B.J.Pol.S. **35,** 635–663 *Copyright © 2005 Cambridge University Press*
doi:10.1017/S0007123405000335 *Printed in the United Kingdom*

Multiparty Competition in Israel, 1988–96

NORMAN SCHOFIELD AND ITAI SENED*

Formal models of voting usually assume that political agents, whether parties or candidates, attempt to maximize expected vote shares. 'Stochastic' models typically derive the 'mean voter theorem' that each agent will adopt a 'convergent' policy strategy at the mean of the electoral distribution. In this article, it is argued that this conclusion is contradicted by empirical evidence. Estimates of vote intentions require 'valence' terms. The valence of each party derives from the average weight, given by members of the electorate, in judging the overall competence or 'quality' of the particular party leader. In empirical models, a party's valence is independent of current policy declarations and can be shown to be statistically significant in the estimation. It is shown here that the addition of valence gives a very strong Bayes factor over an electoral model without valence. The formal model is analysed and shown to be classified by a 'convergence' coefficient, defined in terms of the parameters of the empirical model. This coefficient gives necessary and sufficient conditions for convergence. When the necessary condition fails, as it does in these empirical studies with valence, then the convergent equilibrium fails to exist. The empirical evidence is consistent with a formal stochastic model of voting in which there are multiple local Nash equilibria to the vote-maximizing electoral game. Simulation techniques based on the parameters of the empirical model have been used to obtain these local equilibria, which are determined by the principal component of the electoral distribution. Low valence parties, in equilibrium, will tend to adopt positions at the electoral periphery. High valence parties will contest the electoral centre, but will not, in fact, position themselves at the electoral mean. Survey data from Israel for the elections of 1988, 1992 and 1996 are used to compute the parameters of the empirical model and to illustrate the dependence of equilibria on the electoral principal components. The vote maximizing equilibria do not perfectly coincide with the actual party positions. This divergence may be accounted for by more refined models that either (i) include activism or (ii) consider strategic party considerations over post-election coalition bargaining.

Empirical and theoretical models of representative democracy typically have two distinct components. At the micro-level, individual voting behaviour is modelled as a function of the preferences, or beliefs, of the voters and the policy positions or declarations of political candidates (or agents). It has also generally been assumed that the agents adopt strategies to maximize a utility function defined in terms of the overall vote share of the agent. Other possibilities include maximizing seat share, or some combination of policy outcomes with seat or vote share, or the probability of winning a majority.[1]

The natural formal concept to use in examining political agent strategies is that of Nash equilibrium – the vector of agent strategies with the property that no agent may deviate from the Nash equilibrium strategy and gain anything by doing so. Almost all formal models of voting suggest that political agents, in equilibrium, will adopt 'convergent'

* Center in Political Economy and Department of Political Science, respectively, at Washington University in St Louis. This article is based on research funded by NSF grants SBR 97–30275 and SES 02–41732. The authors thank Andrew Martin, David Nixon, Dganit Ofek and Kevin Quinn for their collaboration on earlier research and Alexandra Shankster and Ben Klemens for assistance in preparing the figures. Detailed comments on the theoretical result by John Duggan and on the empirical model by James Adams and Michael Laver were very helpful. The authors are grateful for remarks made by seminar participants at the Universities of Bonn, Dublin, Konstanz, Mannheim, Siena, Bergamo and Palermo. Schofield wishes to express his appreciation to the Fulbright Commission for the opportunity provided by his tenure of the Fulbright Distinguished Chair at Humboldt University, Berlin, in 2002–3.

[1] John Duggan, 'Equilibrium Equivalence under Expected Plurality and Probability of Winning Maximization' (unpublished, University of Rochester, 2000).

strategies that are located in some central domain of the policy space, as defined by voter preferences or beliefs.[2]

Arguments and evidence that parties do not adopt centrist strategies have been commonplace for decades.[3] Various theoretical models have been devised to account for policy divergence. These include theories based on activist support,[4] directional voting,[5] socio-demographic variables[6] and valence.[7]

Models based on activism have been used to account for the fact that candidates in competitive plurality electoral systems, such as in the United States, must obtain campaign resources from activists who tend to be more radical than the electorate as a whole. We regard this as a second order effect in electoral systems based on proportional representation. Studies based on directional voting with party identification have generally been unable to explain why small extremist parties do not moderate their policies.[8] Models using socio-demographic variables have also indicated that smaller parties could have

[2] Randall L. Calvert, 'Robustness of the Multidimensional Voting Model: Candidates, Motivations, Uncertainty and Convergence', *American Journal of Political Science*, 29 (1985), 69–85; Jeffrey Banks, John Duggan and Michel Le Breton, 'Bounds for Mixed Strategy Equilibria and the Spatial Model of Elections', *Journal of Economic Theory*, 103 (2002), 88–105.

[3] Maurice Duverger, *Political Parties: Their Organization and Activity in the Modern State* (New York: Wiley 1954); David Robertson, *A Theory of Party Competition* (London: Wiley, 1976); Hans Daalder, 'In Search of the Center of European Party Systems', *American Political Science Review*, 78 (1984), 92–109; Keith Poole and Howard Rosenthal, 'U.S. Presidential Elections 1968–1980: A Spatial Analysis', *American Journal of Political Science*, 28 (1984), 283–312; Ian Budge, David Robertson and Derek Hearl, *Ideology, Strategy and Party Change: Spatial Analyses of Post War Election Programmes in 19 Democracies* (Cambridge: Cambridge University Press, 1987); Gary Cox, 'Centripetal and Centrifugal Incentives in Electoral Systems', *American Journal of Political Science*, 34 (1990), 903–35; Gary Cox, *Making Votes Count* (Cambridge: Cambridge University Press, 1997); George Rabinowitz, Stuart Macdonald and Ola Listhaug, 'New Players in an Old Game: Party Strategy in Multiparty Systems', *Comparative Political Studies*, 24 (1991), 147–85; James Adams and Samuel Merrill III, 'Modeling Party Strategies and Policy Representation in Multiparty Elections: Why are Strategies so Extreme?' *American Journal of Political Science*, 43 (1999), 765–91.

[4] John Aldrich, 'A Spatial Model with Party Activists: Implications for Electoral Dynamics', *Public Choice*, 41 (1983), 63–100; John Aldrich, 'A Downsian Spatial Model with Party Activists', *American Political Science Review*, 77 (1983), 974–90; John Aldrich, *Why Parties*? (Chicago: Chicago University Press, 1995); John Aldrich and Michael McGinnis, 'A Model of Party Constraints on Optimal Candidate Positions', *Mathematical and Computer Modelling*, 12 (1989), 437–50.

[5] James Adams, *Party Competition and Responsible Party Government* (Ann Arbor: University of Michigan Press, 2001); Samuel Merrill III and Bernard Grofman, *A Unified Theory of Voting* (Cambridge: Cambridge University Press, 1999); Samuel Merrill III, Bernard Grofman and Scott Feld, 'Nash Equilibrium Strategies in Directional Models of Two-Candidate Spatial Competition', *Public Choice*, 98 (1999), 369–83.

[6] Michael Alvarez, Jonathan Nagler and Shaun Bowler, 'Issues, Economics, and the Dynamics of Multiparty Elections: The British 1987 General Election', *American Political Science Review*, 94 (2000), 131–50; Michael Alvarez, Jonathan Nagler and Jennifer Willette, 'Measuring the Relative Impact of Issues and the Economy in Democratic Elections', *Electoral Studies*, 19 (2000), 237–53.

[7] Donald Stokes, 'Spatial Models and Party Competition', *American Political Science Review*, 57 (1963), 368–77; Donald Stokes, 'Valence Politics', in Dennis Kavanagh, ed., *Electoral Politics* (Oxford: Clarendon Press, 1992) pp. 141–62; Timothy Groseclose, 'A Model of Candidate Location When One Candidate Has a Valance Advantage', *American Journal of Political Science*, 45 (2001), 862–86; Enriqueta Aragones and Thomas Palfrey, 'Mixed Equilibrium in a Downsian Model with a Favored Candidate', *Journal of Economic Theory*, 103 (2002), 131–61; Enriqueta Aragones and Thomas Palfrey, 'Electoral Competition Between Two Candidates of Different Quality: The Effects of Candidate Ideology and Private Information', in David Austen-Smith and John Duggan, eds, *Social Choice and Strategic Decisions* (Heidelberg: Springer, 2005), pp. 93–112.

[8] James Adams and Samuel Merrill III, 'Spatial Models of Candidate Competition in the 1988 French Presidential Election: Are Presidential Candidates Vote Maximizers?' *Journal of Politics*, 62 (2000), 729–56.

gained votes by moving to the centre.[9] Incorporating valence, or the perception in the electorate of a candidate's 'quality', is a plausible way to modify the usual vote models. Recent models with valence have concentrated on adapting the basic Downsian model[10] where the voters 'know with certainty' the location of the candidates.[11] In contrast to the Downsian model, empirical models of voting make the implicit assumption that there is a stochastic component to the individual voter choice.[12] Therefore, it is appropriate to use, as a benchmark for such empirical studies, a formal stochastic model of voting. The 'probabilistic' vote model has been developed to extend the early work of Hinich.[13] Initially focusing on two-candidate competition[14] it has recently been extended to the case of multiparty competition with three or more candidates.[15] The principal result of this work is 'the mean voter theorem', which asserts that parties, in situations where they are motivated to increase vote shares, will adopt convergent strategies at the mean of the electoral distribution.[16] This conclusion is subject to a constraint that the stochastic component is 'sufficiently' important. To date, the relevance of this result for empirical analysis has not been evaluated, because the constraint has not been formulated in a precise enough fashion to be applied to empirical work.

The purpose of this article is to contrast the formal model with empirical analyses so as to determine, first, whether convergence does occur and, secondly, to ascertain whether parties in typical situations where they do not converge, could, in fact, increase their vote shares by moving to the electoral centre.

Our empirical analysis focuses on Israel for a number of reasons. First, most such electoral analyses have examined polities such as the Netherlands, where there are three or four parties of comparable size. Israel has had approximately nineteen parties attaining seats in the Knesset in the last few elections. Moreover, these range in size from small parties with 2 seats or so, to moderately large parties such as Likud and Labor whose seat strengths lie in the range 19 to 44, out of a total of 120 Knesset seats. Because of the proportional electoral system, coalitions must cohere to form government, adding a further

[9] Alvarez, Nagler and Willette, 'Measuring the Relative Impact'; see also Michael Alvarez and Garrett Glasgow, 'Voting Behavior and the Electoral Context of Government Formation', and Michael Alvarez and Garrett Glasgow, 'Modeling voter support in the 1989 and 1994 Dutch Elections' (unpublished, University of California at Santa Barbara). See also Kevin Quinn and Andrew Martin, 'An Integrated Computational Model of Multiparty Electoral Competition', *Statistical Sciences*, 17 (2002), 405–19.

[10] Anthony Downs, *An Economic Theory of Democracy* (New York: Harper and Row, 1957).

[11] See, for example, Stephen Ansolabehere and James Snyder, 'Valence Politics and Equilibrium in Spatial Election Models', *Public Choice*, 103 (2000), 327–36.

[12] Poole and Rosenthal, 'U.S. Presidential Elections'; Kevin Quinn, Andrew Martin and Andrew Whitford, 'Voter Choice in Multiparty Democracies', *American Journal of Political Science*, 43 (1999), 1231–47.

[13] Melvin Hinich, 'Equilibrium in Spatial Voting: The Median Voter Result is an Artifact', *Journal of Economic Theory*, 16 (1977), 208–19.

[14] Peter Coughlin, *Probabilistic Voting Theory* (Cambridge: Cambridge University Press, 1992); James Enelow and Melvin Hinich, *The Spatial Theory of Voting* (Cambridge: Cambridge University Press, 1984).

[15] See Tse Min Lin, James Enelow and Han Dorussen, 'Equilibrium in Multicandidate Probabilistic Spatial Voting', *Public Choice*, 98 (1999), 59–82; James Adams, 'Multiparty Spatial Competition with Probabilistic Voting', *Public Choice*, 99 (1999), 259–74; James Adams, 'Policy Divergence in Multicandidate Probabilistic Spatial Voting', *Public Choice*, 100 (1999), 103–22.

[16] The derivation of this result depends on differentiating the vote share functions, to obtain the first order conditions. It is generally assumed that voters have 'quadratic loss functions', and it is this assumption that gives the mean voter location. If voters have linear loss functions, then a 'median voter location' satisfies the first order condition. Our focus is not on the mean voter result *per se* but on the validity of the argument that convergence is rational for parties.

interesting aspect. Since Likud and Labor compete for dominance of coalition government, one may expect these larger parties to attempt to maximize their seat strength in the Knesset. Because Israel uses a highly proportional electoral system, seat shares and vote shares are in close correspondence. Thus one can consider vote shares as the maximand for these parties. Finally, empirical electoral models are designed to estimate vote shares, and we can therefore use these models to examine the hypothesis of vote share maximization. Excellent survey data are available for the period 1988 to 1996, which facilitates the construction of the electoral model.[17] As is common in such studies, exploratory factor analysis of these data allow for the construction of a low dimensional 'policy space'. To construct an empirical model of voter choice for parties in the Knesset it is necessary to obtain estimates of party positions. This we were able to do for seven of the parties for these elections, by using data from party manifestos. Experts on Israeli politics evaluated the manifestos, using the survey questions, and the party positions in the policy space were estimated using the factor weights.

