Journal of Elections, Public Opinion and Parties

ISSN: 1745-7289 (Print) 1745-7297 (Online) Journal homepage: http://www.tandfonline.com/loi/fbep20

Evaluating Voter–Candidate Proximity in a Non-Euclidean Space

Min Ye , Quan Li & Kyle W. Leiker

To cite this article: Min Ye , Quan Li & Kyle W. Leiker (2011) Evaluating Voter–Candidate Proximity in a Non-Euclidean Space, Journal of Elections, Public Opinion and Parties, 21:4, 497-521, DOI: 10.1080/17457289.2011.609619

To link to this article: http://dx.doi.org/10.1080/17457289.2011.609619

Journal of Elections, Public Opinion and Parties
Vol. 21, No. 4, 497–521, November 2011

Evaluating Voter–Candidate Proximity in a Non-Euclidean Space

MIN YE*, QUAN LI** & KYLE W. LEIKER***
*Coastal Carolina University, United States; **University of Central Florida, United States; ***Kansas University Medical Center, United States

ABSTRACT *When applying the proximity model in electoral studies, scholars face the challenge of estimating voter–candidate proximity when voters' responses to issues/policies in a multidimensional policy space are correlated. In this article, we contend that voters' correlated evaluations can be captured by the structure of a non-orthogonal policy space. After orthogonalizing such a space using the Gram-Schmidt process, we can improve our estimation of the spatial distance between voters and candidates. Moreover, our study suggests that in multidimensional space neither the city-block nor the Euclidean distance is ideal for estimating proximity. We propose to use a generalized parametric Minkowski model and our analysis demonstrates that the most appropriate distance metric for a particular study is an empirical issue that hinges on the particular structure of a dataset.*

Introduction

The proximity model plays a prominent role in the spatial theory of electoral choice. The central component of the model is the idea that electoral choice can be better understood by examining the proximity between candidates and voters' preferences. In other words, voters would choose candidates whose preferred policies are "closer" to theirs. Compared with other approaches of electoral choice, the proximity model contributes to our understanding on at least two aspects. First, it directly applies the fundamental of theories on individual decision making to electoral choice: individuals (voters) make decisions (choosing candidates) according to their preferences (proximity between policy positions). Second, it presents a simple and intuitive link between formal representation of voters' issue-based information processing and their observable candidate-based electoral choice. Therefore, it is not surprising to see that since Downs' (1957) seminal study, the proximity model has swiftly grown into one of the most influential approaches in voting research.

Scholars using the proximity model usually establish a multidimensional policy space, with each dimension corresponding to a pre-selected significant campaigning

Correspondence Address: Dr. Min Ye, Department of Politics, Coastal Carolina University, Hwy 544, Conway, 29528, United States. Email: mye@coastal.edu

ISSN 1745-7289 Print/1745-7297 Online/11/040497–25 © 2011 Elections, Public Opinion & Parties
http://dx.doi.org/10.1080/17457289.2011.609619

issue (Davis & Hinich, 1966). In this Davis-Hinich space, a voter's most preferred policy bundle and her perception of candidates' issue positions are represented by points in the space, denoted as their respective ideal points. Their preference proximity is captured by the spatial distance between these ideal points: the smaller the distance, the more proximate the preferences. Regarding individual voters, preference proximity helps us explain their comparison of candidates: a voter ranks candidates in accordance with the distance from their respective ideal points. Regarding each candidate, preference proximity sheds light on different evaluation scores from voters: variation in their assessments can be explicated by their respective subjective separation from the candidate. Borrowing the terms from Westholm (1997), proximity lays the ground for both an intrapersonal comparison of candidates and an interpersonal comparison across voters.

The above discussion highlights two essential steps to build a proximity model: one from distance to preference proximity, the other from proximity to choice. If we use a utility function U_{ij} to represent voter i's evaluation of candidate j, U_{ij} is a composition of these two relations: $U_{ij} = f \circ d$, where function d captures the voter–candidate distance and f is the relationship between proximity and utility (Eguia, 2009). While the theoretical and empirical relationship between preference proximity and electoral choice has been extensively analyzed in the literature of decision-making, the link between spatial distance and preference proximity receives little attention. Although the advancement of scaling methods (e.g., Aldrich & McKelvey, 1977; Rivers, 1988) helps generate more reliable survey data about respondents' policy preferences, without fully understanding the relationship between distance and proximity, accurate estimates of ideal points do not automatically insure valid measurements of voter–candidate proximity of policy preferences. For instance, current proximity models use either a city-block or (squared) Euclidean distance to gauge the voter–candidate preference proximity. But because the theoretical implications and empirical consequences of these metrics are rarely explored, the choice is made largely at researchers' convenience. As a result, different or even conflicting inferences about voter–candidate preference proximity can be drawn from the same survey data. In this article, we try to fill the gap by giving a systemic examination of the relationship between distance and proximity. In particular, we contend that two issues are essential to gauging voter–candidate proximity of policy preferences: (1) the choice of distance metric and (2) the spatial structure of a policy space within which proximity is measured. Our study suggests that these issues not only have significant theoretical implications for voter choice but also influence the empirical merits of the proximity model.

As pointed out earlier, the proximity model relies on distance to gauge the voter–candidate preference proximity: smaller distance is associated with more proximate policy preferences. In a multidimensional space, however, there are potentially infinite notions of distance and the meaning of spatial distance is not as self-evident as we usually assume. Except for a few brief discussions (Ordeshook, 1986; Westholm, 1997; Grynaviski & Corrigan, 2006), the choice of distance metric has received scant attention. In this study, we contend that various distance metrics reflect different ways whereby voters "penalize" candidates for their policy disagreements. Focusing exclusively on the city-block or Euclidean distance actually assumes away heterogeneous rules voters may use to reach their electoral choice.[1] To resolve this problem, we argue that both the city-block and Euclidean distance are members of a family of parametric distances known as the Minkowski distances. In order to study the choice of a heterogeneous electorate, the proximity model must be open to more options of distance functions and the most appropriate "average" distance norm for the population should be the one that optimizes the model's explanatory power. In particular, we propose a generalized parametric Minkowski model for which the distance norm is treated as a parameter and the selection of the optimal distance metric is determined by model fit.

The second issue we pursue in this article regards the evaluation of proximity when voters' responses are correlated across different policy dimensions. By definition a Davis-Hinich space is orthogonal, that is, any pair of dimensions is perpendicular to one another. The immediate implication of orthogonality is that a voter's responses on any pair of policy dimensions are uncorrelated. However, when a policy space is constructed, orthogonality is not guaranteed. To researchers of public opinion, correlation in responses is unproblematic because it presents evidence for constraint and consistency in voters' preferences. But for proximity models using spatial distance to examine voter choice, non-orthogonality can cause biased assessment. Mathematically, the same distance norm can return different spatial distance depending on the degree of non-orthogonality. Theoretically, computing the spatial distance when responses are correlated means some ideological disagreement is double-counted under the guise of different issues. In this article, we propose to start with a non-orthogonal space in which issue correlation is captured by the structure of space (or the "angles" between policy dimensions), and then use the Gram-Schmidt process to orthogonalize the input data before any spatial distance is computed.[2]

Our analysis suggests a new scaling method to assess voter–candidate proximity of policy preferences: a generalized parametric Minkowski model in an orthogonalized space. On the one hand, the generalized parametric Minkowski distance allows the proximity model to consider more plausible rules voters follow in their assessment of candidates. On the other hand, the non-orthogonal space captures the real structure of the input data and the process of orthogonalization eliminates the adverse impact from the correlation in voters' responses across issues.

To demonstrate the utility of our new approach, we ran a generalized parametric Minkowski model using the ANES 2000, 2004 and 2008 presidential election data. The results show that the optimal distance metric varies from election to election and from candidate to candidate, suggesting that the choice of distance metric is indeed an empirical issue. Additionally, correlation in voters' preferences across issues sways our estimates of electoral choice. After correcting double counted preferences, the process of orthogonalization can generally enhance the explanatory power of the proximity model.

The Choice of Distance Metric

Previous studies usually use the city-block or (squared) Euclidean norm to evaluate voters' displacement from candidates. Grynaviski and Corrigan (2006) provide the following expressions for candidate utility using the city-block norm (Equation [1]) and the (squared) Euclidean norm (Equation [2]):

$$U_{ij} = \alpha - \sum_k \beta_k w_{ik} |v_{ik} - c_{ijk}| \qquad [1]$$

$$U_{ij} = \alpha - \sum_k \beta_k w_{ik} (v_{ik} - c_{ijk})^2 \qquad [2]$$

Where v_{ik} is voter i's ideal point on issue k, c_{ijk} is voter i's perceived ideal point of candidate j on the same issue, and w_{ik} is the weight voter i assigns to issue k. α and β are estimators.

Both the city-block and Euclidean norms are members of a class of parametric distances known as the Minkowski distances, or L_p distances. Expressing them in the following manner helps to illustrate this.

