METRIC MULTIDIMENSIONAL MAPPING OF CANDIDATES IN AN ISSUE SPACE

Melvin J. Hinich, Virginia Polytechnic Institute and State University
Lawrence Cahoon, Carnegie-Mellon University

INTRODUCTION

The spatial theory of electorial competition provided the basis for the development of the methodology presented in this paper. Spatial analysis seeks, under a variety of assumptions to ascertain the policies candidates should adopt. The essential assumption of spatial theory is that each candidate's strategy and each citizen's preference can be represented in an Euclidean space, and that each citizen's loss function can be represented as a metric in this space. For the individual citizen this space is developed over time as he views the changing political situation. The U.S. has a representative form of government in which the candidates are judged to a large extent on their positions on the issues. The election process generates uncertainty concerning the nature of the many issues involved, and concerning the candidates' position on these issues. The theory assumes that the citizens, in order to more readily deal with this situation, condense the issues into a space in which the dimensions represent their underlying preferences. These dimensions then do not necessarily represent the issues directly. Since the citizens are concerned primarily with the future activity of the candidates it is assumed that they use this space to predict the future positions of the candidates. Thus a "spatial picture" at any point in time is a plot of the predicted positions of the candidates relative to the citizens underlying preferences. When asked how he feels about a given candidate the citizen responds in terms of a loss function relative to this "spatial picture".

With this basis we use the feeling thermometer data collected in the University of Michigan Survey Research Center's 1968 Presidential Survey to map the candidate positions and the respondent's ideal points into a joint two dimensional space. The scaling technique is based on the assumption that the thermometer scores given by the respondents for each candidate is a monotonic function of weighted Euclidean distance. The positions of the candidates, the respondents' ideal points, the issue weights, axis orientation, and the mean ideal point of the population are all identified in the model used to develop the methodology, provided some constraints are satisfied.

THE MODEL

The metric which specifies the utility function of the individual respondents is basic to any metric scaling technique. We assume that this utility function is of the form

(1) $-U(\underline{\theta}_j,\underline{X}_i) = [(\underline{\theta}_j-\underline{X}_i)'A(\underline{\theta}_j-\underline{X}_i)]^{1/K}$

when $\underline{\theta}_j = (\theta_{j1},\theta_{2j})'$ is the position of the jth candidate, $\underline{X}_i = (X_{i1},X_{i2})$ is the ith citizen's ideal point, $\underline{A}$ is a diagonal matrix of positive issue weights a_1, a_2, and K is a positive integer. We have constrained ourselves to a two-dimensional space since that is the dimensionality of the space in which the candidate positions will be estimated in the application of the methodology to the 1968 SRC survey data. When K=1(1) is the quadratic loss function common in statistics. We have found that the date suggests a value of K greater the 1 to be more appropriate.

Suppose that each citizen rates each of p+1 candidates on a thermometer scale that varies from 0 to 100. If 100 is the response given by an individual to a candidate whom he likes extremely well the thermometer score takes the form

(2) $T(\underline{X}_i,\underline{\theta}_j) = 100 - [(\underline{\theta}_j-\underline{X}_i)'A(\underline{\theta}_j-\underline{X}_i)]^{1/K}$

For factor analysis to be useful it is necessary to eliminate the nonlinerity in expression (2).

