Political Analysis, 9:3

The Geometry of Multidimensional Quadratic Utility in Models of Parliamentary Roll Call Voting

Keith T. Poole
Department of Political Science,
University of Houston, Houston, TX 77204-3011

The purpose of this paper is to show how the geometry of the quadratic utility function in the standard spatial model of choice can be exploited to estimate a model of parliamentary roll call voting. In a standard spatial model of parliamentary roll call voting, the legislator votes for the policy outcome corresponding to Yea if her utility for Yea is greater than her utility for Nay. The voting decision of the legislator is modeled as a function of the difference between these two utilities. With quadratic utility, this difference has a simple geometric interpretation that can be exploited to estimate legislator ideal points and roll call parameters in a standard framework where the stochastic portion of the utility function is normally distributed. The geometry is almost identical to that used by Poole (2000) to develop a nonparametric unfolding of binary choice data and the algorithms developed by Poole (2000) can be easily modified to implement the standard maximum-likelihood model.

1 Introduction

THE PURPOSE OF this paper is to show how the geometry of the quadratic utility function in the standard spatial model of choice can be exploited to estimate a model of parliamentary roll call voting. The quadratic utility function has a long history. Beginning with the earliest papers of Davis and his colleagues (Davis and Hinich 1966, 1967; Davis et al. 1970), it has played an important role in the spatial theory of voting and elections. The quadratic utility function is analytically simple and has a number of mathematical properties that make it easy to work with for modeling purposes. For example, it is symmetric around, and has a unique maximum at, the individual's ideal point.

In a standard spatial model of parliamentary roll call voting, the legislator votes for the policy outcome corresponding to Yea if her utility for Yea is greater than her utility for Nay. The voting decision of the legislator is modeled as a function of the *difference* between these two utilities. The difference between two quadratic utilities has a simple geometric interpretation that can be exploited to estimate legislator ideal points and roll call parameters in a standard framework where the stochastic portion of the utility function is normally distributed. In particular, the geometry is almost identical to that used by Poole (2000) to

Author's note: I would like to thank Howard Rosenthal, Nolan McCarty, and two anonymous reviewers for their comments on various drafts of this paper. The programs and example files for implementing the QN scaling procedure can be downloaded from the *Political Analysis* web site.

Copyright 2001 by the Society for Political Methodology

develop a nonparametric unfolding of binary choice data. The algorithms developed by Poole (2000) can be easily modified to implement a standard maximum-likelihood model where the deterministic portion of the utility function of the legislators is quadratic and the stochastic portion is normally distributed.

Section 2 defines the problem and explains the notation used in the paper. Section 3 shows the geometry of the quadratic utility model and discusses the plausibility of the three major error distributions—normal, logit, and uniform—that have been used by researchers to estimate the parameters of spatial models. Section 4 briefly discusses the scaling method developed by Poole (2000). The geometry of this scaling method is essentially the same as that shown in Section 3. Section 5 shows how the algorithms developed in Poole (2000) can be used to estimate the parameters of a standard maximum-likelihood model where the deterministic portion of the utility function of the legislators is quadratic and the stochastic portion is normally distributed. Section 6 shows Monte Carlo tests of the scaling method and compares it to other scaling procedures. Section 7 discusses standard errors, and Section 8 concludes.

2 Notation and Definitions

Assume that legislators have Euclidean preferences defined over some multidimensional policy space and that they vote sincerely for the alternative closest to their ideal point. Let p be the number of legislators ($i = 1, \ldots, p$) and s be the number of dimensions ($k = 1, \ldots, s$). The ith legislator's ideal point on the kth dimension is denoted x_{ik}, and the $p \times s$ matrix of legislator ideal points is represented by $\mathbf{X}$. Each roll call vote has two policy points in the space corresponding to the policy consequences of a Yea or Nay vote on the roll call.[1] Let q be the number of roll calls ($j = 1, \ldots, q$) and coordinates for the Yea and Nay outcomes be denoted z_{jky} and z_{jkn}, respectively. Let "c" indicate the outcome (Yea or Nay) chosen by legislator i, and let b indicate the outcome not chosen by legislator i. This notation will considerably simplify the exposition below.

If there is no voting error, a plane can be placed in the space such that it separates all the legislators voting Yea from all the legislators voting Nay. Geometrically, this *cutting plane* is both perpendicular to the line joining the Yea and Nay policy points and passes through the midpoint of the Yea and Nay policy points. Because the normal vector to a plane is perpendicular to the plane, the normal vector to this cutting plane, by definition, is parallel to the line joining the Yea and Nay policy points. Specifically, let $\mathbf{n_j}$ be the $s \times 1$ normal vector for the jth roll call and let $\mathbf{N}$ be the $q \times s$ matrix of normal vectors for the q cutting planes. A plane is defined as the vector equation, $\mathbf{z}'\mathbf{n_j} = \mathbf{v}'\mathbf{n_j}$, where $\mathbf{z}$, $\mathbf{n_j}$, and $\mathbf{v}$ are $s \times 1$ vectors and the plane consists of all points $\mathbf{z}$ such that $(\mathbf{z} - \mathbf{v})$ is perpendicular to the normal vector, $\mathbf{n_j}$, and $\mathbf{v}$ is a *specific* point in the plane. Note that if $\mathbf{v_1}$ and $\mathbf{v_2}$ are both points in the plane, then $\mathbf{v_1}'\mathbf{n_j} = \mathbf{v_2}'\mathbf{n_j} = m_j$, where m_j is a scalar constant. Geometrically, every point in the plane projects onto the same point on the line defined by the normal vector $\mathbf{n_j}$ and its reflection $-\mathbf{n_j}$. Because the midpoint of the Yea and Nay policy points is on the cutting plane, it also projects to point m_j.

Technically, the general equation for a line is

$$Y(t) = A + t(B - A)$$

[1] This model was first proposed by MacRae (1958) and later developed by Poole and Rosenthal (1997) in their NOMINATE procedure.

where A and B are points in the space and t is a scalar. In this instance, A is placed at the origin of the space so that the equation for the line defined by the normal vector $\mathbf{n_j}$ and its reflection $-\mathbf{n_j}$ is simply

$$Y(t) = t\mathbf{n_j} \tag{1}$$

where $-1 \leq t \leq +1$.

To set the scale of the voting space, let the legislator coordinates lie within the s-dimensional unit hypersphere; that is, let

$$\sum_{k=1}^{s} x_{ik}^2 \leq 1, \qquad i = 1, \ldots, p$$

In addition, without loss of generality, the normal vector, $\mathbf{n_j}$, can be constrained to be of unit length; i.e., $\mathbf{n_j'n_j} = 1$.