City-block Distance

$$\sum_k w_{ik} |v_{ik} - c_{ijk}| = \left(\sum_k w_{ik} |v_{ik} - c_{ijk}|^1 \right)^1 = L_1(ij) \qquad [3]$$

Euclidean Distance

$$\sqrt{\sum_k w_{ik} |v_{ik} - c_{ijk}|^2} = \left(\sum_k w_{ik} |v_{ik} - c_{ijk}|^2 \right)^{\frac{1}{2}} = L_2(ij) \qquad [4]$$

The city-block and Euclidean distances are defined as the L_1 and L_2 Minkowski norms, respectively, while the squared Euclidean distance is L_2^2. The generalized Minkowski distance can be defined in the general case as

$$L_p(ij) = \left(\sum_k w_{ik} |v_{ik} - c_{ijk}|^p \right)^{\frac{1}{p}} \qquad [5]$$

where p ($p \geq 1$) is the parameter of the distance.[3]

In a multidimensional space, depending on which norm is used, a voter could be nearer or further from a given candidate. The city-block distance tends to penalize small disagreements more than the Euclidean distance does. However, because the penalty grows linearly with distance, it does not impose severe penalties for significant disagreements in any one dimension. If one believes that voters seek to maximize overall agreement with a candidate across a broad swathe of issues, then the city-block metric is a good choice. By contrast, the Euclidean distance minimizes the impact of small disagreements across a range of issues. But the Euclidean distance – particularly when squared – imposes a large penalty for substantial disagreement on a single dimension (Grynaviski & Corrigan, 2006). Although the squared Euclidean distance will always produce the same preference ordering as the Euclidean distance, when used in models that have both additive and multiplicative terms (e.g. OLS), using the squared Euclidean distance gives yet greater weight to disagreements in single dimensions. Thus, it imposes a larger penalty than the Euclidean distance. If one accepts the logic behind using the squared Euclidean distance then, in effect, one assumes that voters choose candidates with the goal of minimizing their maximum disagreement (analogous to approval voting). In that case, the Chebyshev metric, a special case of the Minkowski distance as $p \to \infty$, is a more appropriate choice. It simply picks the largest distance between two points in any single dimension:

$$L_\infty(ij) = \left(\sum_k w_{ik}|v_{ik} - c_{ijk}|^\infty\right)^{\frac{1}{\infty}} = \max(w_{ik}|v_{ik} - c_{ijk}|) \qquad [6]$$

The interpretation of the weighted Chebyshev distance in a policy space is that voters vote against the candidate with whom they have the most important single disagreement. In the multiparty case, this implies a negative rank-ordering, with the least offensive candidate being selected.

Therefore, in a multidimensional space, the choice of distance metric $L_p(ij)$ is a choice of the parameter p, reflecting different ways by which voters evaluate candidates. To illustrate the influence of different values of p, consider a two-dimensional policy space as an example. Without loss of generality, let the origin O be a voter's ideal point. In the space, this voter would feel indifferent to candidates whose ideal points are equally distant from O. For any given distance, say q, these ideal points compose a closed contour line, an indifference contour of O. As illustrated in Figure 1, the size and shape of the indifference contour may alter dramatically with different values of parameter p.[4] Note that when p equals 1 and 2 we have the usual city-block and Euclidean distance, respectively. As the value of p approaches infinity, the distance is closer to the Chebyshev distance (Equation [6]). Because a voter always prefers candidates within than outside an indifference contour, Figure 1 vividly demonstrates how the choice of distance metric affects voters' evaluation of candidates.[5] More significantly, different distance metric may also lead to reversed preference order of the same voter. This can be illustrated with a simple example. When the distance is equal to q, Figure 2 presents O's indifference contours based on the city-block distance (contour 1) and Euclidean distance (contour 2), respectively. With the city-block distance we would expect O prefers

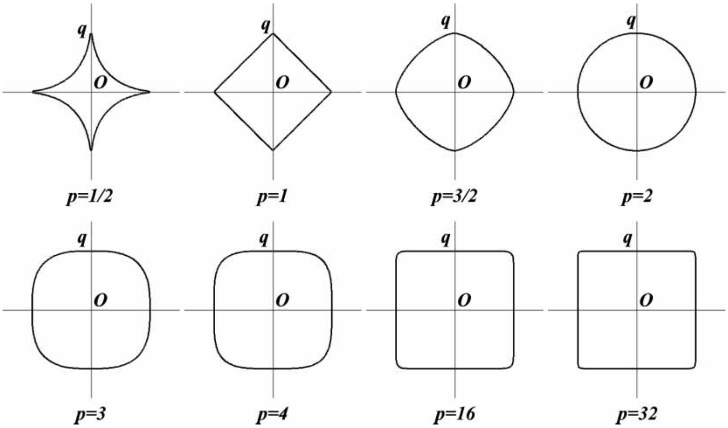

Figure 1. p values and indifference contours.

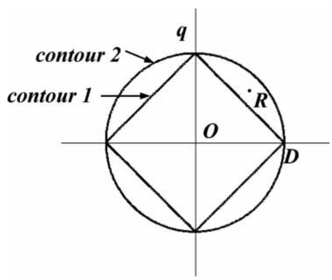

Figure 2. Voter choice with city-block and Euclidean distance.

candidate D to R (because R is located outside contour 1 whereas D is exactly on the contour). Nevertheless, if we use the Euclidean distance, we would predict the same voter choosing R over D because R is "nearer" to O according to contour 2.

If the choice of distance metric seriously affects voters' assessment and electoral choice, how can we determine the appropriate distance metric in our research? It should be noted that unless all voters use identical distance metric, what the proximity model estimates is the "average" distance norm that captures the aggregated perceptions of the population of voters.[6] Therefore, as long as no evidence can be found to justify voters are using homogeneous utility functions to rate candidates, there is no convincing reason for the proximity model to predetermine a distance metric. The selection of distance metric, we believe, should be an empirical question. In particular, we propose a p-norm generalized parametric Minkowski model as specified in Equation [5]. In this model, the distance metric is not predetermined, but a parameter to be decided by input data. When a multidimensional policy space is used, the selection of distance metric boils down to choosing the optimal value of distance parameter p. Our model is open to potentially infinitely many distance norms (including the city-block and Euclidean distance). Because we compare these distance norms with regard to their empirical merits, our model would generate the best prediction of voter choice and candidate evaluations according to issue-based assessments. Since the empirical performance has become the principal ground of the ongoing debate about the proximity model (Lewis & King, 1999), for both proponents and opponents of the proximity model, our model offers a better ground to evaluate the empirical quality of the proximity model.

From Non-orthogonal to Orthogonal Space

So far we have discussed distance metrics in a Euclidean space, where by definition the dimensions are mutually perpendicular. If voters' responses on the policy dimensions in a model were actually uncorrelated with each other, Equation [5] could be used directly to compute the p-norm generalized Minkowski distance between voters and candidates. Unfortunately, such a perfect situation is nearly nonexistent in actual input datasets. What we have to deal with in applied research is always a non-orthogonal input space. The hazard of overlooking non-orthogonality can be illustrated using a simple example. Suppose there is a two-dimensional policy space, as can be seen from Figure 3.

Without loss of generality, let the origin (O) represent a voter's ideal point, and Q (1, 1) her perceived ideal point of a candidate. Figure 3(a) portrays a traditional Euclidean space. It is easy to calculate that the Euclidean distance between O and Q is $||OQ|| = \sqrt{2}$. By contrast, if this voter's responses on dimensions x and y are correlated, the policy space becomes non-orthogonal, as demonstrated in Figure 3(b), where the correlation is reflected into the angle (θ) between the two dimensions. Geometric analysis shows that $||OQ|| = \frac{1}{\cos\frac{\theta}{2}}$, which could be either greater or smaller than $\sqrt{2}$, depending on the value of θ (note that in Figure 3(b), $||OQ|| \to \sqrt{2}$ as $\theta \to 90°$).

Our solution is to orthogonalize the policy space before any distance is calculated. For a two-dimensional policy space as illustrated in Figure 3(b), we simply project the voter's evaluation on dimension y onto a new dimension y' that is perpendicular to x, as pictured in Figure 3(c). In the orthogonalized space, while the x-coordinate of Q

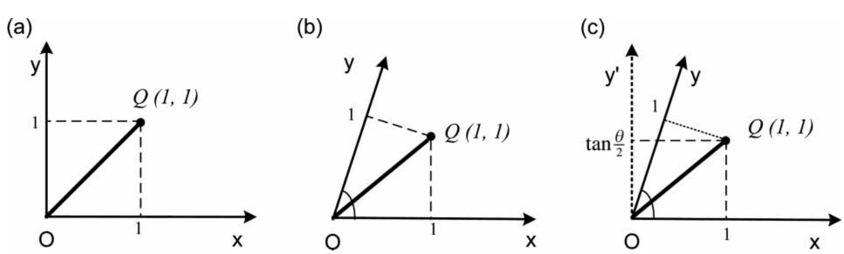

Figure 3. Two-dimensional policy spaces.

remains the same, the *y'*-coordinate of *Q*, after eliminating those double counted responses, now reduces to $\tan\frac{\theta}{2}$. In this orthogonalized $x - O - y'$ space, we can estimate the *p*-norm Minkowski model.

While a two-dimensional space can be easily orthogonalized, orthogonalizing a high-dimensional space could be extremely complicated. In the rest of this section, we present a method that can orthogonalize any *d*-dimensional space. The core of our method is to use the Gram-Schmidt process to subtract out the portion of a dimension that is correlated with another dimension previously entered into the model.

