Modeling Elections with Varying Party Bundles: Applications to the 2004 Canadian Election

Kevin McAlister, Jee Seon Jeon, and Norman Schofield

1 Introduction

Early work in formal political theory focused on the relationship between constituencies and parties in two-party systems. It generally showed that in these cases, parties had strong incentive to converge to the electoral median (Hotelling 1929; Downs 1957; Riker and Ordeshook 1973). These models assumed a one-dimensional policy space and non-stochastic policy choice, meaning that voters would certainly vote for a party. These models showed that there exists a Condorcet point at the electoral median. However, when extended into spaces with more than one dimension, these two-party pure-strategy Nash equilibria generally do not exist. While attempts were made to reconcile this difference, the conditions necessary to assure that there is a pure-strategy Nash equilibrium at the electoral median were strong and unrealistic with regards to actual electoral systems (Caplin and Nalebuff 1991).

Instead of pure-strategy Nash equilibria (PNE) there often exist mixed strategy Nash equilibria, which lie in the subset of the policy space called the uncovered set (Kramer 1978). Many times, this uncovered set includes the electoral mean, thus giving some credence to the median voter theorem in multiple dimensions (Poole and Rosenthal 1984; Adams and Merrill 1999; Merrill and Grofman 1999; Adams

K. McAlister (✉) · J.S. Jeon
Center in Political Economy, Washington University in Saint Louis, 1 Brookings Drive,
Saint Louis, MO 63130, USA
e-mail: kevinmcalister@wustl.edu

J.S. Jeon
e-mail: jsjeon@wustl.edu

N. Schofield
Weidenbaum Center, Washington University in St. Louis, Seigle Hall, Campus Box 1027,
One Brookings Drive, St. Louis, MO 63130-4899, USA
e-mail: schofield.norman@gmail.com

2001). However, this seems at odds with the chaos theorems which apply to multi-dimensional policy spaces.

The contrast between the instability theorems and the stability theorems suggest that a model in which the individual vote is not deterministic is most appropriate (Schofield et al. 1998; Quinn et al. 1999). This kind of stochastic model states that the voter has a vector of probabilities corresponding to the choices available in the election. This insinuates that if the voter went to the polls for the same election multiple times, he might not make the same vote every time. This model is in line with multiple theories of voter behavior and still yields the desirable property of showing that rational parties will converge to the electoral mean given the simple spatial framework.

Using this framework, Schofield (2007) shows that convergence to the mean need not occur given that valence asymmetries are accounted for. In this context, valence is taken to mean any sorts of quality that a candidates has that is independent of his location within a policy space. In general, valence is linked to the revealed ability of a party to govern in the past or the predicted ability of a party to govern well in the future. In recent years, models with a valence measure have been developed and utilized in studies of this sort. Schofield extends upon these models and demonstrates a necessary and sufficient condition for convergence to the mean, meaning that the joint electoral mean is a local pure-strategy Nash equilibrium (LNE) in the stochastic model with valence.

Valence can generally be divided into two types of valence: aggregate valence (or character valence) and individual valence (or sociodemographic valence). Both types of valence are exogenous to the position that a party takes in an election, meaning that these valence measures rely on some other underlying characteristic. Aggregate valence is a measure of valence which is common to all members in an electorate, and can be interpreted as the average perceived governing ability of a party for all members of an electorate (Penn 2003). Individual valence is a bit more specific, where this kind of valence depends upon the characteristics of a voter. This kind of valence differs from individual to individual. For example, in United States elections, African-American voters are very much more likely to vote for the Democratic candidate than they are to vote for the Republican candidate. Thus, it can be said that the Democratic candidate is of higher valence among African-American voters than the Republican candidate is. Both kinds of valence can be important in determining the outcomes of elections and are necessary to consider when building models of this sort.

Recent empirical work on the stochastic vote model has relied upon the assumption of Type-I extreme value distributed errors (Dow and Endersby 2004). These errors, commonly associated with microeconometric models, are typical of models that deal with individual choice, where individual utility is determined by the valence terms and the individual's distance from the party in the policy space. This distance is weighted by β, a constant that is determined by the average weight that individuals give to their respective distances from the parties. The workhorse of individual choice models is the multinomial logit distribution, which is an extension of the dichotomous response logit distribution. This distribution assumes

that the probability that an individual votes for a party follows the Type-I extreme value distribution, thus matching the assumed distribution of the stochastic voting model. This creates a natural empirical partner for the stochastic vote model.

