ESTIMATING THE GEOMETRIC MODEL OF INDIVIDUAL VALUE CHOICES

William G. Jacoby
Professor, Department of Political Science
Michigan State University
jacoby@msu.edu

July 2014

A supplemental report to accompany "Is There a Culture War? Conflicting Value Structures in American Public Opinion."

This report describes the estimation procedure used to construct the geometric model of individual value choices that is analyzed in the manuscript, "Is There a Culture War? Heterogeneous Value Choices and American Public Opinion." The input data consist of n individuals' rank-ordered preferences for k values. This information is contained in the n by k matrix, $\mathbf{X}$. Each row of $\mathbf{X}$ represents an individual and each column corresponds to a value. The entry in cell x_{ij} gives the number of values that individual i believes to be less important than value j. The entries within any row of $\mathbf{X}$, say $\mathbf{x}_i$, usually range from zero through $k-1$. In this manner, the x_{ij} vector represents individual i's full hierarchy of value choices.[1]

The Geometric Model of Value Choices

The geometric model shows the values as k points and the individuals as n vectors within a common m-dimensional space. The relative positions of the value points are determined by the similarities and dissimilarities in the ranks that people assign to the respective values. For example, two values that tend to receive similar rank scores (i.e., individuals who rank one of the values highly tend to rank the other value highly as well, while those who assign one of the values a low rank also tend to assign a low rank to the other value) will be represented by points that are located close to each other within the space. In contrast, two values that tend to receive different rank scores (i.e., people who rank one of the values highly usually assign a low rank to the other value and vice versa) are represented by points

[1]If individual i gives a complete ordering of the values, then the entries within $\mathbf{x}_i$ will be composed of the successive integers from zero to $k-1$. But, there are situations where an incomplete ranking can occur. For example, an individual may be unable or unwilling to rank-order some subset of the values. Or, pairwise comparisons of certain values may be intransitive, thereby precluding a full ranking. In such cases, tied, and possibly non-integer, scores would occur within $\mathbf{x}_i$. If these problems arise either among the values a person believes to be most important, or among those that he or she believes to be least important, then the maximum and minimum cell entries in the vector will be less than $k-1$ or greater than zero, respectively. The dataset used in the present analysis is limited to individuals who provided full rank-orders for the values. Nevertheless, the model and estimation methodology explained here can handle the tied and incomplete value rankings with no difficulties.

that are located far apart from each other within the space. The coordinates for the value points are collected into the k by m matrix, $\mathbf{\Phi}$.

The n vectors are oriented within the space such that, to the greatest extent possible, each person's vector points toward those values that he or she believes to be most important, and away from those values that he or she believes to be least important. Stated differently, the rank-order of the perpendicular projections from the value points onto individual i's vector should correspond to the ranked importance scores that person assigns to the values (i.e., more important values project onto the vector at a location closer to the tip of the vector, while less important values project onto the vector further away from the tip).

The vectors emanate from the origin of the space. The *direction* of each vector is important, because that determines the order in which the value points project onto that vector. But, the specific *length* of each vector is arbitrary; for convenience, we will adjust the vectors to be unit-length. The coordinates for the vector terminal points are collected into the n by m matrix, $\mathbf{\Gamma}$.

The geometric structure just described is sometimes called the MDPREF model, an acronym for "multidimensional preference scaling" (e.g., Carroll 1972; Weller and Romney 1990). The analytic task is to use the information in $\mathbf{X}$ to estimate $\mathbf{\Phi}$ and $\mathbf{\Gamma}$. As we will see, the ordinal nature of the information in $\mathbf{X}$ will require us to use a new n by k matrix, $\mathbf{X}^*$. Each row of $\mathbf{X}^*$ contains a monotonic transformation of the corresponding row in $\mathbf{X}$.[2] The nature of the transformation will be explained below. For now, it is important to emphasize that $\mathbf{X}^*$ provides exactly the same information as the original $\mathbf{X}$ regarding the individuals' relative importance rankings of the respective values.

[2]Assume that $\mathbf{x}_{ia}$ and $\mathbf{x}_{ib}$ are two scalar elements of the vector, $\mathbf{x}_i$, while $\mathbf{x}^*_{ia}$ and $\mathbf{x}^*_{ib}$ are the corresponding elements of $\mathbf{x}^*_i$. If $\mathbf{x}^*_i$ is monotonically related to $\mathbf{x}_i$ then, for all pairs of elements, a and b, if $\mathbf{x}_{ia} < \mathbf{x}_{ib}$ then it must be the case that $\mathbf{x}^*_{ia} \leq \mathbf{x}^*_{ib}$. Thus, monotonicity implies that the ordering of the elements in $\mathbf{x}^*_i$ never contradicts the ordering of the elements in $\mathbf{x}_i$.