Before proceeding to those modifications, let us first make clear a few assumptions which are necessary for the procedure to be applicable to spatial analysis. First, we assume that all citizens have the same perceptions of the candidate. Secondly, the weight a_1 and a_2 are independent of $\underline{X}$. These two assumptions possess a long history in spatial analysis. It must also be admitted that they are restrictive. In an ideal situation it is desirable to allow different perceptives on the part of different people. Rather than continue into an extensive discussion of these assumptions let us simply note that we must begin somewhere so we begin with these assumptions. Our third assumption is that, the covariance matrix, $\underline{\Sigma}$, is known, and is a diagonal matrix. This assumption may seem restrictive. It is not. The reason for making the assumption is that $\underline{A}$ and $\underline{\Sigma}$ are not jointly identifiable. In practice $\underline{A}$ and $\underline{\Sigma}$ jointly provide the same type of information about the space. That is, they both provide information on the relative importance the citizens give to the different dimensions. Since this is true, when $\underline{\Sigma}$ is not known it is sufficient to assume that $\underline{\Sigma}=\underline{I}$ and obtain the information about the relative importance of the dimensions from the estimate of $\underline{A}$ which we will obtain. In requiring that $\underline{\Sigma}$ be diagonal we are simply stating that the axis of the space which will be recovered will be that for which $\underline{\Sigma}$ is diagonal. This is possible since for a general covariance matrix $\underline{\Sigma}$ it is possible to rotate the space such that relative to the rotated set of axis the new covariance matrix $\underline{\Gamma}'\underline{\Sigma}\underline{\Gamma}$ is diagonal. $\underline{\Gamma}'X$ is the new set of observations and $\underline{\Gamma}$ is an orthogonal matrix. With these assumptions it is possible to proceed to the discussion of the methodology.

THE PROCEDURE

Subtracting equation (2) from 100 and allowing for an additive error in the individual observation of the candidates we then have a set of observations of the form,

(3) $D_{ij} = [(\underline{\theta}_j-\underline{X}_i)'\underline{A}(\underline{\theta}_j-\underline{X}_i)]^{1/K} + E_{ij}$

the errors, E_{ij}, are assumed to be independent with unknown variance σ^2.

Briefly, the procedure uses the observations D_{ij} to construct a new set of observations which are linear in X_i and θ_j. These observations, when viewed as a set of observations on the set of candidates will satisfy the assumptions necessary for factor analysis when $K = 1$ or 2. For $K>2$ some modifications are necessary. Thus the covariance matrix of the new set of observations will be constructed. Factor analysis will be applied to that matrix. As usual with factor analysis the factored matrix will be a rotation of the factorization we desire. While factor analysis must in general be satisfied with an arbitrary rotation of the true matrix, we will use the additional information available to us about the observations to construct a regression to identify jointly the rotation and the axis weights. This will in turn allow us to estimate the positions of the candidates. Once that has been done we may then return to the set of observations which we constructed to obtain estimates of the individual's ideal points.

The Procedure was first developed relative to squared Euclidean distance. Thus to eliminate the non-linearities in (2) we first raise the observations to the k^{th} power to obtain,

(4) $D_{ij}^K = \underline{\theta_j'A\theta_j} - 2\underline{\theta_j'AX_i} \pm X_i'\underline{AX_i} + \delta_{ij}$

where δ_{ij} is generic for the error term. From (4) the construction of a set of observations linear in X_i and θ_j is straightforward. First we note that the entire problem is invariant with respect to the origin of the underlying coordinate system. For that reason it is possible to simplify the mathematics of the problem by defining the origin to be located at the position of one of the candidates. Specifically let $\theta_{p+1}=0$. Thus for $j=p+1$ (4) becomes,

(5) $D_{ip+1}^K = \underline{X_i'AX_i} + \delta_{ip+1}$

The set of observations desired is formed by subtracting (5) from (4) and then subtracting the mean $\bar{D}_j^K - \bar{D}_{p+1}^K$ to obtain,

(6) $Y_{ij} = D_{ij}^K - D_{ip+1}^K - (\bar{D}_j^K - \bar{D}_{p+1}^K) = 2\underline{\theta_j'AX_i} + \delta_{ij}$.

Let $\underline{Y_i}=(Y_{i1},\ldots,Y_{ip})'$ and $\underline{Y}$ be the nxp matrix whose jth element is Y_{ij}. The sample covariance matrix of the $\underline{Y_i}$ is $\frac{1}{n-1}\underline{Y'Y}$. It is this matrix which we wish to factor. Since $\frac{1}{n-1}Y'Y$ is the covariance matrix of $\underline{Y_i}$ it estimates