Using this statistical framework and the assumption that individual choice follows this distribution, Schofield (2007) introduced the idea of the convergence coefficient, c, which is a measure of attraction to the electoral mean in an electoral system. This coefficient is unitless, thus it can be compared across models. Low values of this value indicate strong attraction to the electoral mean, meaning that the electoral mean is a local pure-strategy Nash equilibrium (Patty 2005, 2007). High values indicate the opposite. He also lays out a necessary and a sufficient condition for convergence to the electoral mean with regards to the convergence coefficient:

1. When the dimension of the policy space is 2, then the sufficient condition for convergence to the electoral mean is c < 1.
2. The necessary condition for convergence is if c < w, where w is the number of dimensions of the policy space of interest.

When the necessary condition fails, at least one party will adopt a position away from the electoral mean in equilibrium, meaning that a LNE does not exist at the electoral mean. As a LNE must exist for the point to be a pure strategy equilibrium, this implies non-existence of a PNE at the center. Given the definition of the convergence coefficient, the general conclusion is that the smaller β is, the smaller the valence differences are among candidates, and the lower the variance of the electoral distribution is, the more likely there is to be a LNE at the electoral center.

However, this only answers the question where the local Nash equilibria are in the simplest case of having one electoral mean that parties are responding to. This problem can quickly become more complicated. Imagine a country with five parties and two different regions. Four of the parties run in both regions, and are thus attempting to appeal to voters in both regions. However, one of these parties only runs in one of the regions and is only trying to appeal to the voters of this region. Thus, it would be unreasonable for it to position itself with regards to the electoral mean for the entire electorate. Rather, it wants to maximize its vote share within in the region in which it runs. Parties can choose to run in select regions for a variety of reasons. They may run for historical reasons or responsive reasons or even choose not to run in regions where they know they will not do well at all. As parties have limited resources, sometimes this kind of decision must be made.

In order to assess convergence to the electoral mean in this case, one must take into account the electoral centers that parties are responding to. In the above example, convergence to the electoral mean would mean that the first four parties converge to the overall electoral mean, or the mean of all voters in the electorate, while the fifth party would converge to the electoral mean of those individuals in its respective region. Thus, the convergence coefficient would no longer be appropriate, as it is proven only when the position for all parties is equal to zero on all dimensions. Similarly, when there are parties which run in different combinations of

regions, the typical multinomial logit model is no longer appropriate because the underlying assumption of "independence of irrelevant alternatives" (IIA) is no longer met (Train 2003). Given that there are problems with estimation of parameters from the currently utilized empirical methodology and problems with the underlying theoretical mechanism that drives the reasoning behind the convergence coefficient, we are left without the useful information gained about party tendencies in the stochastic model. Under the current framework, researchers can only analyze convergence, valence, and spatial adherence within specific regions. However, in this paper we propose a method for handling more structurally complex electorates.

In this chapter, we introduce methods for analyzing the stochastic vote model in electorates where individuals do not all vote for the same party bundle. First, this chapter will demonstrate that the convergence coefficient first defined by Schofield can be adjusted to handle any vector of party positions. We will determine the first and second order conditions necessary to show that a vector of policy positions is a local Nash equilibrium (LNE). From this, we will show that the convergence coefficient for a more complex electorate can be derived in a similar manner to that used previously. We will also show the necessary and sufficient conditions for convergence. Secondly, we will introduce a method that can be used to estimate the parameters necessary to find equilibria in the model. This empirical model, an extension of the mixed logit model, will utilize the same Type-I extreme value distribution assumptions used previously, but will not rely upon the IIA assumption necessary to use the basic multinomial logit model. This varying choice set logit (VCL: see Yamamoto 2011) will allow for aggregate estimation of parameters to occur while also allowing regional parameters to be estimated. This method of estimation along with the notions of convergence that will allow analysis of the stochastic voting model in more complex situations.

Finally, to illustrate these methods, we will analyze the Canadian elections in 2004. Canada has a regional party which only runs in one region of the country, however, in 2004, the regional party gained seats in the Parliament. As this election is an ideal testing point for these new methods, they can tell us whether or not these new methods give logical results. From this analysis, some insight can be gained as to the way in which parties can organize themselves to maximize the number of votes received.

2 The Formal Stochastic Model

The data in the spatial model is distributed $x_i \in X$ where $i \in N$ represents a member of the electorates's ideal point and n is the number of members in the sample. We assume that X is an open convex subset of Euclidian space, $\mathbb{R}^w$, where w is finite and corresponds to the number of dimensions selected to represent the policy space.