Let the projections of the legislator points onto the line defined by Eq. (1) be

$$\mathbf{Xn_j} = \mathbf{w} \tag{2}$$

Note that the elements in the p-length vector, $\mathbf{w}$, range from -1 to $+1$. The elements in $\mathbf{w}$ all lie on the line defined by Eq. (1) that passes through the origin of the s-dimensional unit hypersphere in the direction of the normal vector with exit points $-\mathbf{n_j}$ and $+\mathbf{n_j}$ respectively. This line is hereafter referred to as the *projection line*.

3 The Quadratic Utility Model

Given these definitions, legislator i's utility for her chosen outcome, c, on roll call j is

$$U_{ijc} = u_{ijc} + \varepsilon_{ijc} = -\sum_{k=1}^{s}(x_{ik} - z_{jkc})^2 + \varepsilon_{ijc} \tag{3}$$

where u_{ijc} is the deterministic portion of the utility function and ε_{ijc} is the stochastic portion.

The probability that legislator i votes for her chosen outcome, c, is

$$P(U_{ijc} > U_{ijb}) = P(\varepsilon_{ijb} - \varepsilon_{ijc} < u_{ijc} - u_{ijb}) \tag{4}$$

3.1 *The Deterministic Portion of the Utility Function*

The difference between the deterministic utilities can be simplified as follows:

$$\begin{aligned} u_{ijc} - u_{ijb} &= -\sum_{k=1}^{s}(x_{ik} - z_{jkc})^2 + \sum_{k=1}^{s}(x_{ik} - z_{jkb})^2 \\ &= -\sum_{k=1}^{s} x_{ik}^2 + 2\sum_{k=1}^{s} x_{ik}z_{jkc} - \sum_{k=1}^{s} z_{jkc}^2 + \sum_{k=1}^{s} x_{ik}^2 \\ &\quad - 2\sum_{k=1}^{s} x_{ik}z_{jkb} + \sum_{k=1}^{s} z_{jkb}^2 \\ &= -2\sum_{k=1}^{s} x_{ik}(z_{jkb} - z_{jkc}) + \sum_{k=1}^{s}(z_{jkb} - z_{jkc})(z_{jkb} + z_{jkc}) \end{aligned} \tag{5}$$

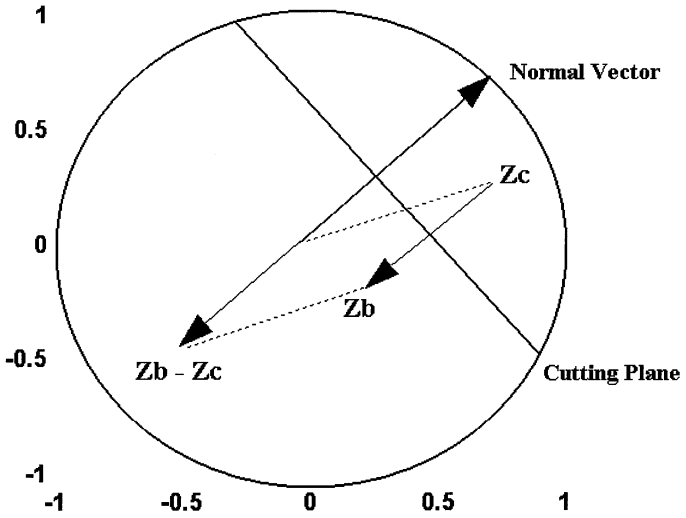

Fig. 1 Normal vector example.

Now, note that the $s \times 1$ vector,

$$\mathbf{z_{jb}} - \mathbf{z_{jc}} = \begin{bmatrix} z_{j1b} - z_{j1c} \\ z_{j2b} - z_{j2c} \\ \vdots \\ z_{jsb} - z_{jsc} \end{bmatrix}$$

is equal to a constant times the normal vector, $\mathbf{n_j}$ (see Fig. 1). Namely,

$$\gamma_{\mathbf{j}}\mathbf{n_j} = \mathbf{z_{jb}} - \mathbf{z_{jc}} \tag{6}$$

where

$$\gamma_j = -\left[\sum_{k=1}^{s}(z_{jkb} - z_{jkc})^2\right]^{\frac{1}{2}} \qquad \text{if} \qquad \mathbf{z'_{jb}n_j} > \mathbf{z'_{jc}n_j} \qquad \text{or}$$

$$\gamma_j = +\left[\sum_{k=1}^{s}(z_{jkb} - z_{jkc})^2\right]^{\frac{1}{2}} \qquad \text{if} \qquad \mathbf{z'_{jb}n_j} < \mathbf{z'_{jc}n_j}$$

$\gamma_{\mathbf{j}}$ is the *directional distance* between the Yea and the Nay outcomes in the space.

The $s \times 1$ vector

$$\mathbf{z_{jb}} + \mathbf{z_{jc}} = \begin{bmatrix} z_{j1b} + z_{j1c} \\ z_{j2b} + z_{j2c} \\ \cdot \\ \cdot \\ z_{jsb} + z_{jsc} \end{bmatrix}$$

divided by 2 is simply the $s \times 1$ vector of midpoints for the Yea and Nay outcomes for roll call j. That is,

$$\mathbf{z_{mj}} = \frac{\mathbf{z_{jb}} + \mathbf{z_{jc}}}{2}$$

This allows Eq. (5) to be rewritten as the vector equation

$$u_{ijc} - u_{ijb} = 2\gamma_j(\mathbf{x_i}'\mathbf{n_j} - \mathbf{z'_{mj}n_j}) = 2\gamma_j(w_i - m_j) \tag{7}$$

where w_i is the projection of the ith legislator's ideal point onto the projection line as defined by Eq. (1), and m_j is the projection of the midpoint of the roll call outcomes onto the projection line. Equation (7) shows that

if $\gamma_\mathbf{j} > 0$ and $w_i > m_j$, or
if $\gamma_\mathbf{j} < 0$ and $w_i < m_j$, then $u_{ijc} > u_{ijb}$.

