*Recovering a Basic Space From a Set of Issue Scales**

Keith T. Poole, *Carnegie-Mellon University*

This paper develops a scaling procedure for estimating the latent/unobservable dimensions underlying a set of manifest/observable variables. The scaling procedure performs, in effect, a singular value decomposition of a rectangular matrix of real elements with missing entries. In contrast to existing techniques such as factor analysis which work with a correlation or covariance matrix computed from the data matrix, the scaling procedure shown here analyzes the data matrix *directly*.

The scaling procedure is a general-purpose tool that can be used not only to estimate latent/unobservable dimensions but also to estimate an Eckart-Young lower-rank approximation matrix of a matrix with missing entries. Monte Carlo tests show that the procedure reliably estimates the latent dimensions and reproduces the missing elements of a matrix even at high levels of error and missing data.

A number of applications to political data are shown and discussed.

1. Introduction

The purpose of this paper is to show a scaling method for estimating latent/unobservable dimensions underlying a set of manifest/observable variables. The data to be analyzed is assumed to be in the form of a rectangular matrix of real elements with missing entries. What the method does, in effect, is to perform a singular value decomposition of the rectangular matrix with missing elements. In contrast to existing techniques such as factor analysis which work with a correlation or covariance matrix computed from the data matrix, the scaling procedure shown here analyzes the data matrix directly without any intervening transformations of the original data.

For example, asking respondents to place themselves and/or stimuli on issue/attribute scales is a common type of data gathered by social scientists. The Center for Political Studies at the University of Michigan has been collecting seven-point scale data in its National Election Studies since 1968. The endpoints of these scales are labeled, and the respondent is asked to place him or herself on the scale (his or her "ideal point") along with a set of political figures and, in some cases, the two political parties and current federal government policy.

*I would like to thank Howard Rosenthal, Nolan McCarty, Tim Groseclose, and three anonymous reviewers for their very helpful comments and suggestions. The software that performs the analyses shown in this article and documentation on how to use the software along with many additional empirical examples of its use (Poole 1997) can be downloaded from http://k7moa.gsia.cmu.edu.

American Journal of Political Science, Vol. 42, No. 3, July 1998, Pp. 954–993

The extent to which a set of issue scale placements arises from common underlying latent (evaluative) dimensions is an important empirical question for political scientists. Consider the matrix of self-placements where the columns correspond to a set of seven-point scales, and each row is a respondent's self-placement on these scales. Suppose political theory suggests that the observed placements are generated by linear mappings from two latent dimensions (e.g., liberal-conservative and racial attitude). If there were no error and no missing data, the solution to this problem is immediate. Because the rank of the matrix is two, the singular value decomposition of the matrix will produce only two singular values and two left and two right singular vectors. The latent dimensions are the two left singular vectors.

Continuing with the example, if there is no missing data and error is present, then, provided the error is generated by a distribution that is symmetric with zero mean, the best two-dimensional approximation to this matrix is given by the famous Eckart-Young (1936) theorem. Namely, given the singular value decomposition of the matrix, set all but the two largest singular values to zero and remultiply (see Section 2 below). The two left singular vectors corresponding to the two largest singular values are the estimates of the two latent dimensions. This is the same solution as the error free case.

In short, if there is no missing data, obtaining estimates of latent dimensions is easy provided certain assumptions are maintained; namely, if the observed data are generated linearly from the latent dimensions and the error process is symmetric with zero mean, then the left singular vectors are the estimates of the latent dimensions.

If there are missing elements in the matrix, however, then things are not so simple. Because each matrix has a unique singular value decomposition,[1] the problem of estimating the latent dimensions cannot be disentangled from the problem of estimating the missing elements of the matrix. The purpose of this paper is to show a solution for this problem.

In Section 2 a simple model of latent/evaluative dimensions is stated, and a procedure for estimating them from a set of issue/attribute scales is developed. The motivation and the examples are all drawn from political science but the technique is a general one. It can be applied to *any matrix of real numbers*. The model outlined in Section 2 is motivated, however, by the spatial theory of choice. In particular, it corresponds to a spatial theory proposed by Ordeshook (1976) and by Hinich and Pollard (1981) and then elaborated by Hinich and his colleagues (Enelow and Hinich 1984; Hinich and Munger,

[1]If two or more singular values are identical, then the corresponding singular vectors can be interchanged. Technically, these are different decompositions but they only differ in the fact that the corresponding pair(s) of dimensions are interchanged.

1994, 1997). In standard spatial theory, each issue is modeled as an ordered dimension of alternatives, and each respondent is assumed to have an ideal point on, and single-peaked preferences over, each issue dimension. If the respondents have highly structured belief systems (Converse 1964), then this means that the issues lie on a low-dimensional hyperplane through the issue space. This low dimensional space was dubbed a *basic space* by Ordeshook (1976) and the *predictive dimensions* by Hinich. More generally, these are latent or evaluative dimensions and in political science work are commonly referred to as *ideological dimensions* (Hinich and Munger 1994). I will refer to these as *basic* dimensions in the discussion below.[2]

The scaling procedure developed below recovers basic dimensions from issue scales. Although 40 years of empirical work by political scientists has shown that the American public is not *strictly* ideological in that most people do not have *highly structured* belief systems, they also are not "ideologically innocent" either (Feldman and Zaller 1992). Ideological consistency, that is, the degree to which issue positions are coherently generated by basic dimensions, clearly varies. Below I find that about half the variance of the individuals' seven-point issue scale positions in the NES 1980 survey is explained by a single basic dimension. In contrast, Poole and Rosenthal (1997) show considerable evidence that members of Congress are quite consistent over long periods of time on one and sometimes two basic dimensions. Although it is beyond the scope of this paper, the empirical examples shown in Section 5 suggest that ideological consistency is a top-down phenomenon. Political elites are more ideologically consistent than the mass public, and it is quite likely that this has an impact on how issues are "packaged" (Hinich and Munger 1997, chap. 9).

The estimation procedure uses an alternating least squares approach, and the method for estimating the missing entries is best thought of as a hybrid of the principal components and the regression methods discussed by Gleason and Staelin (1975).[3] The approach taken here is unique, however, because the observed (nonmissing) elements of the data matrix are used in an alternating least squares procedure to estimate a *lower rank approximation of the entire matrix*. The aim is not to estimate the missing data per se. Rather, the estimates of the missing data are a byproduct of the lower rank

[2]My use of the label "basic dimensions" is also motivated by the fact that Horst (1963) refers to the singular value decomposition of a rectangular matrix as the *basic structure* of a matrix.

[3]Gleason and Staelin test and compare a number of methods for estimating missing data. Their preferred method is principal components. They found little difference, however, between the regression method and the principal components method and preferred the latter in part because of computer time considerations (1975, 245).

approximation. In contrast, traditional approaches to estimating missing data reconstruct the missing data *directly from the observed data.*

For example, suppose the data matrix is 500 respondents' self-placements on fifteen issue scales and that 50% of the data is missing. Suppose further that theory suggests two latent dimensions. The procedure developed in Section 2 estimates a matrix of rank two such that the sum of squared differences between the elements of the estimated matrix and the observed elements of the data matrix is minimized. Geometrically, the data matrix is a set of 500 points in a fifteen dimensional space where, because of the missing entries, most of the points only have coordinates on seven or eight dimensions. The missing data can be thought of as being zeroes on the corresponding dimensions so that the 500 points on average lie on seven or eight dimensional hyperplanes through the fifteen dimensional space. The least squares problem is equivalent to finding a two dimensional plane through this fifteen dimensional space that comes as close as possible to these points. The missing entries are given by the estimated plane.

In contrast, using a traditional approach to estimate missing entries produces a matrix such that, geometrically, the corresponding hyperplane passes through *all the observed coordinates* and is almost certainly of full rank. In terms of the 500 by 15 example in the previous paragraph, this would almost certainly produce a matrix of rank 15. The two dimensional approximation to *this* matrix is *not* the same as the approach developed in Section 2. Indeed, it is distinctly inferior.

Monte Carlo tests of the procedure are discussed in Section 3. The Monte Carlo tests show that the estimation procedure accurately reproduces the true data even with high levels of error and missing data levels as high as 70%. In addition, the procedure accurately reproduces the true missing data.

The relationship of the model stated in Section 2 with other scaling techniques is discussed in Section 4. The scaling procedure developed below is unique in that it works *directly* with the rectangular data matrix (denoted by $\mathbf{X_0}$ below). Most current scaling techniques deal with a covariance or correlation matrix formed from the rectangular data matrix by either list-wise (throwing out any row with missing data) or pair-wise deletion (that is, constructing some version of $\mathbf{X_0}'\mathbf{X_0}$). In contrast to other techniques, the scaling procedure developed here allows the recovery of latent dimensions from very sparse matrices in which *every row* has missing data (for example, overlapping generations data gathered over a long period of time; or split sample data).

Two applications are shown in Section 5. The first is to the self-placements of respondents on a set of issue scales in the 1980 CPS national election study. The second is to W-NOMINATE scores of the House and Senate from 1937 to 1995.

2. The Model

Let x_{ij} be the ith individual's (i = 1, . . . , n) reported position on the jth issue (j = 1, . . . , m) and let $\mathbf{X}_0$ be the n by m matrix of observed data where the "0" subscript indicates that elements are missing from the matrix—not all individuals report their positions on all issues. Let ψ_{ik} be the ith individual's position on the kth (k = 1, . . . , s) basic dimension. The model estimated is:

$$\mathbf{X}_0 = \left[\Psi\mathbf{W}' + \mathbf{J}_n\underline{\mathbf{c}}'\right]_0 + \mathbf{E}_0 \qquad [1A]$$

where Ψ is the n by s matrix of coordinates of the individuals on the basic dimensions, $\mathbf{W}$ is an m by s matrix of weights, $\underline{\mathbf{c}}$ is a vector of constants of length m, $\mathbf{J}_n$ is an n length vector of ones, and $\mathbf{E}_0$ is an n by m matrix of error terms. $\mathbf{W}$ and $\underline{\mathbf{c}}$ map the individuals from the basic space onto the issue dimensions.

I assume that the elements of $\mathbf{E}_0$ are random draws from a symmetric distribution with zero mean.

Without loss of generality, the centroid of the coordinates of the individuals on the basic dimensions may be placed at the origin; that is, $\mathbf{J}_n'\Psi = \underline{\mathbf{0}}'$ (note that this is simply taking the sum of the elements of each column), where $\underline{\mathbf{0}}$ is an s length vector of zeroes. Because $\mathbf{J}_n'\Psi = \underline{\mathbf{0}}'$, if there were no missing data, then $\mathbf{J}_n'[\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}'] = \underline{\mathbf{0}}'$ where $\underline{\mathbf{0}}$ is an m length vector of zeroes.

Equation [1A] can be written as the product of partitioned matrices

$$\mathbf{X}_0 = \left[\Psi|\mathbf{J}_n\right]\left[\frac{\mathbf{W}'}{\underline{\mathbf{c}}'}\right]_0 + \mathbf{E}_0 \qquad [1B]$$

where $\left[\psi|\mathbf{J}_n\right]$ is a n by s+1 matrix, and $\left[\mathbf{W}|\underline{\mathbf{c}}\right]$ is an m by s+1 matrix. If n > m and there is no error or missing data, then the rank of $\mathbf{X}$ is s and the rank of $\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}'$ is less than or equal to s.[4] If m > n and there is no error or missing data, then the rank of $\mathbf{X}$ is s and if $\mathbf{J}_n'\mathbf{X} \neq \underline{\mathbf{0}}$, then the rank of $\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}'$ is s – 1 because $\mathbf{J}_n'[\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}'] = \underline{\mathbf{0}}'$. That is, subtracting off the column means from $\mathbf{X}$ so that the n entries in each of the m columns sum to zero reduces the rank by 1.

If there were no missing data in [1], then it could be estimated quite simply by using singular value decomposition. In particular, the following two well known matrix decomposition theorems can be utilized to solve [1].

[4]In most circumstances, subtracting the column means will leave the rank unchanged. If the entries are squared distances, however, then subtracting the column means will reduce the rank of the matrix by 1.

Theorem I (Singular Value Decomposition)

Let **A** be an n by m matrix of real elements with $n \geq m$. Then there is an n by n orthogonal matrix **U**, an m by m orthogonal matrix **V**, and an n by m matrix Λ such that

$$\mathbf{A} = \mathbf{U}\,\Lambda\,\mathbf{V}' \quad \text{and} \quad \mathbf{U}'\mathbf{A}\mathbf{V} = \Lambda$$

where

$$\Lambda = \begin{bmatrix} \Lambda_{\mathbf{m}} \\ \hline \mathbf{0} \end{bmatrix}$$

where $\Lambda_{\mathbf{m}}$ is an m by m diagonal matrix and **0** is an n – m by m matrix of zeroes. The diagonal entries of $\Lambda_{\mathbf{m}}$ are nonnegative with exactly s entries strictly positive ($s \leq m$).

