COMPARING ALTERNATIVE SCALING MODELS FOR STATE POLICY EXPENDITURES

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 analysis reported in "Variability in State Policy Priorities: An Empirical Analysis" uses a spatial proximity model to represent state programmatic expenditures. This is quite different from the approach used in most other studies, which employ techniques based upon the factor analytic model to represent state policy outputs (spending and otherwise). By definition, models are abstract depictions of real-world phenomena. As such, it is inappropriate to say that any model is "correct" or "incorrect." Instead, a model's adequacy should be judged according to the degree to which the empirical data can be reproduced from the components of the model, itself— a property that is sometimes called the model's "explanatory power." And, given two different models with equal levels of explanatory power, the principle of parsimony is usually applied; that is, the simpler model is preferable (e.g., Kaplan 1964; King, Keohane, Verba 1994).

Using the preceding criterion, the spatial proximity model is generally better than the factor analytic model for representing state policy commitments because it produces lower-dimensioned— and hence, simpler— depictions of the data. This contrast between spatial and factor representations of empirical data has long been recognized in the literature on scaling methods and dimensional analysis (e.g., Coombs 1964; Weisberg 1974; Davison 1983; Jacoby 1991). The difference stems from the nature of the respective models and their associated geometry.

The factor analytic model represents linear correlations between columns (or rows) of the data matrix as angles between vectors. If the data contain nonrandom patterns which are also nonlinear in form, then a factor analysis will almost certainly produce "extra" factors in order to account for the nonlinearities. This is not necessarily the case with the spatial proximity model, which represents entries in the data matrix as distances between points. These distances can readily incorporate a variety of nonlinear data patterns.

The higher dimensionality that is virtually inevitable in the factor analytic approach (under certain circumstances) can be demonstrated quite easily, using some simple, hypothetical data. Table 1 shows a data matrix. The rows represent eleven "states" (labeled s_1 through s_{11}) and the columns represent three "policies" (labeled A, B, and C). The cells contain state policy "expenditures," so that entry x_{ij} is the amount state i spends on policy j.

Table 2 shows the correlation matrix for the policies. Note that the "total variance" in the correlation matrix (represented by the sum of the main diagonal entries) is 3.0, since each of the variables is standardized to a variance of one, and there are three variables. If this correlation matrix is used as input to a factor analysis, it produces two orthogonal factors. Figure 1 shows the resultant factor space and Table 3 gives the factor pattern coefficients, along with the communalities and the eigenvalues. Speaking informally, the communalities give the variance explained in each of the variables, while the eigenvalues give the variance explained by each factor. The sum of the communalities is 3.0, as is the sum of the eigenvalues. Hence, the factor solution accounts for 100% of the variance in the input data. Policies *A* and *C* load at opposite ends of the first factor. Policy *B*, alone, has a large loading on the second factor. For present purposes, the specific configuration of factor coefficients is less important than the fact that it takes two factors (or dimensions) to represent the data.

Now, let us fit a spatial proximity model fitted to the same hypothetical data (i.e., from Table 1). The rule used to construct the model is simple: Fourteen points (representing s_1 through s_{11}, *A*, *B*, and *C*) are arranged along a continuum such that the distances between state points and policy points (in the unidimensional case, distances can be shown as $d_{ij} = |s_i - p_j|$, where i ranges from 1 to 11, and p_j is either *A*, *B*, or *C*) are inversely proportional to the amounts that states spend on the respective policies. That is, $d_{ij} = k - mx_{ij}$, where m is a constant of proportionality and k is a constant that is at least as large as the maximum value in the original data matrix. For the present example, we will set $m = 1$ and $k = 10$. Figure 2 shows one possible dimension that could be constructed using this modeling approach and Table 4 gives the resultant distances between state points and policy points. Now, the correlation between the distances in Table 4 and the input data values in Table 1 is -1.00; the R^2 value is 1.00. So, the spatial proximity model also accounts for 100% of the variance in the data. But, it does so with a single dimension, rather than the two that were required in the factor analytic approach.

Of course, the results presented here occur as a result of the simulated data employed in the illustrative analysis. So, it is important to ask: Under what circumstances will this difference between the two models be manifested in empirical data? One situation is particularly relevant for the present research context:

A data matrix that simultaneously contains bipolarity across certain variables and consensus in others. For example, in Table 1, states that spend a large amount on policy *A* spend very little on policy *C* and vice versa. But, all states devote at least a moderate amount of resources to policy *B* (the smallest expenditure for *B* is five, while the minimum values for *A* and *C* are both zero). This drives down the correlations between spending on *B* and spending on the other policies.