In one dimension, $\mathbf{n_j}$ is equal to 1 and $\gamma_j = z_{jc} - z_{jb}$. Hence, Eq. (1) becomes simply $2(z_{jc} - z_{jb})(x_i - z_{mj}) = 2\gamma_j(x_i - m_j)$. Except for an added "valence" dimension, this is identical to the one-dimensional model developed by Londregan (2000, pp. 40–41). Specifically, in Londregan's notation, $g = (z_{jc} - z_{jb}.)$, $x_v = x_i$, and $m = z_{mj}$.[2] Now note that

if $z_{jc} > z_{mj}$ and $x_i > z_{mj}$, or
if $z_{jc} < z_{mj}$ and $x_i < z_{mj}$, then $u_{ijc} > u_{ijb}$.

If voting is sincere and without error, then $2\gamma_j(w_i - m_j) > 0$ for all i and j, and in one dimension, the legislator ideal points and the roll call midpoints are identified only up to a joint rank ordering. With "perfect" voting in more than one dimension, if a variety of voting coalitions forms among the legislators, then the cutting planes will intersect one another in a myriad of directions, creating a maximum of $\sum_{k=1}^{s}\binom{q}{k}$ regions in the policy space (Coombs 1964, p. 262). Each region corresponds to a unique voting pattern on the q roll calls, e.g., YYYNNYNYNYYY.... Hence, a legislator's ideal point is identified up to a region in the space. Note, however, that as q gets large the number of regions explodes so that the volume of these regions is extremely small. For example, with 500 roll calls, there are a maximum 125,251 regions in two dimensions and a maximum of 20,833,751 in three dimensions. Most of these regions are so small that a typical legislator's point is very precisely pinned down (Poole 2000).[3] Similarly, with a large number of legislators, the cutting plane–defined by the normal vector, $\mathbf{n_j}$, and the midpoint of the roll call, m_j–is also precisely pinned down (Poole 2000).

To reiterate, if voting is perfect, that is, sincere and without error, then in one dimension the legislator ideal points and the roll call midpoints are identified only up to a joint rank ordering. In more than one dimension, legislators are identified up to regions in the space (polytopes) and roll calls are identified up to cone-shaped regions containing the normal

[2] Specifically, the utility function used by Londregan is $U(z, q \mid x_v) = (-\frac{1}{2})(z - x_v)^2 + \alpha q$. The αq picks up a "valence" element of policy.

[3] Although the number of regions is very large, there are 2^q possible voting patterns, and in practical applications this number will greatly exceed the maximum number of regions in the space. Finding the region that best matches the legislator's observed pattern of roll call votes is not easy. For a solution, see Poole (2000). Measuring the volume of these regions is a problem that has not been satisfactorily solved (Best et al. 1979; Poole 2000).

vectors and line segments on the normal vectors for the midpoints. These limits on identification arise because the data are simply Yea and Nay. If legislators could report "thermometer scores" for the alternatives, then the perfect case would have an exact solution.

With quadratic utility, only the location of the cutting hyperplane and the signed distance between the Yea and the Nay alternatives can be estimated. In this framework, there is an infinite number of pairs of Yea and Nay points that produce the same hyperplane and signed distance. In contrast, an advantage to the normal distribution deterministic utility function used in NOMINATE is that it introduces enough nonlinearities in the utility differences to identify actual positions of the Yea and Nay alternatives. These points have to be used with some caution (Poole and Rosenthal 1997, Appendix A) but it is important to note that the use of the quadratic utility function precludes the estimation of *specific* Yea and Nay points.

3.2 *The Stochastic Portion of the Utility Function*

Superficially, it is the stochastic portion of the utility function that allows for more precise solutions for the legislator ideal points and roll call parameters. However, as Londregan (2000) proves, this precision is, to an extent, an illusion. Londregan shows that consistency in its usual statistical sense does not hold in the roll call voting problem outlined above. With nominal choices standard maximum-likelihood estimators that attempt *simultaneously* to recover legislators' ideal points and roll call parameters inherit the "granularity" of the choice data and so cannot recapture the underlying *continuous* parameter space. However, when the number of roll calls and legislators is large, the bias in the estimated parameters is not severe (Londregan 2000).[4]

Turning to the stochastic portion of the utility function stated in Eq. (3), three probability distributions have been used to model the error: the normal (Ladha 1991; Londregan 2000), uniform (Heckman and Snyder 1997), and logit (Poole and Rosenthal 1997). The normal is clearly the best from both a theoretical *and* a behavioral standpoint.

From a statistical standpoint, given the difference between the two random errors, $\varepsilon_{ijb} - \varepsilon_{ijc}$, the standard assumptions are that ε_{ijb} and ε_{ijc} are a *random sample* (independent and identically distributed random variables) from a known distribution. Hence, it is therefore easy to write down the probability distribution of the difference, $\varepsilon_{ijb} - \varepsilon_{ijc}$. From a behavioral standpoint, it seems sensible to assume that the distributions of ε_{ijb} and ε_{ijc} are *symmetric and unimodal* and that ε_{ijb} and ε_{ijc} are uncorrelated. The normal distribution is the only one of the three distributions to satisfy all these criteria. To illustrate, assume that ε_{ijb} and ε_{ijc} are drawn (a random sample of size two) from a normal distribution with mean zero and variance one-half. The difference between the two errors has a standard normal distribution; that is,

$$\varepsilon_{ijb} - \varepsilon_{ijc} \sim N(0, 1)$$

Hence, the probability that legislator i votes for her chosen outcome, c, can be rewritten

$$\begin{aligned} P_{ijc} &= P(U_{ijc} > U_{ijb}) = P(\varepsilon_{ijb} - \varepsilon_{ijc} < u_{ijc} - u_{ijb}) \\ &= \Phi[2\gamma_j(\mathbf{x_i'n_j} - \mathbf{z'_{mj}n_j})] = \Phi[2\gamma_j(w_i - m_j)] \end{aligned} \tag{8}$$

[4]Londregan cites the Monte Carlo work of Lord (1983) and Poole and Rosenthal (1991). Lord's Monte Carlo work was on the Rasch (1961) model used in ability tests [which is isomorphic with the one-dimensional spatial model with quadratic utility (Ladha 1991; Londregan 2000)], and Poole and Rosenthal's work was done on their NOMINATE model. In the NOMINATE model the deterministic utility function is the normal distribution and the stochastic utility is the logit distribution. Both sets of studies indicate that when the questions/legislators are at least 100, the bias is not very large. This is confirmed by Monte Carlo studies by Poole (2000, Appendix).

Heckman and Snyder (1997) assume that $\varepsilon_{ijb} - \varepsilon_{ijc}$ has a uniform distribution. This is an extremely problematic assumption because ε_{ijb} and ε_{ijc} *cannot be a random sample!*[5] Assuming that $\varepsilon_{ijb} - \varepsilon_{ijc}$ has a uniform distribution enables Heckman and Snyder to develop a linear probability model, but the price for this simplicity is that they have no intuitive basis for a behavioral model.