Theorem II (Eckart and Young)

Given an n by m matrix **A** of rank $r \leq m \leq n$ and its singular value decomposition, $\mathbf{U}\Lambda\mathbf{V}'$, with the singular values arranged in decreasing sequence

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \lambda_m \geq 0$$

then there exists an n by m matrix **B** of rank s, $s \leq r$, which minimizes the sum of the squared error between the elements of **A** and the corresponding elements of **B** when

$$\mathbf{B} = \mathbf{U}\Lambda_{\mathbf{s}}\mathbf{V}'$$

where the diagonal elements of $\Lambda_{\mathbf{s}}$ are

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \lambda_s > \lambda_{s+1} = \lambda_{s+2} = \ldots = \lambda_m = 0$$

Theorem I states that every real matrix can be written as the product of two orthogonal matrices and one diagonal matrix.[5] Theorem II states that the

[5] Theorem I was stated by Eckart and Young (1936) in their famous paper but they did not provide a proof. The first proof that every *rectangular* matrix of real elements can be decomposed as shown in Theorem I was given by Johnson (1963). Horst (1963) refers to the decomposition shown in Theorem I as the *basic structure* of a matrix and discusses the mechanics of matrix decomposition in detail in chapters 17 and 18. A more recent treatment can be found in chapters 1 and 2 of Lawson and Hanson (1974).

least squares approximation in s dimensions of a matrix **A** can be found by replacing the smallest m-s roots of Λ with zeroes and remultiplying $\mathbf{U}\Lambda\mathbf{V}'$.[6]

Because the lower n-m rows of Λ are all zeros, it is convenient to discard them and work only with the m by m diagonal matrix Λ_m. In addition, the n-m eigenvectors in **U** corresponding to the n-m lower rows of Λ may also be discarded. Hereafter, unless otherwise noted, Λ will be treated as a square diagonal matrix with the singular values on the diagonal (that is, from now on $\Lambda = \Lambda_m$). Hence **U** is an n by m matrix, Λ is an m by m diagonal matrix, and **V** is an m by m matrix. A decomposition according to Theorem I will be assumed to be in this form.

To solve [1], set $\underline{\mathbf{c}}$ equal to the column means of **X**; that is

$$\mathbf{c}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n} = \bar{x}_j$$

and perform a singular value decomposition of $\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}'$. By Theorem I

$$\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}' = \mathbf{U}\Lambda\mathbf{V}' = \Psi\mathbf{W}'$$

where, as noted above, **U** is an n by m matrix, Λ is an m by m matrix, and **V** is an m by m matrix.

A simple solution for Ψ and **W** is

$$\begin{aligned} \Psi &= \mathbf{U}\Lambda^{\frac{1}{2}} \\ \mathbf{W} &= \mathbf{V}\Lambda^{\frac{1}{2}} \end{aligned} \qquad [2]$$

where the diagonal elements of $\Lambda^{\frac{1}{2}}$ are the square roots of Λ. Let $\mathbf{I}_m$ be the m by m identity matrix. Equation [2] implies that $\Psi'\Psi = \mathbf{W}'\mathbf{W}$. That is:

$$\Psi'\Psi = \Lambda^{\frac{1}{2}}\mathbf{U}'\mathbf{U}\Lambda^{\frac{1}{2}} = \Lambda^{\frac{1}{2}}\mathbf{I}_m\Lambda^{\frac{1}{2}} = \Lambda$$

and

$$\mathbf{W}'\mathbf{W} = \Lambda^{\frac{1}{2}}\mathbf{V}'\mathbf{V}\Lambda^{\frac{1}{2}} = \Lambda^{\frac{1}{2}}\mathbf{I}_m\Lambda^{\frac{1}{2}} = \Lambda$$

[6]Theorem II was never explicitly stated by Eckart and Young. Rather, they use two theorems from linear algebra (Theorem I was the first) and a very clever argument to show the truth of their result. Later, Keller (1962) independently rediscovered the Eckart-Young result (Theorem II).

In addition, as noted above, because $\mathbf{J}_n'[\mathbf{X} - \mathbf{J}_n\underline{\mathbf{c}}'] = \underline{\mathbf{0}}'$, then $\mathbf{J}_n'\mathbf{U} = \mathbf{J}_n'\Psi = \underline{\mathbf{0}}'$, where $\underline{\mathbf{0}}$ is an m length vector of zeros.

When an s < m is preferred, Theorem II may be used in [2] to arrive at solutions for Ψ and $\mathbf{W}$. That is, the s + 1 to m singular values are set equal to zero so that Ψ and $\mathbf{W}$ from [2] are n by s and m by s matrices, respectively.

Because of the presence of missing data, Theorems I and II cannot be used directly. Instead, I shall work with the loss function

$$\xi = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\{[\sum_{k=1}^{s}\psi_{ik}\mathbf{W}_{jk}] + \mathbf{c}_j - \mathbf{x}_{ij}\}^2 \qquad [3]$$

which, if there were no missing data, is the function which is minimized by Theorem II when $c_j = 0$. The notation m_i means that the total of the summation over j may vary from s + 1 to m depending on how many entries there are in the ith row of $\mathbf{X}_0$. That is, each individual must report at least s + 1 issue positions in order to be identified. Furthermore, the number of missing entries in the columns of $\mathbf{X}_0$ must also be restricted. In most practical applications n will be much larger than m. Consequently, I will adopt the convention that there must be at least 2m entries in each column of $\mathbf{X}_0$.

In line with the discussion above, the following two restrictions are applied to the loss function:

$$\Psi'\Psi = \mathbf{W}'\mathbf{W} \text{ and } \mathbf{J}_n'\Psi = \underline{\mathbf{0}}'$$

These restrictions produce the Lagrangean multiplier problem

$$\mu = \xi + 2\underline{\gamma}'[\Psi'\mathbf{J}_n] + \text{tr}[\Phi(\Psi'\Psi - \mathbf{W}'\mathbf{W})] \qquad [4]$$

where $\underline{\gamma}$ is an s length vector of Lagrangean multipliers and Φ is a symmetric s by s matrix of Langrangean multipliers.

In Appendix A, I show that all the Lagrangean multipliers are zero. Intuitively, $\underline{\gamma} = \underline{\mathbf{0}}$ because the columns of Ψ can always be set equal to zero because of the presence of the vector of constants, $\underline{\mathbf{c}}$. $\Phi = \mathbf{0}$ because, given an estimated Ψ and $\mathbf{W}$, Theorem I can be used at any time as shown in equation [2] to produce Ψ and $\mathbf{W}$ such that $\Psi'\Psi = \mathbf{W}'\mathbf{W}$. However, because these constraints are important in the way the model specified in equation [1] is estimated, Appendix A shows a more formal demonstration.

Given that the Lagrangean multipliers are all zero, the partial derivatives of Ψ, $\mathbf{W}$, and $\underline{\mathbf{c}}$ from equations [3] and [4] are identical. In particular:

$$\frac{\partial\mu}{\partial\Psi_{ik}} = 2\sum_{j=1}^{m_i}\left[\left(\sum_{\ell=1}^{s}\mathbf{w}_{j\ell}\psi_{i\ell}\right) + \mathbf{c}_j - \mathbf{x}_{ij}\right]\mathbf{w}_{jk} \qquad [5A]$$

$$\frac{\partial \mu}{\partial w_{jk}} = 2\sum_{i=1}^{n_j}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right]\psi_{ik} \qquad [5B]$$

$$\frac{\partial \mu}{\partial c_j} = 2\sum_{i=1}^{n_j}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right] \qquad [5C]$$

where n_j means that the total of the summation over i may vary from 2m to n depending upon how many entries there are in the ith column of $\mathbf{X}_0$.

Setting [5A] to zero and collecting the s partial derivatives of the ith row of Ψ into a vector and dividing by 2 produces

$$[\mathbf{W}^{*\prime}\mathbf{W}^{*}]\underline{\psi}_i - \mathbf{W}^{*\prime}[\underline{x}_{oi} - \underline{c}_o] = \underline{0}$$

where $\mathbf{W}^*$ is an m_i by s matrix with the appropriate rows corresponding to missing entries in $\mathbf{X}_o$ removed, $\underline{\psi}_i$ is the ith row of Ψ, $\underline{x}_{oi}$ is the ith row of $\mathbf{X}_0$ and is of length m_i, $\underline{c}_o$ is the m_i length vector of constants corresponding to the elements of $\underline{x}_{oi}$, and $\underline{0}$ is an s length vector of zeroes.

If $\mathbf{W}^{*\prime}\mathbf{W}^*$ is nonsingular, then

$$\hat{\underline{\psi}}_i = (\mathbf{W}^{*\prime}\mathbf{W}^{*})^{-1}\mathbf{W}^{*\prime}\left[\underline{x}_{oi} - \underline{c}_o\right] \qquad [6]$$

and the rows of Ψ can be estimated through ordinary least squares.

The s partial derivatives of the jth row of $\mathbf{W}$ from equation [5B] and the partial derivative for $\underline{c}_j$ from [5C] can be collected into the vector

$$\left[\Psi_j^{*\prime}\ \Psi_j^{*}\right]\begin{bmatrix}\underline{w}_j \\ c_j\end{bmatrix} - \Psi_j^{*\prime}\underline{x}_{oj} = \underline{0}$$

where $\Psi_j^* = [\Psi_o|\mathbf{J}_o]$ is an n_j by s+1 matrix (the matrix Ψ with the appropriate rows corresponding to missing data removed and then bordered by ones), $\underline{w}_j$ is the s length vector of the jth row elements of $\mathbf{W}$, c_j is the jth element of $\underline{c}$, $\underline{x}_{oj}$ is the jth column of $\mathbf{X}_o$ and is of length n_j, and $\underline{0}$ is an s+1 length vector of zeroes.

If $\Psi_j^{*\prime}\Psi_j^*$ is nonsingular, then

$$\begin{bmatrix}\hat{\underline{w}}_j \\ \hat{c}_j\end{bmatrix} = \left(\Psi_j^{*\prime}\ \Psi_j^{*}\right)^{-1}\Psi_j^{*\prime}\ \underline{x}_{oj} \qquad [7]$$

and the rows of $\mathbf{W}$ and the elements of $\underline{c}$ can be estimated through ordinary least squares.

The easiest way to estimate $\mathbf{W}$ and Ψ is to select some suitable starting estimate of either matrix and then iterate between [6] and [7] until convergence is achieved. This alternating least squares (ALS) procedure is similar in form to that employed by Carroll and Chang (1970) and Takane, Young, and de Leeuw (1977) for individual differences scaling. The constraints on $\mathbf{W}$ and Ψ can be met at any stage of the iteration by simply setting the column means of $\hat{\Psi}$ equal to zero, forming the matrix product $\hat{\Psi}\hat{\mathbf{W}}'$, and performing a singular value decomposition of $\hat{\Psi}\hat{\mathbf{W}}'$ according to Theorem I. That is,

$$\hat{\Psi}\hat{\mathbf{W}}' = \mathbf{U}\Lambda\mathbf{V}'$$

where Λ is an s by s diagonal matrix containing the s singular values in descending order, and $\mathbf{U}$ and $\mathbf{V}$ are n by s and s by s matrices, respectively, such that $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_s$. Setting $\hat{\Psi} = \mathbf{U}\Lambda^{\frac{1}{2}}$ and $\hat{\mathbf{W}} = \mathbf{V}\Lambda^{\frac{1}{2}}$ as in [2] satisfies the constraints.