(7) $4\underline{\theta}'\underline{A}^2\underline{\theta} + \underline{C} + \underline{\psi}$

where $\underline{\theta} = (\underline{\theta_1},\ldots,\underline{\theta_p})$, $\underline{\psi}$ is the covariance matrix of the vector of errors $(\delta_{i1},\ldots,\delta_{ip})$ and C is the matrix of covariance between $2\underline{\theta_j'AX_i}$ and δ_{ij}, $i,j = 1,\ldots p$. For $K = 1,2$ $\underline{C}$ is a matrix of zeros and $\underline{\psi} = \underline{1}'\underline{1}d+\underline{D}$ where $\underline{1}=(1,\ldots,1)'$, d is a constant and $\underline{D}$ is a diagonal matrix. Thus when $K=1$ or 2 $\frac{1}{n-1}\underline{Y'Y}$ is factored, by the method of factor analysis we obtain an estimate

(8) $\underline{\Lambda} = (2\underline{\theta}'\underline{A},\underline{1})\underline{P}$

where $\underline{P}$ is a 3x3 orthogonal matrix.

If $K>2$ some modifications are necessary. From equation (6) it is possible to specify the exact form of $\underline{C+\psi}$ in (7). If we ignore moments greater than two this can be written as $\underline{E}\sigma^2$. A typical term in the matrix $\underline{E}$ is

$\mathrm{cov}\{(\underline{\theta_j}-\underline{X_i})'\underline{A}(\underline{\theta_i}-\underline{X_i}),[(\underline{\theta_j}-\underline{X_i})'\underline{A}(\underline{\theta_j}-\underline{X_i})]^{\frac{K-2}{K}}\}$.

This can be estimated from the observations as $\mathrm{cov}(D^K,D^{K-2})$. Once the matrix $\underline{E}$ is estimated we use the fact that $4\underline{\theta'A\theta}$ is of rank two to estimate σ^2. That is we search over σ^2 to find the value of σ^2 such that the smallest p-2 eigenvalues of

(9) $\frac{1}{n-1}\underline{Y'Y} - \underline{E}\,\sigma^2$

are near zero. Specifically we choose σ^2 such that the mean of the last p-2 eigenvalues of (9) is zero. Then (9) estimates $4\underline{\theta}'\underline{A}^2\underline{\theta}$. This matrix is then factored by the principle components version of factor analysis to obtain an estimate of $2\underline{\theta}'\underline{A\Gamma}$ where $\underline{\Gamma}$ is a 2x2 orthogonal matrix.

Thus for $K=1,2$ we have obtained an estimate of $(2\underline{\theta}'\underline{A},\underline{1})\underline{P}$ and for $K>2$ we estimate $2\underline{\theta}'\underline{A\Gamma}$. When $K=1,2$ a regression is constructed using the fact that we have a column of 1's in the matrix $2\underline{\theta}'\underline{A},\underline{1}$ to identify two of the three rotational parameters in $\underline{P}$. Once these have been identified the matrix $(2\underline{\theta}'\underline{A},\underline{1})\underline{P}$ is rotated to form an estimate of $(2\underline{\theta}'\underline{A\Gamma},\underline{1})$.

The estimate of $\underline{\theta}'\underline{A\Gamma}$ is next used together with the means $\bar{D}_j^K-\bar{D}_{p+1}^K$ to obtain estimates of $\underline{\Gamma}$ and $\underline{A}$. We begin by noticing that

(10) $\bar{D}_j^K - \bar{D}_{p+1}^K = \underline{\theta_j'A\theta_j} - 2\underline{\theta_j'A\bar{X}} + \delta_{ij}$

If $K>2$ a minor correction is made to (10) to account for the bias introduced into the error term by raising the observations to the K^{th} power. Post multiplying our estimate of $\underline{\theta}'\underline{A\Gamma}$ by the unknown $\underline{\Gamma}'\underline{A}^{-1/2}$ and forming the inner product $(\underline{\theta}'\underline{A\Gamma\Gamma}'\underline{A}^{-1/2})(\underline{\theta}'\underline{A\Gamma\Gamma}'\underline{A}^{-1/2})'$ we obtain an estimate of $\underline{\theta_j'A\theta_j}$. Post-multiplying the estimate of $\underline{\theta_j'A\Gamma}$ by the unknown $\underline{\Gamma}'\underline{\bar{X}}$ we obtain an estimate of $\underline{\theta_j'A\bar{X}}$ the sum of these two provides an estimate of $\bar{D}_j^K - \bar{D}_{p+1}^K$. Expanding this expression we have the regression model