Each of the parties, $j \in P$, where $P = \{1, \ldots, j, \ldots, p\}$ chooses a policy, $z_j \in X$, to declare to the electorate prior to the election. Let $\mathbf{z} = (z_1, z_2, \ldots, z_p)$ be the vector of party positions. Given $\mathbf{z}$, each voter i is described by a vector:

$$u_i(x_i, \mathbf{z}) = \big(u_{i1}(x_i, z_1), u_{i2}(x_i, z_2), \ldots, u_{ip}(x_i, z_p)\big)$$

$$\text{where } u_{ij}(x_i, z_j) = u_{ij}^*(x_i, z_j) + \epsilon_{ij}$$

$$\text{and } u_{ij}^*(x_i, z_j) = \lambda_j - \beta \|z_j - x_i\|^2 + \alpha_{ij}$$

Here, $u_{ij}^*(x_i, z_j)$ is the observable utility for i, associated with party j. λ_j is an exogenous valence term for agent j which is common throughout all members of a population (i.e. party quality).[1] β is a positive constant and $\|.\|$ is the Euclidian distance between individual i and party j.[2] α_{ij} is an exogenous sociodemographic valence term, meaning that this term can be viewed as the average assessment of a party's governing ability to the members of a specific group.[3] The error term, ϵ_{ij} is assumed to be commonly distributed among individuals. In particular, we assume that the cumulative distribution of the errors follows a Type-I extreme value distribution. This is not only the norm in individual choices, it also allows the theoretical model to match the corresponding empirical model, making the transition between the two easier.

Given the stochastic assumption of the model, the probability that i votes for j given z, $\rho_{ij}(z)$ is equal to:

$$\rho_{ij}(\mathbf{z}) = Pr\big[u_{ij}(x_i, z_j) > u_{il}(x_i, z_l), \ \forall l \neq j\big]$$

In turn, we assume that the expected vote share for agent j given $\mathbf{z}$, is $V_j(\mathbf{z})$ where:

$$V_j(\mathbf{z}) = \frac{1}{n} \sum_{\forall i \in N} \rho_{ij}(\mathbf{z})$$

We assume in this model that agent j chooses z_j to maximize $V_j(\mathbf{z})$ given the positions of the other parties. We seek equilibria of the model where each of the parties attempts to maximize vote share.

For the purposes of this paper, when we talk about an equilibria, we refer to a local Nash equilibria (LNE). This definition of equilibrium relies on maximizing the expected vote share gained by a party given the positions of the other parties. A vector of positions, $\mathbf{z}^*$, is said the be a LNE if $\forall j$, z_j^* is a critical point of the

---

[1] This can be conceptualized as an average assessment of the parties quality to govern among all members of the electorate, regardless of sociodemographic identity.

[2] To match up with the empirical applications later in the paper, the utility individual i gains from having party j in office is compared to a base party, $j = 1$. As is normal, we assume this party has a utility of zero and the other utilities are compared to this party. Thus, the utility gained by i by voting for j can also be seen as $u_{ij}^*(x_i, z_j) = \lambda_j - \beta(\sum_{m=1}^w((x_{jm} - x_{im})^2 - (x_1m - x_{im})^2)) + \alpha_{ij}$ where the summation is of the Euclidian distances for each dimension of the policy space. This places our model in line with the latent utility models that are commonly used in microeconometric theory and bridges the gap between our theoretical model and the corresponding empirical model.

[3] In this paper, we assume that this term is common among all members of a specific sociodemographic group. However, we can set up these terms to represent individuals with individual level random effects.

vote function and the Hessian matrix of second derivatives is non-positive, meaning that the eigenvalues are all non positive. More simply put, a vector, $\mathbf{z}^*$, is a LNE if each party locates itself at a local maximum in its respective vote function. This means, that given the opportunity to make moves in the policy space and relocate its platform, no vote-maximizing party would choose to move. We assume that parties can estimate how their vote shares would change if they marginally move their policy position. The local Nash equilibrium is that vector z of party positions so that no party may shift position by a small amount to increase its vote share. More formally a LNE is a vector $\mathbf{z} = (z_1, \ldots, z_j, \ldots, z_p)$ such that each $V_j(\mathbf{z})$ is weakly locally maximized at the position z_j. To avoid problems with zero eigenvalues we also define a strict local Nash equilibrium (SLNE) to be a vector that strictly locally maximizes $V_j(\mathbf{z})$. We typically denote an LNE by $\mathbf{z}(K)$ where K refers to the model we consider. Using the estimated MNL coefficients we simulate these models and then relate any vector of party positions, $\mathbf{z}$, to a vector of vote share functions $V(\mathbf{z}) = (V_1(\mathbf{z}), \ldots, V_p(\mathbf{z}))$, predicted by the particular model with p parties.