Generally speaking, in a *d*-dimensional input space, if we orthogonalize a set of linearly dependent vectors (x_1, x_2, x_3,. . .,x_d) into a set of mutually uncorrelated (or perpendicular) vectors (z_1, z_2, z_3,. . .,z_d), the Gram-Schmidt process is defined by

$$z_k = x_k - \sum_{j=1}^{k-1} proj_{z_j} x_k,\ (k \le d) \tag{7}$$

In particular, for any two vectors **x** and **z**, the projection operator $proj_z x$ is defined as $proj_z x = \frac{|x| \cdot |z| \cos \theta_{x,z}}{|z|^2} \cdot z$, where the norms $|x|$ and $|z|$ are the Euclidean lengths (L_2) of the two vectors, **x** and **z,** and the angle between the vectors is $\theta_{x,z}$.

To illustrate the application of this process, we use a five-dimensional space as an example. The Gram-Schmidt process for such a space is described by the following system of equations:

$$\begin{aligned} z_1 &= x_1 \\ z_2 &= x_2 - proj_{z_1}x_2 \\ z_3 &= x_3 - proj_{z_1}x_3 - proj_{z_2}x_3 \\ z_4 &= x_4 - proj_{z_1}x_4 - proj_{z_2}x_4 - proj_{z_3}x_4 \\ z_5 &= x_5 - proj_{z_1}x_5 - proj_{z_2}x_5 - proj_{z_3}x_5 - proj_{z_4}x_5 \end{aligned} \tag{8}$$

Now consider a unit non-orthogonal input space in five dimensions. We still assume the origin (0, 0, 0, 0, 0) represents the ideal point of a voter. If we use the set of unit vectors $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5)$ to denote the class of the dimensions, then their norms are 1, or $|\hat{x}_1| = |\hat{x}_2| = |\hat{x}_3| = |\hat{x}_4| = |\hat{x}_5| = 1$. Applying the Gram-Schmidt process in Equation [8] to $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5)$ yields their corresponding dimensions $(\hat{z}_1, \hat{z}_2, \hat{z}_3, \hat{z}_4, \hat{z}_5)$ in an orthogonalized space. The norms of these orthogonalized dimensions, that is, $(|\hat{z}_1|, |\hat{z}_2|, |\hat{z}_3|, |\hat{z}_4|, |\hat{z}_5|)$, constitute a translation vector between this non-orthogonal space and its orthogonalized space.

In the non-orthogonal input space, let $\theta_{\hat{x}_i,\hat{x}_j}$ denote the angle between dimensions $\hat{x}_i, \hat{x}_j$. Because cosine of the angle between two dimensions (vectors) is simply the correlation between them,[7] we have

$$\cos\theta_{s,t} = \frac{\sum s_i t_i}{|s|\cdot|t|} = \frac{\sum s_i t_i}{\sqrt{\sum (s_i)^2}\sqrt{\sum (t_i)^2}} = r_{s,t} \quad [9]$$

This finding allows us to cancel the cosine out of the Gram-Schmidt process, considerably simplifying the calculation. The projection operator in Equation [7] can be re-written in a linear equation:

$$proj_z x = \frac{|x|\cdot|z| r_{x,z}}{|z|^2}\cdot z \quad [10]$$

Without loss of generality, let $\hat{x}_1 = (1,0,0,0,0)^T$ and call it the benchmark dimension. Subsequent dimensions are defined as $\hat{x}_2 = (a_2, b_2, c_2, d_2, e_2)^T, \ldots, \hat{x}_5 = (a_5, b_5, c_5, d_5, e_5)^T$. Because these vectors are unit vectors, for each vector $\hat{x}_i (i = 2, \ldots, 5)$, we have $a_i^2 + b_i^2 + c_i^2 + d_i^2 + e_i^2 = 1$. Also, because $\hat{x}_1 = (1,0,0,0,0)^T$, we have $a_i = \cos\theta_{\hat{x}_i,\hat{x}_1} = r_{\hat{x}_i,\hat{x}_1}$. With the Gram-Schmidt process, it is easy to see that $\hat{z}_1 = \hat{x}_1 = (1,0,0,0,0)^T$ and $|\hat{z}_1| = |\hat{x}_1| = 1$; $\hat{z}_2 = (0, b_2, c_2, d_2, e_2)^T$ and $|\hat{z}_2| = \sqrt{1 - r^2_{\hat{x}_1,\hat{x}_2}}$.[8] The process to calculate $|\hat{z}_3|$ to $|\hat{z}_5|$ is rather involved and the details are provided in the Appendix.[9] A key feature of this algorithm is that the norms of $\hat{z}$'s can be expressed as a function of $\cos\theta_{\hat{x}_i,\hat{x}_j}$. In other words, as long as we know the structure of an input space (i.e. angles between dimensions), we can always calculate a unique translation vector of this space. Recall that for an input dataset, the angle between two dimensions can be computed using Equation [9].

With the translation vector, we can orthogonalize input data. Still use the five-dimensional space as the example. Suppose in an input space voter *i*'s ideal point is $O = (v_{i1}, v_{i2}, v_{i3}, v_{i4}, v_{i5})^T$ and her perceived ideal point of candidate *j* is $Q = (c_{ij1}, c_{ij2}, c_{ij3}, c_{ij4}, c_{ij5})^T$. To orthogonalize the input data, simply post-multiply them by their corresponding translation vector. If we use $\hat{O} = (\hat{v}_{i1}, \hat{v}_{i2}, \hat{v}_{i3}, \hat{v}_{i4}, \hat{v}_{i5})^T$ and $\hat{Q} = (\hat{c}_{ij1}, \hat{c}_{ij2}, \hat{c}_{ij3}, \hat{c}_{ij4}, \hat{c}_{ij5})^T$ to denote the ideal points of voter *i* and candidate *j* in the orthogonalized space, we have

$$\begin{aligned} \hat{O} &= (v_{i1}\cdot|\hat{z}_{i1}|, \ldots, v_{i5}\cdot|\hat{z}_{i5}|)^T, \\ \hat{Q} &= (c_{ij1}\cdot|\hat{z}_{ij1}|, \ldots, c_{ij5}\cdot|\hat{z}_{ij5}|)^T \end{aligned} \quad [11]$$

Where $(|\hat{z}_{i1}|, |\hat{z}_{i2}|, |\hat{z}_{i3}|, |\hat{z}_{i4}|, |\hat{z}_{i5}|)$ and $(|\hat{z}_{ij1}|, |\hat{z}_{ij2}|, |\hat{z}_{ij3}|, |\hat{z}_{ij4}|, |\hat{z}_{ij5}|)$ are the translation vectors for *O* and *Q*, respectively.

After the input policy space is orthogonalized, we can then test a generalized parametric Minkowski model as proposed in the previous section, and a search for the most appropriate metric distance can follow suit. Specifically, we define voter *i*'s utility for candidate j in a k-dimensional policy space as follows:

$$U_{ij} = \alpha - \beta\Big(\sum\nolimits_k w_{ik}|\hat{v}_{ik} - \hat{c}_{ijk}|^p\Big)^{\frac{1}{p}} \qquad [12]$$

where w_{ik} is the weight voter i assigns to issue k, $\hat{v}_{ik}$ and $\hat{c}_{ijk}$ are the preferred position of voter i and candidate j on dimension k in an orthogonalized space as defined in Equation [11], α and β are coefficients to be estimated, and p is the distance parameter. Compared to traditional proximity models of candidate evaluation, our model has several improvements. First, it is not completely specified. Rather than predetermining a particular distance norm (i.e. assigning a particular value to p), we treat the distance norm as a parameter, a continuous variable whose optimal value is contingent on particular input datasets. Second, this model takes into consideration voters' correlated responses on different policy issues. By definition, $\hat{v}_{ik}$ and $\hat{c}_{ijk}$ are the orthogonalized stands of voter (i) and candidate (j) on the kth dimension. Their un-orthogonalized counterpart is v_{ik} and c_{ijk} in the primitive input data. Therefore, as long as the correlations between dimensions are known, we can always carry out the Gram-Schmidt process and calculate any p-norm Minkowski distance free of any non-orthogonality concern. Finally, it should be noted that while some proximity models test the additive impact of multiple distances along individual policy dimensions (e.g. Macdonald et al., 1998; Grynaviski & Corrigan, 2006), our model tests the distance in a real spatial context. Consider the utility functions specified in Equations [1] and [2]. The models have k coefficients (β_k) to evaluate the impact of distance in k individual dimensions. Each policy area constitutes its own single dimensional space, and the voter–candidate proximity is the sum of a series of k disjoint subsets of a multidimensional policy space. Although this additive model allows us to examine the empirical influence of each issue, it has at least two major shortcomings. Conceptually it only captures the piecewise disagreements on particular issues rather than the effect of the actual distance between a candidate and a voter in that multidimensional policy space. Our model, by contrast, uses one coefficient (β) and estimates the spatial distances between voters and candidates in the space. More significantly, except for the city-block distance, the piecewise approach does not apply to any other p-norm distance. Note that Equation [2] estimates the *squared* Euclidean distance rather than the Euclidean distance itself. Using the term of the generalized Minkowski distance, the additive proximity model estimates only $L_p^p(ij)$ rather than $L_p(ij)$.[10] By way of contrast, our model has only one distance term and therefore applies to any p-norm distance.

