<td>1194.0</td>
<td>1412.0</td>
<td>16.0</td>
<td>30.0</td>
<td>93.0</td>
</tr>
<tr>
<td></td>
<td>b. 673.5</td>
<td>110.6</td>
<td>1165.6</td>
<td>184.6</td>
<td>-43.6</td>
<td>35.6</td>
<td>774.9</td>
</tr>
<tr>
<td>Middle East</td>
<td>a. 1434.0</td>
<td>619.0</td>
<td>3074.0</td>
<td>505.0</td>
<td>353.0</td>
<td>395.0</td>
<td>424.0</td>
</tr>
<tr>
<td></td>
<td>b. 1786.1</td>
<td>1223.2</td>
<td>2278.2</td>
<td>1297.2</td>
<td>10.69</td>
<td>1077.0</td>
<td>1888.3</td>
</tr>
<tr>
<td>Africa</td>
<td>a. 1078.0</td>
<td>221.0</td>
<td>2319.0</td>
<td>188.0</td>
<td>120.0</td>
<td>129.0</td>
<td>97.0</td>
</tr>
<tr>
<td></td>
<td>b. 860.9</td>
<td>290.8</td>
<td>1353.0</td>
<td>372.0</td>
<td>143.8</td>
<td>151.8</td>
<td>962.4</td>
</tr>
<tr>
<td>Far East and Oceania</td>
<td>a. 359.0</td>
<td>24.0</td>
<td>19.0</td>
<td>14.0</td>
<td>0.0</td>
<td>15.0</td>
<td>1566.0</td>
</tr>
<tr>
<td></td>
<td>b. 554.5</td>
<td>-8.4</td>
<td>1046.6</td>
<td>65.6</td>
<td>-162.6</td>
<td>154.6</td>
<td>655.9</td>
</tr>
<tr>
<td>TOTAL IMPORTS</td>
<td>5556.0</td>
<td>1617.0</td>
<td>9000.0</td>
<td>2135.0</td>
<td>588.0</td>
<td>594.0</td>
<td>6284.0</td>
</tr>
</tbody>
</table>


Source: Annual Energy Review 1990, Table 116. Energy Information Administration

International flows within North America are more than optimal.

Another distortion of international crude oil flow is apparent in the cells recording flows within Eastern Europe and the USSR. The observed values are far in excess of the estimated optimal flow. This is the result of Soviet policy, over four decades, to subsidize satellites. The argument between Japan and Russia over several Kurile Islands annexed by Russia at the end of World War II will continue to depress the flow of oil from the USSR to Japan (Far East) until the argument is resolved. As Russia moves to a market economy, these distortions will mitigate. As constraints to a free market are reduced or eliminated, exports of crude oil from Russia and other sovereign republics (USSR) to the United States and to Japan will both increase.

Vestiges of colonialism are likely reasons for the deficit in traffic from Africa to the Far East. Actual flows of crude oil from Africa to Western Europe are 70% more than the sum of square value. Invisible pulls based on history, emotion, and language (preference arrangements) still influence trade flows three decades and more after independence. As one consequence, flows of African oil to the Far East are only one tenth the least squares optimal estimate.

Modern industrialism is utterly dependant on massive inputs of energy. In fact, the course of modern civilization is a record of energy application to production processes. For this reason energy has been called the ultimate resource. The brief comments based on the least squares estimates strongly suggest that crude oil worldwide commodity flows are far from optimal and that national policies should address such manifest discrepancies from free market benchmarks for the ultimate resource. The logic of neoclassical equilibrium theory attributes fundamental significance to the tendency of an economic system to move toward an optimal solution.

**CONCLUSION**

The method presented is remarkable for two reasons: 1) because the most likely solution (in a least squares sense) is a function of the number of interacting regions, \( k \), that comprise the trading system; and 2) because the exact inverse required for the solutions of appropriate empirical problems, and the exact derived multiplier matrix (distributor matrix) that incorporates the exact inverse can be written from the formula as exact functions of \( k \). The authors are proceeding to explore some of the many ramifications of this approach to spatial modeling. Distance decay effects and commodity specific transport cost will be introduced in future publications by the authors.

**LITERATURE CITED**