Given that we have defined the errors as cumulatively coming from a Type-I extreme value distribution, the probability $\rho_{ij}(z)$ has a multinomial logit specification and can be estimated. For each voter i and party j the probability that i votes for j given z is given by:

$$\rho_{ij}(\mathbf{z}) = \frac{\exp(u_{ij}^*(x_i, z_j))}{\sum_{k=1}^{p} \exp(u_{il}^*(x_i, z_k))}$$

$$= \left[1 + \sum_{k \neq j}^{p} \exp(f_k) \right]^{-1}$$

where $f_k = \sum_{k=1}^{p} \left(u_{il}^*(x_i, z_k) \right) - \left(u_{ij}^*(x_i, z_j) \right).$

Thus $\quad \dfrac{d\rho_j(\mathbf{z})}{dz_j} = 2\beta(z_j - x_i) \left[1 \times \left[1 + \sum_{k \neq j}^{p} \exp(f_k) \right] \right]^{-2} \left[\sum_{k \neq j}^{p} \exp(f_k) \right]$

$$= 2\beta(z_j - x_i) \times \left[\rho_{ij}(\mathbf{z}) \right] \left[1 - \rho_{ij}(\mathbf{z}) \right]$$

in region k, with population, N_k, of size n_k the first order condition becomes

$$\frac{dV_{jk}(\mathbf{z}_k)}{dz_j} \bigg|_{z-j=z} = \frac{1}{n_k} 2\beta_k \sum_{i \in N_k} \rho_{ijk}(1 - \rho_{ijk})(z_j - x_i) = 0, \qquad (1)$$

$$\text{so } z_j = \sum_{i \in N_k} w_{ij} x_i, \qquad (2)$$

$$\text{where } w_{ij} = \frac{\rho_{ijk}(1 - \rho_{ijk})}{\sum_{k=1} \rho_{ijk}(1 - \rho_{ijk})}. \qquad (3)$$

In order to show that points are LNE, we need to show that given **z**, all agents are located at a critical point of their respective vote functions, $V_j(\mathbf{z})$. Thus, we need to show that the first derivative of the vote function, given **z**, is equal to zero. Then we need to show the Hessian matrices at these points and compute their eigenvalues.

In this paper, we make two key departures from previous papers that have used this stochastic vote model. First, and certainly the most important departure, we intend to assess convergence in a model where the position vector of interest does not have all of the parties at the joint aggregate electoral origin. As explained before, in cases where there are regional parties that do not run in all parts of an electorate, there is no incentive for these agents to locate at the overall electoral mean. Rather, in line with other median voter results, these parties have incentives to locate at their respective electoral means, meaning that they position themselves on the ideal point of the average voter that actually has the choice to vote for that party. Thus, should we find that parties in an electoral system converge to the electoral mean in equilibrium, we should find that parties that run in all regions of an electorate converge to the joint electoral mean and regional parties converge to their respective regional electoral means. Previous papers have adjusted the scale of the policy space such that the electoral mean corresponds to the origin of the policy space and this allowed for some convenient cancelation to occur in proofs. For the purposes of this paper, though, we cannot make those cancelations and, thus, we are assessing convergence for a general vector of party positions rather than a zero vector. Second, we assume a second kind of valence, an individual valence, that was not previously included in utility equation. We intend to assess convergence to the mean given these individual valence measures as well, showing proofs including these variables.

The first derivative of $V_j(z)$ with respect to one dimension of the policy space is:

$$\frac{dV_j(\mathbf{z})}{dz_j} = \frac{2\beta}{n} \sum_{i=1}^{n}(z_j - x_i)\rho_{ij}(1 - \rho_{ij})$$

Of course, a LNE has to be at a critical point, so all the set of possible LNE can be obtained by setting this equation to 0. Note that this derivative is somewhat different than that from earlier works as we do not assume that ρ_{ij} equals ρ_j (being independent of i). This is due to the fact that we do not assume that all parties are located at the electoral mean.

This result is important in a couple of ways. First, we see that the first derivative does not rely on λ_j or α_{ij} in any way aside from the calculation of the probability, ρ_{ij}, that an individual i votes for party j. This is an encouraging result because any resulting measures that assess convergence (i.e. the convergence coefficient) will not depend on the individual level valences. Previously, Schofield (2007) only showed that the convergence coefficient could be calculated when we assume a common valence for agent j across all members of an electorate. This finding allows us to expand the convergence coefficient notion to include these individual level valences

as long as they are exogenous of a voter's ideal point. Second, after doing some simple algebra, it is easy to see that when a party locates at its respective electoral mean, the equation always equals zero, meaning that it is always at a critical point. This is also a good result, because it gives further support to the idea that the electoral mean is always a possible LNE.