How would the process of orthogonalization affect the empirical performance of the proximity model? Recall that in Equation [11] we orthogonalize voter and (perceived) candidate policy positions using their respective translation vectors ($|\hat{z}_{ik}|$ and $|\hat{z}_{ijk}|$ for issue k). When these translation vectors are the same (i.e. $|\hat{z}_{ik}| = |\hat{z}_{ijk}|$ for all k), we have $|\hat{v}_{ik} - \hat{c}_{ijk}| = |v_{ik} - c_{ijk}| \cdot |\hat{z}_{ik}|$ in Equation [12]. This means orthogonalization is equivalent to rescaling the voter–candidate distance on each dimension.

In this situation, there will be no significant change to the proximity's empirical performance. The only significant consequence occurs to models that study the particular impact of individual policy dimensions. The relative importance of policy issues would be different: an issue will become less or more significant after orthogonalization. In contrast, when the two translations vectors are different,[11] the process of orthogonalization transforms $|v_{ik} - c_{ijk}|$ into $|\hat{v}_{ik} - \hat{c}_{ijk}|$ non-linearly. For a voter, her spatial distance from a candidate in the orthogonalized space could be either greater or smaller than in the primitive input space. As a result, after we use orthogonalization to get rid of the impact from correlated responses, the actual explanatory power of the proximity model could be either greater or smaller.

Empirical Analysis

To demonstrate the utility of a generalized parametric Minkowski model in an orthogonalized space, we analyzed the ANES data of the 2000, 2004 and 2008 presidential elections. The dependent variable in our analysis is respondents' overall evaluation of the two major party candidates on a 101-point feeling thermometer scale. To estimate voter–candidate proximity of policy preferences, we established a policy space with five dimensions: (1) *Government Services* (the trade-offs between government services and spending), (2) *Defense Spending* (whether defense spending should be increased or decreased), (3) *Living Standard* (whether government should see to it that every person has a job and a good standard of living),[12] (4) *Job and Environment* (should the government protect the environment even if it costs jobs), (5) *Aid to Blacks* (whether government should improve the status of blacks). On each dimension, respondents are asked to report their own positions (v_i) and the positions of two candidates (c_{ij}) on a seven-point scale.[13] For each dimension, the respondent is also required to evaluate the importance of the issue on five-point scale (from extremely important to not important at all).

We first checked the correlation between voters' responses on these five policy dimensions. The correlation coefficients are listed in Table 1, where we report the correlation of voters' self-reported ideal points (row 1), their perceived position of Democratic Party (row 2) and Republican Party (row 3) presidential candidates on these dimensions. Table 1 demonstrates that voters' responses of these five policy dimensions are correlated to varying extents. If we select *Government Services* as the benchmark dimension and rank other dimensions in the sequence listed in the table, we can use the method introduced above to calculate the translation vectors ($|\hat{z}_1|, |\hat{z}_2|, |\hat{z}_3|, |\hat{z}_4|, |\hat{z}_5|$), which are listed at the bottom in Table 1.

After orthogonalizing the data using Equation [11], we applied the generalized parametric Minkowski model specified in Equation [12] and searched for the optimal distance metric for each candidate. Note that when a p-norm distance between a voter and a candidate is calculated, the candidate's mean perceived position is used.[14] In particular, our search started with $p = 1$ and increased to $p = 10$ with increments of 0.01. Therefore, for each candidate, we ran totally 901 different regressions with different distance metrics. The results are summarized in Figure 4,

Table 1. Correlation coefficients and translation vectors of the 2000, 2004 and 2008 presidential elections

	(2) Defense Spending			(3) Living Standard [a]			(4) Job & Environment			(5) Aid to Blacks		
	2000	2004	2008	2000	2004	2008	2000	2004	2008	2000	2004	2008
(1) Government Services												
Self-placement	−0.21	−0.23	−0.01	−0.39	−0.50	−0.04	−0.19	−0.28	−0.10	−0.28	−0.42	−0.08
Democrat[b]	−0.16	−0.13	−0.20	−0.34	−0.43	0.07	−0.31	−0.04	−0.07	−0.36	−0.30	−0.07
Republican[c]	−0.33	−0.01	−0.19	−0.37	−0.40	0.01	−0.25	−0.11	−0.15	−0.36	−0.30	−0.06
(2) Defense Spending												
Self-placement				0.27	0.29	0.08	0.08	0.24	0.06	0.28	0.27	0.10
Democrat				0.19	0.22	0.04	0.13	0.09	0.10	0.31	0.29	0.12
Republican				0.25	0.18	0.04	0.21	0.08	0.17	0.24	0.14	0.06
(3) Living Standard												
Self-placement							0.04	0.21	0.33	0.42	0.54	0.32
Democrat							0.27	0.10	0.34	0.49	0.51	0.21
Republican							0.24	0.11	0.31	0.34	0.49	0.30
(4) Job & Environment												
Self-placement										0.14	0.21	0.34
Democrat										0.35	0.05	0.27
Republican										0.21	0.12	0.27
Translation Vectors ($\lvert\hat{z}_2\rvert$, $\lvert\hat{z}_3\rvert$, $\lvert\hat{z}_4\rvert$, $\lvert\hat{z}_5\rvert$) [d]												

	(2) Defense Spending			(3) Living Standard			(4) Job & Environment			(5) Aid to Blacks		
	2000	2004	2008	2000	2004	2008	2000	2004	2008	2000	2004	2008
Self-placement	0.9783	0.9728	1	0.9008	0.8477	0.996	0.9793	0.9418	0.9375	0.8796	0.8174	0.9106
Democrat	0.9866	0.9909	0.9804	0.9281	0.8881	0.9956	0.9333	0.9928	0.9343	0.8067	0.8381	0.9486
Republican	0.9434	0.9999	0.9814	0.919	0.8987	0.9991	0.9488	0.9889	0.9306	0.8976	0.8603	0.9346

Notes:
[a] In the 2008 election, we use the issue *Health Insurance* (whether there should be a government insurance plan that covers all medical and hospital expenses for every one) due to data availability.
[b] The three Democrat presidential candidates are Al Gore, John Kerry and Barack Obama in 2000, 2004 and 2008 presidential elections.
[c] The three Republican presidential candidates are George Bush, George Bush and John McCain in 2000, 2004 and 2008 presidential elections.
[d] Recall that $|\hat{z}_1|$ is always equal to 1.

where the R^2 of the 901 weighted p-norm generalized Minkowski models ($1 \leq p \leq 10$) of each candidate/election are plotted.

Previous studies report that the city-block distance consistently outperforms the (squared) Euclidean distance in various contexts (Westholm, 1997; Grynaviski & Corrigan, 2006). Nevertheless, our analysis demonstrates that the comparison between the city-block and Euclidean distance is misleading. If voters may use different rules to assess and choose candidates, the optimal system-level distance metric could vary across elections. The choice of distance norm is an empirical issue and is largely determined by the features of particular elections. As illustrated in the figure, if we use R^2 of the proximity model as the standard to determine the value of p, the optimal distance norm should be the one that returns the largest R^2. The optimal distance norm and the corresponding R^2 are marked in Figure 4.[15] Our results confirm that there is no universally optimal distance norm. The optimal distance to gauge system-level preference proximity does vary from election to election. For instance, in the most recent 2008 presidential election, the optimal distance is at $p = 3.94$ for Obama and at $p = 2.17$ for McCain.[16]

Using Westholm's (1997) term, Figure 4 offers an interpersonal comparison of voters' evaluation scores of the same candidate. Another way to check the quality

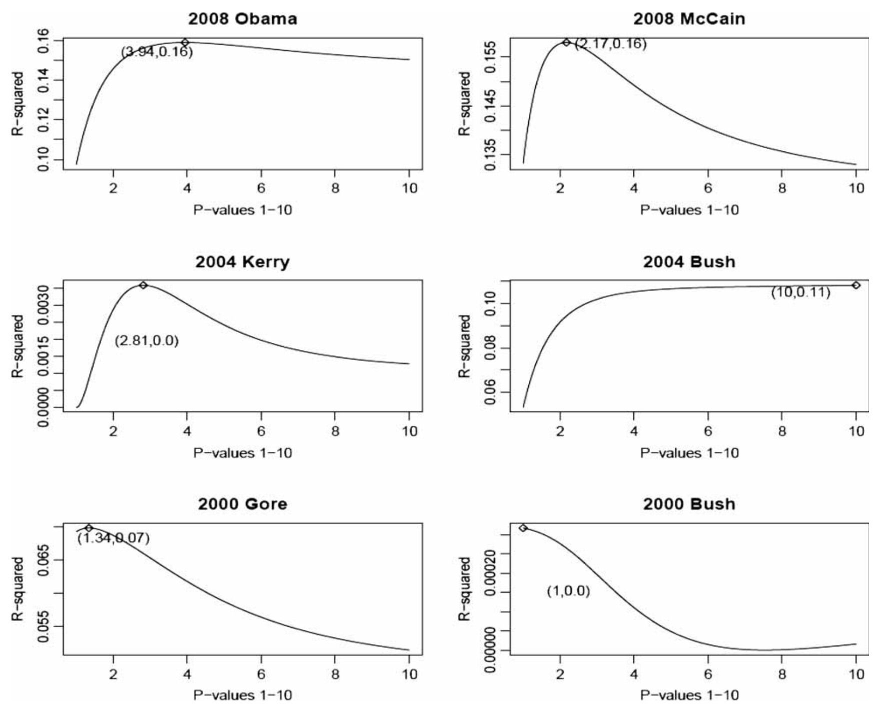

Figure 4. Generalized Minkowski model with optimal distance norm (with orthogonalized data).

of our Minkowski p-norm distance model is what he called an intrapersonal comparison of voters, that is, the model's ability to explain voters' choice between candidates. Because the proximity model postulates that a voter prefers the candidate whose ideal point is closer to her own, we can compare the predictions based on the voter–candidate spatial distance (using different distance norms) with the voter's self-reported evaluation scores (on the 101-point feeling thermometer scale). The outcome is displayed in Table 2, where the percentage of correct predictions with different distance norms is presented. In particular, we report the predictions using the city-block, Euclidean, and our optimal p-norm. As we can see, the optimal p-norm generally outperforms the city-block and Euclidean distance in correctly predicting voters' choice between two candidates in these elections. To isolate the impact of orthogonalization, we ran the same analysis using original non-orthogonalized data and the outcome is displayed in the lower part of the table. Again, the results show that orthogonalization in general improves the predictive power of our models using different distance metrics.

