

OPERATIONALIZING THE SPATIAL PROXIMITY MODEL OF STATE POLICY PRIORITIES

A report to accompany the paper “A Summary
Measure of Yearly State Policy Spending”

William G. Jacoby
Saundra K. Schneider
Michigan State University

March 2005

jacoby@msu.edu
sks@msu.edu

The spatial proximity model employed in “A Summary Measure of Yearly State Policy Spending” provides a geometric representation of state program expenditures. The input data for the spatial proximity model are yearly state spending figures for each of ten different policy areas, expressed as a proportion of total spending across all then policy areas. Each data value, x_{ijt} , represents the i^{th} state’s relative expenditure for policy j in year t . The model parameters are estimated on a yearly basis. In other words, a complete geometric representation of the first year’s data is constructed, then a complete representation is obtained for the second year’s data, and so on, down to the last year in the dataset. Because of this year-by-year strategy, the calculations within any year are completely separate from those for any other year. And, in order to simplify the presentation a bit, we can temporarily drop the “ t ” subscript from the data and parameter values, and treat the algorithm as if we are only estimating the model for a single year.

The model, itself, contains two sets of parameters: A set of 50 points representing the states (designated s_1, s_2, \dots, s_{50}) and another set of 10 points representing the policy areas (designated p_1, p_2, \dots, p_{10}). These two point sets are arrayed along a common unidimensional continuum such that interpoint distances are inversely proportional to state expenditures. Specifically, as state i spends more on policy j , then the distance from the point representing i to the point representing j (designated d_{ij}) gets smaller and vice versa.

For convenience in the discussion below, we will assume that the x_{ij} ’s have been “reflected,” so that higher spending levels actually correspond to *lower* data values. This is merely done so that distances in the spatial proximity model are now directly proportional to the data values. The reflection is easily accomplished by subtracting all of the data values from a constant. In the present case, it is natural to use a value of one as the constant, since the input data are proportions of total state spending and they accordingly sum to one within every state. The reflected data values will be shown as x_{ij}^* below, in order to distinguish them from the original expenditures.

Many specific procedures— usually called “unfolding techniques”— have been developed to estimate the parameters of the spatial proximity model (e.g., Cox and Cox 2000; Borg and Groenen 1997). Our analysis uses a metric, least-squares unfolding method developed by Keith Poole (1984). The overall approach is called “unfolding” because, according to the geometry of the model, a state’s profile of (reflected) spending values can be obtained by “folding” the unidimensional continuum at the location of the state’s point (Coombs 1964). The scaling task is the opposite of this process: We begin with the folded versions of the dimension (i.e., the input data values) and seek to “unfold” them simultaneously across all states, in order to estimate the dimension itself (i.e., the relative positions of the state and policy points). The method is “metric” because it assumes that the input data are measured at the interval level or higher (many unfolding techniques only assume ordinal or even nominal measurement levels). The method is “least-squares” because its immediate analytic objective is to find the set of state and policy point locations such that the squared errors between distances and data values are minimized. That is:

$$\text{Minimize: } \sum e_{ij}^2 = \sum_{i=1}^{50} \sum_{j=1}^{10} (d_{ij} - x_{ij}^*)^2 \quad (1)$$

Alternatively, the procedure seeks to maximize the squared correlation between the interpoint distances and the reflected program expenditures. This squared correlation, itself, can be used as a goodness-of-fit measure for the final scale. If the data fit the model exactly, then the errors will all be zero, the squared correlation will be a perfect 1.0, and the reflected data values will be exactly proportional to the interpoint distances on the unfolded dimension.

Readers are referred to Poole’s original article (Poole 1984) for theoretical background and technical development on this least-squares, metric unfolding technique. The scaling strategy is quite simple, although the notation and computations can be a bit cumbersome. The remainder of this section contains a brief, largely informal, description of the methodology.