According to the model, the transformed individual importance rankings are linear functions of the individual vectors, with the value points providing the coefficients:

$$\widehat{\mathbf{X}}^* = \mathbf{\Gamma}\ \mathbf{\Phi'} \tag{1}$$

In equation (1), the $\widehat{\mathbf{X}}^*$ on the left-hand side is a matrix of predicted importance values that is generated from the estimated model parameters. That is, the entries in any row of $\widehat{\mathbf{X}}^*$, say $\widehat{\mathbf{x}}^*{}_i$, give the projections from the k value points onto i's vector within the m-space. Each element in $\widehat{\mathbf{x}}^*{}_i$, say $\hat{x}^*_{ij}$, is a scalar product between individual i's vector and the point for value j:

$$\hat{x}^*_{ij} = \boldsymbol{\gamma}_i \cdot \boldsymbol{\phi}'_j \tag{2}$$

Where $\boldsymbol{\gamma}_i$ is the i^{th} row of the $\mathbf{\Gamma}$ matrix and $\boldsymbol{\phi}_j$ is the j^{th} row of the $\mathbf{\Phi}$ matrix Of course, $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are constructed in a way that optimizes the correspondence between $\widehat{\mathbf{X}}^*$ and the original data matrix, $\mathbf{X}$. So, the full model can be shown as follows:

$$\mathbf{X}^* = \mathbf{\Gamma}\ \mathbf{\Phi'} + \mathbf{E} \tag{3}$$

Where $\mathbf{E}$ is an n by k matrix of random errors, with $\bar{\mathbf{E}} = 0$ and the variance of the elements in $\mathbf{E}$ as small as possible.

Estimation with Interval-Level Data

Let us assume for the moment that $\mathbf{X}$ contains interval-level data. In that case, Carroll (1972) shows that the model can be estimated very easily, using a singular value decomposition. Begin by standardizing the entries within each row to zero mean and unit variance, producing $\mathbf{X}_{std}$. While not absolutely necessary, this preliminary step is useful because it places the origin of the space at the centroid of the points and vector termini. Next, factor $\mathbf{X}_{std}$ using the Eckart-Young decomposition (Eckart and Young 1936):

$$\mathbf{X}_{std} = \mathbf{UDV}' \tag{4}$$

On the right-hand side of equation (3), $\mathbf{U}$ is the n by q matrix of left singular vectors, $\mathbf{D}$ is the q-order diagonal matrix of singular values (arranged from largest to smallest), and $\mathbf{V}$ is the k by q matrix of right singular vectors. Note that q is the rank of $\mathbf{X}_{std}$, which typically will be $k-1$ (assuming that $k < n$, as will generally be the case), since the scores add to a constant within each row.

The next step is to determine m, the dimensionality of the model space. This is specified by the analyst, but the general objective is to choose m so that it is as small as possible, while still producing a model that provides a sufficiently good fit to the empirical data. Some guidance can be obtained from the fact that the squared singular values give the sums of squares in $\mathbf{X}_{std}$ that are "explained" by each pair of singular vectors. A goodness-of-fit measure for the model in m dimensions can be defined as follows:

$$\mathrm{R}^2 = tr(\mathbf{D}_m^2)/tr(\mathbf{D}^2) \tag{5}$$

Where tr is the matrix trace, or sum of the diagonal elements, and $\mathbf{D}_m^2$ is the diagonal matrix containing the squares of the first m singular values. As with a linear regression model, R^2 is interpreted as the proportion of variance in $\mathbf{X}_{std}$ that is explained by the m-dimensional model. Alternatively, it is the squared correlation between the entries in $\mathbf{X}_{std}$ and the entries in the $\widehat{\mathbf{X}}$ that is produced by the points and vectors in m-space.

After determining the appropriate value of m, it is a simple matter to obtain $\mathbf{\Phi}$ and $\mathbf{\Gamma}$. Take the first m singular vectors and singular values and use them to define the following:

$$\mathbf{\Phi} = \mathbf{V}_m \tag{6}$$

$$\mathbf{\Gamma} = \mathbf{U}_m\mathbf{D}_m \tag{7}$$

These values for $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ comprise the "best" solution in the least-squares sense, because they generate the largest R^2 that is possible for an m-dimensional representation of the data.

Typically, the rows of $\mathbf{\Gamma}$ are normalized to unit length by dividing the entries in each row by the sum of squared entries in that row.

Estimation with Ordinal-Level Data

The preceding estimation procedure assumes that $\mathbf{X}$ contains interval-level data. This implies that, for $i = 1, 2, \ldots, n$, the projections from the k value points onto i's vector, or $\widehat{\mathbf{x}}_i$, are *linearly* related to the entries in the corresponding row of the data matrix, $\mathbf{x}_i$. But, the entries within each row of $\mathbf{X}$ only give the rank-order of each individual's importance judgments about the values. It seems very unlikely either that the differences in importance across successive ranks are always constant, or that these differences are identical from one individual to the next. Therefore, it is more appropriate to specify a model in which the projections from the value points onto the individual vectors are idiosyncratic and row-specific *monotonic* functions of the importance ratings. This is equivalent to an assumption that each row of $\mathbf{X}$ provides strictly *ordinal* and inter-personally incomparable information about a person's value preferences.