A simple way to proceed with the estimation is to exploit the orthogonality of Ψ and estimate one column of Ψ and $\mathbf{W}$ at a time. This is motivated by the fact that if the n_j are close to n, $\Psi_j^{*\prime}\Psi_j^*$ in [7] will be very close to a diagonal matrix. This process begins with computing simple starting estimates of c_j and the first row of $\mathbf{W}$, the m length vector, $\underline{w}_1$, and then using these to obtain starting estimates of Ψ_{i1} from the formula:

$$\hat{\psi}_{i1} = \frac{\sum_{j=1}^{m_i} \hat{w}_{j1}\left(x_{ij} - \hat{c}_j\right)}{m_i} \qquad [8]$$

Consider the m_i terms in the numerator of equation [8]. Intuitively, if the $\hat{w}_{j1}(x_{ij} - \hat{c}_j)$ terms in the numerator all have the same sign, then this maximizes the absolute value of the sum in the numerator and would maximize $\hat{\psi}_{i1}^2$. Note that if this was true for all i, that is, for every row of $\mathbf{X_0} - \mathbf{J_n}\underline{c}'$, then every entry in the m by m covariance matrix, $[\mathbf{X_0} - \mathbf{J_n}\underline{c}']'[\mathbf{X_0} - \mathbf{J_n}\underline{c}']$ would be positive. Given the c_j, the starting estimate of $\underline{w}_1$ is set equal to a vector of plus and minus ones that maximizes the number of positive entries in the covariance matrix (see Appendix B). This is a convenient starting point because it tends to maximize the sum of the $\hat{\psi}_{i1}^2$.[7]

The first step is to obtain starting estimates of the c_j. These are simply taken to be the column means of $\mathbf{X}_0$:

[7]This technique of finding the vector of plus and minus ones that maximizes the number of positive terms in the covariance matrix is very similar to the approach used to speed the convergence of the simple algorithm to find the eigenvectors and eigenvalues of a matrix. See Van de Geer (1971, 273–6) and Horst (1963, chap. 18) for detailed examples.

$$\hat{c}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} = \bar{x}_j$$

To obtain a starting estimate of $\underline{w}_1$, let Γ be an m by m diagonal matrix where the diagonal entries are either +1 or –1. An iterative search is conducted to find a Γ that maximizes the number of positive elements in the m by m covariance matrix

$$\Gamma[\mathbf{X}_0 - \mathbf{J}_n\underline{c}']'[\mathbf{X}_0 - \mathbf{J}_n\underline{c}']\Gamma$$

Appendix B shows a simple numerical example of how to estimate Γ. The diagonal of Γ is used as the starting estimate of $\underline{w}_1$.

This first estimate of the first column of Ψ from equation [8], $\hat{\underline{\psi}}_1^{(1)}$ (where the superscript indicates the iteration number), is returned to [7] to get a second estimate of $\underline{w}_1$, $\hat{\underline{w}}_1^{(2)}$, and a new estimate of $\underline{c}_j$, $\hat{\underline{c}}_j^{(2)}$. These are returned to [6] to reestimate $\underline{\psi}_1$. After each reestimate of $\underline{\psi}_1$, it is adjusted so that its mean is equal to zero and its sum of squares is held constant; namely, at the hth iteration:

$$\sum_{i=1}^{n} \hat{\psi}_{i1}^{(h)^2} = \sum_{i=1}^{n} \hat{\psi}_{i1}^{(1)^2}$$

Going back and forth between [6] and [7] will always reduce the sum of squared error. This process is continued until there is no further improvement in the sum of squared error. This usually takes not more than five iterations. At convergence, the $\hat{\psi}_1^{(h)}$ from equation [6] and the $\hat{\underline{w}}_1^{(h)}$ and $\hat{\underline{c}}_j^{(h)}$ from equation [7] reproduce each other.

The second column of Ψ is computed in the same way as the first, only now $\mathbf{X}_0$ is replaced by the matrix of residuals

$$\mathbf{E}_{01} = \mathbf{X}_0 - \hat{\underline{\psi}}_1\hat{\underline{w}}'_1 - \mathbf{J}_n\hat{\underline{c}}' \qquad [9]$$

Note that the columns of $\mathbf{E}_{01}$ and subsequent residual matrices ($\mathbf{E}_{02}$, . . . , $\mathbf{E}_{0s-1}$) sum to zero due to the standard OLS form of equation [6]. (The rows of the residual matrices do not necessarily sum to zero.) Consequently, unlike the starting estimates of $\hat{\psi}_{i1}$ shown in equation [8] which use the c_j, the starting estimates of $\hat{\psi}_{i2}$ are simply:

$$\hat{\psi}_{i2} = \frac{\sum_{j=1}^{m_i} \hat{w}_{j2}e_{1ij}}{m_i} \qquad [10]$$

where, as before, the $\hat{w}_{j2}$ are plus or minus ones. To obtain a starting estimate of $\underline{w}_2$ an iterative search is conducted to find an m by m diagonal matrix of plus and minus ones, Γ, that maximizes the number of positive elements in the m by m covariance matrix:

$$\Gamma \mathbf{E}'_{01} \mathbf{E}_{01} \Gamma$$

As with the starting estimates for $\underline{\hat{w}}_1$, the diagonal of Γ is used as the starting estimate of $\underline{\hat{w}}_2$. Because $\underline{c}$ has been estimated, $\underline{\hat{\psi}}_2$ does not have to be bordered by ones to construct Ψ_j^* for use in equation [7]. Instead, equation [7] reduces to the simple form:

$$\hat{w}_{j2} = \frac{\sum_{i=1}^{n_j} \hat{\psi}_{i2} e_{1ij}}{\sum_{i=1}^{n_j} \hat{\psi}_{i2}^2} \qquad [11]$$

and equation[6] becomes:

$$\hat{\psi}_{i2} = \frac{\sum_{j=1}^{m_i} \hat{w}_{j2} e_{1ij}}{\sum_{j=1}^{m_j} \hat{w}_{j2}^2} \qquad [12]$$

After each reestimate of $\underline{\hat{\psi}}_2$, it is adjusted so that its mean is equal to zero and its sum of squares is held constant; namely, at the hth iteration:

$$\sum_{i=1}^{n} \hat{\psi}_{i2}^{(h)^2} = \sum_{i=1}^{n} \hat{\psi}_{i2}^{(1)^2}$$

Going back and forth between [11] and [12] will always reduce the sum of squared error. This process is continued until there is no further improvement in the sum of squared error. After convergence is achieved, a second matrix of residuals is computed. That is

$$\mathbf{E}_{02} = \mathbf{X}_0 - \underline{\hat{\psi}}_1 \underline{\hat{w}}'_1 - \mathbf{J}_n \underline{\hat{c}}' - \underline{\hat{\psi}}_2 \underline{\hat{w}}'_2 = \mathbf{E}_{01} - \underline{\hat{\psi}}_2 \underline{\hat{w}}'_2 \qquad [13]$$

To estimate the remaining columns of Ψ and $\mathbf{W}$, equations [10], [11], and [12] are used to obtain the $\underline{\hat{\psi}}_3, \ldots, \underline{\hat{\psi}}_s$, and the $\underline{\hat{w}}_3, \ldots, \underline{\hat{w}}_s$, respectively. $\hat{\Psi}$ computed in this fashion will not be perfectly orthogonal because of the missing data.

Given the full n by s matrix $\hat{\Psi}$, it is bordered by a vector of ones to form the n by s+1 matrix $\hat{\Psi}^*$ which is used in equation [7] to obtain the full m by s matrix $\hat{\mathbf{W}}$ and the m length vector of constants $\hat{\underline{\mathbf{c}}}$. $\hat{\mathbf{W}}$ and $\hat{\underline{\mathbf{c}}}$ are then used in equation [6] to obtain a new estimate of $\hat{\Psi}$. Going back and forth between [6] and [7] will always reduce the sum of squared error. After each pass the columns of $\hat{\Psi}$ are set equal to zero. This process is continued until there is no further improvement in the sum of squared error. Only about five iterations using the entire matrices are usually necessary to achieve final convergence. At convergence, a singular value decomposition of $\hat{\Psi}\hat{\mathbf{W}}$ is performed so that the final estimates are as shown in equation [2].

In summary, the estimation procedure consists of two main ALS phases. In the first, each dimension is estimated one at a time (the columns of Ψ and $\mathbf{W}$ and the elements of $\underline{\mathbf{c}}$). In the second ALS phase, the full Ψ and $\mathbf{W}$ matrices and the vector of constants $\underline{\mathbf{c}}$ are used in equations [6] and [7] until convergence. The estimation procedure is summarized in detail in Table 1. The first ALS phase consists of steps [1] to [13] and the second ALS phase consists of steps [14] to [16].

3. Monte Carlo Tests of the Model

Two sets of Monte Carlo results are reported in this section. First, the ability of the procedure to recover $\Psi\mathbf{W}' + \mathbf{J}_n\underline{\mathbf{c}}'$ from equation [1] will be tested. Second, the ability of the procedure to accurately reproduce the true variance of the error distribution will be tested and how these results relate to the problem of obtaining variances of the estimated parameters will be discussed.[8]

A. Recovery of Ψ, W, and $\underline{\mathbf{c}}$

In order to test the ability of the procedure to recover Ψ, $\mathbf{W}$, and $\underline{\mathbf{c}}$, a number of matrices of varying ranks were created such that $\mathbf{X} = \Psi\mathbf{W}' + \mathbf{J}_n\underline{\mathbf{c}}'$, where $\mathbf{X}$ is an n by m matrix of rank s. To construct $\mathbf{X}$ of rank s, $\mathbf{U}$ and $\mathbf{V}$ were obtained from a singular value decomposition of an n by m matrix of uniform [0,1] random numbers. The s singular values of $\mathbf{X}$ were set so that the latent dimensions—the columns of Ψ—were of approximately equal salience. That is, when the column means are subtracted from $\mathbf{X}$, the s singular values of the resulting matrix, $\Psi\mathbf{W}'$, are approximately equal.[9] The matri-

[8]Additional Monte Carlo work and empirical examples are reported in Poole (1997). Specifically, the ability of the procedure to estimate the Eckart-Young lower rank approximation matrix of an arbitrary matrix of real numbers with missing entries is tested and found to be highly accurate.

[9]Because the sum of the squared singular values of a matrix is equal to the sum of the squared elements of the matrix, if the column means, $\underline{\mathbf{c}}$, are large positive numbers (e.g., a matrix of individual self-placements on seven-point scales), then the first singular value of $\mathbf{X}$ will be quite large compared to the remaining s–1 singular values.

Table 1. Summary of the Estimation Procedure

1) Obtain starting estimates of $\hat{\underline{c}}$, denoted by $\hat{\underline{c}}^{(1)}$, using the column means of $\mathbf{X_0}$. Obtain starting estimates of $\hat{\underline{w}}_1$, denoted by $\hat{\underline{w}}_1^{(1)}$, by finding the vector of plus and minus ones that maximizes the number of positive elements in the covariance matrix $\left[\mathbf{X_0} - \mathbf{J_n}\hat{\underline{c}}'\right]'\left[\mathbf{X_0} - \mathbf{J_n}\hat{\underline{c}}'\right]$ (see Appendix B).

2) Use $\hat{\underline{c}}^{(1)}$ and $\hat{\underline{w}}_1^{(1)}$ in equation [8] to obtain a starting estimate of $\hat{\underline{\psi}}_1$, denoted by $\hat{\underline{\psi}}_1^{(1)}$, and set the mean of $\hat{\underline{\psi}}_1^{(1)}$ equal to zero.

3) Use $\hat{\underline{\psi}}_1^{(1)}$ in equation [7] to obtain a second estimate of $\hat{\underline{c}}$ and $\hat{\underline{w}}_1$—$\hat{\underline{c}}^{(2)}$ and $\hat{\underline{w}}_1^{(2)}$, respectively.

4) Use $\hat{\underline{c}}^{(2)}$ and $\hat{\underline{w}}_1^{(2)}$ in equation [6] to obtain a second estimate of $\hat{\underline{\psi}}_1$, $\hat{\underline{\psi}}_1^{(2)}$. Set the mean of $\hat{\underline{\psi}}_1^{(2)}$ equal to zero and set the sum of squares of $\hat{\underline{\psi}}_1^{(2)}$ equal to the sum of squares of $\hat{\underline{\psi}}_1^{(1)}$; that is $\sum_{i=1}^{n} \hat{\psi}_{i1}^{(2)^2} = \sum_{i=1}^{n} \hat{\psi}_{i1}^{(1)^2}$.

5) Repeat steps (3) and (4) until convergence.

6) Compute $\mathbf{E_{01}} = \mathbf{X_0} - \hat{\underline{\psi}}_1\hat{\underline{w}}'_1 - \mathbf{J_n}\hat{\underline{c}}'$.

7) Obtain starting estimates of $\hat{\underline{w}}_2$, $\hat{\underline{w}}_2^{(1)}$, by finding the vector of plus and minus ones that maximizes the number of positive elements in the covariance matrix $\mathbf{E}'_{01}\mathbf{E}_{01}$.

8) Use $\hat{\underline{w}}_2^{(1)}$ in equation [10] to obtain starting estimates of $\hat{\underline{\psi}}_2$, $\hat{\underline{\psi}}_2^{(1)}$.