The Conditional Global Minimum Criterion

Let us consider a strategy for estimating the positions of one point set (either state or policy points) conditional upon fixed locations of the other point set (policy or state points, respectively). For

this discussion, assume that we are estimating policy points, with fixed state points. Imagine that each state point, s_i has attached to it 10 different vectors— one for each policy. The length of each such vector (say, for the j th policy) is equal to the state's (reflected) spending value for that policy, x_{ij}^* . Since we are considering a unidimensional model, each of these vectors can only point in one of two directions; either to the left or the right of the state point to which it is attached. In either case, the terminal point of each vector, v_{ij} can be calculated as one of the following:

$$v_{ij} = s_i - x_{ij}^* \quad (2a)$$

$$v_{ij} = s_i + x_{ij}^* \quad (2b)$$

Condition (2a) is used when state i 's vector for policy j points to the left of the state point, and condition (2b) is used when the vector points toward the right. The sum of squared errors for any given policy, j can be calculated as:

$$\sum e_j^2 = \sum_{i=1}^{50} (x_{ij}^* - d_{ij})^2 = \sum_{i=1}^{50} (x_{ij}^* - |s_i - p_j|)^2 \quad (3)$$

In equation (3), the x_{ij}^* are the reflected data values and the s_i are fixed by construction; therefore, only the p_j values can be manipulated.

For any given policy, Poole (1984) proved that this sum of squared errors is minimized when the states' vectors are all pointed in the correct direction (i.e., directly toward the location of policy j , rather than away from it) and the policy location is estimated as the centroid of the vector terminal points. That is:

$$p_j = \frac{\sum_{i=1}^{50} v_{ij}}{50} \quad (4)$$

This estimate of the policy point location is called the *conditional global minimum* (CGM). It is a “minimum” in the sense that it is the p_j value which produces the smallest possible value of $\sum e_j^2$. It is “conditional” because the result only holds when the state points, the s_i 's, are held fixed at their current locations. But, within this constraint, it is a “global” minimum: No other value of p_j will produce a smaller squared error.

A Scaling Procedure Based Upon CGM

The preceding result leads to a particularly simple, but exhaustive and computationally-intensive, search procedure for finding the optimal policy location, relative to the currently-fixed state points. The states comprise 50 different locations along the dimension, and their points therefore divide the dimension into 51 distinct intervals. Start by tentatively placing the policy point to the left of the leftmost state point; in this case, all 50 vectors will point to the left, and the v_{ij} values would be calculated according to equation (2a). Use equation (4) to estimate the tentative policy location (designate this \hat{p}_j with the carat indicating that this is only a tentative point location at this stage of the estimation process), and also calculate the variance associated with this tentative position, as follows:

$$\text{var}(\hat{p}_j) = \frac{\sum_{i=1}^{50} (v_{ij} - \hat{p}_j)^2}{50} \quad (5)$$

Next, “move” the policy point to the second interval from the left along the dimension (i.e., between the first and second state point locations). Note that, in doing so, one vector reverses direction—the one that originates from s_1 , the leftmost state point. Calculate a new value for \hat{p}_j and also a new value for $\text{var}(\hat{p}_j)$. If this latter variance is smaller than the first one, then there is less error associated with this second, tentative policy point location. Therefore, the policy point should be moved to this new centroid.

This process continues, “moving” the policy point into each successive interval between an adjacent pair of state points. After the policy point has been tried in each of the 51 intervals (i.e., it is moved all the way to the right of the rightmost state point), the final policy point location estimate, p_j , is the centroid value (\hat{p}_j) that was associated with the smallest value of $\text{var}(\hat{p}_j)$. Again, Poole (1984) proved that this is a global minimum for the amount of error associated with the position of p_j , conditional upon the current, fixed set of ideal points. Hence, p_j is referred to as the “CGM estimate” of the point location.

The “point moving” procedure is repeated for each of the 10 policies in order to obtain the least squares estimate of the point location for each one. Then, the two point sets (policies and states) are interchanged and the search procedure is repeated. In other words, the vectors are now conceived as

originating from the policy points and terminating at the various state points. The policy points are held fixed at their current locations, $p_1, p_2, \dots, p_j, \dots, p_{10}$, and each of the 50 state points are tried in each of the 11 resultant intervals along the dimension. As before, each state's estimated point location is associated with the centroid of vector termini that is associated with the smallest $\mathbf{var}(\hat{s}_i)$ value.