To test if a critical point is a local maximum in the vote function, thus a LNE, we need a second order condition. The Hessian matrix of second derivatives is a $w \times w$ matrix defined as follows:

- Let $v_t = (x_{1t}, x_{2t}, \ldots, x_{nt})$ be the vector of the tth coordinates of the positions of the n voters and let. Let $z_j = (z_{1j}, z_{2j}, \ldots, z_{tj})$ and $\langle v_t - z_{tj}, v_s - z_{sj} \rangle$ be the scalar product, with $\Delta_0 = [\langle v_t - 0, v_s - 0 \rangle]$ the electoral covariance matrix about the origin. Then diagonal entries of the Hessian for candidate j have the following form:

$$\frac{1}{n} \sum_{i=1}^{n} 2\beta(\rho_{ij})(1 - \rho_{ij})\big(2\beta(x_{it} - z_{tj})^2(1 - 2\rho_{ij}) - 1\big)$$

- The off diagonal elements have the following form:

$$\frac{1}{n} \sum_{i=1}^{n} 4\beta^2(x_{is} - z_{sj})(x_{is} - z_{tj})\rho_{ij}(1 - \rho_{ij})(1 - 2\rho_{ij})$$

- where $s \neq t$, and $s = 1, \ldots, w$, and $t = 1, \ldots, w$.

Given this matrix, if all w eigenvalues of the Hessian are negative given **z**, then we can say that the position of interest is a LNE.

Unlike previous models of this sort, there is no characteristic matrix that the Hessian can be reduced to in order to assess whether or not a point is a local Nash equilibria. Thus, for the proper second order test, the eigenvalues of the Hessian must be found. However, as in earlier works, a reduced equation can be used to find a convergence coefficient, a unitless measure of how quickly the second derivative is changing at a given point. This convergence coefficient can be viewed substantively as a measure of how much a rational, vote-optimizing party is attracted to a certain position. As the coefficient becomes large, the party is repelled from the position.

We know that the trace of the Hessian is equal to the sum of the eigenvalues associated with the matrix. In order to be a local maximum, and thus a LNE, the eigenvalues have to all be negative. Thus, the trace of the Hessian must be negative as well in order for the point to be a local maximum. Given the equation for the main diagonal elements, we can see that it relies on β, $\rho_i j$, and the squared distance between the individual's ideal point on one dimension and the party's position on the same dimension. As β and $\rho_i j$ are necessarily positive, the only way in which the second derivative can be negative is if $2\beta(x_i - z_i)^2(1 - 2\rho_{ij})$ is greater than 1. Thus, this is the value of interest when trying to assess whether or not a point is a local maximum. This value can be viewed as the measure of how fast the probability that voter i votes for party j changes as the party makes small moves. We reason

that the mean of $2\beta(x_i - z_i)^2(1 - 2\rho_{ij})$ over all voters is an equivalent concept to the convergence coefficient that does not rely on parties being positioned at the electoral origin. However, this is only for one dimension, so the full definition of the convergence coefficient is:

$$c(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^{w} \sum_{i=1}^{n} 2\beta(x_{it} - z_{tj})^2(1 - 2\rho_{ij})$$

In words, the convergence coefficient is equal to the sum of mean values of

$$2\beta(x_i - z_i)^2(1 - 2\rho_{ij})$$

over all individuals in the electorate for each dimension of the policy space. This notion is supported by the fact that when all parties do locate at the electoral origin, this definition of the convergence coefficient is equivalent to the definition provided in Schofield (2007).

Given this definition of the convergence coefficient, we can derive necessary and sufficient conditions for convergence to a given vector of party positions. Given a vector of party positions, a sufficient condition for the vector being a local Nash equilibrium is that $c(\mathbf{z}) < 1$. If $c(\mathbf{z})$ is less than 1, then we can guarantee that the second derivatives with respect to each dimension are less than 0. This eliminates the possibility that the party is located at a saddle point. A *necessary* condition for convergence to the vector of interest is that $c(\mathbf{z}) < w$. However, for the position to be a LNE, each second derivative has to be negative. Thus, each constituent part of $c(\mathbf{z})$ must be less than 1.

It is important to note that a convergence coefficient can be calculated for each party in the electoral system. Previously, given that all of the parties have been attempting to optimize over the same population, an assumption could be made that the highest convergence coefficient would belong to the party which had the lowest exogenous valence. However, with the slight restructuring of the model to include individual level valences and parties which run in singular regions, as ρ_j can no longer be reduced down to a difference of valences, we can no longer make the assumption that the lowest valence party will be the first to move away from the mean should that be equilibrium behavior. In fact, given that there are multiple definitions of valence in the equation and multiple values of these valences for each region, a notion of lowest valence party becomes very difficult to define. Thus, the convergence coefficient should be calculated for each party to ensure a complete analysis of convergence behavior. Then the party with the highest convergence coefficient represents the electoral behavior of the system. Thus, for an electoral system, the convergence coefficient is:

$$c(\mathbf{z}) = \arg_p c_p(z)$$

In summary, the method for assessing whether or not a vector of party positions is a LNE is as follows:

1. Define z*, or the vector of party positions in the policy space.
2. Check that each party position meets the first order condition given the other party positions:

$$\frac{dV_j(z)}{dz_j} = \frac{2\beta}{n} \sum_{t=1}^{w} \sum_{i=1}^{n} (x_i - z_j)\rho_{ij}(1 - \rho_{ij}) = 0$$

   - Note that each party's respective electoral mean is a position that is always a critical point in the vote function.