9) Use $\hat{\underline{\psi}}_2^{(1)}$ in equation [11] to obtain $\hat{\underline{w}}_2^{(2)}$.

10) Use $\hat{\underline{w}}_2^{(2)}$ in equation [12] to obtain $\hat{\underline{\psi}}_2^{(2)}$. Set the mean of $\hat{\underline{\psi}}_2^{(2)}$ equal to zero and set the sum of squares of $\hat{\underline{\psi}}_2^{2}$ equal to the sum of squares of $\hat{\underline{\psi}}_2^{(1)}$ as in step (4) above.

11) Repeat steps (9) and (10) until convergence.

12) Compute $\mathbf{E_{02}} = \mathbf{X_0} - \hat{\underline{\psi}}_1\hat{\underline{w}}'_1 - \mathbf{J_n}\hat{\underline{c}}' - \hat{\underline{\psi}}_2\hat{\underline{w}}'_2 = \mathbf{E_{01}} - \hat{\underline{\psi}}_2\hat{\underline{w}}'_2$.

13) Repeat steps (7)–(12) to estimate remaining dimensions; that is: $\hat{\underline{w}}_3$ and $\hat{\underline{\psi}}_3$, $\hat{\underline{w}}_4$ and $\hat{\underline{\psi}}_4$, . . . , and $\hat{\underline{w}}_s$ and $\hat{\underline{\psi}}_s$.

14) Use the full n by s matrix $\hat{\underline{\Psi}}$ in equation [7] to obtain the full m by s matrix $\hat{\underline{\mathbf{W}}}$ and the m length vector of constants $\hat{\underline{c}}$.

15) Use $\hat{\underline{\mathbf{W}}}$ and $\hat{\underline{c}}$ in equation [6] to obtain a new estimate of $\hat{\underline{\Psi}}$.

16) Repeat steps (14) and (15) until convergence.

ces were then converted to the form shown in equation [1] by adding random error to the elements of the matrix and then randomly removing some of the elements; that is, $\mathbf{X_0} = [\,\mathbf{X} + \mathbf{E}\,]_0$.

Random error was generated in three ways. First, by sampling from a normal distribution with mean 0 and constant variance. Second, by sampling from a normal distribution with mean 0 and variable variance (heteroskedasticity).[10] And third, by sampling from a uniform [–.5, +.5] distribution. The level of error was controlled by adjusting the standard deviation of the distribution.

To create missing data a number was drawn from the uniform [0,1] distribution for each entry in $\mathbf{X} + \mathbf{E}$ and if the number exceeded a preset threshold, the entry was treated as missing. For example, to create a level of 50% missing data, if the number drawn was greater than .5, the corresponding entry in $\mathbf{X} + \mathbf{E}$ was removed and treated as missing. Every row was required to contain at least s+2 entries. If the number of missing entries resulted in fewer than s+2 entries, a new row was created and the process repeated. This ensured that all the columns in the target matrix, $\mathbf{X_0} = [\ \mathbf{X} + \mathbf{E}\ ]_0$, had approximately the same number of missing entries. This procedure was repeated ten times for each true matrix. The results are shown in Table 2.

The first three columns of Table 2 show the number of rows, the number of columns, and the rank of the true matrix $\mathbf{X}$. The fourth column, r-square with target, shows the average squared Pearson correlation between the elements in $\mathbf{X_0}$ and the reproduced elements formed by $[\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\underline{\hat{\mathbf{c}}}']_0$. The fifth column, r-square with full, displays the squared Pearson correlation between all the nm elements in $\mathbf{X} + \mathbf{E}$ and $\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\underline{\hat{\mathbf{c}}}'$. The sixth column, r-square with true, shows the squared Pearson correlation between all the nm elements in the true $\mathbf{X}$ matrix and $\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\underline{\hat{\mathbf{c}}}'$. The seventh column, r-square with E-Young, shows the squared Pearson correlation between all the nm elements in the Eckart-Young approximation of rank s using Theorem II and $\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\underline{\hat{\mathbf{c}}}'$. That is, Theorem II was applied to the full matrix with error, $\mathbf{X} + \mathbf{E}$, to produce the rank s approximation matrix.

The eighth column shows the average squared Pearson correlation between the s true basic dimensions and the s estimated basic dimensions; that is, the average of the s r-squares computed between each column of the true Ψ matrix, and its corresponding column in $\hat{\Psi}$.[11]

The ninth and tenth columns are concerned with just the missing data. The ninth column shows the squared Pearson correlation between the missing elements in the true $\mathbf{X}$ matrix and their corresponding estimates in $\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\underline{\hat{\mathbf{c}}}'$. Finally, for purposes of comparison, the tenth column shows

[10]The error was generated by a normal with mean zero and variance $\mathbf{k}_{ij}^2\sigma^2$ where $\mathbf{k}_{ij}$ was randomly drawn from a uniform [.5, 1.5] for each nonmissing entry of the matrix.

[11]Because $\hat{\Psi}$ is defined only up to an arbitrary rotation, this was removed before the Pearson r-squares were computed. This is known as the "orthogonal procrustes" problem, and it was solved by Schonemann (1966). His solution was used to remove the arbitrary rotation.

Table 2. Monte Carlo Tests of the Equation [1] Model
(Average of 10 Trials, Standard Deviations in Parentheses)

N	M	S	R^2 With Target	R^2 With Full	R^2 With True	R^2 With E-Young	R^2 With True Ψ	R^2 With True Missing	R^2 With True Miss (Reg)	Error Level	Percent Missing
1000	25	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.00	50
			(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)		
1000	25	2	.951	.939	.988	.993	.988	.986	.889	.25	50
			(.001)	(.000)	(.000)	(.000)	(.001)	(.000)	(.012)		
			.950	.940	.988	.993	.988	.987	.895		
			(.000)	(.000)	(.001)	(.000)	(.001)	(.001)	(.012)		
			.951	.940	.988	.993	.988	.987	.890		
			(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.016)		
1000	25	2	.833	.795	.954	.973	.953	.947	.759	.50	50
			(.002)	(.002)	(.002)	(.002)	(.003)	(.003)	(.016)		
			.832	.794	.953	.974	.953	.947	.765		
			(.003)	(.002)	(.002)	(.001)	(.004)	(.003)	(.036)		
			.831	.797	.956	.975	.956	.951	.784		
			(.002)	(.001)	(.001)	(.001)	(.002)	(.002)	(.038)		
1000	25	2	.698	.632	.903	.944	.901	.890	.696	.75	50
			(.003)	(.002)	(.003)	(.003)	(.006)	(.005)	(.022)		
			.698	.631	.900	.942	.899	.886	.687		
			(.002)	(.003)	(.004)	(.003)	(.008)	(.007)	(.008)		
			.697	.632	.903	.945	.903	.891	.700		
			(.002)	(.003)	(.003)	(.002)	(.005)	(.004)	(.010)		

(continued on next page)

Table 2. Monte Carlo Tests of the Equation [1] Model *(continued)*

N	M	S	R^2 With Target	R^2 With Full	R^2 With True	R^2 With E-Young	R^2 With True Ψ	R^2 With True Missing	R^2 With True Miss (Reg)	Error Level	Percent Missing
1000	25	2	.583	.494	.841	.908	.839	.822	.570	1.00	50
			(.005)	(.003)	(.003)	(.003)	(.007)	(.005)	(.037)		
			.584	.492	.839	.906	.840	.819	.564		
			(.005)	(.005)	(.005)	(.004)	(.007)	(.006)	(.022)		
			.580	.493	.841	.910	.840	.823	.570		
			(.004)	(.003)	(.004)	(.003)	(.007)	(.006)	(.027)		
1000	15	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.00	70
			(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)		
1000	15	1	.842	.782	.946	.962	.946	.944	.811	.50	70
			(.003)	(.002)	(.002)	(.002)	(.002)	(.002)	(.020)		
			.842	.783	.947	.963	.946	.945	.803		
			(.004)	(.002)	(.002)	(.002)	(.002)	(.002)	(.018)		
			.841	.782	.946	.962	.946	.943	.802		
			(.002)	(.002)	(.002)	(.001)	(.001)	(.002)	(.012)		
1000	15	1	.602	.462	.811	.868	.812	.804	.452	1.00	70
			(.008)	(.005)	(.008)	(.006)	(.008)	(.008)	(.021)		
			.607	.465	.816	.869	.815	.809	.453		
			(.009)	(.006)	(.010)	(.006)	(.010)	(.011)	(.015)		
			.601	.465	.816	.871	.817	.808	.454		
			(.007)	(.004)	(.006)	(.006)	(.006)	(.005)	(.026)		

(continued on next page)

Table 2. Monte Carlo Tests of the Equation [1] Model *(continued)*

N	M	S	R^2 With Target	R^2 With Full	R^2 With True	R^2 With E-Young	R^2 With True Ψ	R^2 With True Missing	R^2 With True Miss (Reg)	Error Level	Percent Missing
100	25	3	.860 (.004)	.780 (.007)	.905 (.009)	.940 (.007)	.918 (.018)	.882 (.014)	.456 (.047)	.50	50
250	25	3	.852 (.006)	.788 (.005)	.923 (.004)	.954 (.004)	.928 (.009)	.908 (.006)	.621 (.038)	.50	50
500	25	3	.850 (.004)	.785 (.005)	.922 (.006)	.952 (.005)	.924 (.008)	.905 (.011)	.658 (.028)	.50	50
1000	25	3	.849 (.002)	.786 (.002)	.924 (.002)	.952 (.002)	.923 (.004)	.906 (.004)	.724 (.033)	.50	50
500	10	2	.873 (.003)	.772 (.006)	.871 (.008)	.918 (.007)	.834 (.016)	.824 (.014)	.777 (.022)	.50	50
500	15	2	.853 (.005)	.784 (.005)	.915 (.005)	.947 (.006)	.905 (.013)	.895 (.009)	.762 (.034)	.50	50
500	20	2	.843 (.003)	.791 (.003)	.938 (.003)	.963 (.003)	.935 (.009)	.927 (.005)	.730 (.028)	.50	50
500	25	2	.835 (.004)	.795 (.001)	.954 (.002)	.973 (.001)	.954 (.003)	.947 (.002)	.695 (.029)	.50	50

the squared Pearson correlation between the missing elements in the true **X** matrix and those estimated by the regression method.[12]

The eleventh column shows the level of error introduced. This was computed as

$$\text{Error Level} = \left[\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(e_{ij} - \bar{e}\right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_{ij} - \bar{x}\right)^2} \right]^{\frac{1}{2}} \qquad [14]$$

where e_{ij} is the error added to x_{ij}, $\bar{e}$ is the mean of the error, and $\bar{x}$ is the mean of the x_{ij}. The error level is the ratio of the standard deviation of the error to the standard deviation of the x_{ij}.

Finally, the twelfth column shows the percentage of missing elements in $\mathbf{X_0}$.

Each entry in the fourth through tenth columns in a row of the table represents the average of ten runs using the same true **X** matrix each time but with different error matrices and different patterns of random removal of elements. Below each entry is the standard deviation of the ten runs. In the eighth column, only the largest of the s standard deviations of the dimension by dimension r-squares is shown.

Table 2 is divided into four sections. In the first and second sections of the table the size of the matrix and the level of missing data are held fixed while the level of error is increased. When error is present, results for all three error distributions—normal with constant variance, normal with variable variance, and uniform—are shown. For example, when the true matrix was 1000 by 25 and rank 2, 50% missing, with an error level of .50, then the r-square between the entries of $\mathbf{X_0}$ and the reproduced elements from $[\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\hat{\underline{c}}']_0$ (r-square with target) is .833 for the normal error constant variance case, .832 for the heteroskedastic normal error, and .831 for the uniform error. The r-squares between the true missing entries and the reproduced missing entries from $[\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J_n}\hat{\underline{c}}']_0$ (r-square with true missing) are .947, .947, and .951, respectively.

[12]In the regression method, for each row of the data matrix, the submatrix of the m by m Pearson correlation matrix computed between the columns of the nonmissing entries is inverted and multiplied by the submatrix of the correlation matrix corresponding to the columns of the missing and nonmissing entries. This produces coefficients that are applied to the nonmissing entries to obtain the estimates of the missing entries. These missing entries were then used to get a better estimate of the correlation matrix. Experimentally, I found that iterating this process from 3 to 5 times was optimal.

Three aspects of sections one and two stand out. First, the accuracy of the procedure appears not to be very sensitive to the type of error. The procedure recovers the true data—both missing and nonmissing—equally well for all three types of error. Second, the procedure appears to be very stable. The standard deviations are small. Neither result is surprising. In standard OLS the estimates of the coefficients are unbiased in the presence of heteroskedastic error. Indeed, as long as the error process is symmetric, the principle of least squares will ensure that the true data are recovered with reasonable accuracy. Finally, the regression method of estimating missing entries deteriorates badly as the level of error increases.