(11) $\bar{D}_j^K-\bar{D}_{p+1}^K = \alpha_0\hat{M}_{j1}^2 + \alpha_1\hat{M}_{j2}^2 + \alpha_2\hat{M}_{j1}\hat{M}_{j2}$

$+ \alpha_3\hat{M}_{j1} + \alpha_4\hat{M}_{j2} + \delta_{ij}$

where our estimate of $\underline{\theta}'\underline{A\Gamma}$ is $\hat{M}=(\hat{M}_{ij})$. The coefficients $\alpha_0,\ldots,\alpha_4$ are nonlinear functions of $a_1,a_2,\bar{X}_1,\bar{X}_2$, and the rotation parameter in the matrix $\underline{\Gamma}$. $\underline{\bar{X}} = (\bar{X}_1,\bar{X}_2) = n^{-1}\Sigma_{i=1}^{n}\underline{X_i}$

$\alpha_0,\ldots,\alpha_4$ are then used to solve the system of equations for $a_1,a_2,\bar{X}_1,\bar{X}_2$ and the rotation parameter. It then becomes a single matter to use these estimates together with the estimates of $\underline{\theta}'\underline{A\Gamma}$ to estimate $\underline{\theta}$, the matrix of candidate positions.

The respondents' ideal points can also be estimated. This is done by using the estimates $\underline{\theta}'\underline{A\Gamma}$, and Γ together with the set of observations $Y_{ij} = 2\underline{\theta_j}\underline{AX_i} + \delta_{ij}$ to form a regression to identify the $\underline{X_i}$.

THE DATA

The procedure has been applied to the 1968 SRC thermeter data. Prior to a discussion of the results of that analysis several comments are necessary. We have assumed that all citizens have the same perceptions of the candidates. To provide for a greater certainty in the truth of this assumption the population was divided into three groups: Democrats, Independents, and Republicans. A respondent who leaned towards the Democrats or to the Republicans while belonging to neither group was nontheless included with that group, as well as with the Independents. Thus there is some overlap in the three groups. Also, while the survey question did provide for an answer in case a respondent did not know a candidate, it was also possible to assume a score of 50 could be given to a candidate whom the respondent was unfamiliar with. A quick review of the actual responses given indicates that this did in fact occur. To eliminate distortions caused by this any respondent who gave more than three of the eleven candidates a score of 50 was eliminated from the sample.

For the utility function $U(\theta_j,X_i)$ a value of K equal to 4 was chosen. Such a value provides for a myopic view of the candidates. That this is the case is evident from the data. This effect also shows up in the plots of the candidates positions. The Republicans (Fig. 3) distinguish more between Agnew and Reagon than do the Democrats (Fig. 1).

Figures 1, 2, and 3 show the positions of the candidates as estimated by the procedure. The distances in these plots have been changed to Euclidean distances for ease of interpretation. Wallace and LeMay were eliminated from the study because, after a review of the data, it was felt that many respondents wished to score them beyond the permissable range of the thermometer scale. Also included with these plots are the information concerning the relative axis weights, error variance, predicted and true votes, and the R^2 for the regression used to estimate the rotation and axis weights. a_1 is the weight on the horizontal dimension, a_2 the weight on the vertical dimension, a_3 represents an efficiency weight. That is, it was felt that certain candidates were considered by the citizen's to be clear possibilities as presidential candidate, the others were not, thus the citizens would tend to like (or dislike) those who were more likely to be president at some future date. This weight was specific to Nixon, Kennedy, Johnson and Humphry. Thus these four candidates are viewed as being at $(\underline{\theta}_j,0)$, while the remainder are at $(\theta_j,1)$. The ideal points are at $(\underline{X}_i,0)$. This effect can be considered by a simple modification of the regression. It cannot be treated in the covariance matrix of the $\underline{Y}_i$ because the ideal points do not vary over that dimension. That the weight a_3 is positive in all cases indicates a likeing for Nixon, Kennedy, Johnson, and Humphry which is distinct from how the citizens see the other candidate.