In order to address this issue, a strategy called "alternating least squares, optimal scaling" or ALSOS (Young 1981; Jacoby 1999), is used to perform a nonmetric version of the singular value decomposition. ALSOS does not carry out the analysis on the original data matrix. Instead, the ALSOS routine uses a transformation of $\mathbf{X}$, designated $\mathbf{X}^*$, that contains *optimally-scaled* versions of the original data values. This means that the entries within any row of $\mathbf{X}^*$, say $\mathbf{x}_i^*$, are a monotonic transformation of the entries in the corresponding row of $\mathbf{X}$, or $\mathbf{x}_i$. The specific monotonic transformation is allowed to vary across the n rows of $\mathbf{x}_i^*$. The monotonic transformations are chosen so that they maximize the model's R^2 in a given dimensionality, m. Thus, $\mathbf{x}_i^*$ is an *optimally-scaled* version of the input data in the sense that it is the vector of numeric values that is most highly-correlated with the model-based predicted values (i.e., $\widehat{\mathbf{x}}_i$), subject to the constraint that it is also monotonic with the original $\mathbf{x}_i$.

The only real difference from the interval-level situation is that the model R^2 is now the squared correlation between the entries in $\widehat{\mathbf{X}}$ and the entries in $\mathbf{X}^*$ (rather than $\mathbf{X}_{std}$). Briefly, the steps in an ALSOS version of multidimensional preference scaling are as follows:

1. At the outset, specify m, the dimensionality of the space, initialize R^2 to zero, and initialize $\mathbf{X}^*$ by setting it equal to the original $\mathbf{X}$ matrix. Standardize within rows of $\mathbf{X}^*$ to obtain $\mathbf{X}^*_{std}$

2. Carry out the singular value decomposition on the current version of $\mathbf{X}^*_{std}$ to obtain $\mathbf{U}_m$, $\mathbf{D}_m$, and $\mathbf{V}_m$. Use these matrices to calculate current estimates of $\widehat{\mathbf{X}}$ and R^2.

3. If the current R^2 is larger than the previous value, then continue. If R^2 has not changed from the previous iteration (i.e., its value has converged) then terminate the procedure and go to step 6.

4. For $i = 1, 2, \ldots, n$, use Kruskal's monotonic regression (1964) to find a new estimate of $\widehat{\mathbf{x}}^*_i$, containing values that are maximally correlated with the current model-based predicted values (that is, the entries in the i^{th} row of $\widehat{\mathbf{X}}$, or $\widehat{\mathbf{x}}_i$) but always weakly monotonic to the entries in the original $\mathbf{x}_i$.

5. Return to step 2 and carry out another iteration of the estimation procedure on the new version of $\mathbf{X}^*$ that was obtained in step 4.

6. When R^2 converges, construct the $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ matrices from the singular vectors and values, and use the final R^2 as the goodness-of-fit.

Thus, the nonmetric approach simply estimates the MDPREF model on a transformed (i.e., optimally-scaled) version of the data matrix. Note that separate transformations are obtained for each individual by performing Kruskal's monotonic regression repeatedly, for each row of $\widehat{\mathbf{X}}$, relative to the corresponding row of $\mathbf{X}_{std}$, producing the same row in $\mathbf{X}^*$. By carrying out these row-specific transformations, the model explciitly takes into account the

problem known either as "inter-personal incomparability" (Brady 1985) or "differential item functioning" (King, Murray, Salomon, Tandor 2004).

In practice, the combined ALSOS-MDPREF procedure works very well. It usually converges quickly and it optimizes the appropriate *monotonic*, rather than linear, correspondence between the model elements and the original data values. The ALSOS model still represents the original data, since the various $\widehat{\mathbf{x}}_i^*$ are linear functions of the $\mathbf{x}_i^*$, while the latter are monotonic functions of the $\mathbf{x}_i$. And, since the monotonic function can vary from one row of $\mathbf{X}^*$ to the next, the procedure explicitly recognizes that the entries in the original data matrix, $\mathbf{X}$, are not comparable across the rows. Hence the model provides the best-fitting (in the least-squares sense) m-dimensional representation of the n individuals' rank-ordered importance ratings of the k values.

REFERENCES

Brady, Henry. (1985) "The Perils of Survey Research: Inter-Personally Incomparable Responses." *Political Methodology* 11: 269-291.

Carroll, J. D. (1972) "Individual Differences in Multidimensional Scaling." In R. N. Shepard, A. K. Romney, J. B. Nerlove (Editors) *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Volume I.* New York, NY: Seminar Press.

Eckart, Carl H. and Gale Young. (1936) "The Approximation of One Matrix by Another of a Lower Rank." *Psychometrika* 1: 211-d218.

Jacoby, William G. (1999) "Levels of Measurement and Political Research: An Optimistic View." *American Journal of Political Science* 43: 271-301.

King, Gary; Christopher J. L. Murray; Joshua A. Salomon; Ajay Tandon. (2004) "Correction: Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98: 191-207.

Kruskal, Joseph B. (1964) "Non-metric Multidimensional Scaling: A Numerical Method." *Psychometrika* 29: 115-129.

Weller, Susan C. and A. Kimball Romney. (1990) *Metric Scaling: Correspondence Analysis.* Newbury Park, CA: Sage Publications.

Young, Forrest W. (1981) "Quantitative Analysis of Qualitative Data." *Psychometrika* 46: 357-388.