The regression models for each candidate with their respective optimal distance norm are presented in Table 3. Generally speaking, the proximity model of issue voting offers a valid explanation for voter assessment of candidates. From the intercepts in Table 3, we can see that in each election both candidates were viewed favorably by voters,[17] and the winner does fare better in terms of evaluation scores.[18] Moreover, our proximity coefficient, an estimate of the effect of the overall policy disagreement on voters' evaluation of candidates, indicates how voters penalize candidates for their policy differences. Using the most recent 2008 election for example, our analysis shows that voters were likely to punish Obama more severely for policy differences. For one thing, the absolute value of the distance coefficients is much greater for Obama (-13.61) than for McCain (-9.99). Moreover, the optimal distance norm for Obama ($p = 3.94$) is greater than that of McCain ($p = 2.17$).

To check the effect of orthogonalization, we first ran the generalized parametric Minkowski model (Equation [12]) using non-orthogonalized data of these elections. The outcome, listed in Table 4 and Figure 5, shows that the process of

Table 2. Predictions of proximity models with different distance norms

	City-block	Euclidean	Optimal p-norm
Correct predictions (with orthogonalized data)			
2000	56%	58%	66%
2004	67%	73%	72%
2008	73%	74%	75%
Correct predictions (with non-orthogonalized data)			
2000	50%	51%	59%
2004	67%	71%	73%
2008	70%	73%	73%

Table 3. Proximity models with the optimal distance norm (with orthogonalized data)

	Democratic Party candidate			Republican Party candidate		
	2000 Gore	*2004 Kerry*	*2008 Obama*	*2000 Bush*	*2004 Bush*	*2008 McCain*
Optimal *p*-norm	1.34	2.81	3.94	1	10	2.17
Proximity	−6.78	−3.89	−13.61	−0.74	−18.32	−9.99
	(1.03)	(2.92)	(0.99)	(1.74)	(2.37)	(0.73)
Intercept	82.39	76.74	95.76	72.26	117.91	80.89
R^2	0.07	0.004	0.16	0.0003	0.11	0.16
AIC	6043.824	5882.547	9763.334	6876.145	5616.901	9493.605
BIC	6056.887	5895.154	9778.06	6889.208	5629.509	9508.331

Table 4. Proximity models with the optimal distance norm (with non-orthogonalized data)

	Democratic Party candidate			Republican Party candidate		
	2000 Gore	*2004 Kerry*	*2008 Obama*	*2000 Bush*	*2004 Bush*	*2008 McCain*
p-norm	1.34	2.81	3.94	1	10	2.17
Proximity	−5.65	−3.44	−13.32	−0.74	−18.96	−8.60
	(1.01)	(2.51)	(0.93)	(1.66)	(2.70)	(0.82)
Intercept	78.77	76.55	96.48	72.54	115.10	77.35
R^2	0.04	0.004	0.17	0.0003	0.09	0.10
AIC	6059.036	5882.440	9750.225	6876.127	5626.208	9561.386
BIC	6072.099	5895.047	9764.952	6889.19	5638.815	9576.112

orthogonalization influences the Minkowski model by changing both the selection of the optimal distance metric and the empirical performance of the model. In the former case, comparing Figures 4 and 5, we can see that the optimal distance norm of four out of the six candidates was changed.[19] In terms of model fit, after correcting for the double-counted disagreement, the process of orthogonalization increases three proximity models' performance to explain voters' assessment (after orthogonalization, the R^2 of the model of McCain increases from 0.10 to 0.16, of Obama from 0.04 to 0.07, of Bush (2004) from 0.09 to 0.11). The AIC/BIC of these three models also decreases. As we discussed above, a very important fact about the Gram-Schmidt process is that the consequences of orthogonalization do not depend on the magnitude of correlation. Instead, it is influenced by the variation between the two sets of translation vectors, one for the respondents' self-placements ($|\hat{z}_{i1}|$, $|\hat{z}_{i2}|$, $|\hat{z}_{i3}|$, $|\hat{z}_{i4}|$, $|\hat{z}_{i5}|$) and the other for the perceived positions of candidates ($|\hat{z}_{ij1}|$, $|\hat{z}_{ij2}|$, $|\hat{z}_{ij3}|$, $|\hat{z}_{ij4}|$, $|\hat{z}_{ij5}|$). Observing the translation vectors listed in Table 1, we can see that the less salient distinction of some proximity models using orthogonal and non-orthogonalized data is largely due to the lack of variation between the two sets of translation vectors, which in most

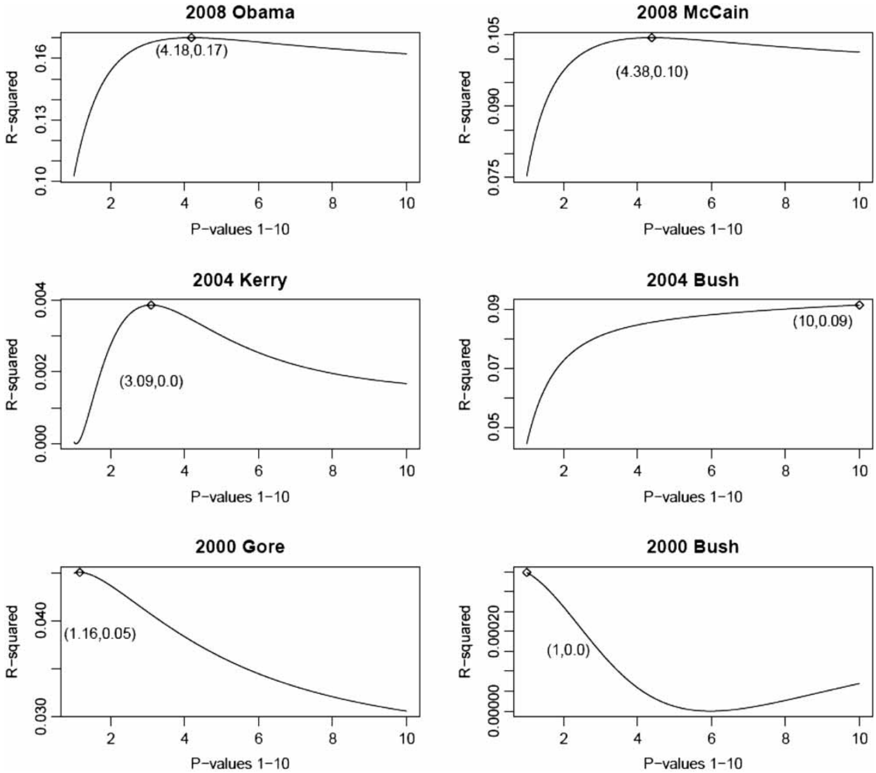

Figure 5. Generalized Minkowski model with optimal distance norm (with non-orthogonalized data).

cases is smaller than a thousandth decimal point.[20] Nevertheless, this does not mean that we are free of the concern when it comes to non-orthogonality in data. On the contrary, since we have no control of the structure of the input dataset, it is always necessary to check the correlations in responses and orthogonalize the data before examining voter–candidate proximity.

To further examine the effect of orthogonalization, let us revisit Table 2 and focus on the predictions of voters' intrapersonal choice between orthogonalized and original non-orthogonalized data. If the process of orthogonalization does eliminate those double-counted responses, the proximity model with orthogonalized data should have a better performance in predicting voter choice as a function of policy distance. The outcome in Tables 2 shows that, even with rather small variation between the two sets of translation vectors, the process of orthogonalization generally increases the predictive power of the proximity model. Comparing the rate of correct predictions using orthogonalized data (upper part) with that using non-orthogonalized data (lower part), we can see the process of orthogonalization increases the correct predictions as much as 7%.

Table 5. Voters with opposite predictions using orthogonalized and non-orthogonalized data

	Non-orthogonalized	Orthogonalized
Optimal p-norm		
2000	28%	68%
2004	53%	42%
2008	34%	56%
Euclidean		
2000	20%	77%
2004	37%	58%
2008	37%	54%
City-block		
2000	23%	74%
2004	46%	50%
2008	33%	56%

More significantly, when we focus exclusively on the difference in their predictions, namely, the specific voters for whom the two types of data give rise to different predictions, the impact of orthogonalization is even more prominent. The outcome is presented in Table 5. For these voters, orthogonalization greatly improves the percentage of correct predictions with respect to their observed vote choice. In 2000, for example, using the Euclidean distance norm and non-orthogonalized data, only 20% of the voters are predicted correctly when compared to their observed vote choice. After orthogonalization, this percentage increases to 77%. Therefore, by eliminating double counting in the calculation of voter–candidate distance, orthogonalization greatly improves the number of voters for whom our model can make correct predictions about their vote choice.