The line used to predict the votes for each candidate is not the bisector of the line joining Nixon and Humphry, but is a line parallel to the bisector. Moving the line was considered justified because the thermometer data was gathered shortly after the election, thus creating some possible post election bias. The line chosen is that which maximizes the sum of the percentages of correctly predicted votes for Nixon and Humphry.

One rather interesting result became evident when the ideal points of the citizens who claimed to support Nixon and Humphry but abstained from voting were plotted (see Fig. 4). The plot indicates that the reason for abstention was not alienation, but that these citizens were on the dividing line between voting for Nixon or Humphry.

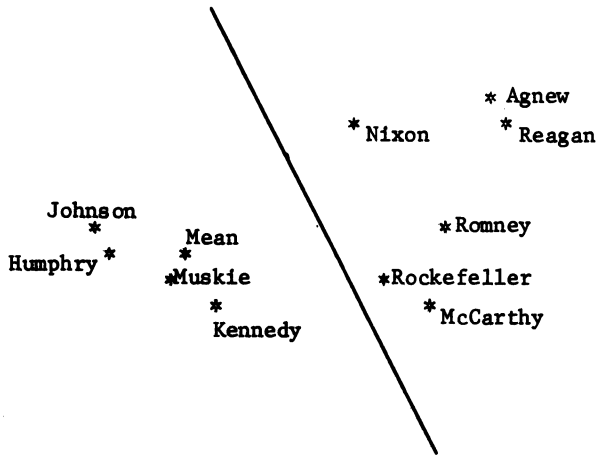
The interpretation of the meaning of the dimensions must be made from the positions of the candidates or from external information. In each of the three groupings it seems that the horizonal dimension is a party dimension. The vertical dimension is more difficult to interpret. Considering Kennedy and McCarthy's anti Viet Nam war stands, and Kennedy's position on the race issue, the vertical dimension could be regarded as a liberal-conservative social action dimension. One hesitates to use terms such as liberal and conservative here because of the many possible meanings these words may have in the political sphere, however it is felt that their use is clear in this instance. The position of Reagan relative to Nixon and Agnew in Fig. 3, (the Republican respondents), is consistant with such an interpretation, as is the position of Nixon relative to the democrats as an entire group. As mentioned before the Democrats, and Independents as well, have some difficulty separating Agnew and Reagan, thus since Agnew was Nixon's running-mate, their positions are determined, in the eyes of the voters, by Nixon's position.

The horizontal dimension was conjectured to be a party dimension. The extreme position of Johnson, and Nixon are factors indicating the truth of that conjecture. It might be asked, then, why are Agnew and Reagan to the right of Nixon, since this would indicate that they are more loyal to the Republican party than Nixon is. This fact is evident in each of the three groupings. It must be remembered that in 1968 Agnew was an unknown political figure, and for the Democrats this seems to be true of Reagan as well. With Nixon being the presidential candidate, the voters then would be very likely to place Nixon closer to their mean position than they would Agnew or Reagan.

Conclusions

We have given a method by which the positions of candidates may be estimated in an Euclidean space. This method has then been applied to real data from the 1968 presidential election. The principle objective here is not simply to secure by some means a spatial map of the candidates that satisfies some intuitive criteria. Rather it is to develop a multidimensional scaling procedure based on spatial theory's assumptions. In this way we hope to link theory to the empirical world.

