Traits versus Issues:
Factor versus Ideal-Point Analysis of
Candidate Thermometer Ratings

Henry E. Brady

Abstract

This article describes a computationally simple, statistically consistent, reasonably efficient, and statistically informative generalized least squares (GLS) estimator for a general class of nonlinear, multidimensional scaling (MDS) models including the "ideal-point" models of voters' and legislators' behavior proposed by Melvin Hinich, Keith Poole, and others. Unlike other methods, the method described in this article provides a statistical framework for testing a wide range of hypotheses about these models including their functional form, their dimensionality, and the values of specific parameters. The Hinich ideal-point model is estimated using this method. It fits the data remarkably well compared to a standard factor analysis model that does not provide a reasonable fit to the data. This has the substantive implication of suggesting that voters base their voting decisions upon ideal-point dimensions like liberalism-conservatism and not upon factor analysis dimensions like competence and leadership.

1. Ideal-Point and Factor Models of Choices

I may vote for a candidate because she is the most competent alternative while you may vote for someone else because he has liberal views closest to your own. We are obviously using very different criteria for our choices. One aspect of this difference is that I base my choice upon a dimension for which everyone agrees that more is always better,[1] whereas you base your choice

---

1. Perhaps Senator Roman Hruska (R-Nebraska) would not agree. In support of Richard Nixon's controversial nomination of G. Harrold Carswell for the Supreme Court, Hruska attained

upon a dimension for which people strongly disagree about the ideal position. Candidate traits are generally the first kind of dimension, and candidate issue positions are usually the second kind.

Voting researchers, not to mention politicans themselves, have often wondered whether traits or issues are more important in voting decisions: Do voters vote for persons or for policies? Standard methods of answering this question rely upon survey responses to questions about the candidates' traits and issue positions. These responses, unfortunately, may be incomplete, or they may suffer from serious projection effects (Page and Jones 1979). It would be useful, and perhaps artful, if we could learn something about the dimensions underlying voting decisions by analyzing just the preferences themselves. This is what I shall do in this article by using factor and ideal-point analysis to explore the dimensionality of the 100 point "feeling thermometers" of the American National Election Studies.

Although the connection may not be obvious, the choice of linear factor analysis or of nonlinear ideal-point analysis commits a researcher to a theoretical position on the nature of the underlying dimensions. This is surprising because the choice between these methods is usually treated as simply a matter of convenience. When confronted with a set of variables that might have some common underlying structure, most survey researchers instinctively apply some readily available factor analysis routine to get at this structure. Yet, in the last twenty years, economic theories of party competition in multidimensional spaces (McKelvey 1975; Enelow and Hinich 1984) and psychometric theories of ideological conceptualization (Shepard 1962a and 1962b; Weisberg and Rusk 1970; Weisberg 1974) have suggested that the nonlinear, ideal-point model might be theoretically more appropriate for many kinds of data—especially preference data.

Both types of models assume that there are dimensions underlying preferences, but they have significantly different implications for our understanding of preferences. When candidate preferences are factor analyzed, the factor model implies that more of each dimension is better so that dimensions are *traits* like intelligence, competence, and leadership ability for which it is generally agreed that more is better.[2] When the ideal-point model is used, however, it is assumed that each person has his or her own ideal notion of the

---

a sort of fame when he said that "there are a lot of mediocre judges and people and lawyers. They are entitled to a little representation, aren't they, and a little chance?" (Barone, Ujifusa, and Matthews 1975, 494).

2. To be precise, the model assumes that there may be two kinds of people. Those for whom more of a specific dimension is better and those for whom less of this same dimension is better. In practice, however, it seems likely that for most dimensions one type of person predominates.

proper amount of each dimension. This means that dimensions are *issues* like liberalism-conservatism and party identification for which there are substantial differences of opinion about the proper amount or correct position.

The ideal-point and factor analysis models are very different models of preferences. Unfortunately, none of the existing work compares them using statistical tests because no such tests have been developed for the ideal-point model. In this article I show how an ideal-point model can be estimated for preferences, and I show how statistical tests can be constructed. Using these techniques, I compare the results for one particular ideal-point model with those from standard factor analysis methods. I resoundingly reject the factor analysis model in favor of the ideal-point model, and I conclude that issue dimensions may be substantially more important than trait dimensions in voters' decision making.

2. Ideal-Point Models

In the one-dimensional case, ideal-point models have the following form:

$$Y_{ij} = -|X_i - \beta_j| + \delta_{ij}, \tag{1}$$

where Y_{ij} is person i's rating or evaluation of candidate j, X_i is the person's location or "ideal-point" on some underlying dimension such as "liberalism-conservatism," β_j is the candidate's location on this dimension, and δ_{ij} represents random or unique factors that affect the rating. The quantity $|X_i - \beta_j|$ represents the distance between the person and the candidate, and as this distance gets bigger, the rating goes down.

A general multidimensional scaling (MDS) model for T dimensions can be written as:

$$Y_{ij} = \alpha_j - \left(\sum_{t=1}^{T} |X_{it} - \beta_{ij}|^\nu \right)^\eta + \delta_{ij}. \tag{2}$$

Parameter α_j summarizes those aspects of a candidate's rating that are not the result of spatial location. The summation is over the T spatial dimensions in the model, and the values of ν and η describe the generalized distance measure that people use in forming their evaluations. Keith Poole (1981) uses the "Euclidian" metric in which $\nu = 2$ and $\eta = 1/2$. The "city-block" metric sets $\nu = \eta = 1$. Both of these are special cases of the general Minkowski metric with $\nu = 1/\eta$. Melvin Hinich and his collaborators (Enelow and Hinich 1984) have suggested a model with $\nu = 2$ and $\eta = 1$.

It is relatively easy to write down a general ideal-point model such as

equation 2, but it is much harder to produce a method for estimating its parameters. Equation 2 poses two major complications: it is nonlinear in its parameters and nonlinear in the unobserved X_{ij}. Estimating nonlinear models with perfectly observed variables requires complicated iterative estimation methods and careful attention to a number of theoretical issues (see Gallant 1987). The estimation of simple linear unobserved or errors-in-variables models (Fuller 1987) is often complicated by identification problems and the need for strong assumptions about the stochastic form of the unobserved variables. It is no surprise, then, that models that are both nonlinear in the parameters and nonlinear in the unobserved variables have appeared intractable and have barely been studied.

Some estimation methods have been proposed for specific models (Weisberg and Rusk 1970; Rabinowitz 1976 and 1978; Cahoon, Hinich, and Ordeshook 1978; Poole 1981 and 1984), but little is known about their statistical properties and they do not provide a general framework for testing one model versus another. In this article, I describe a computationally simple, statistically consistent, reasonably efficient, and statistically informative estimator for general nonlinear, ideal-point models. In the next section I illustrate this method by applying it to the one dimensional Hinich ideal-point model. Section 4 describes a generalized estimator for nonlinear MDS models. Section 5 discusses the statistical properties of the estimator. This might be skipped on an initial reading. In section 6, I discuss the basic equations and identification conditions for the multidimensional Hinich model. I also compare these equations with those for factor analysis. Finally, in section 7, I analyze American National Election Studies data using these models.

3. An Estimator for the Hinich Model

Consider the one-dimensional, Hinich ideal-point model:

$$Y_j = \alpha_j - (X - \beta_j)^2 + \delta_j, \tag{3}$$

for $j = 1, \ldots, J$ where Y_j, X, and δ_j are random variables, α_j and β_j are parameters, δ_j is assumed to be independent of X, and each δ_j is assumed to be uncorrelated with every other δ_k and to have a zero expectation because its mean is folded into α_j. In this model, the random variable X represents the distribution of voters' ideal points along the one dimension. Parameter β_j is the location of candidate j along this dimension, and parameter α_j represents the population's average evaluation of candidate j once variation along the one common dimension has been controlled. Finally, random variable δ_j represents all the unique factors, uncorrelated with the distribution of ideal points X, that affect voters' evaluations of candidate j.

The data, of course, are evaluations Y_{ij}. A natural way to summarize these observations is to form their means and covariances:

$$d_j = \sum_{i=1}^{I} Y_{ij}/I,$$ (4)

and

$$d_{jk} = \sum_{i=1}^{I} (Y_{ij} - d_j)(Y_{ik} - d_k)/I.$$ (5)

Under most standard sampling conditions (certainly under random sampling), the law of large numbers implies that these empirical moments will converge to the true ones. Thus, d_j will be close to $\gamma_j = E(Y_j)$ and d_{jk} will be close to $\gamma_{jk} = \text{Cov}(Y_j, Y_k)$.

These true moments, in turn, will be functions of the parameters of the model in equation 3 and the characteristics of the random variables X and δ_j. One obvious approach to estimating the parameters of the model is to choose estimated values for them that provide the best fit between the sample moments and the true moments. To do this, we must know how the true moments depend upon the parameters.

We start with the expectation of Y_j:

$$\gamma_j = \alpha_j - E(X^2) + 2\beta_j E(X) - \beta_j^2 + E(\delta_j).$$

Because the pair X and β_j and the pair $X + \lambda$ and $\beta_j - \lambda$ always yield the same result in equation 3, the values of X are unidentified up to an additive constant. One way to solve this problem is to choose one value or characteristic of X arbitrarily—we choose to set $E(X) = 0$ because this simplifies the following algebra. With these results and the fact that $E(\delta_j)$ is zero for all j:

$$\gamma_j = \alpha_j - \mu_2 - \beta_j^2,$$ (6)

where μ_2 equals $E(X^2)$.