A final note is that since the process of orthogonalization is dependent on the order of dimensions (especially the selection of the benchmark dimension), we tested the influence of the order on the stability of the coefficient for our composite proximity variable (in Equation [12]). In particular, we tried two different permutations of policy dimensions, with *Aid to Blacks* and *Job and Environment* as the benchmark dimension, respectively. Each permutation generates a new set of orthogonalized data. The results indicate that all the coefficients based on different permutations remain very close to each other with differences at one hundredth decimal point. Therefore, when applying our generalized parametric Minkowski model to orthogonalized data, scholars are largely free to choose the order of orthogonalization.

Conclusion

In his seminal criticism of the Downsian model, Stokes (1963) had envisioned that the (non)-orthogonality issue may arise in the application of multidimensional policy spaces. Unfortunately, while Stokes' criticism of unidimensionality is widely invoked, his caution on multidimensional models is largely overlooked. Scholars tend to assess the voter–candidate proximity in multidimensional policy spaces as if voters have uncorrelated responses on those policy dimensions. As we point out in this article, such an assumption could lead to biased estimation of voter–candidate proximity. In this study, we offer a solution applicable to most survey data. Rather than directly computing the policy distance in an orthogonal Euclidean space, we propose to construct a non-orthogonal policy space to reflect voters' correlated responses on policy dimensions. The process of orthogonalizing such a space is actually to peel off those correlated portions of responses in a dataset. Spatial distances calculated in such an orthogonalized space can more accurately reflect voters' evaluation of candidates.

In addition, we contend that the traditional debate between the city-block and Euclidean distance is misleading. The best distance norm should be sought in a much broader context. Specifically, we posit a generalized parametric Minkowski model. The selection of the most appropriate distance norm should be treated as an empirical issue, and the decision should be based on the features of datasets. With these issues in mind, our empirical analysis generates some encouraging evidence. We confirmed that the optimal distance norm varies across elections.

We believe that our approach moves us one step closer to a better understanding of how voters evaluate candidates. Correlation in responses is not merely confined to the study of electoral choice. It is quite common to all studies that use public surveys and polls. In this sense, while the Gram-Schmidt process we proposed helps offset the adverse influence of correlated responses, a more fundamental solution should be achieved by improving our survey design. For instance, in this article, our investigation of voters' correlated responses is conducted at the aggregate level. Proximity models at this level presume that the preferences of voters, as a group, can be studied using a single policy space. The structure of this policy space lays the foundation for our analysis. This is a very strong presumption, though. The actual preferences of voters are much more heterogeneous. Individual voters differ not only in terms of which policy issues are vital to them, but also in terms of how these issues are related. A thorough understanding must include the examination of a voter's evaluation of candidates in her idiosyncratic policy space. Although technically such individual-level analysis is not demanding, the puzzle is that current surveys do not provide enough information. Even if we can build a policy space for each respondent using those issues she claims as vital, there is no way to check the structure of such a space based on the survey questions. One plausible solution is to add some questions to future surveys. For instance, after respondents identify the most critical policy issues, we can further ask them to report the relationship between those issues. These self-reported correlations, although not completely free of personal biases, can provide us with at least some clues to examine the structure of individual voters' personal policy space.

Notes

1. As we will discuss below, the only plausible reason to predetermine a system-level distance metric is to assume voters are using homogenous utility functions to evaluate candidates. For detailed discussion about this homogeneity assumption, see Rivers (1988).
2. Theoretically, non-orthogonality may also cause another problem: nonseparability across voters' responses on different policy dimensions. That is, voters' responses on one question are conditional on their responses of previous questions (Hinich & Munger, 1997; Lacy, 2001). A major consequence of nonseparability is that present survey questions do not offer adequate information to identify voters' ideal points and the (perceived) ideal points of candidates. However, while non-orthogonality is both sufficient and necessary condition for correlated responses, it is only a necessary (not sufficient) condition for nonseparability of responses. For an illustrating example, consider the situation of asymmetric nonseparability: a voter has nonseparable preferences on one issue from the other but not vice versa (this situation can never occur to correlated responses). Additional tests are requires to diagnose nonseparable responses (Lacy, 2001). The presence of nonseparability would undoubtedly further complicate our research. Since our focus in this study is on correlated responses only, in the remainder of the article we assume separability of responses.
3. By definition, the Minkowski distances only include norms with $p \geq 1$ in Equation [5]. Fractional norm distances ($p < 1$) are not "metrics" because the rule of triangle inequality is violated (Eguia, 2011).
4. For the sake of simplicity, in Figure 1 the two issues are equally weighed. When the issues are assigned different weights, the size and shape would be different.
5. For examples in a three-dimensional space (where the candidates' ideal points compose a closed ball), see Akleman and Chen (1999) who demonstrate how the size and shape of the closed ball change as the parameter p takes different values.
6. This is why in Equations [1] and [2], a common estimator β_k (rather than specific one, β_i) is used for individual respondent i.
7. Recall that for two vectors $s = (s_1, s_2, s_3, \ldots, s_d)^T$ and $t = (t_1, t_2, t_3, \ldots, t_d)^T$, their scalar product is $s \cdot t = \sum s_i t_i = |s||t| \cos \theta_{s,t}$.
8. $\hat{z}_2 = \hat{x}_2 - proj_{\hat{z}_1}\hat{x}_2 = \hat{x}_2 - \frac{|\hat{x}_2| \cdot |\hat{z}_1| \cos \theta_{\hat{x}_2, \hat{z}_1}}{|\hat{z}_1|^2}\hat{z}_1 = \hat{x}_2 - a_2 \cdot \hat{z}_1 = (a_2, b_2, c_2, d_2, e_2)^T - (a_2,0,0,0,0)^T = (0,b_2,c_2,d_2, e_2)^T$
9. To help researchers orthogonalize input datasets, we created a package (in R, Maple, and Excel spreadsheet) and posted it on the authors' website for free download. Based on the computation specified in Appendix, this package can calculate the translation function of a multidimensional (with five or fewer dimensions) policy space using just the first-order correlations between issues.
10. Recall that $L_p^p(ij) = \sum_k w_{ik}|v_{ik} - c_{ijk}|^p$ and $L_p(ij) = (\sum_k w_{ik}|v_{ik} - c_{ijk}|^p)^{\frac{1}{p}}$. Another reason for us not to use squared Euclidean distance or more generally $L_p^p(ij)$ Minkowski norm is that, like fractional Minkowski distance (see footnote 3), they violate the triangle inequality of metrics (Westholm, 2001).
11. When $(|\hat{z}_{i1}|, |\hat{z}_{i2}|, |\hat{z}_{i3}|, |\hat{z}_{i4}|, |\hat{z}_{i5}|) \neq (|\hat{z}_{ij1}|, |\hat{z}_{ij2}|, |\hat{z}_{ij3}|, |\hat{z}_{ij4}|, |\hat{z}_{ij5}|)$, the theoretical implication is voters' preferences across issues and their (perceived) candidates' preferences across issues are correlated in different ways.
12. In the 2008 election, we use the issue *Health Insurance* (whether there should be a government insurance plan that covers all medical and hospital expenses for everyone) due to data availability.
13. See the NES codebook (www.electionstudies.org) for questions and options.
14. This would insure that there exists a fixed location for each candidate in a policy space (Davis et al., 1970; Rabinowitz & Macdonald, 1989; Macdonald et al., 1998).
15. Note that because the Minkowski distances require that $p \geq 1$ (see footnote 3), this is why the optimal p norm for Bush in the 2000 election is 1. If this restriction is relaxed and we consider fractional norm distances ($p < 1$), the optimal p norm would be 0.53.
16. As discussed above, if voters have heterogeneous preferences (and accordingly utility functions), the estimated system-level distance norm reflects the aggregated perceptions of the voters. It should be noted that as long as each voter is using the same utility function to assess the two candidates,

theoretically there should be no differences in the system-level distance norms for two candidates in the same election. In other words, the divergence in the optimal distance norms between two candidates in Table 3 does not suggest each individual voter is using different utility functions to assess the two candidates. We believe that there are two major sources of this divergence. The two distance norms should converge if these concerns are addressed. The first is model specification. Like most issue voting models, our model uses the proximity in issue preferences as the only explanatory variable. However, as Macdonald et al. (1998) pointed out, issues or policies are only one of the factors that affect voter choice. Therefore, whenever the missed factors (such as incumbency, party identification, and demographic factors, etc.) affect voters' assessment of two candidates in an unbalanced way, the estimated system-level distance metric would be tipped in one way or another. To test the effect of specification, we ran the models in Table 3 but with a group of control variables used in previous studies (including voters' demographic information, Party ID, etc.). The divergence between two candidates' optimal distance norms narrows down dramatically. For instance, in 2008 election, the distinction between Obama and McCain diminishes from 1.77 (3.94 vs. 2.17) to 0.04 (2.26 vs. 2.30). The second reason is the systematic error in the survey data. For example, previous studies have discovered the "rationalization" effect in public surveys, that is, respondents tend to place themselves closer to their preferred candidate and farther away from candidates they dislike on issue scales (Kinder, 1978; Sniderman et al., 1991). In other words, the perceived ideal points of candidates may be biased in a systematic manner, which also contributes to the divergence of the optimal distance norms of the two candidates. We thank an anonymous reviewer for insightful comments on this issue.