Poole and Rosenthal (1985, 1991, 1997) assume that ε_{ijb} and ε_{ijc} are a random sample from the log of the inverse exponential distribution. Consequently, $\varepsilon_{ijb} - \varepsilon_{ijc}$ has the logit distribution. The log of the inverse exponential and the logit distribution which is derived from it are unimodal *but not symmetric*.[6] However, they are not too skewed and the distribution function of the logit distribution is reasonably close to the normal distribution function.

A further difficulty with the approaches taken both by Heckman and Snyder and by Poole and Rosenthal is that they assume that the error variance is homoskedastic. A more realistic assumption is that the error variance varies across the roll call votes and across the legislators. For the roll calls, it is impossible to distinguish between the underlying unknown error variance and the distance between the Yea and the Nay alternatives (Ladha 1991; Londregan 2000). The intuition behind this is straightforward. As the distance between the Yea and the Nay alternatives increases, the easier it is for legislators to distinguish between the two policy outcomes and the less likely it is that they make an error. Conversely, if the Yea and Nay alternatives are very close together, then the utility difference is small and it is more likely that voting errors occur. Increasing/decreasing the distance is equivalent to decreasing/increasing the variance of the underlying error.

Because γ_j is picking up the roll call-specific variance, the difference between the two errors for legislator i on roll call j can be modeled as

$$\varepsilon_{ijb} - \varepsilon_{ijc} \sim N\left(0, \sigma_i^2\right),$$

With heteroskedastic error, Eq. (8) becomes

$$P_{ijc} = P\left(\frac{\varepsilon_{ijb} - \varepsilon_{ijc}}{\sigma_i} < \frac{u_{ijb} - u_{ijc}}{\sigma_i}\right) = \Phi\left[\frac{2\gamma_i}{\sigma_i}(w_i - m_{Ji})\right] \tag{9}$$

The corresponding log likelihood function is

$$L = \sum_{i=1}^{p}\sum_{j=1}^{q} \ln P_{ijc} \tag{10}$$

The approach developed in Section 5 allows the error to be heteroskedastic. The optimal classification algorithm developed by Poole (2000) can be used to obtain excellent estimates for the legislator points, the x_i's, the roll call normal vectors, the $\mathbf{n_j}$'s, and the cutpoints, the m_j's. With these fixed, the γ_j and σ_i can be estimated.

[5] Heckman and Snyder (1997) acknowledge that no distribution exists such that the probability distribution of the difference between two random draws has a uniform distribution. For example, if ε_{ijb} and ε_{ijc} are drawn from a uniform distribution, then the distribution of their difference will be a triangle-shaped distribution.

[6] The log of the inverse exponential is $f(\varepsilon) = e^{-\varepsilon}e^{-e^{-\varepsilon}}$, where $-\infty < \varepsilon < +\infty$. The "logit" distribution is the distribution of the difference between two random draws of the log of the inverse exponential. That is, $f(y) = e^{-y}/\left(1 + e^{-y}\right)^2$, where $y = \varepsilon_{ijb} - \varepsilon_{ijc}$ and $-\infty < y < +\infty$ [see Dhrymes 1978, pp. 340–352) for the derivation]. Note that integrating $f(y)$ from $-\infty$ to x yields the distribution function, $F(x) = 1/(1 + e^{-x})$, which, in the parliamentary roll call voting context, is $P_{ijc} = 1/(1 + e^{-(u_{ij} - u_{ijc})})$.

In sum, the normal distribution is the most sensible model of error from both a mathematical standpoint and a behavioral standpoint. Consequently, it is the focus of the model developed in this paper.

4 The Classification Algorithm

Poole (2000) develops a new scaling method for analyzing parliamentary roll call data. The scaling method uses almost exactly the same geometry as that shown for the difference between two quadratic utilities shown above. Given the legislator coordinates, the optimal classification (OC) scaling method estimates cutting planes for each roll call; and given the cutting planes, the method finds the region in the space that best matches the legislator's roll call choices. The OC scaling method is *nonparametric* because no assumptions are made about the probability distribution of the legislators' errors in making choices. The only assumptions made are that the choice space is Euclidean and that individuals making choices behave as if they utilize symmetric, single-peaked preferences.

Strictly speaking, OC is *not* a statistical model. However, as shown in the next section, the algorithms in the OC scaling method can be easily modified to implement a standard maximum-likelihood model where the deterministic portion of the utility function of the legislators is quadratic and the stochastic portion is normally distributed.

The OC algorithm uses the geometry outlined above along with the assumption that preferences are symmetric and single peaked to find estimates of $\mathbf{X}$ and $\mathbf{N}$. The rule for correct classification is as follows.

$$\begin{aligned} \text{If legislator } i \text{ votes } c\text{:}\quad & \delta_{ij} = 1 \text{ if } w_i \geq m_j \text{ and } \mathbf{z'_{jc}n_j} > m_j, \text{ or } w_i < m_j \text{ and } \mathbf{z'_{jc}n_j} < m_j. \\ & \delta_{ij} = 0 \text{ if } w_i < m_j \text{ and } \mathbf{z'_{jc}n_j} > m_j, \text{ or } w_i > m_j \text{ and } \mathbf{z'_{jc}n_j} < m_j. \end{aligned}$$

In other words, if the legislator votes Yea/Nay and her ideal point is on the Yea/Nay side of the plane, the legislator's vote is correctly classified. Note that the assumption of *symmetric* single-peaked preferences means that if a legislator votes Yea and her ideal point is *anywhere on the Yea side of the plane*, then that counts as a correct classification. If preferences are *not symmetric*, then this might not be true.

The total correct classification is therefore

$$\delta(\mathbf{X}, \mathbf{N}) = \sum_{i=1}^{p} \sum_{j=1}^{q} \delta_{ij} \tag{11}$$

In sum, given the number of dimensions, s, the classification problem consists of finding estimates of $\mathbf{X}$ and $\mathbf{N}$ that maximize Eq. (11).

5 An Algorithm to Estimate the Multidimensional Quadratic Utility Model

Note that

if $\delta_{ij} = 1$, then $\frac{2\gamma_j}{\sigma_i}(w_i - m_j) > 0$ and $\Phi[\frac{2\gamma_j}{\sigma_i}(w_i - m_j)] > 0.5$; and
if $\delta_{ij} = 0$, then $\frac{2\gamma_j}{\sigma_i}(w_i - m_j) < 0$ and $\Phi[\frac{2\gamma_j}{\sigma_i}(w_i - m_j)] < 0.5$.