In the same fashion, $\gamma_{jk} = \text{Cov}(Y_j, Y_k)$:

$$\gamma_{jk} = E[(- X^2 + 2\beta_j X + \delta_j + \mu_2)(- X^2 + 2\beta_k X + \delta_k + \mu_2)].$$

By repeatedly using the fact that X is independent of δ_j:

$$\gamma_{jk} = E(X^4) - 2\beta_k E(X^3) - E(X^2)\mu_2 - 2\beta_j E(X^3) + \\ 4\beta_j \beta_k E(X^2) + \psi_{jk} - \mu_2 E(X^2) + \mu_2^2,$$

where $\psi_{jk} = E(\delta_k \delta_k)$. Combining terms and letting $E(X^k) = \mu_k$:

$$\gamma_{jk} = \mu_4 - 2(\beta_j + \beta_k)\mu_3 + 4\beta_j\beta_k\mu_2 + \psi_{jk} - \mu_2^2. \tag{7}$$

By assumption, ψ_{jk} is zero unless $j = k$.

The moments γ_j and γ_{jk} are thus functions of the parameter vector $\theta' = [(\alpha_1, \ldots, \alpha_J), (\beta_1, \ldots, \beta_J), (\psi_{11}, \ldots, \psi_{JJ}), (\mu_2, \mu_3, \mu_4)]$. The most remarkable aspect of this result is that even though the Hinich model is highly nonlinear in the random variable X, only the second, third, and fourth moments about the mean (the variance, skew and kurtosis) of X appear in equations 6 and 7. When we turn to the task of estimating this model, this means that there are only a few "nuisance" parameters that get in the way of obtaining estimates of the parameter vectors $\alpha' = (\alpha_1 \ldots, \alpha_J)$, $\beta' = (\beta_1, \ldots, \beta_J)$, and $\psi' = (\psi_{11}, \ldots, \psi_{JJ})$.

The vector of observed moments D is:

$$D' = [d_1, \ldots, d_j, \ldots, d_J; d_{11}, d_{12}, \ldots, d_{jk}, \ldots, d_{J,J-1}, d_{JJ}],$$

so that the first set of moments consists of the observed means of ratings and the succeeding sets consist of the observed covariances. Parallel to these observed quantities, there is the vector $\Gamma(\theta)'$ consisting of moments fitted to the vector θ of unknown parameters:

$$\Gamma(\theta)' = [\gamma_1, \ldots, \gamma_j, \ldots, \gamma_J; \gamma_{11}, \gamma_{12}, \ldots, \gamma_{jk}, \ldots, \gamma_{J,J-1}, \gamma_{JJ}].$$

The vector of observed moments can then be fitted to the vector of fitted moments by minimizing the following with respect to θ:

$$F(\theta) = [D - \Gamma(\theta)]' (\hat{\Omega})^{-1} [D - \Gamma(\theta)], \tag{8}$$

where $\hat{\Omega}$ is a weighting matrix. If $\hat{\Omega}$ is the identity matrix, then the estimator for θ obtained by minimizing $F(\theta)$ will be consistent (see theorem 1 below), but its sampling distribution will be hard to obtain. The best choice for this weighting matrix is the sample size times the covariances of the observed moments because this allows for a simple application of theorem 2, which describes the conditions for asymptotically normal estimates with a known sampling distribution. The sample size times the covariances of the observed moments are $\omega_{j,k} = I\text{Cov}(d_j, d_k)$, $\omega_{j,kl} = I\text{Cov}(d_j, d_{kl})$, and $\omega_{jk,lm} = I\text{Cov}(d_{jk}, d_{lm})$. These elements can be arranged in Ω to conform to the order of the moments γ_j and γ_{jk} in D and $\Gamma(\theta)$. The covariances are multiplied by the sample size, I, to insure that the elements of Ω remain the same size as the

sample size increases. Because the true covariances are unknown, they must be estimated to form the matrix Ω. The elements of this matrix are denoted by $\hat{\omega}_{jk}$, $\hat{\omega}_{j,kl}$, and $\hat{\omega}_{jk,lm}$. Appendix 2 describes how these quantities can be calculated.

Minimizing equation 8 is a relatively simple task. Once the weighting matrix Ω has been computed, it only has to be inverted once, and then its Choleski decomposition, $\hat{C}'\hat{C} = \Omega^{-1}$, can be obtained. After this has been done, the residual vectors $\tau(\hat{\theta}) = \hat{C}[D - \Gamma(\hat{\theta})]$ can be formed for each potential $\hat{\theta}$ vector. With this transformation, $F(\hat{\theta})$ is a simple inner product, $\tau(\hat{\theta})'\tau(\hat{\theta})$, which can be minimized using a standard algorithm.

One of the computational problems is the inversion and reduction of Ω when J is large. This matrix has dimension $N = J + [J(J - 1)/2]$ so that when J is 20, Ω is 210 by 210. This is a large matrix, but its inversion presents no problem for even a microcomputer, because it only has to be inverted and reduced once.

Most important, as I show in section 5, this estimator generates two important types of test statistics: a χ^2 test for dimensionality computed from the final $F(\theta)$ value, and a variance-covariance matrix $E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)']$ for the estimated parameters approximated by the inverse of the Hessian of $F(\theta)/2$ evaluated at the final $\hat{\theta}$ value. In addition, within the class of those estimators based upon just means and covariances of the observed data, this estimator will provide asymptotically efficient estimates.

4. A Generalized Estimator

Problems Estimating the Poole Model

It is not too hard to see why this simple approach fails with the Poole model. Consider the basic, one-dimensional Poole model:

$$Y_j = \alpha_j - |X - \beta_j| + \delta_j,$$

and take its expectation:

$$E(Y_j) = \alpha_j - E(|X - \beta_j|).$$

The difficulty is the term $E(|X - \beta_j|)$, which depends upon the exact nature of the distribution of X as well as the value of β_j. With the Hinich model, this expectation could be written in terms of a few parameters, namely μ_1, μ_2, and β_j. There does not appear to be any simple way to do this for the Poole model.

Evaluating Complicated Expectations

However, a slightly more complicated estimator will do the trick. Consider the following generalization of the ideal point models described by equation 2:

$$Y_j = g_j(X, \alpha, \beta) + \delta_j, \tag{9}$$

where $g_j(X, \alpha, \beta)$ is some known, nonlinear function in the random vector $X = (X_1, \ldots, X_t, \ldots, X_T)$ and α and β are parameter vectors.

For this general model, the true means and covariances are:

$$\gamma_j = E[g_j(X, \alpha, \beta)] \tag{10}$$

and

$$\gamma_{jk} = E\{g_j(X, \alpha, \beta) + \delta_j - \gamma_j][g_k(X, \alpha, \beta) + \delta_k - \gamma_k]\}.$$

Or,

$$\gamma_{jk} = E[g_j(X, \alpha, \beta)g_k(X, \alpha, \beta)] - \gamma_j\gamma_k + E(\delta_j\delta_k). \tag{11}$$

The only difficulty in the calculations of γ_j and γ_{jk} is the evaluation of two expectations:

$$\xi_j = E[g_j(X, \alpha, \beta)],$$

and

$$\xi_{jk} = E[g_j(X, \alpha, \beta) g_k(X, \alpha, \beta)].$$

If we knew the density function $f(X; \rho)$ of X up to a set of known nuisance parameters $\rho' = (\rho_1, \ldots, \rho_M)$, then we could compute these expectations as:

$$\xi_j = \int g_j(X, \alpha, \beta) f(X; \rho) dX, \tag{12}$$

and

$$\xi_{jk} = \int g_j(X, \alpha, \beta) g_k(X, \alpha, \beta) f(X; \rho) dX, \tag{13}$$

where the integral is over the entire range of X. If there is only a finite number of these nuisance parameters, then ignorance about them is no great

problem. We simply include them in the θ vector and estimate them as we go along.

For example, in the Hinich model, the expectations have the simple forms:

$$\xi_j = \int \sum_{t=1}^{T} (X_t^2 - 2\beta_{tj}X_t + \beta_{tj}^2) f(X;\rho)dX,$$

and

$$\xi_{jk} = \int \left[\sum_{t=1}^{T} (X_t^2 - 2\beta_{tj}X_t + \beta_{tj}^2) \right] \left[\sum_{t=1}^{T} (X_t^2 - 2\beta_{tk}X_t + \beta_{tk}^2) \right] f(X;\rho)dX.$$

These can be evaluated in terms of first-, second-, third-, and fourth-order moments and cross-moments that constitute the nuisance parameters ρ. I dealt with the one-dimensional version of this model in section 3. In that case, there were only three nuisance parameters, $\rho = (\mu_2, \mu_3, \mu_4)$, which were simply folded into θ.

No matter what the model, once we have the fitted quantities in equations 10 and 11, we can try to estimate θ by using the estimator described in the first section that fits the estimated vector $\Gamma(\hat{\theta})$ of means and covariances to the observed vector D. This estimator is theoretically feasible as long as the means and covariances of the δ_j exist, and the density $f(X; \rho)$ and the functions $f_j(X, \alpha, \beta)$ are well-enough behaved so that the expectations in equations 12 and 13 exist.