To construct the electoral model, socio-demographic data of the respondents, together with the vote intentions and survey responses were utilized. Multinomial probit (MNP) and multinomial conditional logit (MNL) models were constructed, and compared using Bayes factors.[18] The model of choice was a MNL version including valence constants for each party. The Bayes factor (log likelihood ratio) of this model over the model without valence was of the order of 250, giving a very high degree of statistical significance. Using the estimates for the parameters of this model, it was then possible to use a 'hill climbing' algorithm to determine the empirical equilibria of the vote-maximizing political game.[19] The result of this simulation exercise contradicted the conclusions of the standard 'mean voter theorem'. Instead, the simulation showed that, with the model parameters, vote maximizing positions for the high valence parties were 'near', but not at the electoral mean, while equilibrium positions for low valence parties were strung along the principal component of the voter distribution.

The positions of the parties obtained from the simulation were not exactly the same as the estimated positions. However, the model of voter choice that we deployed did accurately predict approximately 50 per cent of the individual choice. We draw four conclusions:

(i) The assumptions of the formal stochastic vote model are compatible with actual voter choice.
(ii) Not all parties choose their positions solely with regard to vote maximization.
(iii) The logic of the 'mean voter theorem' is internally incomplete in the sense that the validity of the theorem depends on unrecognized parameter constraints.

[17] We thank Asher Arian and Michal Shamir for allowing us to make an extensive use of their original data, reported, analysed and extensively discussed in Asher Arian and Michael Shamir, *The Election in Israel: 1988* (Albany: SUNY Press, 1990); *The Election in Israel: 1992* (Albany: SUNY Press, 1995); *The Election in Israel: 1996* (Albany: SUNY Press, 1999). Earlier analyses of these data are reported in Norman Schofield, Itai Sened and David Nixon, 'Nash Equilibrium in Multiparty competition with "Stochastic" Voters', *Annals of Operations Research*, 84 (1998), 3–27; and Dganit Ofek, Kevin Quinn and Itai Sened, 'Voters, Parties and Coalition formation in Israel: Theory and Evidence' (unpublished, Washington University in Saint Louis, 1998).

[18] See Quinn, Martin and Whitford, 'Voter Choice in Multiparty Democracies', for a discussion of this technique.

[19] See Samuel Merrill and James Adams, 'Computing Nash Equilibria in Probabilistic, Multiparty Spatial Models with Non Policy Components', *Political Analysis*, 9 (2001), 347–61, for an efficient algorithm for such hill climbing.

(iv) Since the valence terms were statistically significant in the empirical model, it can be inferred that the appropriate constraint for the validity of the theorem involves these valences.

To obtain the constraints that are necessary and sufficient for the validity of the 'mean voter theorem', we have examined the formal vote model in detail. The existence of a 'Nash equilibrium' at the joint mean voter position depends on showing that all party vote share functions are 'concave' in some domain of the party strategy spaces.[20] Concavity of these functions depends on parameters of the model. Moreover, concavity is a global property of the vote share functions, which can be shown to be violated in spatial models.[21] In fact, the simulation exercise made it evident that concavity did fail, so there was no reason to expect the existence of Nash equilibria. We therefore sought existence of the weaker 'local Nash equilibrium' (LNE). This theoretical concept is compatible with the hill-climbing algorithm used in our simulation exercise. Our formal analysis obtains necessary and sufficient conditions (determined by the model parameters) for existence of a LNE at the voter mean. The necessary constraint obtained from the formal model was shown to be violated by the estimated values of the parameters in the empirical model. Since any Nash equilibrium is necessarily a LNE it follows that no Nash equilibrium can be expected at the voter mean. Consequently, our empirical model of vote maximizing parties in the Israeli Knesset could not lead us to expect convergent strategies at the mean electoral position. The formal result that we present is valid in a policy space of unrestricted dimension, but has a particularly simple expression in the uni-dimensional case. The theorem given here, for the case of an arbitrary number of parties, and number of parties appears to be compatible with the results obtained by others for the one-dimensional model.[22]

The necessary and sufficient conditions of the result allow us to determine whether a low valence party would in fact maximize its vote shares at the mean. More precisely, we can determine whether the mean voter position is a best response, under simple vote maximization, for a low valence party, when all other parties are at the mean. In the empirical model we estimated that low valence parties would, in fact, minimize their vote share if they chose the mean position. Further analysis of the formal model leads us to the following conclusions:

(i) All low valence parties, in maximizing vote shares, should adopt positions far from the electoral centre.
(ii) If low valence parties vacate the mean, then the first order necessary conditions for equilibrium, associated with high valence parties at the mean, will also be violated. Consequently, we should expect that it is a non-generic property for any party to occupy the electoral mean in any vote maximizing equilibrium.

In the simulation, based on the empirical estimation for Israel, the two high valence parties, Labor and Likud, were seen to locate themselves on opposite sides of the electoral mean, and on a principal axis drawn through the two-dimensional policy space. Our formal analysis indicates that this is a feature of equilibrium location when there are at most two high valence parties. Indeed, the actual location of these two parties closely corresponded

[20] Jeffrey Banks and John Duggan, 'The Theory of Probabilistic Voting in the Spatial Model of Elections: The Theory of Office-Motivated Candidates', in Austen-Smith and Duggan, eds, *Social Choice and Strategic Decisions*, pp.1–41.

[21] Norman Schofield and Robert Parks, 'Nash Equilibrium in a Spatial Model of Coalition Bargaining', *Mathematical Social Sciences*, 39 (2000), 133–74.

[22] See, e.g., Groseclose. 'A Model of Candidate Location When One Candidate Has a Valence Advantage'.

to their simulated vote maximizing positions. We infer from this that the principal concern of these two high valence parties was to position themselves so as to gain the largest share possible of the popular vote. Clearly, one or other of these two parties would be the post-election formateur of government coalition.

However, in the election of 2003, thirteen parties in total obtained representation in the Knesset. This suggests that with a low electoral threshold (of 1.5 per cent), and many small parties who adopt positions far from the electoral mean, policy disagreements between high valence parties and smaller parties would make coalition formation difficult.

The difference observed between the simulated vote maximizing positions of the low valence parties and their actual policy positions may be due to their awareness that the probability of coalition membership depends not so much on seat strength but on party position. Thus, a small party like Shas is crucial to successful coalition formation by Likud or Labor. By positioning itself appropriately, Shas can ensure itself membership of such a coalition.[23] There may be constraints on policy choice because of activist party members and the ideological commitment of the party elite. However, vote and seat shares are measures of party success, and are an obvious basis for party motivation. We infer from our results that vote maximization is the fundamental factor in party policy choice, particularly for high valence parties. Clearly, optimal party location depends on the valence by which the electorate, on average, judges party competence. Our simulations suggest that if a single party has a significantly high valence, for whatever reason, then it has the opportunity to locate itself near the electoral centre. In comparison, if two parties have high, comparable valence, then our simulation suggests that neither will closely contest the centre. The simulation, and the theory presented in Section 2 shows that the electoral mean is a vote minimizing position for low valence parties. It is this feature that forces them to the 'electoral periphery'.

The formal and empirical analyses presented here are applicable for any polity using an electoral system based on proportional representation. The underlying formal model is compatible with a wide variety of different theoretical political equilibria. The theory is also compatible with the considerable variation of party political configurations found in multiparty systems.[24]

Our analysis of the formal model emphasizes the notion of 'local' Nash equilibrium in contrast to the notion of a 'global' Nash equilibrium usually employed in the technical literature. One reason for this emphasis is that we deploy the tools of calculus and simulation via hill-climbing algorithms to locate equilibria. By definition, the set of local equilibria must include the set of global Nash equilibria. Sufficient conditions for the existence of a global Nash equilibrium are therefore more stringent than for local equilibrium. Indeed, the necessary and sufficient condition for local equilibrium at the electoral centre, in the vote-maximizing game with valence, is so stringent that we regard it as unlikely to obtain in any polity that uses a proportional electoral system. We therefore infer that the existence of a global Nash equilibrium at the electoral centre is generically impossible for all such games. In contrast, the sufficient condition for the existence of a local, non-centrist equilibrium is much less stringent, so much so that we consider it to hold

[23] Norman Schofield and Itai Sened, 'Local Nash Equilibrium in Multiparty Politics', *Annals of Operations Research*, 109 (2002), 193–210.

[24] Michael Laver and Norman Schofield, *Multiparty Government* (Oxford: Oxford University Press, 1990), reprinted (Ann Arbor: University of Michigan Press, 1998).

in 'almost' every polity.[25] Indeed, in each polity there may well be multiple local equilibria. This suggests that the particular configuration of party positions in any polity is either a matter of historical contingency or is due to the action of activists.[26] We discuss this possibility in the conclusion.

1. AN EMPIRICAL ELECTORAL MODEL

We now present the details of the empirical model. It is assumed that the political preferences (or beliefs) of voter i can be described by a 'latent' utility function of the form:

$$u_i(\mathbf{z}) = (u_{i1}(z_1), \ldots, u_{ip}(z_p)). \tag{1}$$

Here $\mathbf{z} = (z_1, \ldots, z_p)$ is the vector of strategies of the collection, P, of political agents (candidates, parties, etc.). For agent j, z_j is a point in a space X that characterizes the agent. As in all empirical studies, we assume X is a compact convex subset of Euclidean space of dimension w. We make no prior assumption that $w = 1$. Each voter, i, is also described by an ideal point x_i, in the same space X, which is used to denote the beliefs or 'ideal point' of voter i. We represent the dependence of u_i on x_i by writing the jth component of i's utility (for each $j \in P$) as $u_{ij}(x_i, z_j)$. We assume:

$$u_{ij}(x_i, z_j) = \lambda_j - \mathrm{A}_{ij}(x_i, z_j) + \theta_j^T \mathrm{s}_i + \varepsilon_j. \tag{2}$$

Here $\mathrm{A}_{ij}(x_i, z_j)$ is some measure of the distance between the vectors x_i and z_j. In the 'Euclidean' model that we deploy, it is assumed that $\mathrm{A}_{ij}(x_i, z_j) = \beta \| x_i - z_j \|^2$ where $\| \ \|$ is the Euclidean norm on X and β is a positive constant. The term λ_j is called *valence* and is discussed further below. The k vector θ_j represents the effect of socio-demographic parameters (such as class, domicile, education, income) on voting for agent j, while s_i is a k-vector denoting in the ith individual's relevant 'socio-demographic' characteristics (we use T to denote transpose, so $\theta_j^T \mathrm{s}_i$ is a scalar). The vector ε_j is a 'stochastic' error term, associated with the jth party. Early models of this kind assume that the elements of the random vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)$ are independently distributed so the covariance matrix only has diagonal components $\{\sigma_j^2\}$ where σ_j^2 denotes the variance of each ε_j. The variates are called 'iid' if they are identically as well as independently distributed.

In their classic study of US presidential elections, Poole and Rosenthal assumed $\{\varepsilon_j\}$ to be iid.[27] More recent empirical analyses have been based on Markov Chain Monte Carlo (MCMC) methods, allowing for estimation of the full covariance matrix.[28] Assuming that the errors are independent and distributed via the log-Weibull distribution, a multinomial logit (MNL) model results. Assuming that the errors are distributed multivariate normal, with general covariance matrix, gives the multinomial probit (MNP) model. Both models require normalization of the error difference covariance matrix $\Omega(\varepsilon_j - \varepsilon_k)$. For the MNP model, the 'stochastic' normalization sets a variance term $\Omega(\varepsilon_p - \varepsilon_l) = 1$. For the MNL model all error variances are set to 1.65.

MNP models are generally preferable because they do not require the restrictive

[25] Schofield and Sened, 'Local Nash Equilibrium'.

[26] As suggested by Aldrich, 'A Spatial Model with Party Activists' and 'A Downsian Spatial Model with Party Activists', and more recently by Gary Miller and Norman Schofield, 'Activists and Partisan Realignment', *American Political Science Review*, 97 (2003), 245–60.

[27] Poole and Rosenthal, 'U.S. Presidential Elections'.

[28] Siddhartha Chib and Edward Greenberg, 'Markov Chain Monte Carlo Simulation Methods in Econometrics', *Econometric Theory*, 12 (1996), 409–31.

assumption of 'independence of irrelevant alternatives' (IIA).[29] We comment further on the difference between MNP and MNL below. However, a comparison of MNP and MNL models suggests that the results are broadly comparable.[30] We shall compare MNP and MNL models in this article, but focus on simulation of the MNL model for three reasons:

(i) We can compare the empirical results with those of a formal model based on iid normal errors.
(ii) It is possible to use simulation of the MNL model with seven parties with some expectation of accuracy. Simulation of the MNP model may be inaccurate.
(iii) Our comparison of Bayes factors suggests that the MNL model is an efficient predictor of vote shares.[31]

A variety of methods have been used to measure the distance or 'policy' component $A_{ij}(x_i, z_j)$. Alvarez, Nagler and Bowler used a National Election Survey for Britain to locate each voter (in a sample, N, of size n) with regard to preferred positions on a large number of policy issues.[32] Each voter was asked to locate the parties and the average across the survey population was used to estimate the position, on this large number of issues, of each party. This has the virtue that data were not lost, but had the disadvantage that no representation of policy issues was possible.