In other words, the classification algorithm, which is intended to maximize Eq. (11), will also tend to maximize Eq. (10). Given estimates of $\mathbf{X}$, $\mathbf{N}$, and the m_j's from the classification algorithm—denoted $\mathbf{X}^*$, $\mathbf{N}^*$, and $m_j{}^*$, respectively—it is a simple matter to estimate the γ_j and σ_i terms because the likelihood function is convex if the roll call *cannot be classified without error*. With a finite number of legislators, there will be roll calls where $|\gamma_j|$ will

be very large because the roll call is so important or the information is so complete that no legislator makes an error. However, if the roll call can be classified without error, then $|\gamma_j| \to +\infty$; that is, the probabilities assigned to the choices of the p legislators will go to 1 on a "perfect" roll call. This does not present a problem since P_{ijc} can be set equal to 1 and its corresponding log-likelihood can be set equal to 0.

Note that the γ_j's and σ_i's are identified only up to a multiplicative positive constant. That is, for any $k > 0$, $(2\gamma_j/\sigma_i)(w_i - m_j) = (2k\gamma_j/k\sigma_i)(w_i - m_j)$. In practice this is not a problem because we are interested primarily in how noisy legislators are relative to one another and how noisy roll calls are relative to one another.

The multidimensional quadratic utility model (QN) can be efficiently estimated in four steps. Within each step the approach taken is to estimate *one parameter at a time with all the remaining parameters being held fixed*. That is, the maximum of the likelihood function is found for the parameter being estimated conditioned on all the remaining parameters being held fixed. The algorithm always moves uphill in a city block-like fashion. At convergence, every parameter is at a maximum conditioned on all the remaining parameters being held fixed. Convergence to the true *global* maximum cannot be guaranteed but the *conditional global* maximum that the algorithm reaches is almost certainly close to the overall global maximum. Although this is *not* an EM algorithm, it is similar in spirit because the log-likelihood *always increases* from one step to the next.

In the first step, the γ_j's are estimated using a simple grid search using $\mathbf{X}^*$, $\mathbf{N}^*$, and the m_j^* from the classification algorithm with the σ_i set equal to 1. With all the other parameters held fixed, the maximum log-likelihood of each of the γ_j's is easily found. Given these γ_j's and $\mathbf{X}^*$, $\mathbf{N}^*$, and the m_j^*, the σ_i's are estimated using a simple grid search. With all the other parameters held fixed, the maximum log-likelihood of each of the σ_i's is easily found. This process continues until there is no meaningful improvement in the log-likelihood. In practice, this takes no more than three repetitions.

In the second step new legislator coordinates $\mathbf{X}$ are estimated using standard gradient techniques. The $\mathbf{N}^*$ and the m_j^* from the classification algorithm and the estimated γ_j's and σ_i's from step 1 are held fixed.

In the third step new projected midpoints, the m_j, are estimated using a simple grid search. The $\mathbf{N}^*$ from the classification algorithm, the estimated γ_j's and σ_i's from step 1 and the estimated legislator coordinates, $\mathbf{X}$, from step 2 are held fixed.

In the fourth step new normal vectors, the $\mathbf{n_j}$, are estimated using standard gradient techniques with the constraints that $\mathbf{n_j}'\mathbf{n_j} = 1$ and that $\mathbf{z_{mj}} = m_j\mathbf{n_j}$. In other words, the point defined by the end of the normal vector is moved along the surface of the unit hypersphere with the position of the projected midpoint held fixed on the normal vector as it is moved. Geometrically, this is equivalent to moving the cutting plane rigidly through the space as its normal vector is moved. The estimated γ_j's and σ_i's from step 1, the estimated legislator coordinates, $\mathbf{X}$, from step 2, and the estimated projected midpoints, the m_j, from step 3 are held fixed.

In summary, the QN algorithm is as follows.

Step 1a: Estimate the γ_j.
Step 1b: Estimate the σ_i.
Step 1c: Repeat 1a and 1b until convergence.
Step 2: Estimate the x_i.
Step 3: Estimate the m_j.
Step 4: Estimate the n_j.
Go to 1a.

Table 1 Monte Carlo tests of quadratic scaling: Recovery of true legislator coordinates (100 legislators and 500 votes)

Number of dimensions	*Trials*	*Average majority margin*[a]	*Average percentage error*[a]	*True vs. quad*	
				Highest R^2[b]	*Lowest* R^2[c]
		Homoskedastic error: $\gamma_j = 1, \sigma_i = 1$			
2	10	67.8	23.1	0.982	0.979
		(0.09)	(0.24)	(0.003)	(0.003)
3	10	67.8	23.7	0.970	0.963
		(0.06)	(0.26)	(0.007)	(0.008)
		Heteroskedastic error: $\gamma_j \in (0, 2), \sigma_i = 1$			
2	10	67.8	23.5	0.983	0.979
		(0.06)	(0.41)	(0.003)	(0.004)
3	10	67.8	23.6	0.969	0.961
		(0.04)	(0.49)	(0.007)	(0.008)
		Heteroskedastic error: $\gamma_j \in (0, 2), \sigma_i \in (0, 2)$			
2	10	67.7	23.1	0.975	0.970
		(0.20)	(0.24)	(0.004)	(0.005)
3	10	67.8	24.1	0.951	0.928
		(0.04)	(0.80)	(0.005)	(0.012)

[a]Average of the 10 trials. Standard error in parentheses.
[b]Average of the highest r^2 values for the 10 trials. Standard error in parentheses.
[c]Average of the lowest r^2 values for the 10 trials. Standard error in parentheses.

In one dimension, given a joint rank ordering of the legislators and roll call midpoints from the classification algorithm, step 4 is not necessary and the legislator coordinates in step 2 can be found through a simple grid search. In practice, only three overall passes through steps 1 to 4 are required for the QN algorithm to converge.

An attractive feature of the QN model is that *if the ideal points were known*, the roll call parameters could be estimated by probit regression rather than steps 3 and 4. A probit specification is logically consistent with the assumptions of the ideal point estimator. For example, in a probit regression of a particular roll call vote on party and the estimated ideal point, the null model in which only the ideal point mattered would be correctly specified.