3. Define the Hessian, $C_j(\mathbf{z})$ for each party position as follows:

   - diagonal entries are

   $$\frac{1}{n} \sum_{i=1}^{n} 2\beta(\rho_{ij})(1 - \rho_{ij})\big(2\beta(x_{it} - z_{tj})^2(1 - 2\rho_{ij}) - 1\big)$$

   where $t = 1, \ldots, w$.
   - The off diagonal elements have the following form

   $$\frac{1}{n} \sum_{i=1}^{n} 4\beta^2(x_{is} - z_{js})(x_{it} - z_{jt})\rho_{ij}(1 - \rho_{ij})(1 - 2\rho_{ij})$$

4. Check the eigenvalues for each Hessian. If all of the eigenvalues are negative, the vector of positions is a local Nash equilibrium.
5. The necessary condition that the eigenvalues all be negative is that trace$(C_j(\mathbf{z})) < 0$. Since $\beta(\rho_{ij})(1 - \rho_{ij}) > 0$ this reduces to: $\sum_{t=1}^{w} \sum_{i=1}^{n} 2\beta(\rho_{ij})(1 - 2\rho_{ij})(x_{itw} - z_{tj})^2 < w$.
6. In two dimensions, the further sufficient condition is that det$(C_j(\mathbf{z})) > 0$, which is equivalent to the condition that $\sum_{t=1}^{w} \sum_{i=1}^{n} 2\beta(\rho_{ij})(1 - 2\rho_{ij})(x_{itw} - z_{iw})^2 < 1$.
7. Calculate the convergence coefficient for each party,

$$c_j(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^{w} \sum_{i=1}^{n} 2\beta(\rho_{ij})(1 - 2\rho_{ij})(x_{itw} - z_{iw})^2$$

The convergence coefficient, labelled $c(\mathbf{z})$, represents the electoral system.

   - If $c(\mathbf{z}) > w$, then we cannot have convergence. If, however $c(\mathbf{z}) < 1$, then the sufficient condition is satisfied, and the system converges to the vector of interest. If $c(\mathbf{z}) \leq w$, check the components of $c_j(z)$ in dimension w, if all are less than 1, then the system converges to $\mathbf{z}$.
   - To compare this general model with the one presented in Schofield (2007), suppose that all parties adopt the same position at the electoral mean $\mathbf{z} = 0$. Then ρ_{ij} is independent of i. We let Δ_0 be the w by w electoral covariance matrix about the origin. Then

-

$$C_j(\mathbf{z}) = (\rho_j)(1 - \rho_j)4\beta^2(1 - 2\rho_j)\Delta_0(1 - 2\beta I)$$

where I is the w by w identity matrix. Since $(\rho_j)(1 - \rho_j)(2\beta) > 0$, we can identify the Hessian with the matrix

$$C_j^*(\mathbf{z}) = \left[2\beta(1 - 2\rho_j)\Delta_0 - I\right]$$

Thus the eigenvalues are determined by the necessary condition $\text{trace}(C_j^*(\mathbf{z})) \leq w$, which we can write as

$$\mathbf{c} = 2\beta(1 - 2\rho_j)\,\text{trace}(\Delta_0) \leq w$$

It can also be shown that the sufficient condition for convergence, in two dimensions, is given by $\mathbf{c} = 2\beta(1 - 2\rho_j)\,\text{trace}(\Delta_0) < 1$.

3 Estimation Strategies Given Varying Party Bundles

In order to utilize the stochastic election model proposed above, we need to have measures of valence, both aggregate and individual, for each party in the system, and an estimation of β along with the data in order to analyze equilibrium positions within the system. Typically, given the assumptions of the model, it is an easy translation of data to conditional logit model to equilibrium analysis. However, this is only true when all of the voters exist in one region. In other words, this only works when all voters vote with the same bundle of alternatives on the ballot. However, as shown in the beginning, when there are regional parties in a country which only run in one region, and are thus on the ballot for only a fraction of members of an electorate, the situation quickly becomes more complicated.

The reason that a new method is necessary is that multinomial logit models are reliant upon the assumption of independence of irrelevant alternatives. Simply put, IIA is a statement that requires that all odds ratios be preserved from group to group, even if the choice sets are different.