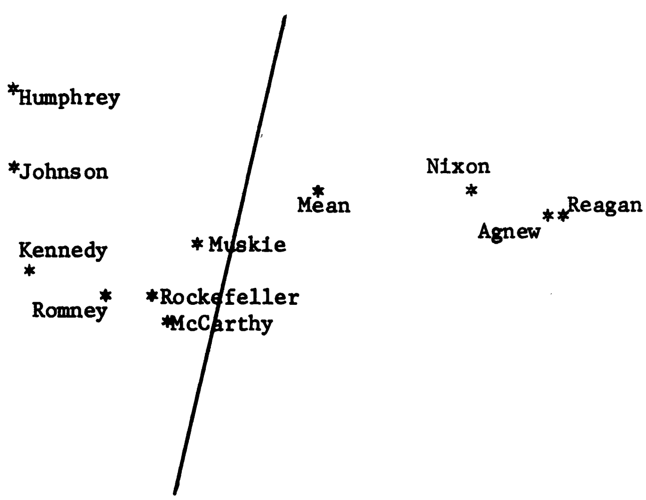
Fig. 1 Predicted Position of Candidate Using Democratic Respondents



Estimated Error Variance $\hat{\sigma}^2$ = 154.7
Ratio of Axis Weights
$\hat{a}_2/\hat{a}_1 = 5.43$
$\hat{a}_3/\hat{a}_1 = 0.39$
Regression $R^2 = 0.928$

Predicted Vote Percentages
Nixon..........22.6%
Humphry........77.4%
True Vote Percentages
Nixon..........21.6%
Humphry........78.4%
Percentage of Votes Correctly
Predicted..........82.0%

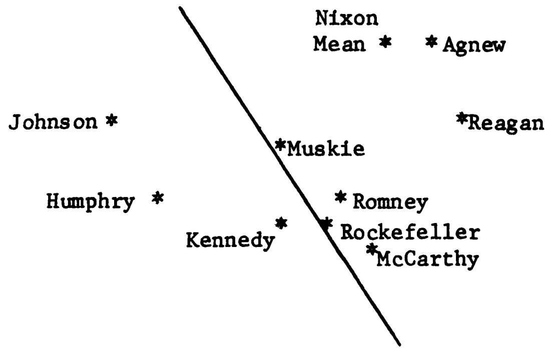
Fig. 2 Predicted Position of Candidate Using Independent Respondents.



Estimated Error Variance $\hat{\sigma}^2$ = 164.4
Ratio of Axis Weights.
$\hat{a}_2/a_1 = 2.44$
$\hat{a}_3/\hat{a}_1 = 1.12$
Regression $R^2 = 0.744$

Predicted Vote Percentages
Nixon..........58.3%
Humphry........41.7%
True Vote Percentages
Nixon..........62.5%
Humphry........37.5%
Percentage of Votes Correctly
Predicted..........83.7%

Fig. 3 Predicted Position of Candidates Using Republican Respondents



Estimated Error Variance $\hat{\sigma}^2 = 171.5$

Ratio Axis Weights

$\hat{a}_2/\hat{a}_1 = 1.74$

$\hat{a}_3/\hat{a}_1 = 0.79$

Regression $R^2 = 0.963$

Predicted Vote Percentages

Nixon..........90.4%

Humphry........ 9.6%

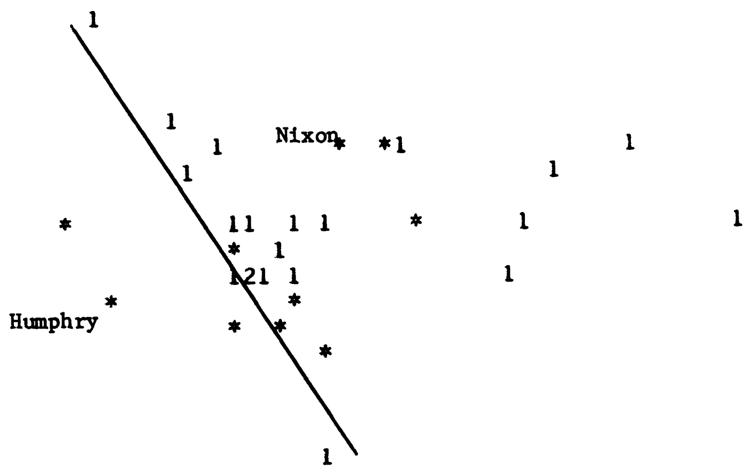
True Vote Percentages

Nixon..........92.6%

Humphry........ 7.4%

Percentage of Votes Correctly Predicted.........93.0%

Fig. 4 Republican Candidate and Position of Respondents who supported Nixon or Humphry but did not vote in the election. The numbers indicate the number of Respondents at any one position. 21 respondents are plotted, 5 more are located in the same position as the candidates, and 1 person is outside the area of the plot. Note the tendency for these respondents to be near the line separating Nixon Voters from Humphry Voters. (Candidates positions are the same as in Fig. 3 above.)