In their study of US presidential elections, Poole and Rosenthal used factor analysis to estimate the distribution of voter ideal points in a two-dimensional policy space, X, and also located presidential candidate positions in the same space.[33] In their analysis, the second non-economic dimension 'capture[ed] the traditional identification of southern conservatives with the Democratic party'.[34] They also noted that there was no evidence that candidates tended to converge to the electoral mean.[35]

There are many possible explanations for non-convergence of candidate positions. For example, primaries may lead to the choice of more radical candidates for each party. In this article, we develop a parsimonious theory based on simple expected vote maximization. We use the standard empirical framework to study party positioning in the complex multiparty electoral environment of Israel. The parameters of the estimated model can then be compared with the results of the formal model to determine whether, in fact, convergence of party positions can be expected.

Figure 1 presents the results of factor analysis of the 1996 survey (of size 794) undertaken by Arian and Shamir for the election that year in Israel. Figure 2 gives the result for the 1992 election.[36] Each respondent is characterized by a point in the resulting two-dimensional policy space, X. The figure presents the estimated density function for the voter ideal points. For example, the outer contour line contains 95 per cent of the estimated distribution. Table 1 presents the factor loadings for the 1996 analysis of the

[29] Michael Alvarez and Jonathan Nagler, 'When Politics and Models Collide: Estimating Models of Multi-candidate Elections', *American Journal of Political Science*, 42 (1998), 55–96.
[30] Quinn, Martin and Whitford, 'Voter Choice in Multiparty Democracies'.
[31] Quinn and Martin, 'An Integrated Computational Model of Multiparty Electoral Competition'.
[32] Alvarez, Nagler and Bowler, 'Issues, Economics, and the Dynamics of Multiparty Elections'.
[33] Poole and Rosenthal, 'U.S. Presidential Elections 1968–1980'.
[34] Poole and Rosenthal, 'U.S. Presidential Elections 1968–1980', p. 287.
[35] As previously argued by Hinich, in 'Equilibrium in Spatial Voting'.
[36] Arian and Shamir, *The Election in Israel: 1992*; Arian and Shamir, *The Election in Israel: 1996*. Notice that the scale of the figures is indeterminate. However, we can fix the scale by a normalization procedure. This is done by fixing the variance of the errors. We comment further on this below.

survey questions. For convenience, we use the terms 'security' and 'religion' for the two factor dimensions. 'Security' involves attitudes to the Oslo Agreement and other peace initiatives. 'Religion' involves the significance of religion in government policy. The axes of Figures 1 and 2 are oriented so that the 'left' on the security axis can be interpreted as supportive of negotiations with the Palestinian Liberation Organization (PLO), while 'north' on the religious axis is indicative of a support for the importance of the Jewish faith in Israel. As in all factor models, it is necessary to normalize the electoral distribution in some fashion. (We shall refer to this as 'electoral normalization'.) The variance on the security axis is normalized to be 1.0. For 1996, the variance on the religion axis is estimated to be 0.73. The covariance between these two axes is 0.59, so $r^2 = 0.477$. In 1992, the second electoral variance is 0.43, while the covariance is 0.45, and $r^2 = 0.471$.

It will be important to note, for the later formal analysis, that the distribution of voter ideal points for 1996 displays a principal component aligned along the axis through the origin and the point (1, 0.8) in Figure 1. Similarly, in Figure 2 there is an evident principal component through the origin and the point (1, 0.55). In the discussion below, we shall refer to the principal component of the electoral data with the greater variance as the major axis, while the second, lower variance principal component will be called the minor axis. In Figures 1 and 2 these two axes are obvious. Although the analysis for 1988 is not reported here, it is worth observing that between 1988 and 1996 the correlation across the two electoral axes has fallen.

Since we are interested in the policy choices of parties, we inferred party locations from expert analysis of party manifestos, using the Arian–Shamir questionnaire as the basis.[37] We implicitly assume that voters have information on party declarations and use this information when making vote choices. Table 2 presents the election results for the Knesset for 1988 to 2003. Since the Israeli electoral system is highly proportional, with a threshold of only 1.5 per cent, seat and vote shares are very close. The variation in seat strengths over 1988–2003 reflects considerable changes in vote shares.

Since the competition between the two major parties, Labor and Likud, is pronounced, it is surprising that these parties do not move to the electoral mean (as suggested by the formal vote model) in order to increase vote and seat shares. Comparing Figures 1 and 2, we note that the vote share for the smaller Sephardic party, Shas, increased between 1992 and 1996, though the move by Shas towards the electoral centre appeared minimal. Our inference is that the shifts of electoral support (indicated in Table 2) are the result of changes in party valence.

To be more explicit, we contend that prior to an election each voter, i, forms a judgement about the relative capability of each party leader.[38] Let λ_{ij} denote the weight given by

[37] Two techniques have been used in the past to determine party locations. Ian Budge and his co-authors have used content analysis of party manifestos, while Michael Laver uses expert opinion. See Budge *et al.*, *Ideology, Strategy and Party Change*; Ian Budge *et al.*, *Mapping Policy Preferences: Estimates for Parties, Electors and Governments* (Oxford: Oxford University Press, 2001); and Michael Laver, ed., *Estimating the Policy Position of Political Actors* (London: Routledge, 2001). We combined these two approaches by using experts on Israeli politics to examine the party manifestos on the basis of the electoral questionnaire.

[38] Of course, in our analysis we cannot distinguish between judgements about the capability of the leader or the capability of the party elite. We focus on the party leader because there is a well-attested relationship between electoral evaluation of leaders and election outcomes. See Anthony King, ed., *Leaders' Personalities and the Outcomes of Democratic Elections* (Oxford: Oxford University Press, 2002); and Harold Clarke, David Sanders, Marianne Stewart and Paul Whiteley, *Political Choice in Britain* (Oxford: Oxford University Press, 2004).

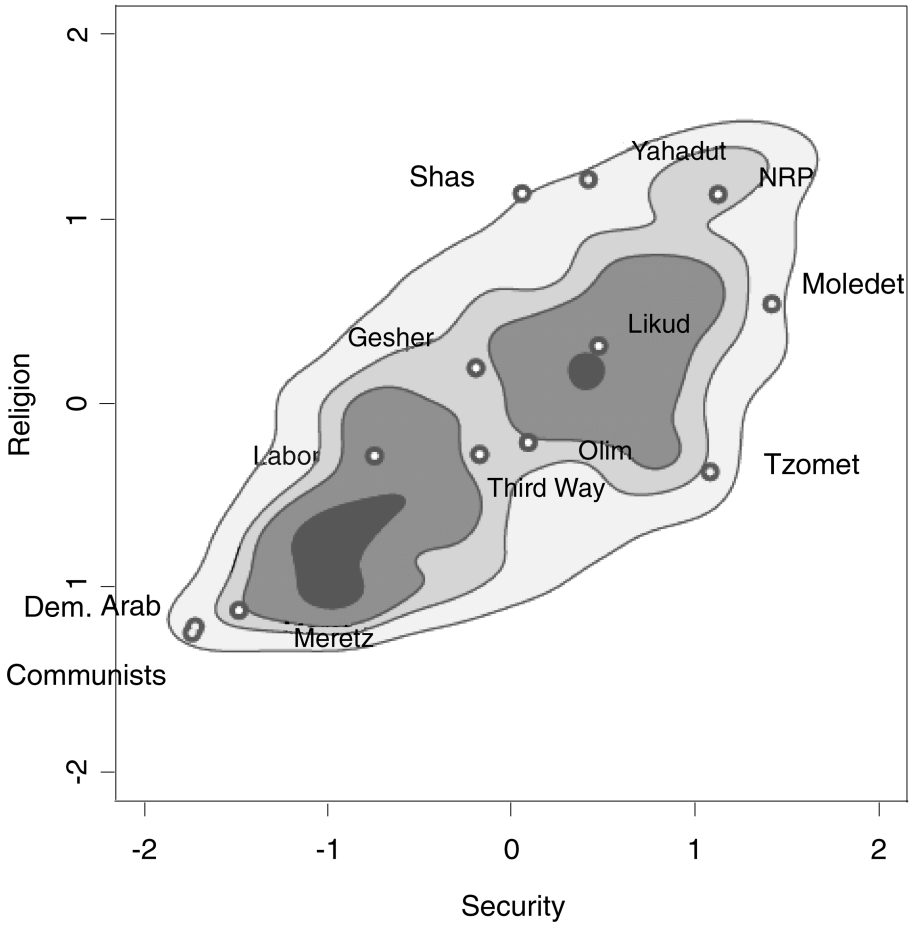

Fig. 1. Party positions in the Israeli Knesset in 1996, showing highest density plots of the sample voter distribution (at the 95, 75, 50 and 10 per cent levels)

voter $i \in N = \{1,\ldots n\}$ to party $j \in P = \{1,\ldots p\}$ in the voter's utility calculation. The model we adopt is given by:

$$u_{ij}(x_i, z_j) = \lambda_{ij} - \beta\|x_i - z_j\|^2 + \theta_j^T \mathrm{s}_i. \tag{3}$$

However, these weights are subjective, and may well be influenced by idiosyncratic characteristics of voters and parties. For empirical analysis, we shall assume $\lambda_{ij} = \lambda_j + \xi_{ij}$, where ξ_{ij} is drawn at random from a probability distribution for the variate ε_j, with expected value 0, and variance σ^2. The expected value of $\{\lambda_{ij}\}_j$ is λ_j, and so we write $\lambda_{ij} = \lambda_j + \varepsilon_j$, thus giving Equation 2. At each election, then, we assume that the *valence*, λ_j, of party j is exogenously determined.

Estimating the voter models given by Equation 2 requires information about sample voter behaviour. It is assumed that data exist about voter intentions: this information is encoded, for each sample voter i by the vector $y_i = (y_{i1},\ldots, y_{ip})$ where $y_{ij} = 1$ if and only if i intends to vote (or did indeed vote) for party j. Given the dataset $\{x_i, \mathrm{s}_i, y_i\}_N$ for the

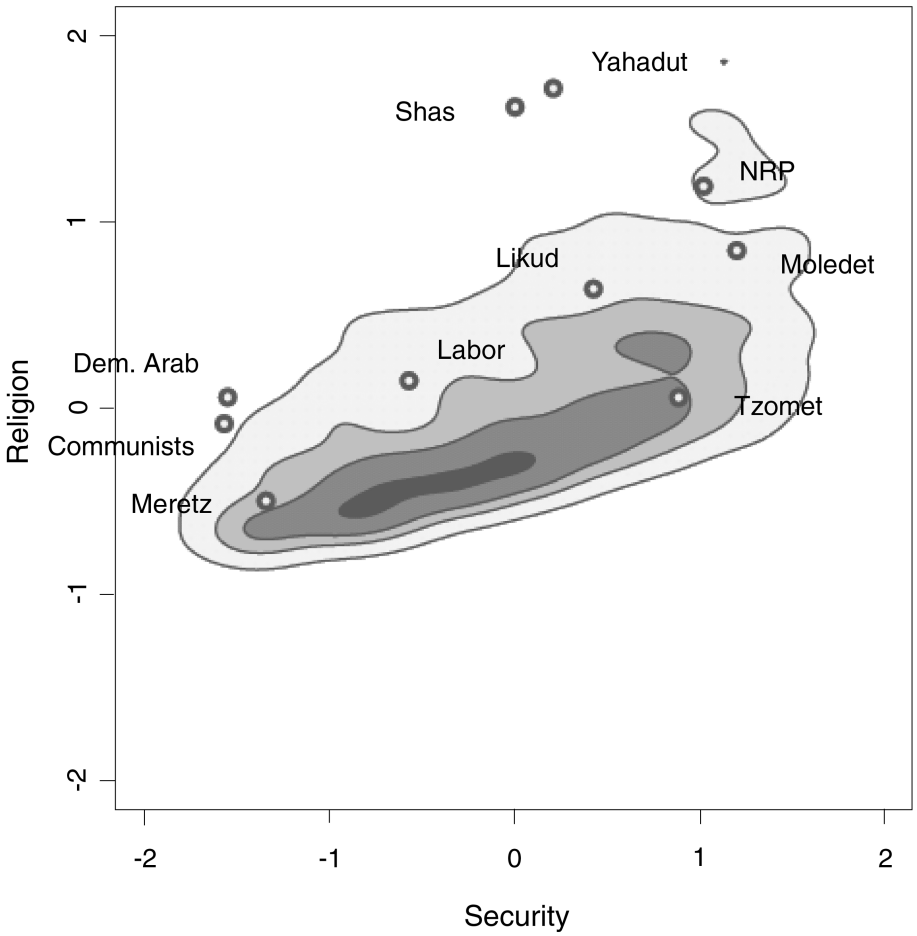

Fig. 2. Party positions in the Israeli Knesset in 1992, showing highest density plots of the sample voter distribution (at the 95, 75, 50 and 10 per cent levels)

sample N (of size n) and $\{z_j\}$, for the political agents, a set $\{\rho^*_i\}_N$ of stochastic variables is estimated. The first moment of ρ^*_i is the probability vector $\rho_i(\mathbf{z}) = \{\rho_{i1}(\mathbf{z}),\ldots, \rho_{ip}(\mathbf{z})\}$. Here $\rho_{ij}(\mathbf{z})$ is the probability that voter i chooses agent j when strategies are given by the vector $\mathbf{z}$.