1. When IIA is violated, the multinomial logit specification is incorrect if we want to do any estimation procedures with this data.

Yamamoto (2011) proposed an appropriate model, called the *varying choice set logit model* (VCL). This model, which follows the same specification as the typical multinomial logit model when Type-I extreme value errors are assumed, is the same as used above to derive the convergence coefficient, that is:

$$\rho_{ij}(\mathbf{z}) = \frac{\exp(u_{ij}^*(x_i, z_j))}{\sum_{k=1}^{p}\exp(u_{ik}^*(x_i, z_k))}$$

Thus the framework of the formal model and the empirical model still match, allowing easy transition from empirical estimations of parameters to analyzing the equilibria of the system given the parameters.

The VCL differs from typical logistic regression models, though, by not relying on the IIA assumption. This is done by allowing there to be individual logistic regression models for each choice set type then aggregating these estimates to make an aggregate estimate of valence for the entire electorate. In this case, each choice set type is seen as a region, as each region has a different bundle of parties offered to voters. In these models, we can assume that parameters are common to all regions in an electorate or that the parameters have values that are region specific. For example, in our model, we assume that β is common to all members of the electorate regardless of region. On the other hand, we assume that both types of valence are individual specific; the VCL is able to accommodate parameters of both types by using a random effects hierarchical structure, meaning that the parameters estimated for each region are assumed to come from some probability distribution, generally a normal distribution. This method of estimation is best done utilizing random effects.

The VCL model uses random effects for the individual choice set types, meaning that for each individual type of choice set in an electorate, we estimate the parameters of interest for the individuals within that choice set. Then, using these estimates, we assume that these individual estimates come from their own distribution, and we use that to determine the best aggregate estimate for a parameter within the model. For our model, we assume the following specification for the observed utility gained by voter i from voting for party j:

$$u_{ij}^*(x_i, z_j) = \lambda_j + \beta \|z_j - x_i\| + \mu_{jr} + \xi_{jrs}$$

where λ_j is the aggregate estimate of the exogenous valence of party j and β and Euclidian distance between voter and party has the same interpretation as within the formal model. μ_{jr} is the added utility over the aggregate valence that the average individual from region r get for voting for party j and ξ_{jrs} is the added utility over μ_{jr} that the average member from sociodemographic group s gets from voting for party j. This clearly hierarchical specification of valence lends itself very well to the VCL model. As with typical logit models, the probability that voter i votes for party j follows the typical logit specification, which states that the probability that the voter votes for party j is the ratio of the exponentiated utility of voting for j to the sum of the utility gained for voting for each party. This model clearly lines up with the formal model specified before and makes the VCL a very attractive choice when attempting to estimate parameters from an electorate with a clear regional structure.

Using the VCL, however, places a few light assumptions on the model, as any estimation procedure does. First, given the structure of the utility equation, we assume that β is common over all members of the electorate, regardless of region or sociodemographic group. This is not a departure from previous papers which have utilized this assumption. This simply means that individuals only differ in how they view each of the parties and not how much weight they apply to the differences between their ideal points and the parties' ideal points. Second, by virtue of the usage

of random effects, this model assumes that each of the regional and sociodemographic group random effects are orthogonal to the other covariates in the model. Simply put, we assume that these random effects for each person are independent of one's position within the policy space. Third, by virtue of our usage of the VCL model, we assume that a party's decision to run in a specific region is exogenous of its perceived success within that region. This assumption can be troublesome in some electoral systems where parties frequently do not remain on the same ballots from year to year. However, many electoral systems with regional parties have parties which are historically bound to one region or another. Thus, when we assume that parties historically choose to run in a region, this model is appropriate. When all three of these assumptions are met by the electorate of interest the VCL is a flexible choice of estimation procedure.

The reason that the varying choice set logit (VCL) is the superior method when handling electorates with multiple regions is that it relaxes the IIA assumption while also providing us with the most information from the model. VCL relaxes IIA by allowing each of the parameters to be estimated within each group and allowing these parameters to derive the aggregate estimation of parameters through the notion of partial pooling. Partial pooling is best achieved through hierarchical modeling and through the use of random effects. VCL can be viewed as a specific kind of mixed logit model, meaning that the mixed logit model can be used to achieve the same aggregate results. However, given the structure of VCL, parameter estimates can be achieved for each choice set type (i.e. region) rather than for each individual, demonstrating a significant efficiency gain over the standard mixed logit model. Similarly, mixed logit does not allow the researcher to estimate choice set specific values of parameters, thus VCL is more efficient and informative. Another alternative is the multinomial probit model, which does not rely on the IIA assumption either. However, the multinomial probit model does not allow the researched to estimate parameters at the level of the individual choice set, as the errors are absorbed in the error matrix and, thus, the IIA itself is absorbed. However, as with the mixed logit, the individual regional values are often of as much interest as the parameter values, so the mixed probit is essentially discarding information that the researcher may find useful. Thus, we opt to use the VCL method when examining the behavior of parties in an electorate with party choice sets that vary over the electorate.