The Density for Unobserved Variables

The major problem is choosing a form for $f(X; \rho)$ that satisfies these theoretical requirements while remaining empirically reasonable and computationally tractable. Empirical suitability depends upon one's notion of the distribution of the ideal points X. This distribution might have a variety of characteristics such as asymmetry or bimodality. This suggests the need for some fairly flexible distributions. Computational tractability depends primarily upon the dimensionality T of the common space and the number of nuisance parameters that must be estimated. The expectations in equations 12 and 13 must be calculated using numerical quadrature, which becomes difficult for integrals with more than three or four dimensions. Luckily, most studies find that the common space has one, two, or three dimensions. The nuisance parameters, whose number usually increase with more flexible distributional forms, have the disadvantage of eating up degrees of freedom and creating computational difficulties.

If the dimensionality of the integrals is greater than three or four, then there is another way of thinking about this problem that seems quite promising. In two recent papers, Ariel Pakes and David Pollard (1989) and Daniel McFadden (1989) have proposed "simulation" estimators for evaluating complicated integrals like those in equations 12 and 13. The basic idea relies upon the fact that these integrals are expectations of nonlinear functions $g_j(X, \alpha, \beta)$ with respect to a multivariate density $f(X; \rho)$, which is known up to a set of parameters ρ. One way to get an approximate value for these integrals given estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\rho}$ of the parameters is to take a random draw of X values from the multivariate density $f(X; \hat{\rho})$, calculate the quantity $\hat{Y}_{ij} = g_j(\bar{X}_i, \hat{\alpha}, \hat{\beta})$ for each "simulated" X_i denoted by $\bar{X}_i$, and then take the average and the cross-product of these simulated $\hat{Y}_{ij}$ values. By making the sample extremely large, it is possible to attain as much precision as desired, but this is computationally difficult and ultimately unnecessary.

Pakes and Pollard and McFadden have shown that, for least-squares methods, estimation can proceed with relatively small simulated samples of $\bar{X}_i$. Moreover, the additional error introduced by the simulation only has the effect of inflating the standard errors of the estimates. These authors show how the adjusted standard errors can be computed, and they suggest that, for a simulation sample about the size of the original sample, the standard errors for the parameters are usually approximately double what they would be if the integrals were calculated directly. The availability of this technique along with the ability to simulate quite complicated multivariate distributions (Johnson, 1987), suggests that our approach to nonlinear errors-in-variables models is computationally feasible for situations with many unobserved independent variables.

5. Properties of the Estimators

I have asserted that the GLS estimation method proposed in this article has reasonable statistical properties. In this section, I discuss these properties in more detail and I provide references to articles which provide the details of the proofs. Those readers not interested in these details should skip to the next section. I rely upon two theorems presented in Brady 1989a. These theorems, in turn, rest upon work by Shapiro (1983), and, ultimately, the basic papers of Chiang (1956) and Ferguson (1958).

In the following, $\hat{s} \rightarrow s$ means that $\hat{s}$ converges in probability to s and θ° refers to the true value of θ. The first theorem is:

THEOREM 1—CONSISTENCY. Consider the quadratic form:

$$Q(\theta) = [\hat{s} - s(\theta)]'\hat{\Psi}[\hat{s} - s(\theta)], \tag{14}$$

where $\hat{s}$ is a vector of observed quantities, $\hat{\Psi}$ is a vector of observed quantities, and $s(\theta)$ is a known function of θ. If the following conditions hold:

(a) Convergence of observed quantities: $\hat{s} \rightarrow s(\theta)$ and $\hat{\Psi} \rightarrow \hat{\Psi}^\circ$,

(b) Regularity: The admissible parameter space Θ is compact and the functions $s(\theta)$ are continuous in θ, and

(c) Identification: Given $\delta > 0$, there is some $\epsilon > 0$ such that:

$$\text{Inf}_{|\theta - \theta^\circ| > \delta}\ [s^\circ - s(\theta)]'\Psi^\circ[s^\circ - s(\theta)] > \epsilon. \tag{15}$$

Then:

1. For each sample size I, there is a non-empty set $\hat{\Theta}_I$ of minimizers $\hat{\theta}_I$ of $Q(\theta)$.
2. A sequence of $\hat{\theta}_I$ from the $\hat{\Theta}_1$ converges in probability to θ°.

This theorem is tailor-made for our kind of problem. The $\hat{s}$ vector corresponds to vector D with elements d_j and d_{jk}, the matrix $\hat{\Psi}$ is $\hat{\Omega}^{-1}$, $\theta = (\alpha, \beta, \psi, \rho)$ (where ρ is the vector of nuisance parameters describing the distribution of ideal points), and $s(\theta)$ is the vector $\Gamma(\theta)$ with elements γ_j and γ_{jk}.

I shall not prove it here, but under the usual sampling schemes, various laws of large numbers can be used to prove that D converges to $\Gamma(\theta)$ and that $\hat{\Omega}^{-1}$ converges to a constant $\Psi^\circ = \Omega^{-1}$ to meet condition a of the theorem. The compactness of the parameter space required in the second condition can be insured by choosing some closed and bounded set for the possible parameter values. If the bounds are set large enough so that no reasonable parameter value could exceed them, then compactness poses no real limitation.

The continuity of the γ_j and γ_{jk} is a somewhat more subtle problem because it involves the properties of the functions $g_j(X, \alpha, \beta)$ and the distribution $f(X; \rho)$ which appear in equations 12 and 13. However, even for a relatively badly behaved $g_j(X, \alpha, \beta)$, taking its expectation helps to smooth things out. For example, consider the one-dimensional Poole model with $g_j(X, \alpha, \beta) = -|X - \beta_j|$. This function is notorious for its failure to have any derivatives at $X = \beta_j$. Yet, its moments are quite smooth with respect to β_j. Consider, for example, its expectation:

$$\xi_j(\beta_j, \rho) = -\int |X - \beta_j| f(X; \rho) dX$$

$$= \beta_j - 2\beta_j F(\beta_j, \rho) + \int_{-\infty}^{\beta_j} Xf(X; \rho) dX - \int_{\beta_j}^{\infty} Xf(X; \rho) dX,$$

where $F(\beta_j, \rho)$ is the cumulative distribution function of $f(X; \rho)$ evaluated at

β_j. Clearly the regularity of ξ_j with respect to β_j and ρ depends entirely upon the properties of $f(X; \rho)$ and $F(X; \rho)$. If X has a standard normal distribution with cumulative distribution function Φ and density ϕ, then:

$$\xi_j(\beta_j; \rho) = \beta_j - 2\beta_j\Phi(\beta_j) - 2\phi(\beta_j),$$

which has derivatives of all orders for any finite value of β_j. Taking expectations seems to cure the ills of some sickly functions such as absolute values, but it is not a panacea. There must be some structure on $g_j(X, \alpha, \beta)$ and $f(X; \rho)$—although it can probably be relatively slight.

The convergence of observed quantities insures that for $\theta = \theta^\circ$, $F(\theta)$ will be zero in probability limit. The identification assumption further requires that $F(\theta)$ will not equal zero in probability limit for any other points in some neighborhood of θ°. As appendix 1 demonstrates for the Hinich ideal-point model, this assumption can be tedious to verify, but with some practice one can become knowledgeable about where the difficulties lie for these types of models. Moreover, the behavior of the minimization routine and the conditioning of the variance-covariance matrix of the parameter estimates are usually good indicators of whether or not this condition holds.

The second theorem insures asymptotic normality:

THEOREM 2—ASYMPTOTIC NORMALITY. Consider the quadratic form in equation 14 and assume that the observed quantities converge as in assumption a in theorem 1. Also assume the following:

(d) Regularity: The admissible parameter space Θ is compact; the derivatives $\partial s'/\partial\theta$ are twice continuously differentiable; and the function $Q(\theta)$ attains its minimum at an interior point of the parameter space.

(e) Invertibility of Quadratic Forms: The quadratic form in θ, $H(\theta) = (\partial\eta'/\partial\theta)\Psi^\circ(\partial\eta/\partial\theta')$, where $\eta = [\hat{s} - s(\theta)]$, is invertible at (θ°).

(f) Asymptotic Normality of Residuals: The square-root of the vector of residuals, $(\eta^\circ)^{1/2} = I^{1/2}[\hat{s} - s(\theta^\circ)]$, is asymptotically normal with mean zero and covariance matrix $(\Psi^\circ)^{-1}$.

Then:

1. $I^{1/2}[\hat{\theta} - \theta^\circ]$ is asymptotically normal with mean zero and covariance matrix $[H(\theta^\circ)]^{-1} = [L(\theta^\circ)\Psi^\circ L(\theta^\circ)']^{-1}$ where $L(\theta^\circ)$ equals $\partial\eta'/\partial\theta$ evaluated at θ°.

2. $IQ(\hat{\theta})$ is asymptotically χ^2 with $N - K$ degrees of freedom where N is the rank of Ψ° and K is the number of free parameters.

Most of these conditions are straightforward. Compactness can be handled as before. The function $Q(\theta)$ will usually attain its minimum at an

interior point of the parameter space, and this can be checked once a result has been obtained. The invertibility of the quadratic form is a stronger form of the identification condition, and it will usually hold as long as some attention is paid to the indeterminancies in the model. Moreover, as with the earlier identification condition, the behavior of the minimization routine is usually a good indicator of whether or not this condition holds. One part of condition f, the asymptotic normality of the residuals, follows from the sampling scheme and central limit theorems such as the multivariate Lindeberg-Levy theorem (Billingsley 1979, 336–37). The covariance matrix for the vector of residuals $(\eta^\circ)^{1/2}$ as defined in condition f will be I times the covariance of each element of $\hat{s}$ with each other element. Since the vector $\hat{s}$ corresponds to the vector D, $(\Psi^\circ)^{-1}$ should be Ω. Indeed, that is exactly what has been assumed above.