6 Monte Carlo Tests and Comparisons with Other Scaling Procedures

Table 1 shows some Monte Carlo tests of the recovery of the legislator coordinates by the QN scaling procedure outlined above. Artificial roll call matrices were constructed with 100 legislators and 500 roll calls with equally salient dimensions and a distribution of roll call margins that closely approximates the average distribution of roll calls in the U.S. Senate.[7] Error was introduced into the choices by making them probabilistic exactly as stated in Eq. (4) above—quadratic deterministic utility with added normal error. The random normal error was adjusted so that approximately 23.5% of the choices were mistakes. This level of error is higher than what is found in empirical work on most legislatures (Poole and

[7]This distribution is 80 roll calls (16%) between 50 and 55% majority margin, 87 roll calls (16.5%) between 56 and 60%, 80 (16%) between 61 and 65%, 75 (15%) between 66 and 70%, 50 (10%) between 71 and 75%, 45 (9%) between 76 and 80%, 25 (5%) between 81 and 85%, 25 (5%) between 86 and 90%, 20 (4%) between 91 and 95%, and 13 (2.5%) between 96 and 97.5%.

Rosenthal 2001). The tests were conducted in two and three dimensions with homoskedastic and heteroskedastic error.

The first section in Table 1 fixes the directional distances and the legislator variances equal to one. The table shows the average squared Pearson correlations between the true and the estimated legislator coordinates. The average of the worst and best r^2 values for the two or three dimensions are shown. Because the legislator configurations are defined only up to an arbitrary rotation and an origin, the QN legislator configuration was rigidly rotated using Schonemann's (1966) method to best fit the corresponding true configuration. The QN scaling procedure does a good job in recovering the true legislator configuration even with the high level of error. The average r^2 values are above 0.96 and the standard errors are quite small, indicating that the procedure is quite stable.

With heteroskedastic error the results are almost as good. The worst result is for three dimensions, where the average lowest r^2 is 0.928, but given the level of error, this is a very good result.

Table 2 shows a comparison of the two-dimensional coordinates estimated by the QN scaling procedure with those estimated by OC, Heckman-Snyder (1997), and W-NOMINATE (Poole and Rosenthal 1997) for all the post World War II Senates. The numbers in the table are Pearson r^2 values between the corresponding dimensions. The QN legislator configuration was rigidly rotated using Schonemann's (1966) method to best fit the corresponding configuration in the table.

All four methods recover essentially the same first dimension. Because the OC configuration is used to start the quadratic scaling, it is not very surprising that the r^2 values between the two are quite high. Overall, the r^2 values with the W-NOMINATE configurations are higher than the Heckman–Snyder configurations. This is almost certainly due to the fact that W-NOMINATE, OC, and QN are all designed to recover *both* legislator and roll call parameters using iterative algorithms that operate on the rectangular roll call matrix. In contrast, the Heckman–Snyder method applies factor analysis to the legislator $\times$ legislator covariance matrix computed over the roll calls to recover a legislator configuration.

There are two interesting anomalies in Table 2—the second dimension r^2 values for the 95th and 104th Senates. In the 95th Senate W-NOM, OC, and QN all recover the same second dimension, while the r^2 between QN and H-S is only 0.06. Further investigation revealed that this was due to the fact that the second dimension recovered by W-NOM, OC, and QN was the third dimension recovered by H-S, and vice versa. The first four eigenvalues of the H-S covariance matrix are 7.0301, 1.0331, 0.9262, and 0.6421, respectively. In OC the second dimension increases correct classification by 2.0% and the third dimension adds 1.5%. Similar results hold for QN and W-NOMINATE. In other words, the second and third dimensions are fairly close together in terms of increasing the fit of the four methods.

The 104th anomaly arises from the fact that the first dimension accounts for so much of the voting that there is little left for higher dimensions to account for. In OC the first dimension correctly classifies 90.0% of the votes. The second dimension increases correct classification by only 0.9%, and the third dimension by 0.8%. The first three eigenvalues of the H-S covariance matrix are 12.8663, 0.7414, and 0.6126, respectively. Beyond the liberal–conservative first dimension there just is not much left to explain.

7 Standard Errors

In a standard maximum-likelihood problem, obtaining the standard errors for the parameters is relatively easy—either the analytical Hessian matrix or the information matrix formed

Table 2 Comparison of two-dimensional quadratic coordinates with optimal classification, Heckman–Snyder, and W-NOMINATE

			Pearson r^2					
			OC		*H-S*		*WNOM*	
Senate	*Senators*	*Roll calls*	*1st*	*2nd*	*1st*	*2nd*	*1st*	*2nd*
106	102[a]	540[b]	0.963	0.883	0.915	0.814	0.935	0.789
105	100	486	0.979	0.896	0.943	0.705	0.967	0.724
104	103	806	0.981	0.925	0.957	0.033	0.967	0.086
103	101	647	0.981	0.889	0.940	0.856	0.984	0.853
102	102	482	0.987	0.926	0.943	0.865	0.977	0.930
101	101	500	0.968	0.907	0.955	0.906	0.981	0.931
100	101	636	0.967	0.911	0.938	0.825	0.964	0.891
99	101	661	0.974	0.927	0.954	0.892	0.978	0.957
98	101	578	0.972	0.957	0.960	0.934	0.979	0.967
97	101	819	0.976	0.946	0.962	0.925	0.976	0.968
96	101	928	0.978	0.944	0.966	0.929	0.978	0.960
95	104	1038	0.976	0.816	0.951	0.060	0.981	0.952
94	100	1144	0.979	0.933	0.954	0.911	0.976	0.970
93	101	987	0.966	0.942	0.951	0.890	0.972	0.955
92	102	783	0.984	0.949	0.967	0.942	0.986	0.979
91	102	557	0.985	0.960	0.965	0.934	0.983	0.973
90	101	518	0.986	0.957	0.937	0.939	0.983	0.965
89	102	441	0.976	0.924	0.934	0.903	0.963	0.919
88	102	505	0.976	0.958	0.951	0.923	0.927	0.907
87	105	400	0.964	0.911	0.881	0.880	0.954	0.922
86	103	360	0.972	0.938	0.948	0.957	0.962	0.962
85	98	255	0.980	0.948	0.935	0.934	0.955	0.952
84	99	184	0.969	0.900	0.935	0.901	0.967	0.948
83	103	242	0.955	0.877	0.920	0.807	0.946	0.901
82	96	208	0.979	0.903	0.935	0.837	0.965	0.937
81	102	447	0.968	0.923	0.954	0.929	0.944	0.958
80	97	237	0.956	0.926	0.928	0.849	0.958	0.887

[a]Number of Senators may exceed two times the number of states because of within-congress replacements.
[b]Number of roll calls with at least 2.5% voting, paired, or announced, on losing side.

from summing the outer product of the vector of first derivatives can be inverted and the square roots of the diagonal entries are the standard errors. Unfortunately, using either approach for the QN model in more than one dimension is extremely difficult because of the presence of the constraints $\left(\sum_{k=1}^{s} x_{ik}^2 \leq 1 \text{ and } \mathbf{n_j}'\mathbf{n_j} = 1\right)$ and the size of the matrix to be inverted [there are $p(s+1)+q(s+2)$ parameters]. For 100 legislators and 500 roll calls in two dimensions the matrix would be 2300×2300. In one dimension the only constraint is the range of the legislator ideal points, and the problem is easy to deal with. Londregan and Poole (2001) show analytical Hessian results for one dimension and discuss in detail various approaches for the multidimensional problem.