17. On the feeling thermometer, 50–100 indicate favorable rating.
18. The only exception is the controversial 2000 election, in which Gore (who won the popular vote) has a higher score than Bush.
19. Again, in the 2000 election, if we allow factional norm distance, Bush's optimal p-norm would be 0.1, which is also changed from its orthogonalized version (0.53).
20. To explore the influence of orthogonalization, we tried hypothetically different translation vectors with the same dataset. The simulations show that when there are significant variations between the translation vectors of voters' self-placement and that of perceived candidate positions, the empirical performance of the proximity model (R-squared) will be dramatically altered.

References

Akleman, E. & Chen, J. (1999) Generalized distance functions. *Proceedings of the International Conference on Shape Modeling and Applications*, [SMI 1999: 72–79].

Aldrich, J. & McKelvey, R. (1977) A method of scaling with applications to the 1968 and 1972 presidential elections. *American Political Science Review*, 71(1), pp. 111–130.

Davis, O. & Hinich, M. (1966) *Mathematical Applications in Political Science II* (Dallas: Southern Methodist University Press).

Davis, O., Hinich, M. & Ordeshook, P. (1970) An expository development of a mathematical model of the electoral process. *American Political Science Review*, 64, pp. 426–448.

Downs, A. (1957) *An Economic Theory of Democracy* (New York: Harper).

Eguia, J. (2009) Utility representations of risk neutral preferences in multiple dimensions. *Quarterly Journal of Political Science*, 4(4), pp. 379–385.

Eguia, J. (2011) Foundations of spatial preferences. *Journal of Mathematical Economics*, 47(2), pp. 200–205.

Grynaviski, J. & Corrigan, B. (2006) Specification issues in proximity models of candidate evaluation. *Political Analysis*, 14, pp. 393–420.

Hinich, M. & Munger, M. (1997) *Analytical Politics* (New York: Cambridge University Press).

Kinder, D. (1978) Political person perception: the asymmetrical influence of sentiment and choice on perceptions of presidential candidates. *Journal of Personality and Social Psychology*, 36, pp. 859–871.

Lacy, D. (2001) Nonseparable preferences, measurement error, and unstable. survey responses *Political Analysis*, 9, pp. 95–115.

Lewis, J. & King, G. (1999) No evidence on directional vs. proximity voting. *Political Analysis*, 8, pp. 21–33.

Macdonald, S., Rabinowitz, G. & Listhaug, O. (1998) On attempting to rehabilitate the proximity model: sometimes the patient just can't be helped. *Journal of Politics*, 60, pp. 653–690.

Ordeshook, P. (1986) *Game Theory and Political Theory* (Cambridge: Cambridge University Press).

Rabinowitz, G. & Macdonald, S. (1989) A directional theory of issue voting. *American Political Science Review*, 83, pp. 93–121.

Rivers, D. (1988) Heterogeneity in models of electoral choice. *American Journal of Political Science*, 32(3), pp. 737–757.

Sniderman, P., Brody, R. & Tetlock, P. (1991) *Reasoning and Choice: Explorations in Political Psychology* (Cambridge: Cambridge University Press).

Stokes, D. (1963) Spatial models of party competition. *American Political Science Review*, 57, pp. 368–377.

Westholm, A. (1997) Distance versus direction: the illusory defeat of the proximity theory of electoral choice. *American Political Science Review*, 91, pp. 865–883.

Westholm, A. (2001) On the return of epicycles: some crossroads in spatial modeling revisited. *Journal of Politics*, 63, pp. 436–481.

Appendix

The five unit vectors are defined as $\hat{\mathbf{x}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \hat{x}_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \\ e_2 \end{pmatrix}, \hat{x}_3 = \begin{pmatrix} a_3 \\ b_3 \\ c_3 \\ d_3 \\ e_3 \end{pmatrix}, \hat{x}_4 = \begin{pmatrix} a_4 \\ b_4 \\ c_4 \\ d_4 \\ e_4 \end{pmatrix}, \hat{x}_5 = \begin{pmatrix} a_5 \\ b_5 \\ c_5 \\ d_5 \\ e_5 \end{pmatrix}$, where $a_i^2 + b_i^2 + c_i^2 + d_i^2 + e_i^2 = 1$ and $a_i = \cos\theta_{\hat{\mathrm{x}}_\mathrm{i},\hat{\mathrm{x}}_1} = r_{\hat{\mathrm{x}}_\mathrm{i},\hat{x}_1}$.

Because the process involves some revolved computation, we use some notations to facilitate the process. In the following table, each Greek letter denotes a function of some correlations in the input space.

α	$1 - r^2_{\hat{x}_1,\hat{x}_2} = 1 - a_2^2$	ϑ	$r_{\hat{\mathrm{x}}_2,\hat{\mathrm{x}}_4} - a_2\, a_4 = r_{\hat{\mathrm{x}}_2,\hat{\mathrm{x}}_4} - r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_2} \cdot r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_4}$
β	$1 - r^2_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_3} = 1 - a_3^2$	μ	$r_{\hat{\mathrm{x}}_3,\hat{\mathrm{x}}_4} - a_3\, a_4 = r_{\hat{\mathrm{x}}_3,\hat{\mathrm{x}}_4} - r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_3} \cdot r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_4}$
γ	$1 - r^2_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_4} = 1 - a_4^2$	π	$r_{\hat{\mathrm{x}}_2,\hat{\mathrm{x}}_5} - a_2\, a_5 = r_{\hat{\mathrm{x}}_2,\hat{\mathrm{x}}_5} - r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_2} \cdot r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_5}$
δ	$1 - r^2_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_5} = 1 - a_5^2$	ρ	$r_{\hat{\mathrm{x}}_3,\hat{\mathrm{x}}_5} - a_3\, a_5 = r_{\hat{\mathrm{x}}_3,\hat{\mathrm{x}}_5} - r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_3} \cdot r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_5}$
ε	$r_{\hat{\mathrm{x}}_2,\hat{\mathrm{x}}_3} - a_2\, a_3 = r_{\hat{\mathrm{x}}_2,\mathrm{x}_3} - r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_3} \cdot r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_2}$	σ	$r_{\hat{\mathrm{x}}_4,\hat{\mathrm{x}}_5} - a_4\, a_5 = r_{\hat{\mathrm{x}}_4,\hat{\mathrm{x}}_5} - r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_4} \cdot r_{\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_5}$

Applying the Gram-Schmidt process to these vectors we have

(1) $\hat{z}_1 = \hat{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. Obviously, $|\hat{z}_1| = |\hat{x}_1| = 1$.

(2) $\hat{z}_2 = \hat{x}_2 - \text{proj}_{\hat{z}_1} \hat{x}_2 = \hat{x}_2 - \dfrac{|\hat{x}_2| \cdot |\hat{z}_1| \cos\theta_{\hat{x}_2, \hat{z}_1}}{|\hat{z}_1|^2} \hat{z}_1 = \hat{x}_2 - a_2 \cdot \hat{z}_1 =$
$\begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \\ e_2 \end{pmatrix} - \begin{pmatrix} a_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ b_2 \\ c_2 \\ d_2 \\ e_2 \end{pmatrix}$.

So we can calculate that $|\hat{z}_2| = \sqrt{1 - a_2^2} = \sqrt{\alpha}.$
(3) $\hat{z}_3 = \hat{x}_3 - \text{proj}_{\hat{z}_1} \hat{x}_3 - \text{proj}_{\hat{z}_2} \hat{x}_3 = \hat{x}_3 - a_3 \hat{z}_1 - \dfrac{\cos\theta_{\hat{x}_3, \hat{z}_2}}{\sqrt{1 - a_2^2}} \hat{z}_2.$

Note because $\cos\theta_{\hat{x}_3, \hat{z}_2} = \dfrac{b_2 b_3 + \cdots + e_2 e_3}{\sqrt{1 - a_2^2}}$, also because $\cos\theta_{\hat{x}_3, \hat{x}_2} = r_{\hat{x}_2, \hat{x}_3}$, we have $\cos\theta_{\hat{x}_3, \hat{z}_2} = \dfrac{r_{\hat{x}_2, \hat{x}_3 - a_2 a_3}}{\sqrt{1 - a_2^2}} = \dfrac{\varepsilon}{\sqrt{\alpha}}$. For simplicity, let $\dfrac{\cos\theta_{\hat{x}_3, \hat{z}_2}}{\sqrt{1 - a_2^2}} = \dfrac{\varepsilon}{\alpha} = f$.

So $\hat{z}_3 = \hat{x}_3 - a_3 \hat{z}_1 - f\,\hat{z}_2 = \begin{pmatrix} a_3 \\ b_3 \\ c_3 \\ d_3 \\ e_3 \end{pmatrix} - \begin{pmatrix} a_3 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ fb_2 \\ fc_2 \\ fd_2 \\ fe_2 \end{pmatrix} = \begin{pmatrix} 0 \\ b_3 - fb_2 \\ c_3 - fc_2 \\ d_3 - fd_2 \\ e_3 - fe_2 \end{pmatrix}$.