There are standard procedures for estimating the model given by Equation 2. The technique is to choose estimators for the coefficients so that the estimated probability is:

$$\rho_{ij}(\mathbf{z}) = \text{Prob}[u_{ij}(x_i, z_j) > u_{il}(x_i, z_l) \text{ for all } l \in P\backslash\{j\}]. \tag{4}$$

Here, u_{ij} is the jth component of estimated latent utility function for i. The estimator for the choice is $\bar{y}_{ij} = 1$ if and only if $\rho_{ij} > \rho_{il}$ for all $l \in P\backslash\{j\}$. The procedure minimizes the errors between the n by p matrix $[\mathbf{y}]$ and the n by p estimated matrix $[\bar{\mathbf{y}}]$. The expected vote share function, $V_j(\mathbf{z})$, of party j, given the vector $\mathbf{z}$ of strategies, is defined to be:

$$V_j(\mathbf{z}) = (1/\text{n})\Sigma_i\rho_{ij}(\mathbf{z}). \tag{5}$$

TABLE 1 *Factor Analysis Results for Israel for the Election of 1996*

Issue question	Security	Religion
Chance for peace	0.49	–
	(0.02)	
Land for peace	0.87	–
	(0.01)	
Religious law vs. democracy	− 0.29	− 0.61
	(0.04)	(0.05)
Must stop peace process	− 0.66	–
	(0.02)	
Agreement with Oslo accord	− 0.84	–
	(0.01)	
Oslo accord contributed to safety	− 0.80	–
	(0.02)	
Personal safety after Oslo	− 0.76	–
	(0.02)	
Should Israel talk with PLO?	0.85	–
	(0.02)	
Opinion of settlers	0.89	–
	(0.02)	
Agree that Palestinians want peace	− 0.75	–
	(0.02)	
Will peace agreement end Arab–Israeli conflict	− 0.75	–
	(0.02)	
Agreement with Palestinian state	− 0.79	–
	(0.02)	
Should encourage Arabs to emigrate	0.62	–
	(0.02)	
What must Israel do to prevent war?	− 0.84	–
	(0.02)	
Settlements 1	− 0.71	–
	(0.01)	
Settlements 2	0.86	–
	(0.01)	
National security	0.55	–
	(0.02)	
Equal rights for Arabs and Jews in Israel	0.77	–
	(0.02)	
More government spending on religious institutions	− 0.89	–
	(0.04)	
More government spending on security	− 0.53	− 0.21
	(0.05)	(0.07)
More government spending on immigrant absorption	0.34	− 0.47
	(0.05)	(0.07)
More government spending on settlements	− 0.60	− 0.23
	(0.04)	(0.06)
More government spending in Arab sector	0.68	–
	(0.02)	
Public life should be in accordance with Jewish tradition (constant)	1.00	–
Views towards an Arab minister	− 0.75	–
	(0.02)	
Var(Security) (constant)	1.00	
Var(Religion)	0.73	
	(0.04)	
Cov(Security, Religion)	0.59	
	(0.02)	

Notes: The table is based on the Arian–Shamir survey and shows the factor loadings for Security and Religion, with standard errors in parentheses beneath each value.

TABLE 2 *Elections in Israel*

Party	Knesset seats				
	1988	1992	1996	1999	2003
Left					
Labor	39	44	34	28	19
Meretz	–	12	9	10	6
Others	9	–	–	–	3
Democrat	1	2	4	5	2
Communists	4	3	5	3	3
Balad	–	–	–	2	3
Subtotal	*53*	*61*	*52*	*48*	*36*
Centre					
Olim	–	–	7	6	2*
Third Way	–	–	4	–	–
Center	–	–	–	6	–
Shinui	2	–	–	6	15
Subtotal	*2*		*11*	*18*	*17**
Right					
Likud	40	32	30	19	38†
Gesher	–	–	2	–	–
Tsomet	2	8	–	–	–
Yisrael	–	–	–	4	7
Subtotal	*42*	*40*	*32*	*23*	*45*
Religious					
Shas	6	6	10	17	11
Yahadut	7	4	4	5	5
NRP	5	6	9	5	6
Moledet	2	3	2	4	–
Techiya	3	–	–	–	–
Subtotal	*23*	*19*	*25*	*31*	*22*
TOTAL	*120*	*120*	*120*	*120*	*120*

* Olim joined Likud to form one party immediately after the last election.
† Likud with Olim has forty seats.

Table 3 gives the results of the MNL estimation of Equation 3, for 1996, under the assumption that the errors follow a log Weibull distribution, with identical variance σ^2. (That is, the errors are assumed iid, while the log Weibull distribution is essentially a normal distribution with truncated tails.) As we have noted above, this log Weibull assumption on the errors results in the IIA property. In this instance it means that for any voter, i, and any two parties j,k, the ratio ρ_{ij}/ρ_{ik} is independent of any third party. Note that the electoral and stochastic normalizations allow us to identify the spatial coefficient, β, and the valence differences.

Instead of giving standard errors in Table 3, it is more useful for our analysis to give 95 per cent confidence intervals on all parameters. For example, Table 3 shows how the socio-demographic coefficients, θ_j, affect voting. Because it is only possible to identify valence differences between parties, we are at liberty to normalize the valences by choosing one party to have valence zero, and identifying the valences for the other parties appropriately. We chose to normalize with respect to Meretz and set the λ of Meretz to be 0. Meretz does not appear with respect to parameter estimates in Table 3, because

TABLE 3 *Multinomial Logit Analysis of the 1996 Elections in Israel*

	Party	Posterior mean	95% Confidence Interval	
			Lower bound	Upper bound
Spatial distance β	–	1.117	0.974	1.278
Constant λ	Shas	− 2.960	− 7.736	1.018
	Likud	3.140	0.709	5.800
	Labor	4.153	1.972	6.640
	NRP	− 4.519	− 8.132	− 1.062
	Moledet	− 0.893	− 4.284	2.706
	III Way	− 2.340	− 4.998	0.411
Ashkenazi	Shas	0.066	− 1.831	1.678
	Likud	− 0.625	− 1.510	0.271
	Labor	− 0.219	− 0.938	0.492
	NRP	1.055	− 0.206	2.242
	Moledet	0.819	− 0.560	2.185
	III Way	− 0.283	− 1.594	1.134
Age	Shas	0.014	− 0.058	0.086
	Likud	− 0.025	− 0.063	0.008
	Labor	− 0.041	− 0.077	− 0.012
	NRP	− 0.064	− 0.111	− 0.020
	Moledet	− 0.026	− 0.088	0.026
	III Way	0.014	− 0.034	0.063
Education	Shas	− 0.377	− 0.693	− 0.063
	Likud	− 0.032	− 0.180	0.115
	Labor	0.011	− 0.099	0.120
	NRP	0.386	0.180	0.599
	Moledet	0.050	− 0.219	0.305
	III Way	− 0.067	− 0.298	0.150
Religious observation	Shas	3.022	1.737	4.308
	Likud	0.930	0.270	1.629
	Labor	0.645	0.077	1.272
	NRP	2.161	1.299	3.103
	Moledet	0.897	− 0.051	1.827
	III Way	0.954	0.031	1.869
Percentage correctly predicted	Shas	0.309	0.210	0.414
	Likud	0.707	0.672	0.740
	Labor	0.717	0.681	0.752
	NRP	0.408	0.324	0.493
	Moledet	0.078	0.046	0.115
	III Way	0.029	0.017	0.043
	Meretz	0.285	0.226	0.349
Entire model		0.638	0.623	0.654

Note: n = 794.

to identify the model all coefficients for Meretz are set to 0. It is important for our analysis that the confidence intervals for the valences of Labor and Likud are in the positive domain, while that for NRP is negative. (It would be more accurate to say that the identified valence difference between Labor or Likud and Meretz is positive, while that between NRP and Meretz is negative.)

Note also from Table 3 that the socio-demographic coefficients (θ_j) are statistically relevant. For example, 'religion' strongly affects the likelihood of voting for Shas or NRP. We denote this empirical model 'Joint MNL'. Table 3 reports the 'correct prediction' of the model to be approximately 64 per cent, with particularly high correct predictions for Likud (71 per cent) and Labor (72 per cent). Here a correct prediction is made if the voter chooses the party with the highest estimated probability.

To validate the model and to verify that the implicit IIA assumption was not unduly restrictive, we compared the model with a MNP model, where the IIA restriction does not apply. A virtue of using the general voting model of Equation 2 is that the Bayes factors (or difference in log likelihoods) can be used to determine which of various possible models is statistically superior.[39] We compared the following models:

(i) A pure 'spatial MNP' model, with $\{\theta_j\} = 0$ and $\{\lambda_j\} = 0$.
(ii) A pure 'spatial MNL' model, called MNL1 with $\{\theta_j\} = 0$ and $\{\lambda_j\} = 0$.
(iii) A 'spatial valence MNL' model, called MNL2 with $\{\theta_j\} = 0$ but non-zero valence.

Table 4 gives the usual interpretation of the significance of the Bayes factors, or log likelihood ratios, while Table 5 gives the Bayes factors for the comparisons we performed.

It is evident that MNL2 is strongly preferred to MNL1. Valence adds considerable predictive power. Adding the socio-demographic components to give the 'joint MNL' increases the power further. It is interesting that the more general MNP model without valence is beaten by the MNL model with valence.

TABLE 4 *Interpretation of Evidence Provided by the Bayes Factor B_{jk}*

$\ln(B_{jk})$	B_{jk}	Evidence in favour of M_j
0–1	1–3	Not worth more than a mention
1–3	3–20	Positive
3–5	20–150	Strong
>5	>150	Very strong

TABLE 5 *Bayes Factor* $[\ln(B_{jk})]$ *for Model j vis-à-vis Model k, Election 1996*

		M_k			
		Spatial MNP	Spatial MNL1	Spatial MNL2	Joint MNL
M_j	Spatial MNP	n.a.	238***	− 16	− 49
	Spatial MNL1	− 238	n.a.	− 254	− 287
	Spatial MNL2	16***	254***	n.a.	− 32
	Joint MNL	49***	287***	32***	n.a.

***Very strong support for M_j.

[39] See Quinn, Martin and Whitford, 'Voter Choice in Multiparty Democracies', as well as Robert Kass and Adrian Raftery, 'Bayes Factors', *Journal of the American Statistical Association*, 90 (1995), 773–95, for a justification of this method.

TABLE 6 *National and Sample Vote Shares and Valence Coefficients for Israel, 1988–96*

	1988		1992		1996		1988	1992	1996
	National	Sample	National	Sample	National	Sample	λ*	λ*	λ*
Labor	30.7	28.5	36.0	35.8	27.5	44.0	− 0.30	0.91	4.15
Meretz	4.4	8.3	10.0	11.9	7.6	6.0	0	0	0
Dem. Arab	1.0	–	1.6	–	3.0	–	–	–	–
Communist	2.8	–	2.5	–	4.3	–	–	–	–
Olim	–	–	–	–	5.8	–	–	–	–
Third Way	–	–	–	–	3.2	1.8	–	–	− 2.34
Likud	31.8	49.7	26.2	30.2	25.8	43.0	2.84	2.73	3.14
Tzomet	2.0	2.5	6.7	9.6	–	–	− 0.75	0.38	–
Shas	4.8	3.6	5.2	3.6	8.7	2.0	− 5.78	− 4.67	− 2.96
Yahadut	–	–	3.3	–	3.3	–	–		–
NRP	4.0	2.9	5.2	4.6	8.0	5.1	− 3.00	− 0.44	− 4.52
Moledet	1.5	–	2.5	4.4	2.4	1.8	–	0.37	− 0.89
Techiya	3.2	4.4	–	–	–	–	0.39	–	–
Others	8.0	–	0.8	–	0.4	–	–	–	–
β coefficients							1.32	1.25	1.12
Log marginal likelihoods							− 497	− 834	− 465
Sample size							505	781	794
Percentage of vote correctly predicted							52	46	64

* 0 indicates Meretz was chosen to have zero valence in all models.

We constructed similar 'joint MNL' models for 1988 and 1992. Table 6 presents the results of the estimation for the joint MNL model for the three elections of 1988, 1992 and 1996, giving the valence coefficients $\{\lambda_j\}$ and spatial coefficients β. While the estimation only obtained results for seven parties, the high marginal log likelihoods suggested that the joint MNL model provided a close approximation to voter response in Israel. Since one of our purposes in constructing the empirical model was to examine the mean voter theorem, and the proof of this result assumes iid errors, it was appropriate to maintain the assumption of independent errors.