These results make it relatively easy to obtain the standard errors for the parameters. Consider the second derivatives of $F(\theta)/2$:

$$\frac{\partial^2 F(\theta)/2}{\partial\theta\partial\theta'} = -(D - \Gamma)'\Omega^{-1}\frac{\partial^2\Gamma}{\partial\theta\partial\theta'} + \frac{\partial\Gamma'}{\partial\theta}\Omega^{-1}\frac{\partial\Gamma}{\partial\theta'}$$

The first term will get smaller and smaller as the sample size increases so that for a fairly large sample size these second derivatives are approximately equal to $H(\theta)$ defined in condition e. The first result of theorem 2 indicates that the inverse of $H(\theta)$ evaluated at the true parameter value is the asymptotic covariance matrix for the parameters. Hence, the inverse of the second derivatives of $F(\theta)/2$ evaluated at $\hat{\theta}$ is a good approximation to the covariance matrix for the parameters.

The most problematic of these conditions is the smoothness of the derivatives of $s(\theta)$. This depends, as did the continuity of $s(\theta)$, on the form of $g_j(X, \alpha, \beta)$ and $f(X, \rho)$. Once again, the expectation operator will usually do a lot to smooth things out and insure that this condition will hold.

In summary, the researcher must:

- check whether the sampling scheme for the data insures the convergence of observed quantities and the asymptotic normality of residuals,
- check on the identification of the model by trying various combinations of parameters and by monitoring the behavior of the minimization routine, and
- check on the regularity of the model as we did for the one-dimensional Euclidian model.

Although they may pose some algebraic drudgery, these tasks will pose no real problems for most sampling schemes and most models. It is worth noting,

however, that virtually none of the standard scaling literature in political science discusses these issues.

6. Application to the Hinich Model

The Basic Equations

To use the estimator described above for a T-dimensional Hinich ideal point we must calculate γ_j and γ_{jk}. Remembering that $E(X_s) = 0$ for all s as an identification condition, some simple algebra yields for the means:

$$\gamma_j = \alpha_j - \sum_{t=1}^{T} [E(X_t^2) + \beta_{tj}^2]. \tag{16}$$

We must endure some rather tedious algebra for the covariances:

$$\gamma_{jk} = E\left[\left\{\alpha_j - \sum_{t=1}^{T} (X_t^2 - 2\beta_{tj}X_t + \beta_{tj}^2) + \delta_j - \alpha_j + \sum_{t=1}^{T} [E(X_t^2) + \beta_{tj}^2]\right\}\right.$$

$$\left. \times \left\{\alpha_k - \sum_{t=1}^{T} (X_t^2 - 2\beta_{tk}X_t + \beta_{tk}^2) + \delta_k - \alpha_k + \sum_{t=1}^{T} [E(X_t^2) + \beta_{tk}^2]\right\}\right]$$

$$= E\left\{\left[-\sum_{t=1}^{T} (X_t^2 - 2\beta_{tj}X_t) + \delta_j + \sum_{t=1}^{T} E(X_t^2)\right]\right.$$

$$\left. \times \left[-\sum_{t=1}^{T} (X_t^2 - 2\beta_{tk}X_t) + \delta_k + \sum_{t=1}^{T} E(X_t^2)\right]\right\}$$

Now using the fact that every δj is independent of every X_t, we can write this as:

$$\gamma_{jk} = E\left\{\left[\sum_{t=1}^{T} (X_t^2 - 2\beta_{tj}X_t)\right]\left[\sum_{t=1}^{T} (X_t^2 - 2\beta_{tk}X_t)\right]\right\}$$

$$- E\left[\sum_{t=1}^{T} E(X_t^2)\right]\left[\sum_{t=1}^{T} (X_t^2 - 2\beta_{tk}X_t) - \sum_{t=1}^{T} (X_t^2 - 2\beta_{tj}X_t)\right]$$

$$+ E\left[\sum_{t=1}^{T} (E(X_t^2))\right]^2 + E[\delta_j \delta_k].$$

With some more algebra and letting $E(\delta_j \delta_k) = \psi_{jk}$:

$$\gamma_{jk} = \psi_{jk} - \sum_{s=1}^{T} \sum_{t=1}^{T} E(X_s^2 X_t^2) + \sum_{s=1}^{T} E(X_s^2) \sum_{t=1}^{T} E(X_t^2)$$

$$- \sum_{s=1}^{T} \sum_{t=1}^{T} [2\beta_{sj} E(X_s X_t^2) + 2\beta_{tk} E(X_t X_s^2) - 4\beta_{tj}\beta_{sk} E(X_s X_t)]. \quad (17)$$

The terms $E(X_t^2)$ are variances, the $E(X_s^2 X_t^2)$ are kurtoses when $s = t$ and cross-kurtoses when $s \neq t$, the $E(X_s X_t^2)$ are skews when $s = t$ and cross-skews when $s \neq t$, and the $E(X_s X_t)$ are variances when $s = t$ and covariances when $s \neq t$. A solution to the model is a set of these parameters plus a set of ψ_{sj} and ψ_{jk} that satisfy equations 16 and 17.

Can equations 16 and 17 be estimated using the method described in the preceding section? The method can only be used if the regularity, convergence, and identification conditions of theorems 1 and 2 are satisfied. It is easy to prove that equations 16 and 17 have the regularity required by conditions *b* and *d* of theorems 1 and 2. The convergence conditions *a* and *f* will be satisfied for any standard survey such as the American National Election Study. This leaves the identification conditions *c* and *e*.

Identifying the Parameters

As we might expect, many of these parameters are not identified when we use just the first two moments of the observed distributions. The problem occurs when, for any solution to equations 16 and 17, there is always another solution within any arbitrarily small "neighborhood." This is called a failure of "local" identification. An equivalent version of this problem is the existence of some continuous transformation other than the identity transformation which transforms one solution into another set of parameters which is also a solution to equations 16 and 17. If such a transformation exists, it is easy to see that the identification condition, equation 15, will not hold.

Although there is no formulaic approach to insuring identification for nonlinear models like this one, one can proceed by trying out various continuous transformations and seeing whether or not they transform one solution into another solution. In appendix 1, I investigate a variety of transformations, and I impose restrictions on the parameters to limit these transformations to the identity. This goes a long way toward insuring identification, but the final test of identification occurs when we estimate the model. If the model is not identified, then the variance-covariance matrix of the parameter estimates will be singular or nearly singular.

Appendix 1 develops these identification conditions for all s and t:

- *Cross-Kurtoses A Function of Variances*—$E(X_s^2 X_t^2) = E(X_s^2)E(X_t^2)$,
- *Kurtoses A Function of Variances*—$E(X_t^4) = 3[E(X_t^2)]^2$,
- *Cross-Skews Zero*—$E(X_s X_t^2) = 0$,
- *Skews Zero*—$E(X_t^3) = 0$,
- *Covariances Zero*—$E(X_s X_t) = 0$,
- *Variances Equal to One Another*—$E(X_s^2) = E(X_t^2)$, and
- *Means Zero*—$E(X_t) = 0$.

These produce the following equation:

$$\gamma_{jk} - \psi_{jk} = 2T[E(X^2)]^2 + 4E(X^2) \sum_{s=1}^{T} \beta_{sj}\beta_{sk}. \tag{18}$$

This could have been obtained more directly had I just imposed the (very strong) requirement that the X_t have a multivariate normal distribution with zero means, zero covariances, and equal variances. For all *practical* purposes, our identification requirements amount to this assumption, but they are *theoretically* much weaker because they make no assumptions about moments beyond the fourth of the distribution of ideal points.

Do these results mean that, in principle, it is impossible to know anything about the distribution of ideal points beyond a common measure of variance? Not at all. If I wanted to use higher order moments of the observed data such as $\gamma_{jkl} = E\{[Y_j - E(Y_j)][Y_k - E(Y_k)][(Y_l - E(Y_l)]\}$, then we could probably obtain this kind of information. However, using third- and higher order empirical moments typically leads to very tricky estimation problems.

There is one final identification problem. As in standard factor analysis models, the values of β_{sk} are only unique up to a rotation. Consequently, for models with more than one dimension, the β_{sk} values must be restricted to preclude rotations. For solutions with three or fewer dimensions, the easiest way to do this is to set enough of the "diagonal" values of β_{sj} equal to zero to "pin down" the solution. For a two-dimensional solution, β_{11} is set to zero, and for a three-dimensional solution, β_{11}, β_{22}, and β_{33} are set to zero. This approach does not work for four-dimensional solutions.

In general, $T(T-1)/2$ independent conditions must be imposed on the matrix of β_{sk} values. The easiest way to do this is to set the first $T-1$ values of β_{1j} equal to zero, to set the first $T-2$ values of β_{2j} equal to zero, and so on down to set $\beta_{T-1,1}$ equal to zero. This imposes $T(T-1)/2$ independent conditions on the β_{ij}, which is just enough to "pin down" the solution.

This discussion and appendix 1 makes it clear that identification is not a trivial problem with these models. It is somewhat surprising, then, to find that

none of the standard treatments discuss the problem. One explanation is that the estimation methods used by other authors avoid it. This may be true, but it seems highly unlikely. It seems more likely that other methods either make hidden assumptions to insure identification or they estimate unidentified models. In any case, a clear discussion of the issue would be useful.