We can calculate $|\hat{z}_3| = \sqrt{1 - r_{\hat{x}_1, \hat{x}_3}^2 - \dfrac{(r_{\hat{x}_2, \hat{x}_3} - r_{\hat{x}_1, \hat{x}_3} . r_{\hat{x}_1, \hat{x}_2})^2}{1 - r_{\hat{x}_1, \hat{x}_2}^2}} = \sqrt{\beta - \dfrac{\varepsilon^2}{\alpha}}\cdot$

(4) $\hat{z}_4 = \hat{x}_4 - \text{proj}_{\hat{z}_1} \hat{x}_4 - \text{proj}_{\hat{z}_2} \hat{x}_4 - \text{proj}_{\hat{z}_3} \hat{x}_4 = \hat{x}_4 - a_4 \cdot \hat{z}_1 - \dfrac{\cos\theta_{\hat{x}_4, \hat{z}_2}}{\sqrt{1 - a_2^2}} \hat{z}_2 - \dfrac{\cos\theta_{\hat{x}_4, \hat{z}_3}}{|\hat{z}_3|} \hat{z}_3.$
Note that $\cos\theta_{\hat{x}_4, \hat{z}_2} = \dfrac{r_{\hat{x}_2, \hat{x}_4 - a_2 a_4}}{\sqrt{1 - a_2^2}} = \dfrac{\vartheta}{\sqrt{\alpha}}$, and

$\cos\theta_{\hat{x}_4, \hat{z}_3} = \dfrac{b_3 b_4 + \ldots + e_3 e_4 - f(b_2 b_4 + \ldots + e_2 e_4)}{|\hat{z}_3|}$, which can be further reduced to

$\dfrac{r_{\hat{x}_3, \hat{x}_4 - a_3 a_4 - f(r_{\hat{x}_2, \hat{x}_4} - a_2 a_4)}}{|\hat{z}_3|} = \dfrac{\mu - f\vartheta}{|\hat{z}_3|}$. Let $\dfrac{\cos\theta_{\hat{x}_4, \hat{z}_2}}{\sqrt{1 - a_2^2}} = \dfrac{\vartheta}{\alpha} = g$ and

$\frac{\cos\theta_{\hat{x}_4,\hat{z}_3}}{|\hat{z}_3|}=\frac{\mu-f\vartheta}{|\hat{z}_3|^2}=\frac{\alpha\mu-\varepsilon\vartheta}{\alpha\beta-\varepsilon^2}=h$, and

$$\hat{z}_4=\hat{x}_4-a_4\cdot\hat{z}_1-g\hat{z}_2-h\hat{z}_3=\begin{pmatrix}a_4\\b_4\\c_4\\d_4\\e_4\end{pmatrix}-\begin{pmatrix}a_4\\0\\0\\0\\0\end{pmatrix}-\begin{pmatrix}0\\gb_2\\gc_2\\gd_2\\ge_2\end{pmatrix}-\begin{pmatrix}0\\h(b_3-fb_2)\\h(c_3-fc_2)\\h(d_3-fd_2)\\h(e_3-fe_2)\end{pmatrix}=$$

$$\begin{pmatrix}0\\b_4-(g-fh)b_2-hb_3\\c_4-(g-fh)c_2-hc_3\\d_4-(g-fh)d_2-hd_3\\e_4-(g-fh)e_2-he_3\end{pmatrix}$$

We can calculate that $|\hat{z}_4|=\sqrt{r+h^2\beta+(g-fh)^2\alpha-2(g-fh)\vartheta-2h\mu+2(g-fh)h\varepsilon}$

which can be reduced to $|\hat{z}_4|=\sqrt{\frac{\alpha\beta\gamma-\alpha\mu^2-\beta\vartheta^2+2\varepsilon\vartheta\mu-\gamma\varepsilon^2}{\alpha\beta-\varepsilon^2}}$.

(5) $\hat{z}_5 = \hat{x}_5 - \mathrm{proj}_{\hat{z}_1}\hat{x}_5 - \mathrm{proj}_{\hat{z}_2}\hat{x}_5 - \mathrm{proj}_{\hat{z}_3}\hat{x}_5 - \mathrm{proj}_{\hat{z}_4}\hat{x}_5$

$$= \hat{x}_5 - a_5\cdot\hat{z}_1 - \frac{\cos\theta_{\hat{x}_5,\hat{z}_2}}{\sqrt{1-a_2^2}}\hat{z}_2 - \frac{\cos\theta_{\hat{x}_5,\hat{z}_3}}{|\hat{z}_3|}\hat{z}_3 - \frac{\cos\theta_{\hat{x}_5,\hat{z}_4}}{|\hat{z}_4|}\hat{z}_4$$

Note that $\cos\theta_{\hat{x}_5,\hat{z}_2}=\frac{r_{\hat{x}_2,\hat{x}_5-a_2a_5}}{\sqrt{1-a_2^2}}=\frac{\pi}{\sqrt{\alpha}}$, and $\cos\theta_{\hat{x}_5,\hat{z}_3}=\frac{r_{\hat{x}_3,\hat{z}_5}-a_3a_5-f(r_{\hat{x}_2,\hat{x}_5}-a_2a_5)}{|\hat{z}_3|}$
$=\frac{\rho-f\pi}{|\hat{z}_3|}$, and $\cos\theta_{\hat{x}_5,\hat{z}_4}$
$=\frac{r_{\hat{x}_4,\hat{x}_5-a_4a_5-(g-fh)}(r_{\hat{x}_2,\hat{x}_5}-a_2a_5)-h(r_{\hat{x}_3,\hat{x}_5}-a_3a_5)}{|\hat{z}_4|}$
$=\frac{\sigma-(g-fh)\pi-h\rho}{|\hat{z}_4|}$. Let $\frac{\cos\theta_{\hat{x}_5,\hat{x}_2}}{\sqrt{1-a_2^2}}=\frac{\pi}{\alpha}=k$, $\frac{\cos\theta_{\hat{x}_5,\hat{z}_3}}{|\hat{z}_3|}=\frac{\rho-f\pi}{|\hat{z}_3|^2}$
$=\frac{\alpha\rho-\varepsilon\pi}{\alpha\beta-\varepsilon^2}=l$, and $\frac{\cos\theta_{\hat{x}_5,\hat{z}_4}}{|\hat{z}_4|}=\frac{\sigma-(g-fh)\pi-h\rho}{|\hat{z}_4|^2}$
$=\frac{\alpha\beta\sigma-\sigma\varepsilon^2-\beta\vartheta\pi+\varepsilon\mu\pi+\alpha\mu\rho+\varepsilon\vartheta\rho}{\alpha\beta\gamma-\alpha\mu^2-\beta\vartheta^2+2\varepsilon\vartheta\mu-\gamma\varepsilon^2}=p$, then

$$\hat{z}_5 = \hat{x}_5 - a_5 \cdot \hat{z}_1 - k\hat{z}_2 - l\hat{z}_3 - p\hat{z}_4$$

$$= \begin{pmatrix} a_5 \\ b_5 \\ c_5 \\ d_5 \\ e_5 \end{pmatrix} - \begin{pmatrix} a_5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ kb_2 \\ kc_2 \\ kd_2 \\ ke_2 \end{pmatrix} - \begin{pmatrix} 0 \\ l(b_3 - fb_2) \\ l(c_3 - fc_2) \\ l(d_3 - fd_2) \\ l(e_3 - fe_2) \end{pmatrix} -$$

$$\begin{pmatrix} 0 \\ pb_4 - p(g - fh)b_2 - phb_3 \\ pc_4 - p(g - fh)c_2 - phc_3 \\ pd_4 - p(g - fh)d_2 - phd_3 \\ pe_4 - p(g - fh)e_2 - phe_3 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ b_5 - pb_4 + (ph - l)b_3 + (pg - pfh - k + lf)b_2 \\ c_5 - pc_4 + (ph - l)c_3 + (pg - pfh - k + lf)c_2 \\ d_5 - pd_4 + (ph - l)d_3 + (pg - pfh - k + lf)d_2 \\ e_5 - pe_4 + (ph - l)e_3 + (pg - pfh - k + lf)e_2 \end{pmatrix}$$ We can calculate that

$$|\hat{z}_5| = \Big[\delta + p^2\gamma + (ph - l)^2\beta + (pg - pfh - k + lf)^2\alpha - 2p\sigma$$
$$+2(ph - l)\rho + 2(pg - pfh - k + lf)\pi - 2p(ph - l)\mu$$
$$-2p(pg - pfh - k + lf)\vartheta + 2(ph - l)(pg - pfh - k + lf)\varepsilon\Big]^{\frac{1}{2}}$$

which can be reduced to

$$|\hat{z}_5| = \sqrt{\frac{\begin{array}{c}\alpha\beta\delta\gamma - \alpha\gamma\rho^2 - \beta\gamma\pi^2 + 2\gamma\varepsilon\pi\rho - \gamma\delta\varepsilon^2 \\ -\alpha\beta\sigma^2 + 2\alpha\sigma\rho\mu - \alpha\delta\mu^2 - \beta\delta\vartheta^2 + 2\beta\sigma\pi\vartheta + \mu^2\pi^2 + 2\delta\vartheta\varepsilon\mu \\ -2\sigma\pi\varepsilon\mu - 2\pi\vartheta\mu\rho - 2\rho\sigma\varepsilon\vartheta + \rho^2\vartheta^2 + \sigma^2\varepsilon^2\end{array}}{\alpha\beta\gamma - \alpha\mu^2 - \beta\vartheta^2 + 2\vartheta\varepsilon\mu - \gamma\varepsilon^2}}$$