REFERENCES

Anderson, T.W., (1950) The Asymptotic Distribution of Certain Characteristic Roots and Vectors. Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley and Los Angeles, 103-130.

Anderson, T.W., and Rubin, Herman, (1956). Statistical Inference in Factor Analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, University of California Press, Berkeley and Los Angeles, 111-150.

Anderson, T.W., (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Anderson, T.W., (1963). Asymptotic Theory for Principle Components Analysis. Ann. Math. Statist. 34 122-148.

Clarke, M.R.B., (1970). A Rapidly Convergent Method For Maximum-Likelihood Factor Analysis. British Journal of Math. Stat. Psychology. 25 43-52.

Daalder, Hans, and Rusk, Jerrold G. (1972). Perceptions of Party in the Dutch Parliaments. In S.C. Patterson and J.D. Wahlke, eds., Comparative Legislative Behavior. Wiley, New York.

Draper, N.R., Smith, H. (1966). Applied Regression Analysis. Wiley, New York.

Harman, (1967). Modern Factor Analysis. Univ. of Chicago Press. Chicago.

Johnston, J. (1960). Econometric Methods. McGraw Hill. St. Louis.

Jöreskog, K.G., (1967). Some Contributions to Maximum Likelihood Factor Analysis. Psychometrica. 32 443-482.

Karns, David A., (1974). Legislative Roll Call Analysis: The Individual Differences Scaling Approach. Presented at Midwest Political Science Meetings, Chicago 1974.

Lawley, D.N., (1967). Some New Results in Maximum-Likelihood Factor Analysis. Journal of the Royal Soc. of Edinburgh. 67 256-264.

Lawley, D.N., and Maxwell, A.E., (1971). Factor Analysis As A Statistical Method. 2nd ed. Butterworth, London.

Morrison, Donald F., (1967). Multivariate Statistical Methods. McGraw-Hill, New York.

Morrison, Richard, (1972). A Statistical Model of Legislative Roll Call Analysis. Journal of Mathematical Sociology 2 235-48.

Page, Benjamin I., and Brody, Richard A. (1973). Indifference Alienation and Rational Decisions: The Effects of Candidate Evaluations on Turnout and Vote. Public Choice.

Page Benjamin I. (1974). Presidential Campaigning, Party Cleavages, and Responsible Parties. Working Paper, Department of Political Science, University of Chicago.

Rae, Douglas, and Taylor, Michael, (1971). Decision Rules and Policy Outcomes. British Journal of Political Science. 1 71-90.

Rao, C.R., (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Riker, William H., and Ordeshook, Peter C., (1973). An Introduction to Positive Political Theory. Prentice-Hall, Englewood Cliffs.

Rabinowitz, George B. (1973). Spatial Models of Electoral Choice: an Empirical Analysis. University of North Carolina. Chapel Hill.

Rosenthal, Howard, and Sen, Subrata, (1973). Electoral Participation in the French Fifth Republic. American Political Science Review. 67 29-54.

Ross, John, and Cliff, Norman, (1964) A Generalization of the Interpoint Distance Model. Psychometrica. 29 167-176.

Rusk, Jerrold G., and Weisberg, Herbert, (1970) Dimensions of Candidate Evaluation. American Political Science Review. 64 1167-85.

Rusk, Jerrold G., and Weisberg, Herbert. (1972) Perceptions of Presidential Candidates. Midwest Journal of Political Science. 16 388-410.

Shepard, Roger N., Romney, A. Kimbal, and Nerlov, Sara Beth, eds., (1972). Multidimensional Scaling. Seminar Press. New York.

Wendell, R.E., and Thorsen, S.J., (1973). Some Generalizations of Social Decisions Under Majority Rule. Presented at Midwest Political Science Meetings, Chicago 1973.