Given the computational difficulties in more than one dimension, an attractive alternative to obtain standard errors is a bootstrap analysis. For example, to get standard errors for the legislator parameters the q roll calls can be treated as the sample and can be drawn with

replacement. In the 90th Senate there were 518 roll calls with minority percentages of 2.5% or better. From these, 100 samples of 518 roll calls were drawn, with replacement. The QN algorithm in two dimensions was run on each of the 100 samples, and the standard deviation of the 100 estimates of the two-dimensional coordinates and variance term for each senator was computed. These are shown in Figs. 2 and 3.

Figure 2 displays histograms of the first- and second-dimension standard errors for 100 of the 101 senators who served in the 90th Senate [the omitted senator, Charles Goodell (R-NY), is discussed below]. The values along the horizontal axis represent midpoints of 0.01 ranges of the standard errors. For example, 31 senators had standard errors between 0.025 and 0.035 on the first dimension, while only 6 senators had standard errors in this range on the second dimension. For the first dimension, 93 of the 101 senators had standard errors of 0.06 or less. Given that the diameter of the space is 2 units, the estimates of the first dimension coordinates are reasonably precise. The results for the second dimension are somewhat weaker but not bad considering the fact that the second dimension is considerably less important than the first.

These standard errors may in fact overstate the imprecision of the estimates. The 100 senator configurations produced by the bootstrap trials are all quite close to the estimated configuration. The lowest r^2 between the estimated first dimension coordinates and the 100 bootstrap first dimensions was 0.97. The lowest r^2 for the second dimension was 0.90. Simply put, it would be difficult to visually distinguish a randomly chosen bootstrap trial configuration from the estimated configuration.

These standard errors are roughly on par with the level of uncertainty found by Jackman (2000) and Clinton et al. (2000) using Markov chain Monte Carlo (MCMC). Their spatial model in one dimension is identical to the one used here. In an MCMC framework, uncertainty in the legislator ideal points can be assessed by examining the posterior distribution. The geometry used to estimate the QN scaling procedure in more than one dimension

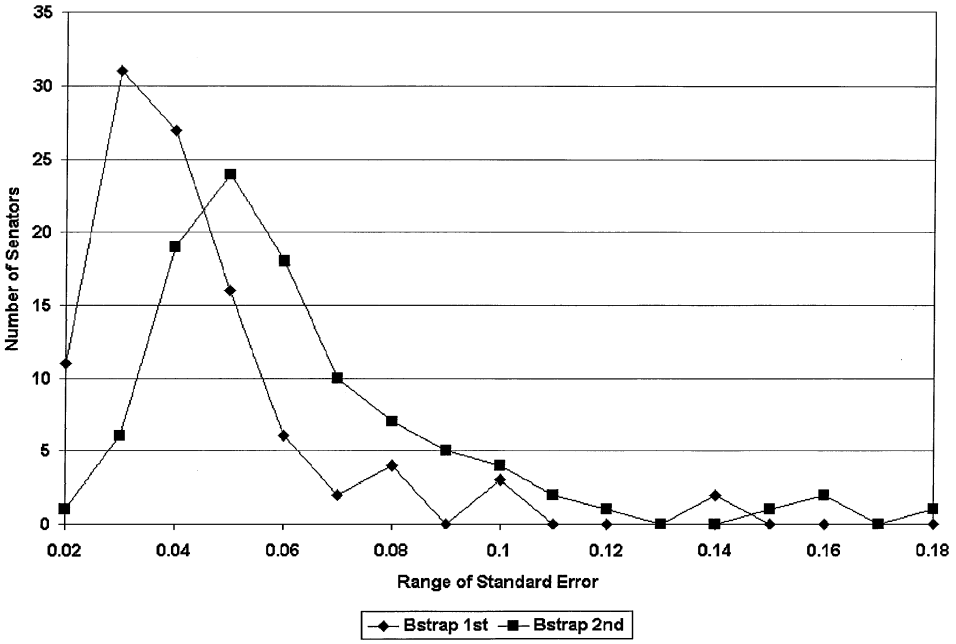

Fig. 2 Standard errors for the 90th Senate.

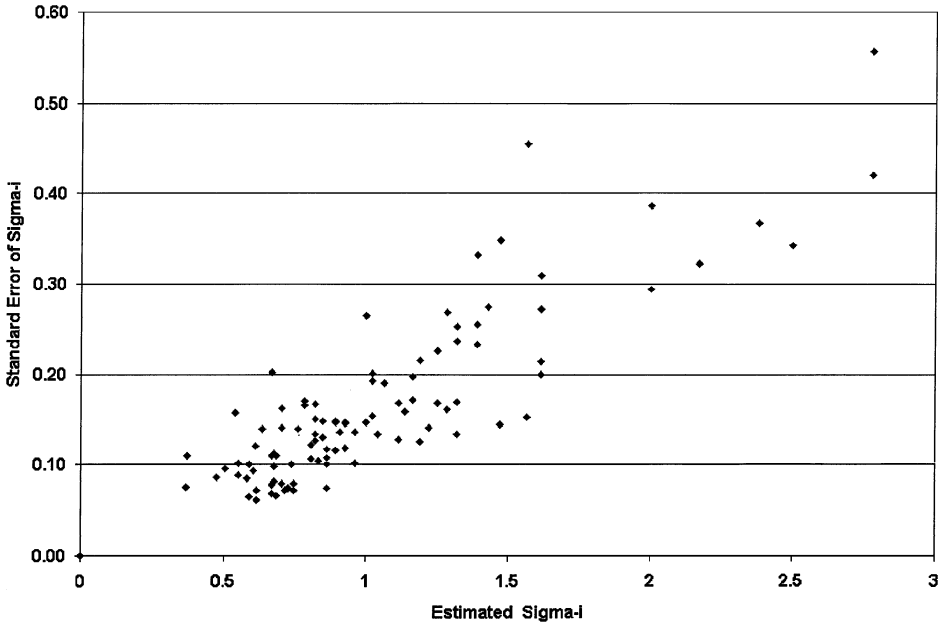

Fig. 3 σ_i vs. standard error.

could also be used to implement MCMC in higher-dimensional spaces. This would be a valuable check on both approaches.