Comparison with Factor Analysis Models

One of the most interesting characteristics of equation 18 is its relationship to the standard linear factor analysis model. It is well known (Jöreskog and Sörbom 1979, 18) that the covariances for this model are:

$$\gamma_{jk} = E(X^2) \sum_{s=1}^{T} \beta_{sj}\beta_{sk} + \psi_{jk}, \qquad (19)$$

with ψ_{jk} equal to zero when $j \neq k$ and with $E(X^2)$ equal to some constant for identification. In the next section, I estimate the factor analysis model using this equation and the GLS method described above.

Of more interest at the moment, however, is the fact that this equation for γ_{jk} is nearly identical to equation 18 except for the addition of the term $2T[E(X^2)]^2$ and the estimation of $E(X^2)$ instead of setting it equal to an arbitrary constant. A moment's reflection suggests that these two differences are related, and that the addition of the term $2T[E(X^2)]^2$ makes it possible to estimate $E(X^2)$. In general, equations 18 and 19 are subcases of the following equation:

$$\gamma_{jk} = 2T\tau^2 + 4\lambda \sum_{s=1}^{T} \beta_{sj}\beta_{sk} + \psi_{jk}. \qquad (20)$$

Estimating equation 20 would seem to provide a simple test of one model versus the other, but this equation cannot be estimated directly because the transformation $\beta_{sj} = \phi\beta_{sj}^*$, $\lambda = \lambda^*/\phi^2$, and $\tau = \tau^*$ provides another solution. One approach to this identification problem is to require that $\tau = \lambda$ in general, so that we must have $\tau^* = \lambda^*$ for the asterisked solution. Because $\tau = \tau^*$ by the transformation, we must have $\lambda = \lambda^*$, which implies that $\phi = 1$. In short, it is always possible to set $\tau = \lambda$ by making suitable adjustments in the β_{sj}.

Consequently, we might as well set $\lambda = 1/4$ and estimate equation 20 with this identifying constraint. Then there are two subcases:

1. If $\tau = 0$, then we have the standard T-dimensional factor analysis model with $4\lambda = E(X^2)$ so that $E(X^2) = 1$, and

2. If $\tau \neq 0$, then the T-dimensional factor model is rejected in favor of some alternative model which might be the ideal-point model.

Thus one test for the T-dimensional linear factor model is to estimate equation 20 for T dimensions with λ fixed. If τ is significantly different from zero, then the T-dimensional linear factor model must be rejected. From this perspective, when we use just the first two moments of the empirical distribution to fit these models, the factor analysis model can be treated as just a special case of a larger set of models which includes the T-dimensional Hinich model. This test has the advantage of testing the factor analysis model against a very wide range of alternatives, but it does not provide much insight into which alternative might provide a suitable fit.

Another test does this more directly. Because the T-dimensional linear factor model is a subcase of the T-dimensional Hinich ideal-point model, we can directly compare the two models by comparing the the the χ^2 statistic for a factor analysis model estimated by GLS with the χ^2 statistic for a Hinich ideal-point model estimated in the same way. It seems remarkable that the relatively slight restriction imposed by the factor analysis model would make a large difference in the results, but, in the next section, I show that it does.

There is still another way to think of equation 20. For a $(T - 1)$ dimensional Hinich model, equation 20 with λ equal to $1/4$ is:

$$\gamma_{jk} = 2(T - 1)\tau^2 + \sum_{s=1}^{T-1} \beta_{sj}\beta_{sk} + \psi_{jk}.$$

This can be written as:

$$\gamma_{jk} = \sum_{s=1}^{T} \beta_{sj}\beta_{sk} + \psi_{jk}$$

if we require that $\beta_{Tj} = \beta_{Tk}$ for all j and k so that we can choose τ to satisfy:

$$\beta_{Tj} = [2(T - 1)\tau^2]^{1/2}.$$

This implies that for the estimation method described in this article which uses the first two moments of the observed data, the $(T - 1)$–dimensional Hinich model can be thought of as a constrained T-dimensional factor model. This should make it possible to estimate the model using LISREL (Jöreskog and Sörbom 1979). In addition, it helps to explain the standard finding that linear factor models in T dimensions provide at least as good a fit as ideal-point models in one fewer dimension.
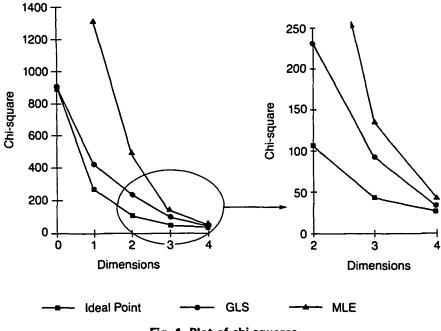
Fig. 1. Plot of chi-squares

7. Application to 1980 NES Data

The method described in this article has been tested using data from the 1980 American National Election Study, a cross-sectional survey of the U.S. electorate. The National Election Studies routinely ask people to rate political figures on 100-point feeling thermometers. In 1980, over one thousand respondents—1,015 to be exact—provided ratings for eight different political figures and the two political parties. The political figures were John Anderson, Jerry Brown, George Bush, Jimmy Carter, Gerald Ford, Edward Kennedy, Walter Mondale, and Ronald Reagan. This list includes four Democrats (Brown, Carter, Kennedy, and Mondale), three Republicans, and one Republican/Independent (Anderson). A GAUSS program was written to implement the estimation method for the Hinich model.

The Best Fitting Model

Figure 1 plots χ^2 values for the goodness-of-fit versus the number of dimensions for three different models: the Hinich ideal point model estimated by the GLS method described in this article, the linear factor analysis model esti-

mated by the same GLS method using GAUSS, and the linear factor analysis model estimated using maximum likelihood on SPSS. Four things are clear.

First, the maximum likelihood factor analysis method always yields very large χ^2 values compared to the values for the GLS linear factor analysis technique. This may be surprising until it is remembered that the maximum likelihood estimation method makes much stronger assumptions than the GLS method. Consequently, these χ^2 statistics can be interpreted to mean that maximum likelihood's additional assumptions, such as multivariate normality, are not acceptable for these data. If the linear factor analysis model is correct, then it can only be correct under some weaker assumptions.

Second, with just one additional parameter, the T-dimensional Hinich ideal-point model always yields a significantly smaller χ^2 statistic than the T-dimensional linear factor model. For example, the χ^2 difference for the four dimensional GLS factor and Hinich models is 6.12 with one degree of freedom which has a probability value of .0128. It is unlikely that this difference occurred by chance if the GLS model is correct. Another test of the adequacy of the GLS linear factor model can be obtained by estimating equation 20 with λ fixed, and testing whether τ is significantly different from zero. This yields a studentized value of 54.56, which decisively rejects the factor analysis model. These tests strongly suggest that we must reject the linear factor model in favor of some other model—perhaps the Hinich ideal-point model.

Third, the T + 1–dimensional GLS factor model always produces a better fit than the T-dimensional Hinich model. This should be true because the T-dimensional Hinich model is a special case of the T + 1–dimensional factor model. What is surprising is that the extra dimension does not help that much. For example, the four-dimensional GLS factor model has a χ^2 of 32.80 with 11 degrees of freedom compared to 42.82 for the three-dimensional Hinich model with 17 degrees of freedom. Hence the χ^2 improvement is only 9.98 for six degrees of freedom for a large probability value of .1807. If the three dimensional Hinich model is correct, then 18 percent of the time the four-dimensional linear factor model with six extra free parameters would produce up to this much improvement in the χ^2 value by sheer chance. This suggests that the Hinich ideal-point model has some substantial virtues.

Fourth, none of the models produces a statistically insignificant χ^2. From a strict hypothesis testing perspective, all of the models are rejected—even the four-dimensional Hinich model with a χ^2 of 26.7 with 10 degrees of freedom which has a probability value of .0014—but the Hinich models are not rejected with the same finality as the GLS or MLE factor analysis models. The four-dimensional factor model, for example, yields a χ^2 of 32.80 with eleven degrees of freedom which has a probability value of .00013.

Although the Hinich models do the best among the models we have considered, there must be still other models that could do even better. Yet the Hinich model does fairly well, and it would be nice to be able to go on and to

make some decision about the dimensionality of the thermometer ratings. How can we do this?

The first step is to realize that it is expecting too much to hope that a model will fit well enough to satisfy an overall χ^2 test for goodness of fit. These tests are omnibus tests against all sorts of specification errors. There are many ways that a model might fail. In the Hinich model, I have made assumptions about functional form, the number of dimensions, the diagonality of the unique factor covariance matrix, and the orthogonality of the common and unique factors. Any one of these assumptions might be false. It might be true, as McDonald and Marsh (1989, 26) have argued, "that no restrictive model fits the population, and *per impossible,* the investigator given the population covariance matrix would still have to choose a restrictive model that approximates the population well enough." Even if one rejects McDonald and Marsh's strong statement about the relationship between models and reality, it is probably best to treat the results of the χ^2 test as a sober caution against getting carried away with all aspects of a model, rather than as a rejection of the entire enterprise.

How should we proceed? Although McDonald and Marsh (1989, 30) end their discussion of "Choosing a Multivariate Model" with the observation that the conflicting needs of parsimony and goodness of approximation require an "essentially nonstatistical resolution," they still recommend using some indices of fit and parsimony. They obviously believe that the theoretical knowledge of the researcher is aided by the crutch of summary measures that can be used to compare models and methods.