In the third and fourth sections of the table only normal error with constant variance is used and the level of error and the percentage of missing data are both held constant at .50 and 50%, respectively. In section three n is increased in stages with m and s held fixed, and in section four m is increased in stages while n and s are held fixed. These two sections show that, holding everything else fixed, increasing the size of the matrix results in more accurate estimates of the true data—both observed and missing. Once again, the ability of the regression method to recover the missing entries is poor.

Although Table 2 is by no means exhaustive, it is apparent that the procedure outlined in Section 2 is stable and will reliably reproduce the basic dimensions and the overall matrix even at high levels of error and missing data.

B. Estimating Standard Errors

The purpose of the experiments in this subsection is to assess how good a job the scaling procedure does in estimating the variance of the error distribution and the variance of the estimated parameters. The former is relatively easy, the latter is not.

The standard error of the estimate is:

$$\hat{\sigma} = \left[\frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \left[\sum_{k=1}^{s} \hat{\psi}_{ik} \hat{w}_{jk} \right] + \hat{c}_j - x_{ij} \right\}^2}{q - s(m+n) - m} \right]^{\frac{1}{2}} = \left[\frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} e_{ij}^2}{q - s(m+n) - n} \right]^{\frac{1}{2}} \quad [15]$$

where q is the number of observed (nonmissing) elements in $\mathbf{X_0}$ and s(m+n)+m is the number of parameters estimated by the scaling procedure. Table 3 shows some Monte Carlo tests using the normal distribution with mean zero and constant variance and the uniform distribution with mean

Table 3. Monte Carlo Tests of the Standard Error of the Estimate (Average of 10 Trials, Standard Deviations in Parentheses)

N	M	S	Error Level	True σ	Normal: Standard Error of Estimate	Uniform: Standard Error of Estimate	Percent Missing
1000	25	2	.50	.1484	.1484 (.0008)	.1485 (.0007)	50
1000	25	2	.75	.2226	.2225 (.0012)	.2227 (.0011)	50
1000	25	2	1.00	.2856	.2823 (.0013)	.2873 (.0015)	50
1000	25	2	.50	.1436	.1437 (.0013)	.1438 (.0010)	70
1000	25	2	.75	.2226	.2223 (.0014)	.2225 (.0014)	70
1000	25	2	1.00	.2968	.2962 (.0014)	.2965 (.0018)	70
250	25	3	.50	.3227	.3220 (.0040)	.3242 (.0033)	50
250	25	3	.75	.4841	.4824 (.0061)	.4857 (.0050)	50
250	25	3	1.00	.6455	.6422 (.0080)	.6466 (.0065)	50

zero and constant variance. Recall that the value of the true σ is set as a fraction of the standard deviation of the x_{ij}. This fraction is shown under the "Error Level" heading in the Table.

The entries in Table 3 were computed in the same fashion as those for Table 2. The entries are the average and standard deviation of ten runs with the same true **X** matrix and true error variance, σ^2, but different draws from the error distribution and different patterns of missing data. Within each row the same true **X** matrix was used so that the results for the two error distributions could be compared.

The scaling procedure does a good job in recovering the standard error of the estimate. All the $\hat{\sigma}$'s in the table are very close to the true σ, and the standard deviations are small. Consistent with the results of Table 2, the standard deviations increase with the overall error level and the level of missing data.

With respect to the variance of the estimated parameters, that is, the $\hat{\psi}$'s, $\hat{\mathbf{w}}$'s, and $\hat{\mathbf{c}}$'s, recall that in ordinary least squares, the variance of the estimated regression coefficients is:

$$\mathbf{Var}\left(\hat{\beta}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

If the true Ψ were known and it was used in equation [7] to obtain estimates of $\mathbf{W}$ and $\underline{\mathbf{c}}$, then the variances of the $\hat{\mathbf{w}}$'s and the $\hat{\mathbf{c}}$'s would be:

$$\mathbf{Var}\left(\hat{\mathbf{w}}_{jk}\right) = \frac{\sigma^2}{\lambda_k}, \; j = 1, \ldots, m, \; k = 1, \ldots, s,$$

$$\text{and } \mathbf{Var}\left(\hat{\mathbf{c}}_j\right) = \frac{\sigma^2}{\mathbf{n}}, \; j = 1, \ldots, m \qquad [16]$$

where λ_k is the kth singular value of the true $\Psi\mathbf{W}'$. To see this, recall that $\Psi'\Psi = \Lambda$ and let $\Psi^* = [\Psi|\mathbf{J}_n]$. Note that $\Psi^{*\prime}\Psi^*$ will be a s+1 by s+1 diagonal matrix with Λ in the upper left s by s submatrix and the s+1st diagonal entry will be n.

Conversely, if the true $\mathbf{W}$ and $\underline{\mathbf{c}}$ were known, the variances of the estimated $\hat{\psi}$'s would also be σ^2/λ_k by using equation [6] and the fact that $\mathbf{W}'\mathbf{W} = \Lambda$.

Table 3 shows that $\hat{\sigma}$ is a good estimate of the true σ and if the amount of missing data is low, then the estimated singular values will be quite close to the true singular values. Consequently it is tempting to use $\frac{\hat{\sigma}^2}{\hat{\lambda}_k}$ to calculate standard errors. Because Ψ, $\mathbf{W}$, and $\underline{\mathbf{c}}$ are all being estimated, however, these variance formulas can only be used as a *lower bound.* In the empirical examples below, I compute the standard errors using a simple bootstrap analysis. In most practical applications, the rows of the data matrix will correspond to individuals and the columns to their responses. Consequently the individuals can be sampled with replacement to perform a bootstrap analysis. That is, the rows of the actual $\mathbf{X}_0$ are sampled with replacement to form a pseudo $\mathbf{X}_0$ matrix. This pseudo matrix is then analyzed by the procedure to obtain an estimate of $\hat{\mathbf{W}}$ and $\hat{\underline{\mathbf{c}}}$. This process is repeated 100 times, and the standard errors are obtained by computing the sum of squared differences between the actual $\hat{\mathbf{W}}$ and $\hat{\underline{\mathbf{c}}}$ and the 100 $\hat{\mathbf{W}}$'s and $\hat{\underline{\mathbf{c}}}$'s from the bootstrap trials, dividing by 100, and taking the square root. The bootstrap variances will be, on average, larger than $\frac{\hat{\sigma}^2}{\hat{\lambda}_k}$.

4. Relationship With Other Scaling Procedures

The fundamental difference between the scaling procedure developed in Section 2 and other commonly used scaling methods is its generality. The procedure estimates a lower rank approximation to any matrix of real numbers with missing data.[13] The only assumption made about the true data matrix, $\mathbf{X}$, is that it is of rank s. By Theorem I, any matrix of real numbers can be written as a linear product. Hence, the equation $\mathbf{X} = \mathbf{U}\Lambda\mathbf{V}' = \Psi\mathbf{W}' + \mathbf{J}_n\underline{\mathbf{c}}'$ is not an assumption but a *property* of any real matrix. What is important, however, is that the decomposition in the form of equation [1] serves as a useful *model* of substantive phenomenon of interest to political scientists. Two examples of this are shown in Section 5.

Most existing scaling techniques that analyze rectangular data matrices are designed for distance data. For example, a very common type of data matrix of interest to political scientists is preferential choice data which can be treated as Euclidean distances between individuals/respondents and stimuli. Geometrically, a matrix of squared distances computed between points in an s-dimensional space will have rank s+2. If there were no error but the matrix had missing entries, then the procedure developed in Section 2 will exactly reproduce the squared distances by estimating $\Psi\mathbf{W}'$ of rank s+1.[14] Since the elements of $\Psi\mathbf{W}' + \mathbf{J}_n\underline{\mathbf{c}}'$ are squared distances they must be analyzed as such and the decomposition of this matrix, $\mathbf{U}\Lambda_{s+2}\mathbf{V}'$, cannot be directly used to solve for the points that produce the squared distances.[15] In addition, in the realistic case where error is present, the scaling procedure developed here is not an appropriate model of noisy distance data.

Examples of distance data are interest group ratings of members of Congress, thermometer ratings of politicians by survey respondents, and congressional roll call votes.[16] Interest group ratings and thermometer scores have been analyzed by a number of scaling techniques. The usual

[13] This aspect of the procedure is discussed in detail in Poole (1997). The Monte Carlo analysis shows that $\hat{\Psi}\hat{\mathbf{W}}' + \mathbf{J}_n\hat{\underline{\mathbf{c}}}'$ will be a good fit to the Eckart-Young approximation matrix $\mathbf{U}\Lambda_s\mathbf{V}'$ of any rectangular matrix of real numbers with missing entries.

[14] To see this, let $\mathbf{A}$ be an n by s matrix of coordinates and $\mathbf{B}$ be a m by s matrix of coordinates. Let $\mathbf{diag(AA')}$ and $\mathbf{diag(BB')}$ be the n length and m length vector of diagonal terms of $\mathbf{AA}'$ and $\mathbf{BB}'$, respectively. The matrix of squared distances can be written as: $\mathbf{diag(AA')J}_m' - \mathbf{2AB}' + \mathbf{J}_n\mathbf{diag(BB')}'$. This is the product of the two matrices $\mathbf{[diag(AA') \mid -2A \mid J}_n\mathbf{]}$ and $\mathbf{[J}_m \mid \mathbf{B} \mid \mathbf{diag(BB')]}$. Note that $\mathbf{diag(AA')J}_m' - \mathbf{2AB}'$ is equivalent to $\Psi\mathbf{W}'$ with rank s+1 and $\mathbf{J}_n\mathbf{diag(BB')}'$ is equivalent to $\mathbf{J}_n\underline{\mathbf{c}}'$.

[15] The points producing the squared distances can be found using Schonemann's (1970) solution to the metric unfolding problem.

[16] Seven-point scale data can also be interpreted as distance data. For an interesting analysis along these lines, see Jacoby (1993).

approach is to treat the ratings/thermometer scores as inverse distances and then estimate a set of points corresponding to the politicians/respondents and a set of points corresponding to the interest groups/political figures such that the reproduced distances match the inverse ratings/scores as closely as possible. This is known as an *unfolding* problem (Coombs 1964). Unfolding a distance matrix is a form of decomposition but the data are not being decomposed into a linear product like the model developed here. Rabinowitz (1974), Cahoon, Hinich, and Ordeshook (1978), and Brady (1990) have developed statistical procedures to solve this problem. Rabinowitz uses his line-of-sight method to produce a matrix of dissimilarities between pairs of stimuli from the preference data and then uses a nonmetric scaling to obtain estimates of the two sets of points. Cahoon, Hinich, and Ordeshook first transform the data by eliminating the squared terms in the distance equation and then produce a covariance matrix of the cross-product terms. The covariance matrix is then decomposed and the decomposition is used to obtain estimates of the two sets of points. Brady develops a generalized least squares estimation method and applies it to the quadratic utility model developed by Hinich (Enelow and Hinich 1984) to recover the candidate configuration. In Brady's method only the parameters of the distribution of the respondent ideal points are estimated.[17]

The Poole and Rosenthal (1997) NOMINATE procedure is an unfolding method for roll call data. Like the Rabinowitz and Cahoon, Hinich, Ordeshook scaling methods it produces a set of points corresponding to the members of Congress. It differs from the other scaling procedures in that it produces *two* points for each roll call—one for Yea and one for Nay.[18]

The scaling procedure developed here is probably most similar to factor analytic work by Aldrich and McKelvey (1977), Enelow and Hinich (1994), and Hinich and Munger (1994) on seven-point scale data. The key difference is that the factor analytic work, by definition, analyzes covariance matrices formed from the rectangular data matrix. The Aldrich-McKelvey scaling method is a one-dimensional version of the model expressed in [1B] only the model is applied to a *transposed* matrix where

[17]Earlier work by Weisberg and Rusk (1970) and Rusk and Weisberg (1972) computed correlation matrices from the rectangular matrix of thermometer scores and then analyzed the correlation matrix with nonmetric multidimensional scaling to obtain a configuration of political candidates. Given the candidate configuration, respondent locations can be estimated using OLS. Locating the individuals via OLS with reference to a given stimulus configuration is known as an external unfolding analysis (Carroll 1972).