This kind of situation is very likely to occur in the context of "real" state policy expenditures. States have some discretion to allocate resources selectively across certain kinds of policies. But, there are also certain services that *all* states must provide, regardless of other political or socioeconomic factors and preferences. For example, education comprises the largest segments within every state's budget (Winter 1999). Thus, state policy spending does exhibit a combination of bipolar and consensual patterns. For this reason, the spatial proximity model is preferable to the factor analytic model for representing state policy spending. Again, it is not that the former is "right" and the latter is "wrong." Instead, the spatial proximity model requires fewer dimensions— i.e., it is a simpler representation— than the factor analytic model.

REFERENCES

Coombs, Clyde H. (1964) *A Theory of Data*. New York: John Wiley.

Davison, Mark L. (1983) *Multidimensional Scaling*. New York: John Wiley.

Jacoby, William G. (1991) *Data Theory and Dimensional Analysis*. Newbury Park, CA: Sage Publications.

Kaplan, Abraham. (1964) *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler Publishing.

King, Gary; Robert O. Keohane; Sidney Verba. (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Weisberg, Herbert F. (1974) "Dimensionland: An Excursion into Spaces." *American Journal of Political Science* 18: 743-776.

Winters, Richard. (1999) "The Politics of Taxing and Spending," In *Politics in the American* States (7th Edition), ed. Virginia Gray; Russell L. Hanson; Herbert Jacob. Washington, DC: Congressional Quarterly Press.

Table 1: Hypothetical Data on Eleven States' Expenditures across Three Policy Areas.

	Policies		
	A	B	C
s_1	10	5	0
s_2	9	6	1
s_3	8	7	2
s_4	7	8	3
s_5	6	9	4
s_6	5	10	5
s_7	4	9	6
s_8	3	8	7
s_9	2	7	8
s_{10}	1	6	9
s_{11}	0	5	10

Table 2: Correlation Matrix for Hypothetical State Expenditures on Three Policy Areas.

	A	B	C
A	1.00	0.00	-1.00
B	0.00	1.00	0.00
C	-1.00	0.00	1.00

Table 3: Factor Pattern Coefficients (or "Factor Loadings"), Communalities, and Eigenvalues from Factor Analysis of Hypothetical State Expenditure Data.

	Factors		
	Factor 1	Factor 2	Communalities
A	1.00	0.00	1.00
B	0.00	1.00	1.00
C	-1.00	0.00	1.00
Eigenvalues	2.00	1.00	3.00

Note: These factor results are based upon a principal axis factor analysis. Prior communality estimates were set to one for all three variables. The solution is not rotated, but this factor configuration is identical to a varimax rotation. Factor scores for the eleven observations are not shown.

Table 4: Interpoint Distances Obtained from the Spatial Proximity Model of Hypothetical State Expenditures.

	Policy Points:		
State Points:	A	B	C
s_1	0	5	10
s_2	1	4	9
s_3	2	3	8
s_4	3	2	7
s_5	4	1	6
s_6	5	0	6
s_7	6	1	5
s_8	7	2	4
s_9	8	3	3
s_{10}	9	4	2
s_{11}	10	5	1

Note: Each cell entry in Table 4 represents the distance between a state point and a policy point in Figure 2. For example, the distance from the s_1 point to the A point is 0, the distance from the s_1 point to the B point is 5, and so on.

Figure 1: Factor Space Obtained from Factor Analysis of Hypothetical State Expenditure Data.

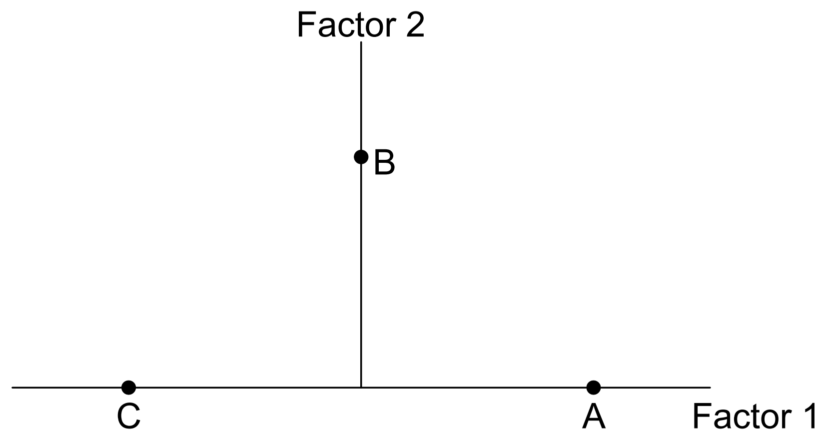

Figure 2: Spatial Proximity Model of Hypothetical State Expenditure Data.