It should also be observed that it was important to include the socio-demographic component (SD) in our model because the purpose of the exercise was to estimate valence values as accurately as possible. In general, the models with a SD give lower valence coefficients than the model without a SD. Because the socio-demographic component of the model was assumed independent of party strategies, we could use the estimated parameters of the joint MNL model to simulate party movement in order to increase the expected vote share of each party. That is, keeping the socio-demographic component fixed, we could determine how vote shares varied as party positions were changed.[40] 'Hill-climbing' algorithms were used for this purpose. Such algorithms involve small changes in party position, and are therefore only capable of obtaining 'local' optima for each party.[41] Consequently, a vector $\mathbf{z}^* = (z_1^*, \ldots, z_p^*)$ of party positions that results from such a search is what we call a 'local pure strategy Nash equilibrium'. We give formal definitions of this local notion below, and contrast it with the more usual idea of Nash equilibrium.[42]

Figure 3 shows one local equilibrium resulting from the simulation exercise for the joint MNL model for 1996. The simulation for 1996 found five distinct local vote-maximizing vectors, all essentially similar to Figure 3, and differing from Figure 3 only by permutation of the positions of the low valence parties. For 1992 just two vote-maximizing vectors of party positions were located. Both had the positions of Likud and Labor almost identical to their positions in Figure 2.

It has been presumed that vote-maximizing locations for the MNL model would occur at the 'electoral mean', namely the point (0,0) in Figures 1 and 2. None of the equilibria obtained by simulation was characterized by all parties adopting the mean voter position. This demonstrates that a MNL model does not satisfy the assumptions of the 'mean voter theorem'.[43] The simulation also suggests that the inferences about convergence to the electoral mean in the formal model are unfounded.[44]

We make some brief remarks about the equilibria located in Figure 3, and then elaborate on the formal model in the next section. The analysis to follow in that section explains the simulation result in terms of the differing valences of the parties as shown in Table 6.

[40] This is a standard assumption in such simulation exercises. For example, see Quinn and Martin, 'An Integrated Computational Model'.

[41] See Merrill and Adams, 'Computing Nash Equilibria'.

[42] Schofield and Sened, 'Local Nash Equilibrium in Multiparty Politics'.

[43] The assertion that the MNL model satisfies the conditions of the 'mean voter theorem' can be found in Adams, *Party Competition*. A numerical example suggesting otherwise has been offered by Jeremy Staum, 'Ideological Platforms and Probabilistic Voting Equilibria' (unpublished, Cornell University, 2001).

[44] Hinich, 'Equilibrium in Spatial Voting'; Enelow and Hinich, *The Spatial Theory of Voting*; Coughlin, *Probabilistic Voting Theory*; Lin *et al.*, 'Equilibrium in Multicandidate Probabilistic Spatial Voting'.

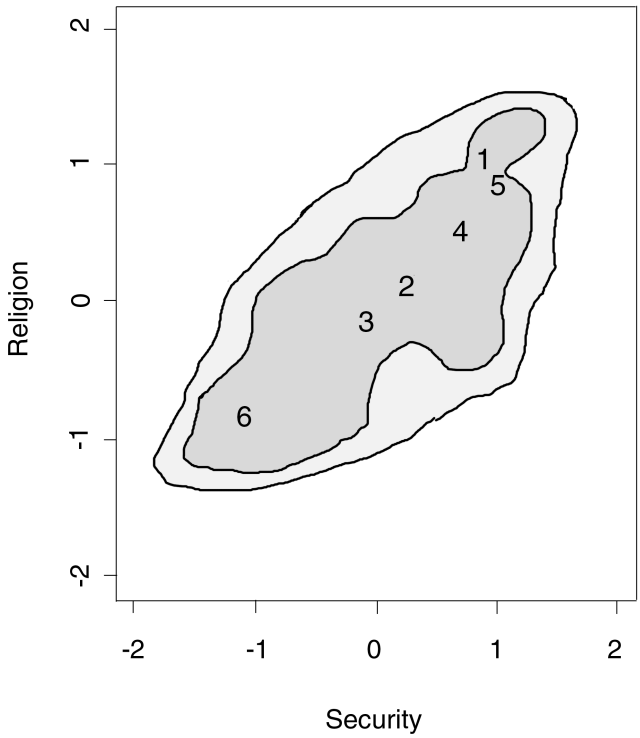

Fig. 3. Simulated local equilibrium party positions in the Israel Knesset, 1996
Key: Shas = 1, Likud = 2, Labor = 3, NRP = 4, Molodet = 5, Meretz and Third Way = 6.

Conclusion from the Simulation

(i) The two high valence parties, Labor and Likud, are seen to occupy positions in the equilibria in Figures 3 that are identical, or almost identical, to their estimated positions in Figure 1. Although we do not report the results of the simulation exercise for 1992 and 1988, the same is true for the models and simulations for these earlier elections. We infer that the vote maximizing assumption is valid for these two parties and for the elections that we have studied. However, the parties are not at identical positions at the mean, but symmetrically located about the mean, on the major principal axis of the voter distribution.

(ii) In the equilibrium figure, the low valence parties (such as Shas or NRP) are scattered apparently at random along the major principal axis. Parties with very low valence are positioned furthest from the electoral mean. The two equilibrium pictures for 1992 and the five for 1996 are essentially all permutations of each other.

(iii) The estimated actual positions of some of the low valence parties in Figures 1 and 2 do not correspond with the simulated equilibrium locations in Figure 3 for 1996, nor with the equilibrium positions we obtained for 1992.

In the next section we examine the formal vote model, and show why we would expect (i) and (ii) under a vote maximization model. This leads us to propose extensions of the vote maximization model to account for (iii).

2. THE FORMAL ELECTORAL MODEL

Our purpose in introducing the formal model is to examine conditions under which the mean voter theorem holds. The distribution assumptions of the formal model and of the MNL model are not identical. Nonetheless, simulation of the empirical model showed that convergence did not occur. We therefore expect, with the parameters given in Table 6, that divergence will occur in the model. We first give formal definitions of the equilibrium concepts that we use. These definitions are applicable both to an empirical model, of the kind constructed above, or to a formal model. In these definitions, a Nash equilibrium is the most restrictive. A weaker equilibrium notion is that of local equilibrium. It is evident from the definitions that a Nash equilibrium must also be a local equilibrium. Since vote shares are differentiable, we can consider critical points of these functions. A local equilibrium satisfies the first-order condition, and must therefore be what we call a 'critical Nash equilibrium'. In addition, it must satisfy so-called second-order conditions on the Hessian.

DEFINITION 1. A *Pure Strategy Nash Equilibrium* (PNE), under expected vote maximization, is a vector $\mathbf{z}^* = (z_1^*, \ldots, z_p^*)$, such that, for each $j \in P$, $V_j(z_1^*, \ldots, z_j^*, \ldots, z_p^*) \geq V_j(z_1^*, \ldots, z_j, \ldots, z_p^*)$ for every $z_j \in \mathbf{X}$. We say that z_j^* is a best response to the vector $\mathbf{z}_{-j}^* = (z_1^*, \ldots, z_{j-1}^*, z_{j+1}, \ldots, z_p^*)$.

DEFINITION 2. Assume that each expected vote function V_j is twice differentiable.

(i) A vector $\mathbf{z}^* = (z_1^*, \ldots, z_p^*)$ is a *critical Nash equilibrium* (CNE), under expected vote maximization, if, for each $j \in P$, the vector $\mathbf{z}^*$ satisfies the first order condition $dV_j/dz_j = 0$ at $z_j = z_j^*$ (under the restriction that z_k^*, for all $k \neq j$, are kept fixed).

(ii) A vector $\mathbf{z}^* = (z_1^*, \ldots, z_p^*)$ is a *local Nash equilibrium* (LNE) if $\mathbf{z}^*$ is a CNE and, in addition, for each $j \in P$, there exists an open neighbourhood U_j of z_j^* in X such that $V_j(z_1^*, \ldots, z_j^*, \ldots, z_p^*) \geq V_j(z_1^*, \ldots, z_j, \ldots, z_p^*)$ for all z_j in U_j. Thus, z_j^* is a weak best response to $(z_1^*, \ldots, z_{j-1}^*, z_{j+1}^*, \ldots, z_p^*)$ in the neighborhood, U_j. We say z_j^* is a *weak local* best response.

(iii) A vector $\mathbf{z}^*$ is a *strict* LNE, written LSNE, if and only if it is an LNE and, for all j, all the eigenvalues of the Hessian of V_j are negative. We say z_j^* is a *strict local* best response.

Obviously, a PNE must be a LNE, but not conversely. We use the notion of LSNE because it avoids problems with zero eigenvalues. If we strengthen the idea of Nash equilibrium by requiring that equilibrium strategies are strictly better than other strategies, then we obtain strict PNE, written PSNE. Obviously a PSNE must be a LSNE. One condition that is sufficient to guarantee that a LNE is a PNE for the electoral game is *concavity* of the estimated vote functions.

DEFINITION 3. A real valued function f: $\mathrm{X} \rightarrow \Re$ is *concave* if and only if, for any real number a, and any x, y in X, $f(\mathrm{a}x + (1 - \mathrm{a})\, y) \geq \mathrm{a}f(x) + (1 - \mathrm{a})\, f(y)$.

Concavity of the 'payoff' functions $\{V_j\}$ in the jth strategy z_j, together with continuity in z_j and compactness and convexity of X is sufficient for existence of PNE.[45]

[45] Banks and Duggan, 'The Theory of Probabilistic Voting'.

We now discuss the 'mean voter theorem' of the formal model. As mentioned above, this theorem asserts that the vector $z^* = (x^*,\ldots, x^*)$, where x^* is the mean of the distribution of bliss points, is a PNE for the vote maximizing formal game.[46] It is usual to assume that the vote functions are concave, thus guaranteeing existence of PNE.

In an empirical model, the motivation is to provide an understanding of voter response to the vector of party positions. In principle, the various parties could use opinion polls and the like to model electoral response, and then modify their party declarations in order to increase vote share. The point of a formal model is to assume that the electoral response is common knowledge. However, electoral information is fundamentally local, in the sense that knowledge of the relationship between vote share and party position cannot, in fact, be known across all possible vectors of positions. This suggests that parties only consider small changes in the positions that they adopt. This provides the motivation for our focus on the local Nash equilibrium concept.

As indicated, conditions for existence of a PNE in such a formal model emphasize continuity and concavity of each V_j. Because of the assumptions made in constructing both the formal and empirical models, the voter utility functions $u_i(x_i, -)$: $X^P \rightarrow \Re^P$ are differentiable in all variables, $\{z_j\}$. Therefore, so are $\{\rho_{ij}\}$, and consequently, so are $\{V_j\}$.

To discuss the mean voter theorem, first let $x^* = (1/n)\ \Sigma x_i$ denote the mean of the voter ideal points. Then the mean voter theorem for the formal stochastic model asserts that $\mathbf{z}^* = (x^*,\ldots, x^*)$ is a PNE (under certain restrictions that are usually assumed to be relatively weak).[47] The presumed restriction is that the stochastic variance, σ^2, be 'sufficiently large'. Now concavity of the vote share function is equivalent to the property that its second differential (its Hessian) is negative semi-definite everywhere. We can therefore test for concavity by examining the Hessian when all parties are at the electoral mean. This allows us to obtain necessary and sufficient conditions for the 'joint mean vector', $\mathbf{z}^*$, to be a LNE. If the necessary condition fails then this vector cannot be a LNE, and thus certainly cannot be a PNE. The condition on the vote share functions, induced by the Hessian condition when all parties are at the mean, we shall call 'local concavity'.

To state the theorem we first choose a system of orthogonal axes indexed by $t = 1, \ldots, w$ where w is the dimension. Now let $\chi_t = (\ldots, x_{it},\ldots)$ be the n vector whose components are the coordinates of the voter ideal points in dimension t. Without loss of generality we can choose the coordinate system so that $\Sigma_i(x_{it}) = 0$, for $t = 1, \ldots, w$. As before, we use x^* to denote the vector mean $x^* = (\ldots\ (1/n)\ \Sigma_i(x_{it}), \ldots)$, so that in the new coordinate system $x^* = (0, \ldots, 0)$. We shall write this vector more simply as 0. Then the voter distribution is given by the vector $(\chi_1, \ldots, \chi_t, \ldots, \chi_w)$. Now define (χ_t, χ_s) to be the scalar product of the two vectors χ_t and χ_s and consider the w by w matrix, D, whose entry in the *(t,s)* position is (χ_t, χ_s). Then $v_t^2 = (1/n)\Sigma_i(x_{it})^2) = (1/n)(\chi_t, \chi_t) = (1/n)\|\chi_t\|^2$ is the variance of the voter ideal points about the origin on the t axis. This is simply the diagonal term in position t in the matrix $\Delta = (1/n)D$. The off-diagonal term in Δ is clearly the covariance between the vectors χ_t and χ_t. The matrix Δ is termed the *voter variance-covariance matrix*. The electoral normalization in empirical models sets the variance on the first axis, v_1^2, to be

[46] Hinich, 'Equilibrium in Spatial Voting'; Enelow and Hinich, *The Spatial Theory of Voting*; Coughlin, *Probabilistic Voting Theory*; Lin *et al.*, 'Equilibrium in Multicandidate Probabilistic Spatial Voting'.