[18]Earlier work by MacRae (1970) on roll call votes applied factor analysis to correlation (Yule's Q and other measures) matrices to recover configurations of legislators and/or roll calls.

m > n. The scaling procedure developed here can be used to estimate more than one latent dimension underlying an issue scale. An example is shown in Poole (1997). [19]

The two applications of the procedure shown in the next Section both assume that the true data matrix, **X**, has the form shown in equation [1A].

5. Empirical Applications

A. Recovering a Basic Space from the 1980 Issue Scales

The first application of the procedure is to fourteen issue scales from the 1980 NES cross-sectional survey. These scales are listed in Table 4. The equal rights amendment and the two abortion questions were four point scales, the tax cut question was a five point scale, and the remaining issue questions were all seven point scales. Table 4A shows the fits for one, two, and three dimensions, and the three singular values of $\hat{\Psi}\hat{\mathbf{W}}'$. Table 4B shows the estimated $\hat{\mathbf{c}}$'s and $\hat{\mathbf{w}}$'s and the fits by issue for one, two, and three dimensions. Bootstrapped standard errors are shown in parentheses below the coefficient estimates. The standard errors were computed as described in the previous section.[20]

Somewhat over half the variance (r-square of .512) of the individuals' issue scale positions is explained by a single dimension. This primary dimension is clearly the classic liberal/conservative continuum familiar to students of American politics. The pre and post liberal/conservative seven point scales fit highly with this dimension as well as the government services, government help for minorities, and government jobs seven point scales. These latter three scales are at the heart of what is meant by liberal/conservative in American politics—government intervention in the economy.

The second dimension picks up the two abortion questions, inflation, and the women's equal role seven point scale. Adding the second dimension raises the r-square for the women's equal role scale to .77. The r-squares for the two abortion questions increase to about .30. Inflation, the odd issue in

[19] The method developed by Groseclose, Levitt, and Snyder (1997) to analyze interest group ratings is closely related to the Aldrich-McKelvey model. Aldrich and McKelvey assume that the stimuli hold fixed positions on an underlying true issue dimension and that a respondent's perception of the stimuli issue scale positions are simple linear mappings from the true stimuli positions. In the Groseclose, Levitt, and Snyder model, the legislators are assumed to hold fixed positions on a single dimension over time. They estimate a linear mapping for each time period for the interest group so that the interest group's scale varies over time. They assume that the interest group is external to the legislator configuration (cf., Poole and Rosenthal 1997, chap. 8). These linear mappings are used to produce the inflation-adjusted ADA scores.

[20] Because $\hat{\mathbf{W}}$ is defined only up to an arbitrary rotation, this was removed before the standard error computation. That is, each $\hat{\mathbf{W}}$ from the bootstrap was rotated to best fit the actual $\hat{\mathbf{W}}$ before the squared difference was computed. See note 11 above.

this bunch, rises to about .40. The third dimension picks up the relations with Russia scale, and more weakly, the defense spending scale. With the third dimension, the r-square for the Russia scale rises to .70. The five-point tax cut scale doesn't fit well with any of the three dimensions. The problem with it is that everyone appears to favor tax cuts.

If conventional t-tests were applied using the estimated standard errors, then all the $\hat{\mathbf{w}}$'s for the first dimension, seven of the fourteen $\hat{\mathbf{w}}$'s for the second dimension, and only two of the fourteen $\hat{\mathbf{w}}$'s for the third dimension are statistically significant. If Ψ were known with certainty, then the standard errors for the $\hat{\mathbf{w}}_1$'s would be $\mathbf{1.356}\sqrt{\mathbf{111.567}} = \mathbf{.128}$, which for reasons explained above, is, on average, smaller than the bootstrapped standard errors. The standards errors for the $\hat{\mathbf{w}}_2$'s and $\hat{\mathbf{w}}_3$'s under the certainty assumption would be .144 and .145, respectively. Both are much smaller than the bootstrapped standard errors.

Figure 1 shows the distribution across the first basic dimension of the Carter, Reagan, and Anderson voters for the 1980 election. The Carter and Reagan voters are differentiated ideologically with the center splitting between them. Anderson was clearly drawing his support from Carter's "territory." On a –1 to +1 scale, the mean locations for the Reagan, Carter, and Anderson voters were .165, –.165, and –.113, respectively. The corresponding standard deviations were .229, .288, and .248. The Reagan voters were clearly more ideologically coherent than the Carter voters (the smaller standard deviation). This coherence is also reflected in the fact that a larger percentage of Reagan's voters answered all fourteen issues questions than did Carter's voters—32.1% versus 26.1%, respectively. However, fully 41.3% of Anderson's voters were informed. Given Anderson's failure to "break through" to widespread mass awareness of his candidacy and what it stood for, it is not surprising that those who *did* support him were more informed than other voters. Third party candidates in the United States typically attract more intense supporters as a percentage of total support than the two major parties (for a history and discussion of this point, see Sundquist 1983).

B. Fitting Together Coordinate Configurations

The scaling procedure for estimating equation [1] is also a solution for what is known in psychometrics as an "orthogonal procrustes" problem (Schonemann 1966; Schonemann and Carroll 1970). Suppose the columns of $\mathbf{X_0}$ are the scaled coordinates of individuals on the same issue scale or attribute dimension at n different times. With s inherently equal to one, $\hat{\psi}$ will be the n by 1 vector of best fitting *average* coordinates for the individuals over the time span. This is so because if the elements of the coordinate matrix $\mathbf{X_0}$ are replaced by $(x_{ij} - \hat{c}_j) / \hat{w}_j$, this minimizes the sum of squares

Table 4A. Overall Fit Statistics for 14 1980 Issue Scales

S	N	% Missing	R^2	Standard Error of Estimate	Singular Values of $\hat{\psi}\hat{W}'$
1	1270	14.1	.512	1.356	111.567
2	1270	14.1	.611	1.272	78.536
3	1270	14.1	.680	1.217	70.679

Table 4B. Fit Statistics by Issue (Bootstrapped Standard Errors in Parentheses)

Issue	n_j	$\hat{c}_j$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	R^2 1	R^2 2	R^2 3
Liberal /Conservative	875	4.29	–2.98	0.72	–1. 10	.414	.424	.448
		(0.05)	(0.18)	(0.49)	(0.68)			
Defense Spending	1163	5.20	–1.81	1.52	2.46	.123	.147	.315
		(0.04)	(0.45)	(1.67)	(3.20)			
Government Services	1119	4.41	4.29	3.67	2.87	.451	.617	.778
		(0.07)	(0.30)	(1.17)	(2.54)			
Inflation	816	4.17	2.00	3.25	1.06	.159	.393	.451
		(0.05)	(0.31)	(0.91)	(1.62)			
Abortion	1238	2.86	0.50	–2.02	1.13	.030	.290	.312
		(0.03)	(0.15)	(0.36)	(0.65)			

(continued on next page)

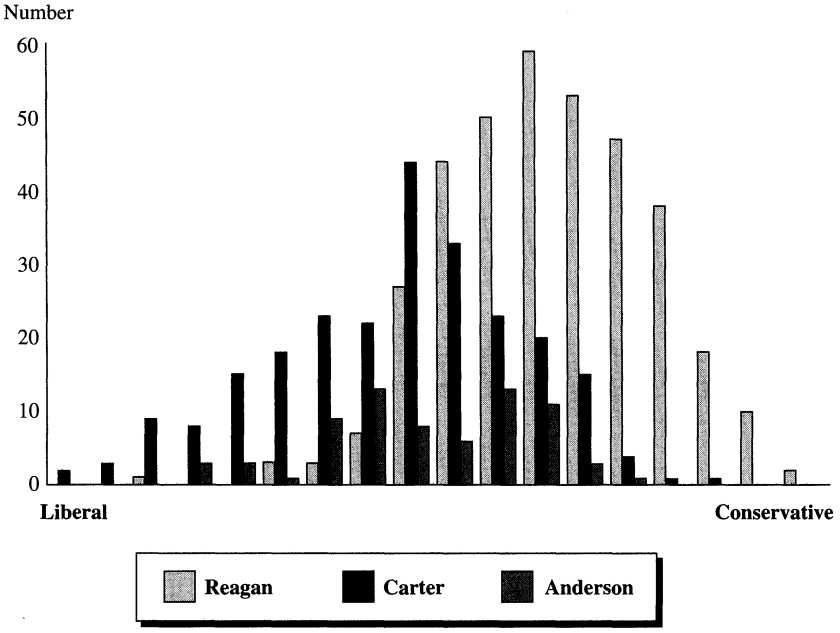
Table 4B. Fit Statistics by Issue *(continued)*

Issue	n_j	$\hat{c}_j$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	R^2 1	R^2 2	R^2 3
Tax Cuts	836	2.81 (0.05)	–1.05 (0.23)	–0.85 (0.61)	–0.76 (1.03)	.055	.071	.091
Liberal/Conservative[a]	949	4.37 (0.04)	–2.79 (0.13)	0.26 (0.38)	–0.53 (0.54)	.414	.423	.437
Govt Help Minorities[a]	1160	4.53 (0.05)	–3.46 (0.28)	0.10 (0.84)	0.86 (1.43)	.412	.424	.440
Relations w/Russia[a]	1152	3.83 (0.10)	–3.26 (0.44)	1.66 (1.14)	5.83 (2.92)	.231	.248	.696
Women's Equal Role[a]	1223	2.89 (0.05)	–2.37 (0.19)	5.61 (0.58)	–2.90 (1.57)	.204	.771	.805
Government Jobs[a]	1131	4.34 (0.05)	–4.64 (0.27)	–2.21 (0.90)	1.31 (1.27)	.518	.635	.649
Equal Rights Amendment[a]	1144	2.68 (0.05)	–3.16 (0.12)	1.87 (0.41)	–2.33 (0.96)	.381	.491	.561
Busing[a]	1219	6.04 (0.05)	–2.82 (0.54)	0.28 (1.44)	1.15 (2.68)	.256	.263	.306
Abortion[a,b]	1246	2.68 (0.03)	0.59 (0.13)	–1.99 (0.26)	0.88 (0.48)	.047	.318	.329

[a]Post-Election question.

[b]Different wording than pre-election question.

Figure 1. 1980 Voters on First Basic Dimension



between the common elements in each pair of configurations. In effect, the m configurations are "squeezed together" as tightly as possible when they are transformed by $(x_{ij} - \hat{c}_j) / \hat{w}_j$, and the mean configuration around which they are "squeezed" or "targeted" is $\hat{\bar{\psi}}$.

In order to test the model's performance as an orthogonal procrustes procedure, the combined set of W-NOMINATE (Poole and Rosenthal 1997) scores for members of the House and Senate for Congresses 75 to 104 (1937 through December, 1995) will be fitted together. The scores range from –1.0 to +1.0 and are based upon all the nonunanimous roll call votes taken during a Congress. The first dimension score measures the degree of liberalism/conservatism of the legislator.

It is possible to analyze the House and Senate W-NOMINATE scores together because 151 individuals served in both chambers for at least five Congresses during the 1937 to 1995 period. If legislators are assumed to maintain a fixed position on the underlying liberal/conservative basic dimension when they move from the House to the Senate or vice versa, then the model shown in equation [1B] and the procedure shown in Table 1 can be modified so that a joint House and Senate configuration can be estimated.

In this application, $\mathbf{X_0}$ has thirty columns—one for each Congress—and a row for each person serving in at least five Congresses from 1937 to 1995. A total of 1,431 legislators served in at least five Congresses during this period—1,106 in the House only, 174 in the Senate only, and 151 served in both chambers.

To see how this is accomplished, let n_h be the number of legislators who served in the House only, let n_u be the number of legislators who served only in the Senate, n_b be the number of legislators who served in both chambers, and let $\underline{\Psi}_h$, $\underline{\Psi}_u$, and $\underline{\Psi}_b$ be the corresponding vectors of underlying liberal/conservative basic coordinates of length n_h, n_u, and n_b, respectively. This produces the equations:

$$\mathbf{X_{0h}} = \begin{bmatrix} \underline{\Psi}_h \\ \underline{\Psi}_b \end{bmatrix} \mathbf{W}_h' + \mathbf{J}_{h+b}\underline{c}'_h \qquad [17A]$$

$$\mathbf{X_{0u}} = \begin{bmatrix} \underline{\Psi}_u \\ \underline{\Psi}_b \end{bmatrix} \mathbf{W}_u' + \mathbf{J}_{u+b}\underline{c}'_u \qquad [17B]$$

where $\mathbf{X_{0h}}$ and $\mathbf{X_{0u}}$ are the $n_h + n_b$ by m and $n_u + n_b$ by m matrices of W-NOMINATE scores for the House and Senate, respectively.