A single iteration of the CGM procedure consists of two complete sets of point movements. The first time through, one set of point locations must be specified by the researcher. In the present case, we will fix the initial state points at certain positions and estimate the policy point positions accordingly; the exact procedure for setting the initial state configuration is discussed in the next paragraph. The CGM algorithm proceeds by moving the policy points contingent upon the fixed state points, and then moving the state points contingent upon the newly-fixed policy points, and so on. The procedure is completed whenever the total sum of squared errors (calculated across all states and policies) converges to a value that does not change across iterations.

The procedure for fixing the initial positions of the state points is straightforward. And, this is where we return to the time-series structure of the data. On the first iteration for the first year's data (i.e., $t = 1$), the state points are simply located at random positions along the dimension. The state points will be moved during the course of the unfolding procedure, so these initial estimated locations are not taken very seriously (and, tests indicate that the initial locations have little, if any, discernible impact on the final scaling solution). For all subsequent years (i.e., $t > 1$), the state points are initially located at their terminal positions from the previous year. That is, the initial estimate of state i 's location at time t is simply the s_i that was estimated at time $t-1$.

This yearly initialization strategy is not only convenient for the scaling algorithm; it is also nicely consistent with the substantive nature of governmental spending. The budgetary decisionmaking process within the states always begins with the previous years' allocation of expenditures (Thompson 1987; Hansen 1990; Kearns 1993). And, that is precisely what the preceding year's state point represents in the spatial proximity model.

A Simple Example

Let us consider a simple example of the CGM unfolding procedure, using hypothetical data on three states (labeled A , B , and C) and two policies (labeled M and N). Table 1 shows the input data matrix, \mathbf{X}^* . The entries in this matrix are reflected proportionate policy expenditures. For example, state A spends 73% of its budget on policy M ; therefore, the entry in the first cell, x_{11}^* , is $1.00 - 0.73 = 0.27$. Similar calculations are carried out on the expenditure figures for all states and policies, in order to produce matrix \mathbf{X}^* . Next, three random numbers (from the uniform distribution on the interval from zero to one) are generated as starting values for the state points: In this case, $s_A = 0.20$, $s_B = 0.29$, and $s_C = 0.60$. These three points divide the dimension into four intervals, as shown in Figure 1.

Table 2 shows the point-moving procedure used to estimate the policy points relative to the current locations of the state points. For each of the two policies, the column labeled “ \hat{p}_j ” gives the centroid of the vector terminal points obtained when the policy point is placed within each of the four intervals along the dimension. The column labeled $\text{var}(\hat{p}_j)$ gives the error variance associated with each of the tentative point locations. The point locations that minimize the error variances are -0.110 for M and 0.890 for N ; therefore, these are taken as the scaled values for the respective policies.

Next, the point-moving process is repeated, with the roles of the state and policy points reversed. Figure 2 shows how the current policy point locations divide the dimension into three intervals. In Table 3, the column labeled “ \hat{s}_i ” gives the centroid of the vector terminal points obtained when the state point is placed within each of the intervals along the dimension. The column labeled $\text{var}(\hat{s}_i)$ gives the error variance associated with each of the tentative point locations. The point locations that minimize the error variances are 0.160 for A , 0.180 for B , and 0.750 for C ; therefore, these are taken as the scaled values for the respective states.

Finally, Table 4 shows $\hat{\mathbf{X}}^*$, the predicted data matrix obtained from the model-based estimates. Each entry in this matrix is an interpoint distance. For example, with state A located at 0.160 and policy M at -0.110 , the distance between the two is 0.27 ; therefore, this is the value in the first cell of the matrix,

\hat{x}_{11}^* . The other cell entries are calculated in the same manner. The goodness of fit for this unfolded scale of state and policy points is obtained by correlating the corresponding entries in the original data matrix and in the model-predicted data matrix. In this case, $r = 1.00$, reflecting a perfect fit between the data and the spatial proximity model; stated differently, the original data values can be predicted without error from the scaled interpoint distances. Therefore, no further point movements are necessary and the unfolding procedure terminates here, after only a single iteration.