Carmines and McIver (1981) have suggested using cutoff levels of two or three for the ratio of χ^2 with its degrees of freedom (the C^2 statistic). Using this criterion, a three-dimensional Hinich model yields a C^2 of 2.52 and a four-dimensional Hinich model yields a C^2 of 2.67, while the lowest C^2 for the GLS factor model is 2.98 for the four-dimensional case. Others have suggested choosing that model for which C^2 is lowest. This is the three-dimensional Hinich model.

Akaike (1987) and others (Cudek and Browne 1983) have recommended using the minimum of the Akaike Information Criterion (AIC). For the methods of this paper, the AIC is the χ^2 value plus twice the number of parameters in the model. The AIC values for the GLS factor models, starting with the one-dimensional model, are 483.18, 313.26, 188.23, and 142.80. The corresponding values for the Hinich models are 331.06, 186.74, 138.85 and 136.88. Strictly speaking, this leads to the four-dimensional Hinich solution, but the three-dimensional version certainly seems close enough. Taken together, these criteria suggest that a three-dimensional Hinich solution is quite adequate. These criteria also strongly imply that one- and two-dimensional ideal-point solutions are inadequate. The χ^2 statistics strongly reject these solutions and the AIC and many other criteria lead to the same result.

Other articles that have analyzed thermometer data (e.g., Poole and Rosenthal 1984; Rabinowitz 1978; Weisberg and Rusk 1970) have emphasized two-dimensional solutions. Weisberg and Rusk argue that nonmetric, multidimensional scaling yields "a 'good' solution in two dimensions with a stress of .050" (1970, 1174). Rabinowitz analyzes 1968 and 1972 data and reports that "Stress using Kruskal's formula 2 is .139 and .108 for these two-dimensional configurations, both of which fall in the good range (1978, 795). Unfortunately, there is no statistical basis for the stress measure, and the appelation "good" for stress measures of this size was simply made up by Joseph Kruskal (1964). Kruskal tells his reader that "since [stress] will turn out to be a 'residual sum of squares,' it is positive, and the smaller the better" (1964, 3). He then introduces the terms *poor, fair, good, excellent,* and *perfect* with the following explanation: "Our experience with experimental and synthetic data suggests the following verbal evaluation" (1964, 3). In short, there is absolutely no statistical justification for these two-dimensional solutions.[3]

Poole and Rosenthal also emphasize two-dimensional solutions, but they note that "most of the variation of the thermometers is accounted for by three dimensions" (1984, 288). The only measure of fit they provide is a "squared Pearson correlation coefficient between the actual and reproduced thermometer" (1984, 285), and they provide no justification for choosing the two-dimensional fit based upon this measure. Indeed, a glance at their table of unfolding results indicates that the third dimension typically improves their measure of fit by about 20 percent. Is this not enough improvement to suggest that a three-dimensional fit is necessary? It is hard to know for sure. In fact, an R^2 by itself is of little use in choosing the correct dimensionality. I have shown elsewhere (Brady 1989b; Brady 1990) that R^2 statistics like that used by Poole and Rosenthal can be very large when the underlying model fits the data very poorly.

The sad truth is that the scaling literature almost invariably fails to provide any tests of the dimensionality of the space, the adequacy of functional forms, or the relative merits of one model versus another. This article provides those kinds of tests.

The Pattern of Candidate Positions

Table 1 lists the values of the parameters and their standard errors for the three-dimensional ideal-point model. The identification conditions for the β_{ij} placed Carter at zero on the first and second dimensions and Reagan at zero on

---

3. Brady 1985 presents some statistical models and methods for nonmetric, multidimensional scaling.

TABLE 1. Parameter Estimates for Three-Dimensional Hinich Ideal-Point Model

Candidate	α_j	ψ_j	β_{1j}	β_{2j}	β_{3j}	Percentage Explained	Distance from Median Voter
Anderson	75.951	14.659	0.066	−3.545	0.819	54.16	3.639
	(2.27)	(0.97)	(0.54)	(0.32)	(0.27)		
Brown	58.404	16.764	1.045	−1.533	−0.200	30.47	1.866
	(1.06)	(0.49)	(0.35)	(0.27)	(0.22)		
Bush	70.952	13.322	−0.777	−0.123	2.164	48.35	2.302
	(1.01)	(0.49)	(0.36)	(0.28)	(0.20)		
Carter	95.206	16.503	0.0	0.0	5.483	64.15	5.483
	(2.36)	(0.72)	Fixed	Fixed	(0.23)		
Ford	77.045	18.779	−0.603	0.632	2.552	34.83	2.697
	(1.28)	(0.51)	(0.36)	(0.30)	(0.24)		
Kennedy	85.469	18.657	4.958	−1.188	−1.718	57.64	5.380
	(8.44)	(2.92)	(0.90)	(0.55)	(0.34)		
Mondale	73.832	13.936	−0.280	−1.185	−2.859	54.03	3.108
	(1.25)	(0.51)	(0.30)	(0.20)	(0.22)		
Reagan	95.945	13.190	0.0	3.460	4.201	72.85	5.442
	(2.17)	(0.81)	Fixed	(0.35)	(0.25)		
Democratic party	92.177	12.326	1.414	0.075	−4.517	70.96	4.734
	(1.83)	(0.68)	(0.32)	(0.26)	(0.21)		
Republican party	81.847	13.801	−0.737	2.234	2.905	58.72	3.738
	(1.44)	(0.54)	(0.26)	(0.30)	(0.24)		

Note: Standard errors in parentheses.

the second. The standard errors are all relatively small, although they are especially small for the α_j and ψ_{jk} values. Table 1 also indicates the percentage of the variance in the thermometers explained by the model. This R^2 measure is defined as $R^2 = 1 - (\psi_j^2/d_{jj})$. In all cases, the model seems to explain a substantial fraction of the variance although this measure, as noted previously, should not be taken too seriously.

Table 1 also includes the unique factor means, α_j. Ideally, we might like all of these to equal one another so that any variation in preference for the candidates would be entirely explained by the variation in β_{ij} instead of the much more nebulous values of the α_j. However, it is obvious that the α_j are substantially different from one another. Probably the best way to interpret these is as the amalgamation of some very basic trait or performance variables about which there is virtually complete agreement in the population.

Table 1 does not include the value of $E(X^2)$, or its square-root, the standard deviation of the ideal points. This standard deviation is 1.83 with a miniscule standard error of .034. This means that about 68 percent of the voters have ideal points inside a three-dimensional sphere with a radius of

1.83 drawn around the median voter, and about 95 percent have their ideal points within double this radius.

The distance from the median voter is also included in table 1. Because the identification conditions for the ideal points are similar to assuming a three-dimensional normal distribution, we can reasonably assume that the median voter is at the vector mean of this distribution, which is (0,0,0) by assumption. The distance reported in the table is the simple Euclidian distance of the candidate from the median voter. A glance at these distances indicates that none of the candidates, with the possible exception of Jerry Brown, is within a standard deviation of the median voter. Other authors (e.g., Poole and Rosenthal 1984), have noted the surprising dispersion of candidates with respect to the median voter. The convergence of candidates and parties predicted by spatial models of party competition does not seem to hold.

Jerry Brown's location near the center of the distribution might seem somewhat surprising. One possible explanation is that more respondents may guess at their candidate evaluation when they know little about the candidate. A guess would probably typically lead to a value of 50 degrees on the thermometer, with little more than random variation around the center of the distribution of voters. This would lead to an apparently central location for the candidate, but it would also lead to a value near 50 for α_j and a comparatively low R^2 value, suggesting a poor fit. All of these things are true for Jerry Brown. Unfortunately, there is no way to know whether respondents have been guessing, and the method described in this article, like all the other methods used in the literature, cannot adjust for this difficulty.
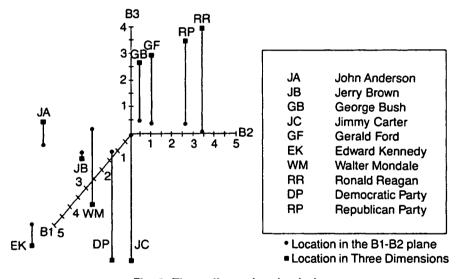

Fig. 2. Three-dimensional solution

Figure 2 displays the eight candidates, two parties, and the location of the median voter (at zero) in the three-dimensional space. Each candidate or party is represented by a vertical line with a solid circle at one end and a solid square at the other end. The circles are the two-dimensional projections of the locations in the first and second dimensions. The squares are the locations in the three-dimensional space.

There are several obvious features of this picture. First, there seems to be a strong "liberal-conservative" dimension with Ronald Reagan at one end and Ted Kennedy at the other. Second, the median voter is approximately at the center of the candidates, although some of the candidates, such as Ronald Reagan, are surprisingly far away from the median voter. Third, the parties are substantial distances from the median voter even though, with the rather interesting exception of John Anderson, both parties are close to candidates who bear their label. Fourth, the presidential candidates for 1980 (Carter and Reagan) are quite far away from the median voter.

8. Discussion

The estimation method described here is computationally feasible and statistically defensible. Unlike other methods for estimating ideal-point models, the method described in this article is statistically identified, the conditions for identification are clearly stated, and the method provides a comprehensive set of statistical tests for the dimensionality, functional form, and specific parameters of the model. For example, in the analysis of the NES data, the method makes it possible to reject the linear factor analysis model in favor of the ideal-point model and to make some statistically informed decisions about the dimensionality of the data. Other methods, such as those that rely upon the "stress" or other ad hoc measures of fit, do not allow for this.