The procedure to estimate equations [17A] and [17B] is the same as that shown in Table 1, only now some bookkeeping is necessary when the legislator coordinates are estimated. Given $\hat{\mathbf{W}}_h$, $\underline{\hat{c}}_h$, $\hat{\mathbf{W}}_u$, and $\underline{\hat{c}}_u$, the coordinate for a legislator who only served in the House can be estimated by using $\hat{\mathbf{W}}_h$, and $\underline{\hat{c}}_h$, in equation [6]; similarly, if a legislator served only in the Senate, then $\hat{W}_u$, and $\underline{\hat{c}}_u$ are used in equation [6]. If a legislator served in both chambers then a $\mathbf{W}$ matrix can be formed from $\hat{\mathbf{W}}_h$,and $\hat{\mathbf{W}}_u$, by using the rows corresponding to the chamber served in and a $\underline{c}$ vector can be constructed from $\underline{\hat{c}}_h$,and $\underline{\hat{c}}_u$ by using the entries corresponding to the chamber served in. Given estimates of the legislator coordinates, $\underline{\hat{\psi}}_h$, $\underline{\hat{\psi}}_u$, and $\underline{\hat{\psi}}_b$ respectively, then it is a simple matter to estimate $\mathbf{W}_h$ and $\underline{c}_h$ and $\mathbf{W}_u$ and $\underline{c}_u$ using equation [7]. This process can be repeated as many times as desired. The results are shown in Table 5.

Table 5 shows the fit statistics for the W-NOMINATE first dimension scores for the House and Senate in the same fashion as Table 4. Because this is overlapping generations data, the requirement of five entries in each row of $\mathbf{X_0}$ reduced the number of members included in the analysis from the early and late Congresses (the n_j). That is, to be included in the column for the 75th Congress (the first column of $\mathbf{X_0}$), a member would have had to serve five Congresses *beginning* with the 75th Congress (service does not have to be consecutive). In contrast, to be included in the column for the 85th Congress, a member could have served in the 81st through the 85th, or

Table 5A. Fit Statistics for Combined House and Senate W-NOMINATE Scores

N	% Missing	Min. Number of Congresses	R^2	Standard Error of Estimate	Singular Values of $\hat{\psi}\hat{W}'$
1431	70.5	5	.918	.156	97.5

Table 5B. Fit Statistics by Congress for Combined House and Senate W-NOMINATE Scores (Bootstrapped Standard Errors in Parentheses)

	Senate				House			
Congress	n_j	$\hat{c}_j$	$\hat{w}_j$	R^2	n_j	$\hat{c}_j$	$\hat{w}_j$	R^2
75	58	–0.078	1.666	.786	175	–0.135	2.011	.890
		(0.050)	(0.175)			(0.020)	(0.104)	
76	72	–0.089	1.703	.795	240	–0.244	1.954	.911
		(0.045)	(0.137)			(0.019)	(0.072)	
77	81	–0.129	1.697	.725	266	–0.195	1.882	.811
		(0.057)	(0.140)			(0.023)	(0.085)	
78	90	0.017	1.423	.868	312	–0.046	1.801	.925
		(0.035)	(0.097)			(0.014)	(0.050)	
79	100	0.007	1.429	.829	342	–0.063	1.883	.928
		(0.028)	(0.095)			(0.015)	(0.047)	
80	103	–0.118	1.809	.835	343	–0.029	1.808	.899
		(0.040)	(0.137)			(0.018)	(0.069)	
81	107	–0.087	1.637	.882	354	–0.037	1.776	.932
		(0.030)	(0.108)			(0.020)	(0.056)	

(continued on next page)

Table 5B. Fit Statistics by Congress for Combined House and Senate W-NOMINATE Scores *(continued)*

	Senate				House			
Congress	n_j	$\hat{c}_j$	$\hat{w}_j$	R^2	n_j	$\hat{c}_j$	$\hat{w}_j$	R^2
82	100	–0.054	1.764	.848	370	0.050	1.719	.894
		(0.034)	(0.123)			(0.019)	(0.055)	
83	102	–0.095	1.818	.892	381	–0.197	1.647	.886
		(0.029)	(0.079)			(0.015)	(0.051)	
84	95	–0.063	1.916	.852	400	–0.030	1.963	.904
		(0.036)	(0.074)			(0.017)	(0.052)	
85	100	–0.104	1.967	.894	401	–0.041	1.828	.917
		(0.035)	(0.089)			(0.018)	(0.053)	
86	102	–0.002	1.891	.901	394	–0.156	1.894	.907
		(0.029)	(0.088)			(0.019)	(0.050)	
87	102	0.181	1.735	.897	417	–0.072	1.793	.889
		(0.026)	(0.074)			(0.018)	(0.046)	
88	103	0.024	1.759	.703	418	–0.025	1.839	.867
		(0.033)	(0.143)			(0.017)	(0.039)	
89	106	0.015	1.790	.887	395	0.074	1.824	.889
		(0.023)	(0.096)			(0.017)	(0.051)	
90	113	0.105	1.712	.789	402	0.132	1.732	.881
		(0.023)	(0.096)			(0.014)	(0.047)	
91	108	0.074	1.766	.899	400	0.108	1.697	.845
		(0.022)	(0.073)			(0.014)	(0.051)	
92	106	0.066	1.659	.902	394	0.200	1.614	.867
		(0.022)	(0.072)			(0.015)	(0.049)	
93	108	0.000	1.808	.924	385	0.091	1.520	.903
		(0.025)	(0.066)			(0.014)	(0.039)	

(continued on next page)

Table 5B. Fit Statistics by Congress for Combined House and Senate W-NOMINATE Scores *(continued)*

	Senate				House			
Congress	n_j	$\hat{c}_j$	$\hat{w}_j$	R^2	n_j	$\hat{c}_j$	$\hat{w}_j$	R^2
94	116	0.045	1.684	.912	386	0.138	1.538	.910
		(0.024)	(0.051)			(0.015)	(0.040)	
95	113	0.075	1.739	.914	377	0.088	1.766	.937
		(0.020)	(0.084)			(0.016)	(0.048)	
96	115	0.009	1.875	.931	382	0.017	1.790	.952
		(0.025)	(0.080)			(0.016)	(0.051)	
97	111	0.059	1.925	.928	382	–0.053	1.771	.958
		(0.030)	(0.051)			(0.017)	(0.047)	
98	113	–0.072	1.861	.949	415	–0.024	1.935	.965
		(0.025)	(0.067)			(0.019)	(0.045)	
99	110	–0.060	1.828	.950	412	–0.030	1.786	.961
		(0.023)	(0.066)			(0.018)	(0.047)	
100	105	–0.163	1.952	.966	420	–0.069	1.691	.961
		(0.023)	(0.054)			(0.018)	(0.044)	
101	95	–0.112	1.811	.972	392	–0.183	1.734	.966
		(0.021)	(0.055)			(0.018)	(0.045)	
102	91	–0.124	1.969	.973	360	–0.056	1.858	.961
		(0.024)	(0.057)			(0.019)	(0.047)	
103	78	–0.141	2.010	.954	264	–0.094	1.809	.950
		(0.022)	(0.089)			(0.019)	(0.046)	
104	69	–0.007	2.492	.945	219	0.050	2.196	.951
		(0.033)	(0.096)			(0.023)	(0.058)	

the 85th through the 89th. Finally, to be included in the 104th column, a member would have had to serve in five Congresses *ending* with the 104th.

The overall fit of the model stated in equation [17] is a respectable r-square of .918 indicating that members of Congress are very stable in their location on the liberal/conservative dimension over time. The standard errors were computed using the same approach as that for the issue scale data shown in Table 4. If conventional *t*-tests were applied using the estimated standard errors (recall that the "certainty" standard errors are smaller than the bootstrapped standard errors), all the $\hat{\mathbf{w}}$'s for both the House and Senate are statistically significant.

The missing data in this example are qualitatively different from the missing data in the 1980 issue scale example. In the issue scale example, there is missing data because of nonresponses. Here the data is missing simply because legislators were not in Congress. Both are treated the same by the scaling procedure. Substantively, if a respondent does not understand a question and does not give an answer, then equation [1] may not be the appropriate model for that respondent's answers. There is no easy solution for this problem. However, if the amount of missing data is low as it is in 1980 issue scale application, it is probably not a serious problem.

Given the $\hat{\psi}$'s for members serving in at least five Congresses, coordinates for members serving in less than five Congresses can be estimated using the $\hat{\mathbf{w}}$'s and $\hat{\mathbf{c}}$'s by applying the transformation: $(x_{ij} - \hat{c}_j) / \hat{w}_j$ and taking the average over the total number of Congresses served. That is:

$$\hat{\psi}_1^* = \frac{\sum_{j=1}^{m_i} \frac{\left(x_{ij} - \hat{c}_j\right)}{\hat{w}_j}}{m_i} \qquad [18]$$

where $\mathbf{m_i}$ is the total number of Congresses served and ranges from one to four, and the $\hat{\mathbf{w}}$'s and $\hat{\mathbf{c}}$'s can be a combination of those from the House and the Senate if the member served in both chambers.

Figure 2 shows the means for the political parties over the 1937 to 1995 period combining the $\hat{\psi}_i$'s estimated by the procedure and the $\hat{\psi}_i$'s from equation [18]. The two chambers are very similar in their patterns over time. In the latter part of the New Deal, voting on minimum wages opened up a split between the northern and southern wings of the Democratic party. During World War II, voting on issues related to the right of Blacks to vote in federal elections exacerbated the split (Poole and Rosenthal 1997, chap. 5). Southern Democrats move to the right in both chambers until just after the Civil Rights era of the mid to late 1960s and then begin moving back to the left during the 1970s and 1980s. Republicans shift to the left from the late

Figure 2A. House of Representatives
1st Dimension of Joint Space

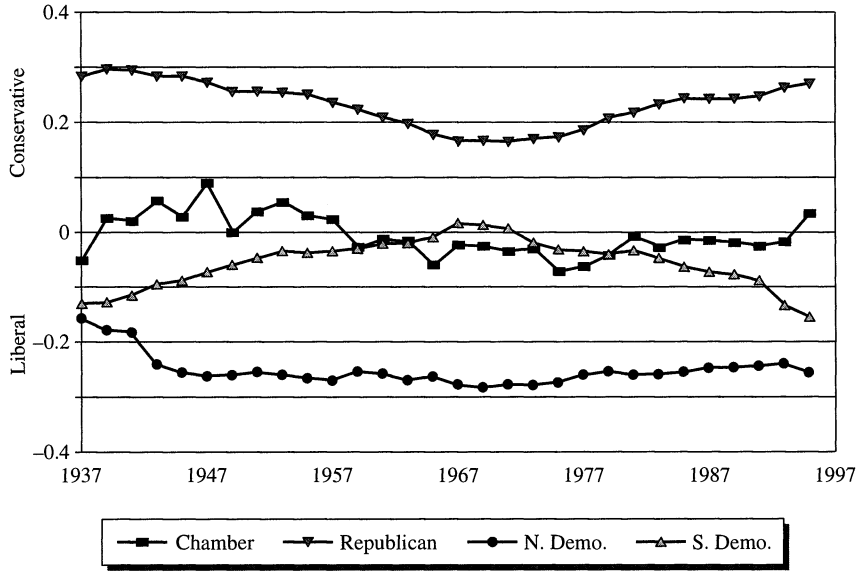
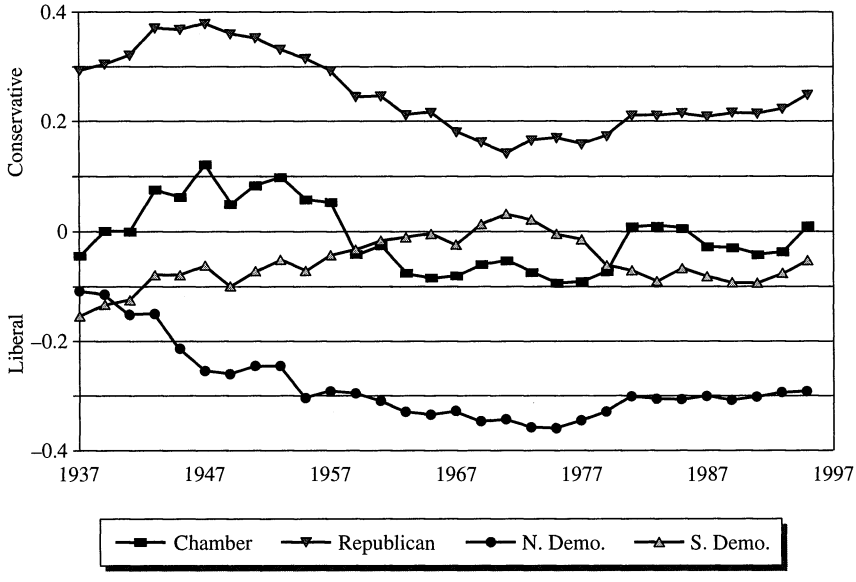


Figure 2B. Senate
1st Dimension of Joint Space



1940s and then reverse course after the Civil Rights era. Republicans in both chambers have been moving to the right since the late 1970s (Poole and Rosenthal 1997; McCarty, Poole, and Rosenthal 1997). Note that because each member is assumed to have the same position throughout his career, these shifts in the various party means are due to *replacement* not conversion. The correlation between the chamber means is .90 indicating that whatever forces are at work in American politics tend to work on both chambers equally regardless of their very different constituencies, terms, and internal rules and procedures. Viewed over a long period of time, there is no pat answer to the question: is the Senate more liberal than the House.[21] The corresponding correlations for Republicans, northern Democrats, and southern Democrats are .81, .84, and .76, respectively.