Figure 3 shows the estimated σ_i's graphed against their corresponding bootstrapped standard errors. The bootstrapped standard error tends to increase with the magnitude of the estimated legislator variance ($r = 0.87$). That is to be expected. Relative to the estimated σ_i's, the bootstrapped standard errors are quite small. The lowest ratio of σ_i to its bootstrapped standard error is 3.36.

The zero–zero point corresponds to Charles Goodell (R-NY), who was appointed to the senate after Robert Kennedy's assassination. Goodell voted only 31 times and was perfectly classified. The results for Goodell nicely illustrate the differences among σ_i, its standard error, and the standard errors for the estimated coordinates. With perfect spatial voting legislators are defined only up to a region in the space. Because Goodell voted on only 31 roll calls, the region in the space corresponding to perfect classification on those 31 roll calls could be quite large in comparison to that of other legislators who voted hundreds of times. The bootstrapped standard errors for Goodell's first- and second-dimension coordinates were 0.274 and 0.356, respectively, by far the largest of the 101 senators.

The correlations between the bootstrapped standard error for the σ_i and the corresponding bootstrapped standard errors for the first and second dimensions were 0.252 and 0.366, respectively. In addition, the correlations between the estimated σ_i and the corresponding bootstrapped standard errors for the first and second dimensions were 0.243 and 0.360, respectively. Finally, the σ_i are not strongly related to how far a legislator is from the origin. The correlation between the σ_i and the distance of a legislator from the origin is only 0.281. In sum, the estimated σ_i are not strongly related to the standard errors of the legislator coordinates or how far a legislator is from the center of the space.

Goodell also illustrates the fact that the correlation between the σ_i and the corresponding correct classifications for the legislator must be negative. If the legislator is perfectly classified—that is, like Goodell, he makes no errors—then σ_i must be equal to zero. If

the correct classification for the legislator is 99%, then σ_i will be small. As the correct classification falls, σ_i will tend to increase. The correlation between the estimated σ_i and the percentage correct classifications for the 90th Senate is -0.84. The two do not necessarily move in lockstep, however. For the 89th Senate the correlation is -0.78, and for the 90th House it is only -0.52. Clearly the σ_i are capturing something beyond simple misclassification. This is an area for future research.

8 Conclusion

The purpose of this paper was to show a method of estimating a multidimensional spatial voting model with quadratic utility and normally distributed errors. The geometry of the multidimensional quadratic utility function can be exploited to estimate legislator ideal points and roll call cutting planes in a standard framework where the stochastic portion of the utility function is normally distributed. The QN scaling procedure is quite stable and does a good job in recovering the true legislator configuration in Monte Carlo tests.

The QN scaling procedure produces legislator configurations that are very similar to those produced by other scaling algorithms. A unique feature of QN is that it allows the variance of the error to vary across legislators. Although the legislator error variances are related to the level of correct classification, more than just correct classification is at work.

The geometry shown in Sections 2 and 3 is very general and could be the basis for other types of estimation methods. For example, the geometry can easily be extended to generalize the standard Rasch (1961) model used in educational testing to a multidimensional framework. In addition, because obtaining good-quality standard errors for the parameters of spatial models of parliamentary voting is very difficult, embedding the geometry in the innovative MCMC work of Jackman (2000) would produce an alternative method of assessing uncertainty in the multidimensional spatial model.

References

Best, Alvin M., Forrest W. Young, and Robert G. Hall. 1979. "On the Precision of a Euclidean Structure." *Psychometrika* 44:395–408.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2000. "The Statistical Analysis of Legislative Behavior: A Unified Approach." Paper presented at the Southern California Area Methodology Program, University of California, Santa Barbara, May 12–13.

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

Davis, Otto A., and Melvin J. Hinich. 1966. "A Mathematical Model of Policy Formation in a Democratic Society." *In Mathematical Applications in Political Science II*, ed. J. Bernd. Dallas, TX: Southern Methodist University Press.

Davis, Otto A., and Melvin J. Hinich. 1967. "Some Results Related to a Mathematical Model of Policy Formation in a Democratic Society." In *Mathematical Applications in Political Science III*, ed. J. Bernd. Charlottesville: University of Virginia Press.

Davis, Otto A., Melvin J. Hinich, and Peter C. Ordeshook. 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64:426–448.

Dhrymes, Phoebus J. 1978. *Introductory Econometrics*. New York: Springer-Verlag.

Heckman, James J., and James M. Snyder. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics* 28:142–189.

Jackman, Simon. 2000. "Estimation and Inference are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation." *Political Analysis* 8(4):307–332.

Ladha, Krishna K. 1991. "A Spatial Model of Legislative Voting with Perceptual Error." *Public Choice* 68:151–174.

Londregan, John B. 2000. "Estimating Legislators' Preferred Points." *Political Analysis* 8(1):35–56.

Londregan, John B., and Keith T. Poole. 2001. "Estimating Standard Errors for Spatial Models of Parliamentary Voting," Manuscript. Houston, TX: University of Houston.

Lord, F. M. 1983. "Unbiased Estimates of Ability Parameters, of Their Variance, and of Their Parallel Forms Reliability." *Psychometrika* 48:477–482.

MacRae, Duncan, Jr. 1958. *Dimensions of Congressional Voting*. Berkeley: University of California Press.
Poole, Keith T. 2000. "Non-parametric Unfolding of Binary Choice Data." *Political Analysis* 8(3):211–237.
Poole, Keith T., and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29:357–384.
Poole, Keith T., and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35:228–278.
Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
Poole, Keith T., and Howard Rosenthal. 2001. "D-NOMINATE After 10 Years: A Comparative Update to *Congress: A Political-Economic History of Roll Call Voting." Legislative Studies Quarterly* 26:5–26.
Rasch, G. 1961. "On General Laws and the Meaning of Measurement in Psychology." *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability* 4:321–333.
Schonemann, Peter H. 1966. "A Generalized Solution of the Orthogonal Procrustes Problem." *Psychometrika* 31:1–10.