The method does, of course, rely upon a variety of assumptions. The most important are:

- the functional form of the ideal point model,
- the independence of the unique factors δ and the ideal points X, and
- the distribution of the ideal points $f(X; \rho)$.

There is no obvious a priori reason for choosing one form of the ideal-point model over another so that the simplest model, say the Hinich form, might serve, as it has here, as a convenient starting place. If other forms seem likely, they can be considered using the framework described in this article, and goodness-of-fit statistics can be used to choose the best model. Indeed, there is nothing stopping some brave soul from including the values of the exponents ν and η in equation 2 in the parameter vector. Consequently, the functional form assumption can be tested, and it is not very restrictive.

The independence of δ and X may be the most inflexible and worrisome assumption. If independence is not assumed, then some distributional form for δ must be chosen as well, and the computational problem becomes much more severe. There are some reasons, suggested by the psychological literature, for believing that the perceptual errors in each δ_j might depend upon the distance between X and β_j. On the other hand, there may be even better reasons for believing that the major component of δ_j involves unique considerations about each candidate that do not depend upon X.

The form of $f(X; \rho)$ could be a problem if a very restrictive distribution is chosen. But there is no reason to do this. For relatively low dimensional problems, very flexible distributional forms can be estimated as long as the vector of nuisance parameters does not get too large. For higher dimensional problems, the simulation methods described here appear very promising.

The application of the method to the Hinich ideal-point model is very promising and even surprising. The results strongly suggest that the standard factor analysis method, although it provides a spatial pattern somewhat like the ideal-point model, does not fit the data at all, whereas the ideal-point model provides a strikingly good fit. From a substantive perspective, this suggests that the dimensions underlying preferences are ideal-point dimensions like liberalism-conservatism and not traits like competence, intelligence, and leadership ability.

REFERENCES

Akaike, H. 1987. "Factor Analysis and AIC." *Psychometrika* 52:317–32.

Barone, Michael, Grant Ujifusa, and Douglas Matthews. 1975. *The Almanac of American Politics 1976*. New York: E. P. Dutton.

Billingsley, Patrick. 1979. *Probability and Measure*. New York: Wiley.

Brady, Henry E. 1985. "Statistical Consistency and Hypothesis Testing for Nonmetric Multidimensional Scaling." *Psychometrika* 50: 509–37.

Brady, Henry E. 1989a. "Factor and Ideal Point Analysis for Interpersonally Incomparable Data." *Psychometrika* 54:181–202.

Brady, Henry E. 1989b. "Multidimensional Scaling in Political Science." Revision of a paper presented at the annual Political Methodology Conference, Los Angeles.

Brady, Henry E. 1990. "Statistical Properties of Metric Multidimensional Scaling." Paper presented at the annual meeting of the American Political Science Association, San Francisco.

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

Carmines, Edward G., and John P. McIver. 1981. "Analyzing Models with Unobserved Variables: Analysis of Covariance Structures." In *Social Measurement:*

Current Issues, ed. George Bohrnstedt and Edgar F. Borgatta. Beverly Hills: Sage.

Chiang, Chin Long. 1956. "On Regular Best Asymptotically Normal Estimates." *Annals of Mathematical Statistics* 27:336–51.

Cudek, R., and M. W. Browne. 1983. "Cross-Validation of Covariance Structures." *Multivariate Behavioral Research* 18:147–67.

Enelow, J., and M. Hinich. 1984. *The Spatial Theory of Voting: An Introduction.* Cambridge: Cambridge University Press.

Ferguson, Thomas S. 1958. "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities." *Annals of Mathematical Statistics* 29:1046–62.

Fuller, Wayne. 1987. *Measurement Error Models.* New York: Wiley.

Gallant, A. R. 1987. *Nonlinear Statistical Models.* New York: Wiley.

Johnson, Mark E. 1987. *Multivariate Statistical Simulation.* New York: Wiley.

Jöreskog, Karl G., and Dag Sörbom. 1979. *Advances in Factor Analysis and Structural Equation Models.* Cambridge, Mass.: Abt Books.

Kruskal, Joseph B. 1964. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika* 29:1–27.

McDonald, Roderick P., and Herbert W. Marsh. 1989. "Choosing a Multivariate Model: Noncentrality and Goodness of Fit." Paper presented at the annual meeting of the Psychometric Society, Los Angeles.

McFadden, D. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica* 57:995–1026.

McKelvey, R. 1975. "Policy Related Voting and Electoral Equilibrium." *Econometrica* 43:815–43.

Page, Benjamin I., and Calvin C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties, and the Vote." *American Political Science Review* 73:1071–89.

Pakes, Ariel, and David Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57:1027–57.

Poole, Keith T. 1981. "Dimensions of Interest Group Evaluation of the U.S. Senate, 1969–1978." *American Journal of Political Science* 25:49–67.

Poole, Keith T. 1984. "Least Squares Metric, Unidimensional Unfolding." *Psychometrika* 49:311–23.

Poole, Keith T., and Howard Rosenthal. 1984. "U.S. Presidential Elections 1968–80: A Spatial Analysis." *American Journal of Political Science* 28:282–312.

Rabinowitz, George. 1976. "A Procedure for Ordering Object Pairs Consistent with the Multidimensional Unfolding Model." *Psychometrika* 45:349–73.

Rabinowitz, George. 1978. "On the Nature of Political Issues: Insights from a Spatial Analysis." *American Journal of Political Science* 22:793–817.

Shapiro, A. 1983. "Asymptotic Distribution Theory in the Analysis of Covariance Structures (a Unified Approach)." *South African Statistical Journal* 17: 33–81.

Shepard, Roger. 1962a. "Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function I." *Psychometrika* 27:125–40.

Shepard, Roger. 1962b. "Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function II." *Psychometrika* 27:219–46.

Weisberg, H. F. 1974. "Dimensionland: An Excursion Into Spaces." *American Journal of Political Science* 18:743–76.

Weisberg, H. F., and J. G. Rusk. 1970. "Dimensions of Candidate Evaluation." *American Political Science Review* 74:1167–85.

APPENDIX 1.

Identification of the Hinich Model

Only the covariance terms present identification problems. The γ_j are consistent with *any* values for the $E(X_i^2)$ and the β_{jt}, because the α_j can always be adjusted to compensate for changes in these other parameters.

Turning to the covariance terms, I shall proceed by first considering the identification of separate terms of equation 17, and then deal with identification of the whole model. Consider the terms for kurtoses and cross-kurtoses in 17:

$$A = \sum_{s=1}^{T} \sum_{t=1}^{T} E(X_s^2 X_t^2).$$

If we set $E(X_s^2 X_t^2) = E(X_s^2 X_t^2)^* + \phi_{st}$ and if $\Sigma_s \Sigma_t \phi_{st} = 0$, then the value of A will be unchanged and $E(X_s^2 X_t^2)^*$ will be an admissible set of parameters because:

$$A = \sum_{s=1}^{T} \sum_{t=1}^{T} E(X_s^2 X_t^2)^*.$$

Thus, there is a continuous transformation of the original parameters into the "asterisked" parameters, which are also a solution. This means these parameters are not locally identified, so we must place some restrictions on them.

If the dimensions were independent of one another, then the cross-kurtoses could be written as follows:

$$E(X_s^2 X_t^2) = E(X_s^2)E(X_t^2)$$

for all $s \neq t$, so that the cross-kurtoses would be written in terms of variances. Hence, although I do not assume independence, I will assume that for $s \neq t$:

$$E(X_s^2 X_t^2) = E(X_s^2)E(X_t^2).$$

This means that:

$$A = \sum_{s=1}^{T} \sum_{t \neq s}^{T} E(X_s^2)E(X_t^2) + \sum_{t=1}^{T} E(X_t^4).$$

Another set of terms in equation 17 is:

$$B = -\sum_{s=1}^{T} \sum_{t=1}^{T} E(X_s^2)E(X_t^2)$$

$$= -\sum_{s=1}^{T} \sum_{t \neq s}^{T} E(X_s^2)E(X_t^2) - \sum_{t=1}^{T} E(X_t^2)E(X_t^2).$$

A piece of equation 17, then, is:

$$A + B = \sum_{t=1}^{T} E(X_t^4) - \sum_{t=1}^{T} [E(X_t^2)]^2.$$

The first term is still not identified because all that is required is that $\sum_{t=1}^{T} E(X_t^4)$ remain constant. This means, as before, that we can define another solution by $E(X_t^4) = E(X_t^4)^* + \phi_t$ as long as $\sum_t \phi_t$ is zero.