[47] See Lin *et al.*, 'Equilibrium in Multicandidate Probabilistic Spatial Voting'; and James Adams, 'Multiparty Spatial Competition with Probabilistic Voting', *Public Choice*, 99 (1999), 259–74; James Adams, 'Policy Divergence in Multicandidate Probabilistic Spatial Voting', *Public Choice*, 100 (1999), 103–22.

1.0. We use $v^2 = \Sigma_t\, v_t^2$ to denote the total of these voter variance terms in all w dimensions. This is the sum of the diagonal terms in Δ, and is also known as the trace of the matrix Δ. What we previously called the 'principal components' of the electoral data are embodied in Δ. Given the system of known parameters, and assuming that the errors $\{\varepsilon_j\}$ are identically independently and normally distributed (iind) with variance-covariance matrix, $\mathbf{I}\sigma^2$ (where $\mathbf{I}$ is the p by p identity matrix), we define the formal vote model in analogous fashion to the empirical model.

DEFINITION 4. The formal vote model $\mathrm{M} = \mathrm{M}(\beta,\lambda : \sigma^2, \Delta)$ is obtained in the following way:

(i) $$u_{ij}(x_i, z_j) = \lambda_j - \beta \,\|x_i - z_j\|^2 + \varepsilon_j, \text{ for each } i,j$$

(ii) $$\rho_{ij}(\mathbf{z}) = \mathrm{Prob}[u_{ij}\,(x_i, z_j) > u_{il}(x_i, z_l) \text{ for all } l \in P\backslash\{j\}]$$

(iii) $$V_j(\mathbf{z}) = (1/\mathrm{n})\Sigma_i\, \rho_{ij}(\mathbf{z}) \text{ for each } j \tag{6}$$

(iv) The valence terms are given by the vector $\lambda = (\lambda_p, \lambda_{p-1}, \ldots, \lambda_1)$ and ranked $\lambda_p \geq \lambda_{p-1} \geq \ldots \geq \lambda_1$

(v) The errors $\{\varepsilon_j\}$ are iind, with variance-covariance matrix, $\mathbf{I}\sigma^2$. The 'corrected' variance of the model is given by $\kappa^2 = \mathrm{p}\sigma^2/[\mathrm{p} - 1]$.

We will also consider the more general model with multivariate normal errors, characterized by an error difference variance-covariance matrix Ω. This more general formal model we denote as $\mathrm{M}(\beta,\lambda{:}\Omega,\Delta)$.The Appendix shows that the following theorem also holds for this model as long as κ^2 is defined in terms of the full error covariance matrix.

We now seek conditions under which the *joint origin* $\mathbf{z}_0^* = (0,\ldots,0) \in \mathrm{X}^{\mathrm{p}}$ will be a LSNE.

LEMMA: The joint origin, $\mathbf{z}_0^*$, is a CNE of the model $\mathrm{M}(\beta,\lambda : \sigma^2,\Delta)$.

The second order conditions can be expressed in terms of a 'convergence coefficient'.

DEFINITION 5. Given the vector $\lambda = (\lambda_p, \lambda_{p-1}, \ldots, \lambda_1)$ of exogeneous valences, define $\lambda_{av(1)} = [1/(\mathrm{p} - 1)]\ \Sigma_{j=2}\ \lambda_j$ and let $\Lambda = \lambda_{av(1)} - \lambda_1$ be the *valence difference*.
For the model $\mathrm{M}(\beta,\lambda,\sigma^2,\Delta)$, define the *convergence coefficient* $\mathrm{c} = \mathrm{c}\ (\beta,\lambda,\sigma^2,\Delta)$ by

$$\mathrm{c}\ (\beta,\lambda,\sigma^2,\Delta) = 2\beta\Lambda\ v^2/[\kappa^2]. \tag{7}$$

ELECTORAL THEOREM. Existence of a local equilibrium at the joint origin in the model $\mathrm{M}(\beta,\lambda_2,\sigma^2,\Delta)$. Suppose that the policy space X is a closed bounded domain in Euclidean space of dimension w. Then, $\mathbf{z}_0^* = (0, \ldots, 0)$ is a LSNE if:

$$\mathrm{c}\ (\beta,\lambda : \sigma^2, \Delta) < 1 \text{ and is a LSNE only if } \mathrm{c}\ (\beta,\lambda : \sigma^2, \Delta) < w.$$

Note that a LSNE is an LNE, so the given sufficient condition is also sufficient for $\mathbf{z}_0^*$ to be a LNE. It can also be shown that $w \geq \mathrm{c}$ is a necessary condition for $\mathbf{z}_0^*$ to be a LNE.

It is important to note that the expression given in Equation 7 is dimensionless. In empirical applications, the scale of the model is indeterminate, so it is natural to normalize by fixing the stochastic variance at some value σ^2. Because only the variances of the error *differences* are identifiable, and these are given by $2\sigma^2$, this sets the scale of the model. This stochastic normalization, together with the electoral normalization $v_1^2 = 1.0$ sets the scale for Figures 1, 2 and 3. It is also the case that the valences can only be identified up

to a constant. Only valence differences are relevant, so with these normalizations, the product $\beta\Lambda$ is identifiable. One way to interpret the theorem is that the model is 'classified' by the identifiable product $2[\beta\Lambda][v/\kappa]^2$. Only if this product is bounded above by the dimension w can the 'local concavity condition' be satisfied at the joint origin. This is the necessary condition for the 'mean voter theorem' to be valid.[48]

Notice that in the case studied by Lin *et al.*, with all $\lambda_j = 0$, the local concavity condition is always satisfied.[49] It follows that the joint origin can always be assured of being a LNE. This does not, of course, imply that the joint origin is a PNE. In the case with non-zero valences, if the product, $\beta\Lambda$, is 'large', or if σ^2 is 'small' relative to the voter variance v^2, then the origin cannot be a LNE and therefore cannot be a PNE. Note also that as the number of parties, p, increases, then the ratio p/(p − 1) will decrease, and c will increase, so *ceteris paribus* the likelihood of satisfaction of the local concavity condition diminishes.

It is obvious that as $\sigma \to 0$, then $\kappa^2 \to 0$, so the origin will generally be neither a LNE nor a PNE. Banks and Duggan have also observed that, for the case $w \geq 2$, as $\sigma \to 0$, then the condition sufficient for existence of PNE, in two-party competition, will fail.[50] In a sense the Electoral Theorem gives a generalization of this two-party result, since the necessary condition of the theorem imposes an upper bound on the identifiable product $2[\beta\Lambda][v/\kappa]^2$. Thus the theorem indicates that, when the voter ideal points are not restricted to a relatively small domain of the policy space, then the centrist PNE will fail to exist. Since there are many possible solutions to the first-order condition (and therefore many CNE), the theorem suggests that many local, but non-centrist, equilibria can be found.

The principal difficulty in the proof of necessary and sufficient conditions (as shown in the Appendix) is that the expression for the probability given in Equation 6 is a multivariate normal integral. The Appendix shows that there is an orthogonal transformation that facilitates differentiation. This technique also shows that the analysis can be performed in the very general case when the errors are multivariate normal with general covariance matrix Ω. Thus, if the matrix Ω is estimated by MCMC methods in a MNP model, we can obtain an analogous convergence coefficient for the formal model by defining κ^2 in terms of the sum of terms in the error difference covariance matrix.

In the case of iind errors, the Appendix shows that the Hessian for the lowest valence party (at the joint origin) is given by the matrix:

$$C(\beta,\lambda : \sigma^2, \Delta) = 2\beta\Lambda\ \Delta\ /[\kappa^2] - I. \tag{8}$$

Here I is the w by w identity matrix.

The Appendix also indicates how the analysis for iind errors can also be developed for the model $M(\beta,\lambda{:}\Omega,\Delta)$. In principle there is little difference between the conditions for the two models.[51]

The condition that the matrix $C(\beta,\lambda : \sigma^2, \Delta)$ has negative eigenvalues imposes a sufficient condition from the determinant, and a necessary condition from its trace. When these

[48] For the proof of the theorem, see Norman Schofield, 'Equilibrium in the Spatial "Valence" Model of Politics', *Journal of Theoretical Politics*, 16 (2004), 447–81; and Norman Schofield, 'Local Political Equilibria', in Austen-Smith and Duggan, eds, *Social Choice and Strategic Decisions*, pp. 57–91.

[49] Lin *et al.*, 'Equilibrium in Multicandidate Probabilistic Spatial Voting'.

[50] Banks and Duggan, 'The Theory of Probabilistic Voting in the Spatial Model of Elections'.

[51] For the general model with multivariate normal errors, it is necessary to modify the expression in Equation 7 and use the average of the covariances of the error differences. See Norman Schofield, 'Social Choice and Elections', forthcoming in James Alt, John Aldrich and Arthur Lupia, eds, *A Positive Change in Political Science* (Ann Arbor: University of Michigan Press).

conditions are satisfied for the lowest valence party, then they will be satisfied for all parties. The eigenvalues can be readily calculated from this matrix, using the electoral data encoded in the covariance matrix Δ. As we have noted, it is clear from Figures 1 and 2 that in each case there is a major axis, or principal component, and a minor axis, along which the electoral variance is much smaller. The analysis of the Appendix shows how the eigenvalues of C can be computed from information about the variances on both axes, and the covariance between them. For example, if the covariance between the two axes is very high, then typically the eigenvalue on this 'major axis' will be large and positive, while the eigenvalue on the minor axis will be small in value, whether positive or negative. The gradient of the vote share function will then be aligned along the major axis, so that a vote maximizing party should then move on the major axis, away from the origin. Whether the party should move up or down the major axis is not determined.

Computation of LNE then becomes relatively simple, since all parties should position themselves at different positions on the major axis. We saw this phenomenon in the simulation exercise in the previous section. Although the distribution assumptions of the formal model as presented here are not identical to the distribution assumptions of the empirical MNL model, it is clear that there is agreement between the predictions of the formal model as regards location of LNE, and the simulation of the empirical model. We now examine this agreement in detail.

3. COMPARISON OF THE FORMAL AND EMPIRICAL MODELS FOR ISRAEL

An examination of the electoral data for 1996, given in Figure 1, shows that the normalized electoral variance on the security axis is 1.0, while on the religion axis the variance is 0.732. The covariance is 0.591, so, r^2 is 0.477. Figure 1 shows there is a principal component, or major axis aligned at 45° to the security axis. To determine whether the eigenvalues are positive or negative, we compute the convergence coefficient c to be 31.4. As in Definition 5, let $\Lambda = \lambda_{av(1)} - \lambda_1$, where we use '1' to denote the lowest valence party, NRP, as obtained from the MNL model for 1996. Computation readily shows that $\Lambda = 4.8$. To determine whether the eigenvalues are positive or negative, we compute the convergence coefficient c to be 9.5. Thus the *necessary condition* for the mean voter theorem fails. We can use the technique of the Appendix to determine the eigenvalues of the Hessian for the lowest valence party, the NRP, when all parties are at the origin. The two eigenvalues can be calculated to be $+7.05$ and $+0.45$. In particular, the 'eigenvectors' of the two eigenvalues are (1, 0.8) for the major eigenvalue and (-1, 1.25) for the minor eigenvalue. Essentially, these eigenvectors correspond to what we have called the major principal axis, and the orthogonal minor axis of the electoral data. What this means is that the NRP vote function increases most rapidly when the NRP moves up (or down) the principal axis, away from the origin. All other low valence parties should also move away from the origin in this direction. Notice that the simulation Figure 3 shows this to be the case.

In 1992, as Table 3 indicates, the lowest valence party was Shas ($\lambda_1 = -4.67$) and the highest Likud ($\lambda_5 = 2.34$). This gives $\Lambda = 5.3$. Since the total electoral variance is 1.435, with a variance of 1.0 on the security axis, and 0.435 on the religion axis, the convergence coefficient can be calculated to be 9.72. Again, the necessary condition of the theorem fails. Using the fact that the covariance between the two axes is 0.453 and the correlation coefficient, r^2, is 0.471, the eigenvalues for Shas can be calculated to be $+7.47$ and 0.24. A slight difference between 1992 and 1996 is that the major eigenvector for Shas in 1992 is (1.0, 0.55). A way to interpret this is that the most preferred direction for Shas to take

is along the principal axis through the origin and the policy position (1.0, 0.55). Again, the simulation for 1992 shows that all the parties were estimated to adopt equilibrium positions on this principal axis.

Notice that the calculations show that the vote share function for NRP in 1996 not only has a *local minimum* in 1996 if it and all other parties were at the origin, but its vote share function has a global minimum in this situation. The NRP can move away from the origin either up or down the gradient axis. A similar conclusion can be drawn for Shas and the other low valence parties in 1992.