Of course, the rapid convergence and perfect fit in the preceding example are due to the small size and simulated content of the data matrix. But, extensive Monte Carlo testing and several applications to real-world data indicate that Poole's unfolding technique works quite well with larger and more realistic datasets. The scaling algorithm based upon the CGM criterion converges to a solution very quickly and it provides highly accurate estimates of the point locations. An easy-to-use SAS/IML Macro for performing this type of least-squares metric unfolding is available from the authors, upon request.

REFERENCES

- Borg, Ingwer, and Patrick Groenen. 1997. *Modern Multidimensional Scaling: Theory and Applications*. Springer-Verlag.
- Coombs, Clyde H. 1964. *A Theory of Data*. New York: John Wiley.
- Cox, Trevor F., and Michael A. A. Cox. 2000. *Multidimensional Scaling (Second Edition)*. New York: Chapman and Hall/CRC.
- Hansen, Susan B. 1990. "The Politics of State Taxing and Spending." In *Politics in the American States*, ed. Virginia Gray, Herbert Jacob, and Robert B. Albritton. Glenview, IL: Scott, Foresman/Little, Brown Higher Education.
- Kearns, Paula S. 1993. "Determinants of State Budget Periodicity: An Empirical Analysis." *Public Budgeting and Finance* (Spring): 40-58.
- Poole, Keith J. 1984. "Least Squares, Metric, Unidimensional Unfolding." *Psychometrika* 49: 311-323.
- Thompson, Joel A. 1987. "Agency Requests, Gubernatorial Support, and Budget Success in State Legislatures Revisited." *Journal of Politics* 49: 756-779.

Table 1: Hypothetical data matrix, X^* . Cell entries are reflected proportionate spending values for three states (A , B , and C) on two policies (M and N).

	M	N
A	0.27	0.73
B	0.29	0.71
C	0.86	0.14

Table 2: Point-moving procedure to estimate locations of policy points.

Interval:	Policy M		Policy N	
	\hat{p}_i	$var(\hat{p}_i)$	\hat{p}_i	$var(\hat{p}_i)$
1	-0.110	0.012	-0.163	0.196
2	0.070	0.091	0.323	0.313
3	0.263	0.139	0.797	0.058
4	0.837	0.196	0.890	0.012

Table 3: Point-moving procedure to estimate locations of state points.

Interval:	State <i>A</i>		State <i>B</i>		State <i>C</i>	
	\hat{p}_j	$\text{var}(\hat{p}_j)$	\hat{p}_j	$\text{var}(\hat{p}_j)$	\hat{p}_j	$\text{var}(\hat{p}_j)$
1	-0.110	0.097	-0.110	0.084	-0.110	0.740
2	0.160	0.000	0.180	0.000	0.750	0.000
3	0.890	0.533	0.890	0.504	0.890	0.020

Table 4: Model-based predicted data matrix, \hat{X}^* . Cell entries are distances between the scaled state points and the scaled policy points.

	<i>M</i>	<i>N</i>
<i>A</i>	0.27	0.73
<i>B</i>	0.29	0.71
<i>C</i>	0.86	0.14

Figure 1: Random starting values for state points and resultant intervals along dimension.

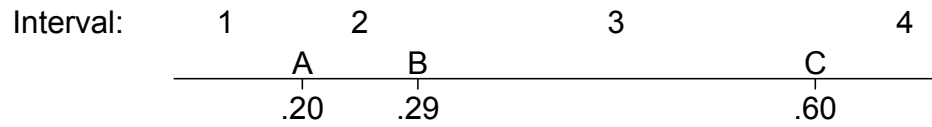


Figure 2: Scaled policy point locations and resultant intervals along dimension.