6. Conclusion

The scaling procedure shown in this paper performs, in effect, a singular value decomposition of a rectangular matrix of real elements with missing entries. In contrast to existing techniques such as factor analysis which work with a correlation or covariance matrix computed from the data matrix, the scaling procedure shown here analyzes the data matrix *directly* without any intervening transformations of the original data. It is a general-purpose tool that can be used to estimate latent/unobservable dimensions underlying a set of manifest/observable variables or as a method to obtain an Eckart-Young lower-rank approximation matrix of a matrix with missing entries. Monte Carlo tests show that the procedure does a good job of reproducing the missing elements of a matrix even at high levels of error and missing data.

Manuscript submitted 21 January 1997.
Final manuscript received 16 September 1997.

APPENDIX A

The partial derivatives of equation [4] are

$$\frac{\partial \mu}{\partial \psi_{ik}} = 2\sum_{j=1}^{m_i}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right] w_{jk} + 2\gamma_k + 2\phi_{kk}\psi_{ik} + 2\sum_{\ell\neq k}^{s}\phi_{k\ell}\psi_{i\ell} \qquad [A1A]$$

$$\frac{\partial \mu}{\partial w_{jk}} = 2\sum_{i=1}^{n_j}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right] \psi_{ik} - 2\phi_{kk}w_{jk} - 2\sum_{\ell\neq k}^{s}\phi_{k\ell}w_{j\ell} \qquad [A1B]$$

[21] See Froman (1971) and Kernell (1973) for a discussion of House-Senate differences.

$$\frac{\partial \mu}{\partial c_j} = 2\sum_{i=1}^{n_j}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right] \qquad \text{[A1C]}$$

$$\frac{\partial \mu}{\partial \gamma_k} = 2\sum_{i=1}^{n_j}\psi_{ik} \qquad \text{[A1D]}$$

$$\frac{\partial \mu}{\partial \phi_{k\ell}} = \begin{cases} \sum_{i=1}^{n}\psi_{ik}^2 - \sum_{j=1}^{m} w_{jk}^2 \text{ if } k=\ell \\ \sum_{i=1}^{n}\psi_{ik}\psi_{i\ell} - \sum_{j=1}^{m} w_{jk}w_{j\ell} \text{ if } k\neq\ell \end{cases} \qquad \text{[A1E]}$$

where n_j means that the total of the summation over i may vary from 2m to n depending upon how many entries there are in the ith column of $\mathbf{X}_0$.

To see that $\underline{\gamma} = \underline{0}$, note that

$$\sum_{i=1}^{n}\frac{\partial \mu}{\partial \psi_{ik}} = \sum_{j=1}^{m} w_{jk}\frac{\partial \mu}{\partial c_j} + 2n\gamma_k + 2\phi_{kk}\sum_{i=1}^{n}\psi_{ik} + 2\sum_{\ell\neq k}^{s}\phi_{k\ell}\sum_{i=1}^{n}\psi_{i\ell} = 0$$

The first term on the right is equal to zero because $\sum_{i=1}^{n}\sum_{j=1}^{m_i} = \sum_{j=1}^{m}\sum_{i=1}^{n_j}$ and $\frac{\partial \mu}{\partial c_j} = 0$,

and the third and fourth terms are equal to zero because $\sum_{i=1}^{n}\psi_{ik} = 0$. Hence, the second term, $2n\gamma_k$, must equal zero. Therefore, $\gamma_k = 0$ because $n > 0$.

To see that the ϕ's are all zero, consider the two equations:

$$\frac{1}{2}\sum_{i=1}^{n}\frac{\partial \mu}{\partial \psi_{ik}}\psi_{ih} = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right]w_{jk}\psi_{ih} - \phi_{kh}\lambda_h = 0$$

and

$$\frac{1}{2}\sum_{j=1}^{m}\frac{\partial \mu}{\partial w_{jk}}w_{jh} = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[\left(\sum_{\ell=1}^{s} w_{j\ell}\psi_{i\ell}\right) + c_j - x_{ij}\right]w_{jk}\psi_{ih} - \phi_{kh}\lambda_h = 0$$

because $\sum_{i=1}^{n}\sum_{j=1}^{m_i} = \sum_{j=1}^{m}\sum_{i=1}^{n_j}$ the first term in both equations are identical. Hence $\phi_{kh}\lambda_h = -\phi_{kh}\lambda_h$. Now, since λ_h is the hth singular value in Λ and by assumption the s singular values in Λ are greater than zero, then $\phi_{kh} = -\phi_{kh}$ which is possible only if $\phi_{kh} = 0$.

APPENDIX B

To maximize the number of positive elements in the m by m covariance matrix

$$[\mathbf{X}_0 - \mathbf{J}_n\underline{c}'][\mathbf{X}_0 - \mathbf{J}_n\underline{c}']$$

a vector of plus and minus ones must be found so that when they are placed on the diagonal of an m by m diagonal matrix Λ^*, the number of positive elements in the m by m matrix:

$$\Lambda^*[\mathbf{X}_0 - \mathbf{J}_n\underline{c}'][\mathbf{X}_0 - \mathbf{J}_n\underline{c}']\Lambda^*$$

will be maximum.

For example, suppose the covariance matrix is

$$\begin{bmatrix} 1 & -2 & 3 & -4 \\ -2 & 1 & 5 & -6 \\ 3 & 5 & 1 & 7 \\ -4 & -6 & 7 & 1 \end{bmatrix}$$

Clearly, changing the signs of the fourth column and row will reduce the number of negative elements in the matrix. That is,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & -2 & 3 & -4 \\ -2 & 1 & 5 & -6 \\ 3 & 5 & 1 & 7 \\ -4 & -6 & 7 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 3 & 4 \\ -2 & 1 & 5 & 6 \\ 3 & 5 & 1 & -7 \\ 4 & 6 & -7 & 1 \end{bmatrix}$$

Any further changes of the signs of the columns/rows will increase the number of negative elements in the matrix. Note that another solution for the diagonal of Λ^* is (–1,1,1,1). This also reduces the number of negative elements to four.

A simple algorithm to produce a diagonal for Λ^* is to find the row in the covariance matrix with the most negative elements and change its sign and that of the corresponding column (if no row has more than (m – 1)/2 negative elements, than no sign changes are necessary). This process is then repeated. The row with the next most negative elements is found and its sign is changed. This process can be continued until all rows have (m – 1)/2 or less negative entries.

REFERENCES

Aldrich, John, and Richard D. McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *American Political Science Review* 71:111–30.

Brady, Henry. 1990. "Traits versus Issues: Factor versus Ideal-Point Analysis of Candidate Thermometer Ratings." *Political Analysis* 2:97–129.

Cahoon, Lawrence, Melvin J. Hinich, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting." In *Graphical Representation of Multivariate Data*, ed. P. C. Wang. New York: Academic Press.

Carroll, J. Douglas. 1972. "Individual Differences and Multidimensional Scaling." In *Multidimensional Scaling*, ed. Roger N. Shepard, A. Kimball Romney, and Sara Beth Nerlove. New York: Seminar Press.

Carroll, J. Douglas, and Jih-Jie Chang. 1970. "Analysis of Individual Differences in Multidimensional Scaling Via an N-way Generalization of "Eckart-Young" Decomposition." *Psychometrika* 35:283–319.

Converse, Phillip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent,* ed. David E. Apter. New York: Free Press.

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

Eckart, Carl, and Gale Young. 1936. "The Approximation of One Matrix By Another of Lower Rank." *Psychometrika* 1:211–18.

Enelow, James, and Melvin J. Hinich. 1984. *The Spatial Theory of Voting*. New York: Cambridge University Press.

Enelow, James, and Melvin J. Hinich. 1994. "A Test of the Predictive Dimensions Model in Spatial Voting Theory." *Public Choice* 78:155–69.

Feldman, Stanley, and John Zaller. 1992. "Political Culture of Ambivalence: Ideological Responses to the Welfare State." *American Journal of Political Science* 36:268–307.

Froman, Lewis A., Jr. 1971. "Differences Between the House and Senate." In *Readings On Congress*, ed. Raymond Wolfinger. Englewood Cliffs, NJ: Prentice-Hall.

Gleason, Terry C., and Richard Staelin. 1975. "A Proposal for Handling Missing Data." *Psychometrika* 40:229–52.

Groseclose, Tim, Steven D. Levitt, and James M. Snyder, Jr. 1997. "An Inflation Index For ADA Scores." Ohio State University. Typescript.

Hinich, Melvin J., and Michael Munger. 1994. *Ideology and the Theory of Political Choice.* Ann Arbor, MI: University of Michigan Press.

Hinich, Melvin J., and Michael Munger. 1997. *Analytical Politics.* New York: Cambridge University Press.

Hinich, Melvin J., and Walker Pollard. 1981. "A New Approach to the Spatial Theory of Electoral Competition." *American Journal of Political Science* 25:323–41.

Horst, Paul. 1963. *Matrix Algebra for Social Scientists.* New York: Holt, Rinehart and Winston, Inc.

Jacoby, William. 1993. "Testing the Effects of Paired Issue Statements on the Seven-Point Issue Scales." *Political Analysis* 5:61–112.

Johnson, Richard M. 1963. "On a Theorem Stated by Eckart and Young." *Psychometrika* 28:259–63.

Keller, Joseph B. 1962. "Factorization of Matrices by Least-Squares." *Biometrika* 49:239–42.

Kernell, Sam. 1973. "Is the Senate More Liberal than the House?" *Journal of Politics* 35:332–63.

Lawson, Charles L., and Richard J. Hanson. 1974. *Solving Least Squares Problems.* Englewood Cliffs, NJ: Prentice-Hall Inc.

MacRae, Duncan. 1970. *Issues and Parties in Legislative Voting: Methods of Statistical Analysis.* New York: Harper and Row.

McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 1997. "Income Redistribution and the Realignment of American Politics." *AEI Studies on Understanding Economic Inequality.* Washington, DC: AEI Press.

Ordeshook, Peter C. 1976. "The Spatial Theory of Elections: A Review and a Critique." In *Party Identification and Beyond* ed. Ian Budge, Ivor Crewe, and Dennis Farlie. New York: Wiley.

Poole, Keith T. 1997. "How to Use the Black Box." GSIA, Carnegie-Mellon University. Typescript.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Rabinowitz, George. 1974. *Spatial Models of Electoral Choice: An Empirical Analysis*. Chapel Hill, NC: Institute for Research in Social Science.

Rusk, Jerrold, and Herbert Weisberg. 1972. "Perceptions of Presidential Candidates." *Midwest Journal of Political Science* 16:388–410.

Schonemann, Peter H. 1966. "A Generalized Solution of the Orthogonal Procrustes Problem." *Psychometrika* 31:1–10.

Schonemann, Peter H. 1970. "On Metric Multidimensional Unfolding." *Psychometrika* 35:349–66

Schonemann, Peter H., and R. M. Carroll. 1970. "Fitting One Matrix to Another Under Choice of a Central Dilation and Rigid Motion." *Psychometrika* 35:245–56

Sundquist, James L. 1983. *Dynamics of the Party System.* Washington, DC: Brookings Institution.

Takane, Yoshio, Forrest W. Young, and Jan de Leeuw. 1977. "NonMetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features." *Psychometrika* 42:7–67.

Van de Geer, John P. 1971. *Introduction to Multivariate Analysis for the Social Sciences*. San Francisco: W. H. Freeman and Company.

Weisberg, Herbert, and Jerrold Rusk. 1970. "Dimensions of Candidate Evaluation." *American Political Science Review* 64:1167–85.