In both years, the theorem shows that the joint origin (0, …, 0) can be neither a PNE nor a LNE. Clearly, if a high valence party like Likud or Labor occupies the mean voter position, then the low valence parties would find that their vote share functions could be increased by vacating the electoral centre. The formal model does not specify whether the shift of position should be up or down the major axis. However, any movement by low valence parties means that the first-order condition for Labor or Likud to occupy the electoral centre would not be satisfied

The simulation of vote maximizing positions given in Figure 3 is compatible with these observations. All parties were able to increase vote shares by moving away from the origin, along the major axis. Because their eigenvalues are large and positive at the origin, low valence parties, such as the NRP and Shas, must move far from the electoral centre. As Figure 1 indicates, and the model suggests, the very large positive eigenvalue for the NRP on the major axis did indeed drive it to a vote maximizing position in the 'north-east' quadrant.

All other parties were then able to increase vote shares by moving away from the origin, along the principal axis. Because the valences of Labor and Likud were much higher than other parties, their optimal positions would be relatively close to, but not identical to, the electoral mean.

Figure 3 also suggests that every party, in local equilibrium, would adopt a position that maintained a minimum distance from every other party. Our formal analysis as well as the simulation exercise, suggests that this minimum distance depends on the valences of the neighbouring parties. In particular, as the theorem indicates, what is relevant is the valence difference of the party (that is, the average of the valences of the other parties minus the party's own valence). Obviously, the configuration of equilibrium party positions will fluctuate as the valences of the large parties, particularly, change in response to exogenous shocks. The logic of the model is that the low valence of some parties obliges them to adopt radical positions, simply in order to maximize their vote shares.

In comparing the configurations of party positions for 1996 (Figure 1) and 1992 (Figure 2), it is evident that while there is some change in the positions of the parties, the only significant transformation is the *appearance* of small parties such as Olim and the Third Way in 1996. We included the Third Way in the estimation for 1996. As Table 6 indicates, the valence of this party was very low. Parties representing new political interests may enter the political contest near the electoral mean, in the hope of offering centrist policies to the electorate, or in the hope of securing perquisites from government membership. As indicated by the example of the Third Way, such parties may be short lived.

We are now in a position to use the formal model in order to draw out the differences between the estimated positions given in Figures 1 and 2, and the simulated equilibria given in Figure 3:

(i) The locations of the high valence parties, Labor and Likud, in any vote maximizing equilibrium should be close to the principal axis, and 'symmetrically' positioned about the origin. All the simulations match this prediction of the formal model (and Figure 3 illustrates this symmetry). The estimated positions of the two parties in Figures 1 and 2 correspond to this prediction.

(ii) In general, the smaller the valence of a party, the further away from the origin it should be located. For very low valence parties, such as NRP and Shas, their second eigenvalue is positive, so they may move off the principal axis. The NRP position is compatible with this. However, because the eigenvalue on the minor axis will be much smaller than the eigenvalue on the major axis, and possibly negative, we can expect all local equilibrium positions to be on, or very close to, the major axis. This provides an explanation for the positions of Meretz, the Democratic Arab Party, the Communist Party and NRP. One of the simulations did show Meretz at an equilibrium position far up the principal axis, in the 'north-east' quadrant, rather than its actual position in the 'south-east' quadrant. The model itself cannot predict whether parties will move up or down the major axis.

(iii) The low valence parties, Gesher and Third Way, can be seen in Figures 1 and 2 to be at centrist positions. It follows immediately from the analysis that these positions cannot maximize vote shares. Because the valences of these parties are low, their locations, close to Likud and Labor, mean they will obtain very low vote shares.

(iv) The four low valence parties, Shas, Yahadut, Molodet and Tzomet, are off the principal axis in 1992 and 1996. Although the estimates for the eigenvalues are positive (but small in value) on the minor axis in 1996, the simulation that we performed suggests that it is unlikely that pure vote maximizing could account for the locations of Shas in particular.

We conclude that the simulation exercise, taken together with the computation of the eigenvalues, can give a very detailed indication of the predictive power of the formal model, $M(\beta,\lambda,\sigma^2,\Delta)$. We can use differences between the formal predictions and the estimated locations to construct more refined second-order theories to account for deviation from the model.

4. CONCLUSION

These observations suggest that the formal model can be extended to account for the differences between the simulation and the estimated location of parties. We offer two possible extensions to the formal theory given here.

Extension 1: Activists

By the very nature of the estimation, the empirical model assumes that valence is a constant. However, by their efforts, activists provide support for a party and their contributions and effort may allow the party to change the overall judgement by the electorate of the party's

quality.[52] In this case the party's valence will be indirectly affected by the party's position. In essence, the party will occupy a political 'niche'. Indeed, these activist valence functions may be sufficiently concave, so that the induced 'equilibrium' positions are stable Nash equilibria. This can be formally analysed by including activist valence terms in the determination of the Hessian. If the Hessian of an activist valence function is 'sufficiently' negative-definite, then the Hessian of the vote share function will also have negative eigenvalues. Under these conditions a PNE may exist. In principle, the location of parties such as Meretz and Tzomet can be explained using this extension.

Extension 2: Strategic Coalition Behaviour

If party leaders care about policy, then they will regard policy declarations as more than just strategic choices used for the purpose of garnering votes. For low valence parties, there is no general or obvious relationship between maximizing vote or seat strength and the ability to influence government policy. Even so, the close correspondence in our empirical analysis between an estimated LNE and the actual political configuration suggests that the appropriate utility function for party j has the form $U_j(\mathbf{z}) = V_j(\mathbf{z}) + \delta_j(\mathbf{z})$, where $\delta_j(\mathbf{z})$ incorporates strategic calculations by the party leader over the likelihood of joining different possible coalition governments.[53]

For example, Shas is pivotal in the coalition game played between Likud and Labor. By adopting a position off the principal axis, Shas could potentially bargain effectively to influence the policy adopted by the government coalition of which it was a member.

The construction of a formal model incorporating these second-order components, combined with the tools of simulation and empirical analysis, would then permit the study of equilibrium characteristics of multiparty democracies. By this method we may better understand the great variety of democratic political configurations.[54]

APPENDIX: PROOF OF THE LEMMA AND THE ELECTORAL THEOREM

The argument depends on a consideration of the vote share function, $V_1(\mathbf{z})$, for the lowest valence party. However, this function involves a multivariate integral whose variates are correlated. We therefore make a linear transformation, so that the covariance matrix of the new random variables are independent. This allows us to construct a vote share function, denoted $V_{1\mathrm{p}}(\mathbf{z})$ with which we can examine the properties of $V_1(\mathbf{z})$. The function $V_{1\mathrm{p}}(\mathbf{z})$ is a univariate integral and models a specific contest between the lowest valence agent, 1, and an imaginary agent whose valence is the average of the valences of the other $(\mathrm{p}-1)$ agents. First-order and second-order conditions on this function then allow us to infer the same conditions on $V_1(\mathbf{z})$.

Let i be any voter. The probability that i picks 1 is

$$\rho_{\mathrm{i}1}(\mathbf{z}) = \mathrm{Prob}[\lambda_1 - \beta\|x_i - z_1\|^2 + \varepsilon_1 > \lambda_k - \beta\|x_i - z_k\|^2 + \varepsilon_k\text{: for all } k \neq 1]. \tag{A1}$$

[52] Norman Schofield, 'Valence Competition in the Spatial Stochastic Model', *Journal of Theoretical Politics*, 15 (2003), 371–83. See also Miller and Schofield, 'Activists and Partisan Realignment in the United States'.

[53] Schofield and Sened, 'Local Nash Equilibria'; and Itai Sened, 'A Model of Coalition Formation: Theory and Evidence', *Journal of Politics*, 58 (1996), 350–72.

[54] This technique has been used to study the Netherlands, Britain and Italy. See Norman Schofield and Itai Sened, 'Modelling the Interaction of Parties, Activists and Voters: Why is the Political Centre so Empty?' *European Journal of Political Research*, 16 (2004), 483–515; Daniella Giannnetti and Itai Sened, 'Party Competition and Coalition Formation: Italy 1994–1996', *Journal of Theoretical Politics*, 44 (2005), 355–90. See also Norman Schofield and Itai Sened, *Multiparty Democracy* (New York: Cambridge University Press, forthcoming).

This expression can be written as $\Phi(g_{i1}(\mathbf{z}))$, where

$$g_{i1}(\mathbf{z}) = [[\lambda_1 - \beta\|x_i - z_1\|^2 - \lambda_2 + \beta\|x_i - z_2\|^2], \ldots, [\lambda_1 - \beta\|x_i - z_1\|^2 - \lambda_p + \beta\|x_i - z_p\|^2]]. \quad \text{(A2)}$$

Here $g_{i1}(\mathbf{z})$ is a (p − 1) dimensional variable, while Φ is the cumulative probability function (cpf) of the (p − 1) dimensional variate $e(1) = [. .\varepsilon_j - \varepsilon_1 . .: j = 2, \ldots p]$. Here we use the notation $\Phi(d) = \text{Prob}[e(1) < d]$, where d is a (p − 1) vector. The error difference variance-covariance matrix, Ω, of this variate is a (p − 1) by (p − 1) matrix which can be written $\Omega = \sigma^2 FF^T$, where F^T stands for the transpose of the (p − 1) by p difference matrix. It is obvious that Ω is only diagonal in the case p = 2 in which case the variance is $2\sigma^2$. In the more general case with p greater than or equal to 3, the matrix Ω has diagonal terms $2\sigma^2$ and off-diagonal terms σ^2. Consequently, the components of e(1) are correlated and this makes it difficult to compute $\rho_{i1}(\mathbf{z})$. However, since e(1) is multivariate normal, by the assumption on the errors, we can utilize a matrix transformation, B, chosen so that the transformed variate $u(1) = B(e(1))$ has covariance matrix $B\Omega B^T = \sigma^2 \mathbf{I}$ (again B^T stands for the transpose of B, while $\mathbf{I}$ is the identity matrix). Because the transformed variates are independent we may now write:

$$\Phi(g_{i1}(\mathbf{z})) = \Phi(B_1\ g_{i1}(\mathbf{z}))\ .\ \Phi(B_2\ (g_{i1}(\mathbf{z})))\ .\ \Phi(B_{p-1}\ (g_{i1}(\mathbf{z}))). \quad \text{(A3)}$$

In this expression each linear transformation, B_j, is represented by the jth row of B, while the integral $\Phi(B_j(g_{i1}(\mathbf{z})))$ is associated with the jth component of u(1) Thus, each $B_j\ (g_{i1}(\mathbf{z}))$ gives the upper bound of the integral. Because of the number of degrees of freedom associated with this transformation, we can choose each vector B_j, for $j = 1, \ldots, p - 2$, so that its coordinates sum to zero. This has the consequence that the bounds, $B_j(g_{i1}(\mathbf{z}))$, for $j = 1, \ldots, p - 2$, are independent of z_1. Moreover, we can also choose, B_{p-1} so that all its coordinates are identical. Note also that such a transformation can be found in the more general case where the errors are correlated. Now write:

$$h_{i1p}(\mathbf{z})) = [\lambda_1 - \beta\|x_i - z_1\|^2] + (1/(p - 1)\ \Sigma_j\ [(-\lambda_j + \beta\|x_i - z_j\|^2]. \quad \text{(A4)}$$

Then the last component of Equation A1 can be written $\Phi^*(h_{i1p}(\mathbf{z}))$ where we use Φ^* to denote the cpf of the variate $u = (1/(p - 1)\Sigma_{j=2}\ (\varepsilon_j - \varepsilon_1)$.

It is easy to see that the variance of u is $[1/(p - 1)^2]\ [(p - 1)^2 + (p - 1)]\ \sigma^2 = p/(p - 1)\ \sigma^2$. We write κ^2 for this 'corrected' variance. We can now prove the Lemma.

Proof of Lemma

From Equation A3:

$$V_1(\mathbf{z}) = (1/n)\ \Sigma_i\ \Phi(g_{i1}(\mathbf{z})) = (1/n)\Sigma_i\ [\Phi(B_1\ (g_{i1}(\mathbf{z}))), \ldots, \Phi(B_{p-2}\ (g_i(\mathbf{z}))), \ldots, \Phi^*\ (h_{i1p}(\mathbf{z}))]. \quad \text{(A5)}$$

The terms involving $B_1, B_2 \ldots, B_{p-2}$ are all independent of z_1, but they do depend on $\{x_i\}$ and on $\mathbf{z}_{-1} = (z_2, \ldots, z_p)$. We seek to show that $z_1 = 0$ is a local best response to $\mathbf{z}_{-1} = (0, \ldots, 0)$. At $\mathbf{z}_{-1} = (0, \ldots, 0)$, all the terms involving $B_1\ (g_{i1}(\mathbf{z})), \ldots, B_2\ (g_{i1}(\mathbf{z})), \ldots, B_{p-2}\ (g_{i1}(\mathbf{z}))$, are independent of x_i, but they do depend on $\{\lambda_j\}$. We may write α for the product of these first (p − 2) terms and, for purposes of differentiation, treat it as a positive constant. Thus $V_1(\mathbf{z}) = V_1(z_1, \mathbf{z}_{-1}) = (\alpha/n)\Sigma_i\ \Phi^*(h_{i1p}(\mathbf{z}))$. Differentiating this expression gives the first-order condition:

$$\Sigma_i\ \alpha\ [\varphi^*(h_{i1p}(\mathbf{z}))\ d\ h_{i1p}(\mathbf{z})dz_1] = 0. \quad \text{(A6)}$$

Here $\varphi^*(h_{i1p}(\mathbf{z}))$ is the value of the probability density function(pdf) of the variate u at $h_{i1p}(\mathbf{z})$, given $\mathbf{z}_{-1} = (0, \ldots, 0)$. This is positive. Moreover at $z_1 = 0$ this term is independent of x_i. Finally $dh_{i1p}/dz_1 = -2\beta(z_1 - x_i)$, so $\Sigma_i\ dh_{i1p}/dz_1$ only involves terms in $\beta\Sigma_i\ (z_1 - x_i)$. Thus, Equation A6 reduces to the condition that $z_1 = (1/n)\ \Sigma_i x_i$. But we have normalized coordinates so that $\Sigma_i x_i = 0$).