One way to provide some identification is to require that each $E(X_t^4)$, up to a proportionality constant δ, equal the value of the kurtosis for normally distributed dimensions. For a normal distribution, the kurtosis can be written in terms of the variances. Thus, I assume that $E(X_t^4)$ equals $3\delta[E(X_t^2)]^2$ where δ is chosen so that $3\delta\sum_{t=1}^{T}[E(X_t^2)]^2$ equals $\sum_{t=1}^{T}E(X_t^4)$. This implies that:

$$A + B = 3\delta \sum_{t=1}^{T} [E(X_t^2)]^2 - \sum_{t=1}^{T} [E(X_t^2)]^2$$

$$= (3\delta - 1) \sum_{t=1}^{T} [E(X_t^2)]^2.$$

Now consider the terms in equation 17 with skews and cross-skews:

$$C = -2 \sum_{s=1}^{T} \sum_{t=1}^{T} [\beta_{sj}E(X_s X_t^2) + \beta_{tk}E(X_t X_s^2)]$$

$$= -2 \sum_{s=1}^{T} (\beta_{sj} + \beta_{sk}) \sum_{t=1}^{T} E(X_s X_t^2).$$

For each s, all that is required for a solution is that $\sum_t E(X_s X_t^2)$ equal a constant. Consequently, these cross-skews and skews are not identified. If the X_t were independent or had a symmetric distribution, then all of the cross-skews, all $E(X_s X_t^2)$ for $s \neq t$, would be equal to zero. Hence, as an identification condition, we will assume that $E(X_s X_t^2) = 0$ for $s \neq t$. This implies that:

$$C = -2 \sum_{s=1}^{T} (\beta_{sj} + \beta_{sk}) E(X_s^3).$$

This is not quite enough, however, because there are many sets of $E(X_s^3)$ that will produce a constant value of C. Moreover, these skews only occur in this term, so they are only identified if this term is identified. Therefore, I impose the additional identification condition that $E(X_s^3) = E(X_s^3)$ for all s and t. Hence:

$$C = -2E(X^3) \sum_{s=1}^{T} (\beta_{sj} + \beta_{sk}),$$

where $E(X^3)$ is the common value of the skews.

The last major term involves covariances and variances:

$$D = 4 \sum_{s=1}^{T} \sum_{t=1}^{T} \beta_{sj} \beta_{sk} E(X_s X_t).$$

For the same reason that we usually assume uncorrelated factors in standard factor analysis, the covariances, which only appear in this term, are unidentified here. Consequently, I assumed that $E(X_s X_t) = 0$ for all $s \neq t$:

$$D = 4 \sum_{s=1}^{T} \beta_{sj} \beta_{sk} E(X_s^2)$$

Note that the expressions for the means are unaffected by all that has been done so far because the cross-kurtoses, cross-skews, skews, and covariances do not appear in that expression.

Putting all of this together, we have:

$$\gamma_{jk} - \psi_{jk} = A + B + C + D$$

$$= (3\delta - 1) \sum_{t=1}^{T} [E(X_t^2)]^2 - 2E(X^3) \sum_{s=1}^{T} (\beta_{sj} + \beta_{sk})$$

$$+ 4 \sum_{s=1}^{T} \beta_{sj} \beta_{sk} E(X_s^2). \tag{21}$$

Now consider this transformation of the parameters for all s, j, and k:

$$\delta = \frac{3\delta^* - 1 + \phi^4}{3\phi^4} \tag{22}$$

$$\beta_{sj} = \beta_{sj}^* / \phi$$

$$E(X_s^2) = \phi^2 E(X_s^2)^*$$

$$E(X^3) = \phi E(X^3)^*$$

Substituting these in equation 21, we obtain:

$$\gamma_{jk} - \psi_{jk} = (3\delta^* - 1) \sum_{t=1}^{T} [E(X_t^2)^*]^2 - 2E(X^3)^* \sum_{s=1}^{T} (\beta_{sj}^* + \beta_{sk}^*)$$

$$+ 4 \sum_{s=1}^{T} \beta_{sj}^* \beta_{sk}^* E(X_s^2)^*.$$

This is identical to equation 21 except for the asterisks.

The basic problem here is that we can use a rather complicated transformation of δ to make up for the transformation by ϕ of the variances of the ideal-point dimensions and of the locations of the candidates. A simple approach to this identification problem is to require that δ have some specific value. The obvious choice is that δ always equals one. Hence, $\delta = \delta^* = 1$. With this choice, equation 22 implies that ϕ^4 equals one.

This means that we have:

$$\gamma_{jk} - \psi_{jk} = 2 \sum_{t=1}^{T} [E(X_t^2)]^2 - 2E(X^3) \sum_{s=1}^{T} (\beta_{sj} + \beta_{sk})$$

$$+ 4 \sum_{s=1}^{T} \beta_{sj} \beta_{sk} E(X_s^2). \tag{23}$$

Consider this transformation of the parameters for all s, j, and k:

$$\beta_{sj} = \phi_s \beta_{sj}^*,$$

$$E(X_s^2) = E(X_s^2)^* / \phi_s^2,$$

$$E(X^3) = \frac{\sum_s (\phi_s^{-4} - 1)[E(X_s^2)^*]^2 + E(X^3)^* \sum_s (\beta_{sj}^* + \beta_{sk}^*)}{\sum_s \phi_s (\beta_{sj}^* + \beta_{sk}^*)}$$

It is not at all obvious, but substituting these into equation 23 yields an asterisked version of the equation. This is an excellent example of the nonobvious transformations that create identification problems in nonlinear models.

A major part of the problem is that the variances can be transformed by ϕ_s. The obvious identification condition is to require that $E(X_s^2)$ equal $E(X_t^2)$ for all s and t. Then, $E(X_s^2)^*$ must equal $E(X_t^2)^*$ as well. By the transformation of the variances, $E(X_s^2)$ must equal $E(X_s^2)^*/\phi_s^2$ so that $E(X_s^2)$ must equal $\phi_t^2 E(X_t^2)$ as well as $\delta_t^2 E(X_t^2)$. This implies that ϕ_s equals ϕ_t. I shall call their common value ϕ.

With this result, the transformation for $E(X^3)$ becomes:

$$E(X^3) = \frac{(\phi^{-4} - 1)T E(X^2)^* + E(X^3)^* \sum_s (\beta_{sj}^* + \beta_{sk}^*)}{\phi \sum_s (\beta_{sj}^* + \beta_{sk}^*)}.$$

It seems reasonable to place a restriction on $E(X^3)$ to restrict ϕ to the identity transformation. This can be done very simply by requiring zero skew, so that $E(X^3) = E(X^3)^*$ = 0. Then a little algebra indicates that ϕ^4 must equal one.

The final version of equation 17 is as follows:

$$\gamma_{jk} - \psi_{jk} = 2T[E(X^2)]^2 + 4E(X^2) \sum_{s=1}^{T} \beta_{sj}\beta_{sk}.$$

This could have been more directly obtained had I just imposed the (very strong) requirement that the X_s have a multivariate normal distribution with zero means, zero covariances, and equal variances.

APPENDIX 2.
Calculation of Covariances

We wish to calculate the covariances among the d_j, d_k, d_{jk}, and the d_{lm} in order to form Ω. In this appendix I illustrate the calculation for one case, and I present the results for the other cases. I then discuss how these quantities can be estimated to obtain $\hat{\Omega}$.

One of the elements of Ω is $\omega_{jk} = I\text{Cov}(d_j, d_k)$. We assume that each observation Y_{ij} is independently and identically distributed:

$$\omega_{jk} = I\, E[(d_j - \gamma_j)(d_k - \gamma_k)]$$

$$= I\, E\left[\left(\frac{\sum_{i=1}^{I} Y_{ij}}{I} \right) \left(\frac{\sum_{m=1}^{I} Y_{mk}}{I} \right) \right] - I\gamma_j\gamma_k$$

$$= E(Y_jY_k) + (I - 1)E(Y_j)E(Y_k) - I\gamma_j\gamma_k$$

$$= E(Y_jY_k) - \gamma_j\gamma_k.$$

where the first line is simply the definition of covariance (remembering that $E[d_j] = \gamma_j$), the second line uses equation 4 from the text, the third line uses the fact that all Y_{ij} are IID so that the subscript i can be dropped from them, and the fourth line uses the fact that $E(Y_j) = \gamma_j$.

The corresponding formulas for $\omega_{j.kl}$ and $\omega_{jk.lm}$ are:

$$\omega_{j.kl} = E(Y_jY_kY_l) - \gamma_kE(Y_jY_l) - \gamma_lE(Y_jY_k) \\ + 2\gamma_j\gamma_{kl} + \gamma_jE(Y_kY_l)$$

$$\omega_{jk.lm} = E(Y_jY_kY_lY_m) - \gamma_lE(Y_jY_kY_m) - \gamma_mE(Y_jY_kY_l) \\ - \gamma_kE(Y_jY_lY_m) - \gamma_jE(Y_kY_lY_m) + 2\gamma_l\gamma_mE(Y_jY_k) \\ + \gamma_k\gamma_lE(Y_jY_m) + \gamma_k\gamma_mE(Y_jY_l) + \gamma_j\gamma_lE(Y_kY_m) \\ + \gamma_j\gamma_mE(Y_kY_l) + 2\gamma_j\gamma_kE(Y_lY_m) - 4\gamma_j\gamma_k\gamma_l\gamma_m \\ - E(Y_jY_k)E(Y_lY_m)$$

With these theoretical expressions for the elements of Ω, the problem is to estimate them. One straightforward approach is to estimate each theoretical quantity by its method of moments estimator. Hence, we have:

$$\dot{\gamma}_j = \frac{\sum_{i=1}^{I} Y_{ij}}{I}$$

$$\widehat{E(Y_j Y_k)} = \frac{\sum_{i=1}^{I} Y_{ij} Y_{ik}}{I}$$

$$\widehat{E(Y_j Y_k Y_l)} = \frac{\sum_{i=1}^{I} Y_{ij} Y_{ik} Y_{il}}{I}$$

$$\widehat{E(Y_j Y_k Y_l Y_m)} = \frac{\sum_{i=1}^{I} Y_{ij} Y_{ik} Y_{il} Y_{im}}{I}$$

These formulas can be substituted into the preceding equations to get consistent estimates of the elements of $\hat{\Omega}$.