Thus $z_1 = 0$ solves the first-order condition for maximizing $V_1(\mathbf{z})$ subject to the constraint that $\mathbf{z}_{-1} = (0, \ldots, 0)$. But the same argument can be carried out for each vote function $V_j(\mathbf{z})$, $j = 1, \ldots, p$, proving that $\mathbf{z}^* = (0, \ldots, 0)$ is a CNE. Q.E.D.

To determine whether $\mathbf{z}^* = (0, \ldots, 0)$ is an LNE requires differentiating the LHS of Equation A6. Because Equation A1 involves a multivariate integral, this cannot readily be computed. However, after the transformation, B, we can define $V_{1p}((z_1, \mathbf{z}_{-1}) = (1/n)\ \Sigma_i\ \Phi^*(h_{i1p}(\mathbf{z}))$. At $\mathbf{z}_{-1} = (0, \ldots, 0)$ we can write $V_1(z_1, \mathbf{z}_{-1}) = \alpha\ V_{1p}(z_1, \mathbf{z}_{-1})$, where α is independent of z_1. Thus, we can use $V_{1p}(z_1, \mathbf{z}_{-1})$ as a proxy for $V_1(z_1, \mathbf{z}_{-1})$.

Proof of the Electoral Theorem

The Hessian of $V_{1p}(z_1, \mathbf{z}_{-1})$ is given by

$$H_1\ (z) = (1/n)\ \Sigma_i\ H_{i1}\ (z). \tag{A7}$$

Here H_{i1} (z) denotes the Hessian of $\Phi^*(h_{i1p}(\mathbf{z}))$. Because this is a univariate integral, it can be shown to be given by:

$$H_{i1}\ (z) = \phi\ (h_{i1p}(z)\ [-(p-1)\ [D_{i1}]\ h_{i1p}(z)/p\sigma^2 + [d^2h_{i1p}/dz_1^2]]. \tag{A8}$$

The term $[D_{i1}]$ is a $w \times w$ symmetric matrix involving the differentials dh_{i1p}/dz_1, while $[d^2h_{i1p}/dz_1^2] = -2\beta\mathbf{I}$ is the negative-definite Hessian of h_{i1p}. (Here we use **I** to denote the w by w Identity matrix.) At the joint origin $(z_1, \mathbf{z}_{-1}) = (0, \ldots, 0)$, it is clear that $h_{i1p}((z_1, \mathbf{z}_{-1}) = \lambda_1 - [1/(p-1)]\Sigma_{j=2}\ \lambda_j$.

Now the matrix $D = \Sigma_i\ [D_{i1}]$ is a quadratic form, whose diagonal entry in the (t,t) cell is $\Sigma_i\ x_{it}^2$.

Let $\chi_t = (\ldots, x_{it}, \ldots)$ be the n vector of the t coordinates of the bliss points. The diagonal terms in D are of the form $\|\chi_t\|^2$, while the off-diagonal terms in the (t,s) cell are scalar products (χ_t, χ_s). Note that for each coordinate axis $v_t^2 = (1/n)\Sigma_i\ x_{it}^2 = (1/n)\|\chi_t\|^2$ is the electoral variance on the t coordinate axis, while $v^2 = \Sigma_t\ v_t^2$ is the total electoral variance. Now using $\lambda_{av(1)} = [1/(p-1)]\ \Sigma_{j=2}\ \lambda_j$ and taking $\Lambda = (\lambda_{av(1)} - \lambda_1)$ as given in Definition 5, then from Equation A8 the condition that that all eigenvalues of the Hessian of V_{1p} be negative gives the same requirement on the matrix $C = [\beta\Lambda]\{(p-1)/[p\sigma^2]\}[4D] - 2nI$.

Writing $\Delta = (1/n)D$ for the *voter covariance matrix* gives the Hessian matrix $C(\beta,\lambda : \sigma^2, \Delta) = 2\beta\Lambda/[\kappa^2]\Delta - I$, as in Equation 8.

We now need to obtain the conditions under which the matrix C has negative eigenvalues. The argument is easiest in the two-dimensional case, but can be pursued in any dimension. If we can show that the determinant, det[C] of C is positive and the trace, trace[C] of C, is strictly negative then both eigenvalues will be strictly negative.

Now let $A = \beta\Lambda/[\kappa^2]$, so$(1/2n)C$ has diagonal terms $(2A/n)\|\chi_t\|^2 - 1$ and off diagonal terms $(2A/n)\ (\chi_t, \chi_s)$. Thus $\det(H_1)$ can be determined from:

$$(1/2n)^2\ \det[C] = (2A/n)^2)\ [\|\chi_s\|^2\|\chi_t\|^2 - (\chi_t, \chi_s)^2] + 1 - (2A/n)\ [\|\chi_t\|^2 + \|\chi_s\|^2]. \tag{A9}$$

The first term is positive, by the triangle inequality. Thus det[C] is strictly positive if $(2A/n)\ [\|\chi_t\|^2 + \|\chi_s\|^2] < 1$.

Now $v_t^2 = (1/n)\Sigma_i\ x_{it}^2 = (1/n)\|\ \chi_t\ \|^2$ and $v^2 = [v_t^2 + v_s^2]$ so this sufficient condition is:

$$2\beta\ \Lambda\ v^2 < \kappa^2. \tag{A10}$$

Moreover $(1/2n)\text{trace}([C]) = (2A/n)[\|\chi_t\|^2 + \|\chi_t\|^{2\cdot}] - 2$, and this is negative if:

$$2\beta\Lambda\ v^2 < 2\kappa^2. \tag{A11}$$

Obviously the sufficient condition (Equation A10) implies the necessary condition (Equation A11).By Definition 5, $c = c\ (\beta,\lambda : \sigma^2, \Delta) = 2\beta\Lambda\ v^2/[\kappa^2]$.Thus, the condition $c < 1$ is *sufficient* while the condition $c < 2$ is *necessary* for C to have negative eigenvalues. Indeed, we have shown that the two eigenvalues a, b of $(1/2n)C$ satisfy the condition $a + b = c - 2$. Moreover $ab = (1/2n)^2\ \det[C]$, and this is positive if $c < 1$. Obviously when Equation A10 is satisfied then both a,b are strictly negative, so this condition is sufficient for the LSNE condition to be satisfied for agent 1. When Equation A10 fails, it is still possible for the eigenvalues to be negative. However, if Equation A11 fails, then both of the eigenvalues cannot be negative.

We can proceed in the same way for agent 2. Let $\lambda_{av(2)} = [1/(p-1)]\ [\lambda_1 + \Sigma_{j=3}\ \lambda_j]$. Since $\lambda_2 \geq \lambda_1$, it follows that $(\lambda_{av(1)} - \lambda_1) \geq \lambda_{av(2)} - \lambda_2)$. Therefore, if $c < 1$, then the determinant of the Hessian for agent 2 will also have negative eigenvalues. Thus, $z_2 = 0$ will be a local best response to $\mathbf{z}_{-2} = (0, \ldots, 0)$. For agents $j = 3, \ldots, p$ we can proceed in exactly the same way to show that the condition $c < 1$ is sufficient for $z_j = 0$ to be a local best response to $\mathbf{z}_{-j} = (0, \ldots, 0)$. Thus the condition $c < 1$ is sufficient for an LSNE at the origin.

Suppose now that Equation A11 fails, so that $2\beta\ \Lambda\ v^2 \geq 2\kappa^2$. This is identical to the condition that $c \geq 2$. Then $a + b \geq 0$, so at least one eigenvalue, say a, must be non-negative. Consequently, one of the eigenvalues of the Hessian of $V_1(\mathbf{z})$ must be non-negative at the joint origin, so $z_1 = 0$ cannot be a strict local best response to $\mathbf{z}_{-1} = (0, \ldots, 0)$.

Thus $\mathbf{z}^* = (0, \ldots, 0)$ cannot a LSNE. This proves the condition $c < 2$ is necessary in two dimensions. Indeed it is easy to see that the condition $2 \geq c$ is also necessary for $\mathbf{z}^*$ to be a LNE. The higher dimensional

case follows in precisely the same fashion, by considering the determinant and trace of the w by w matrix C. Q.E.D.

Discussion, and Computation of Eigenvalues when $w = 2$

We have called the number $c = 2\beta\Lambda\ v^2/[\kappa^2]$ the 'convergence coefficient'. Notice that in empirical applications the electoral and stochastic normalizations imply that both the product $\beta\Lambda$ as well as the ratio $[v/\kappa]^2$ can be identified, so that c is certainly identifiable. When this coefficient is strictly less than 1, then we may say that 'strict local concavity' is satisfied at the origin. We can be sure that both eigenvalues are negative, and the origin will be a LSNE. Conversely, as we have shown, in the two-dimensional case, when the coefficient is greater than 2, then $z_1 = 0$ cannot be a strict best response to (0, …, 0) and the origin cannot be a strict LNE.

In the range [1,2], both eigenvalues may be negative. We can estimate the eigenvalues by using the product moment correlation coefficient, r^2, between the vectors χ_t, χ_s representing the voter ideal points on the two axes. By definition $\|\chi_s\|^2\|\chi_t\|^2 - (\chi_t, \chi_s)^2 = (1 - r^2)\|\chi_s\|^2\|\chi_t\|^2$. Using this we can solve the quadratic expression to obtain a, b. Thus, if we let $ab = 1 - c + y$, and $a + b = c - 2$, we can calculate the eigenvalues to be:

$$a = (1/2)[c - 2 + [c^2 - 4y]^{1/2}] = A\{[v_t^2 + v_s^2] + [[v_t^2 - v_s^2]^2 + 4\ r^2\ v_t^2 v_s^2]^{1/2}\} - 1$$

$$b = (1/2)[c - 2 - [c^2 - 4y]^{1/2} = A\{[v_t^2 + v_s^2] - [[v_t^2 - v_s^2]^2 + 4\ r^2\ v_t^2 v_s^2]^{1/2}\} - 1. \qquad \text{(A12)}$$

Obviously for r^2 close to 1, the two vectors will be highly correlated, and y will be small. Then b will be close to -1, while a will be close to $2A[v_t^2 + v_s^2] - 1 = c - 1$. As an illustration, in the Israel example for 1992, $c = 9.72$, while the eigenvalues for the lowest valence party, Shas, can be computed to be $a = 7.47$ and $b = 0.24$. Thus the origin is a minimum for the vote share function of Shas. The large positive eigenvalue is associated with a particular eigenspace, the major principal component of the electoral distribution.

When the two vectors χ_t, χ_s are uncorrelated, so that r^2 is close to 0, then as Equation A12 shows, the term $+ [c^2 - 4y]$ in the expression for a involves the electoral variance difference $[v_t^2 - v_s^2]$. Indeed in this case $a = 2A\ v_t^2 - 1$ and $b = 2Av_s^2 - 1$. When this difference term is small, then the two eigenvalues will be almost identical. When the difference between these two variance terms is sufficiently large, then the term in y will be large. So the eigenvalue on the high variance axis will be positive and larger than the eigenvalue on the second axis.

The one-dimensional case is the clearest, because the local concavity condition is both necessary and sufficient.

The analysis can be extended to the general case of w dimensions, but conditions for obtaining negative eigenvalues follow in analogous fashion.

Notice, in the case that valence differences are all zero, then the local concavity condition is always satisfied. It follows that the joint origin can always be assured of being a LSNE. However, this does not guarantee that the joint origin is a Nash equilibrium.

The two examples that we have just considered show that there are two separate situations.

(i) If the product moment correlation is close to 1, then the origin will be a saddle or a minimum.
(ii) If the product moment correlation is close to 0, then the origin will either be a maximum or a minimum depending on the relative size of the electoral variances on the two axes.

Extension to the Case of Multivariate Normal Errors

When the errors have a non-diagonal covariance matrix , Ξ, then instead of seeking a solution to the matrix equation $B\Omega B^T = \sigma^2\ \mathbf{I}$ where $\Omega = \sigma^2\ FF^T$ we have $\Omega = F\ \Xi\ F^T$, so we must find a solution to the equation $BF\ \Xi\ (BF)^T = G$, where G is a diagonal matrix and B has the properties as above. The matrix B will of course now depend on the error covariances, but it still can be found because of the number of degrees of freedom associated with this matrix equation. In this more general case, let var(Ω) be the sum of the terms in the matrix Ω. Then it can be shown that the results of the iind case carry through by redefining κ^2, in the above equations, to be $\text{var}(\Omega)/[p - 1]^2$.