The structure of the VCL lends itself to Bayesian estimation methods very easily. While random effects can be estimated in a frequentist manner, as is demonstrated with Yamamoto's (2011) expectation-maximization algorithm for estimation using the VCL, the implementation of the estimation procedure is much easier in a Bayesian hierarchical setting. Assuming that each of the parameters of interest (both random effects and fixed effects) come from commonly used statistical distributions, generally those within the Gamma family, a Gibbs sampler is easily set up and can be utilized to garner estimates of the parameters of interest.

For applications to this model, we make a few assumptions about the underlying distributions of the parameters of interest. We assume that β, λ_j, and the random effects all have underlying normal distributions. Further, we assume that all of these distributions are independent of one another. This assumption follows from

our assumptions that the variables, and thus the draws in the Gibbs sampler, are all orthogonal. We could easily assume that each level of the hierarchy (aggregate, region, sociodemographic) comes from a multivariate normal within itself. However, time spent with this model has shown that this assumption is taxing computationally, adding to the amount of time it takes the Gibbs sampler to converge and yielding results that are virtually indiscernible from those garnered when independence is assumed. However, it is unreasonable to assume that the orthogonality assumption is perfectly met. For example, in some cases, region and location within the policy space are correlated (as in Canada). This assumption violation will lead to biased estimators. While the bias is not large, it is certainly a cause for some concern. However, this problem is easily fixed.

Gelman et al. (2008) utilize a method to rid random effects of the collinearity which causes the estimates to be biased. They propose that the problem is solved very simply by adding the mean of the covariate of interest as a predictor a level lower in the hierarchy than the random effect of interest. In this case, given a specific party, the mean of its regional level random effects and the mean of its sociodemographic level random effects are indeed situated at the respective mean of the difference of Euclidian differences between the party of interest and the base party. Given that this is the covariate that will theoretically be correlated with sociodemographic group and region, this is the mean that we need to include as a predictor in the random effects. In doing this, the researcher controls for the discrepancy as if it is an omitted variable and allows the random effect to take care of its own correlation. The normal priors in this case can still be diffuse, but the mean needs to be at the specified value to fix the problem.

One practical note is necessary regarding the time necessary to achieve convergence within the model. Convergence of the VCL can be quite slow given a large number of choice set types and individual observations. Similarly, as random effects are estimated for each party, the number of parties and the number of sociodemographic groups can slow down the rate at which samples are derived from the Gibbs sampler. Though it is a time consuming method, the sheer amount of information gained from the VCL is, thus, the best choice when it is necessary to use a discrete choice model which does not rely on IIA.

4 Application to Canadian Elections

In recent history, Canadians have elected at least three different parties to the Federal legislature and 2004 was no different. However, the 2004 election in Canada was significant because it yielded the first minority government for Canada since 1979. The Liberal Party gained the most seats (135 seats) and the largest percentage of the vote (36.7 percent), however it failed to gain a majority of the seats in Parliament and needed to form a coalition government in order to control the legislature. Paul Martin and the Liberals initially formed a coalition with the New Democratic Party (NDP), a liberal party whose support increased from the 2000 elections, in

Table 1 Actual and sample vote percentages

	Actual	Sample—All	Sample—Quebec
Liberal	36.71	34.34	25.13
NDP	15.65	18.45	8.02
Conservative	29.66	31.55	9.01
Green	4.29	3.71	2.68
BQ	12.42	11.95	55.08

order to control government (19 seats, 15.7 percent). The Liberal Party's main opponent was the newly formed Conservative Party of Canada, the party formed by the merger of the Alliance Party and the Progressive Conservative party, which significantly chipped into the Liberal's vote share. After splitting support in the 2000 elections, the merger of the two parties gave the Conservative Party hope of controlling the Canadian government. Given exposure of scandal within the Liberal Party, the Conservative Party and the Liberal Party were neck and neck in the weeks leading up to the elections. However, the relative inexperience of the new party led to key mistakes prior to the elections and the Conservative Party was not able to garner a seat majority and was not able to form a coalition to control government.

Perhaps the most interesting aspect of the 2004 Canadian elections was Quebec's regional party, Bloc Quebecois (BQ). The BQ only ran in Quebec and, thus, was only on the ballot for approximately twenty percent of Canadians. However, their support within the region was overwhelming, with nearly fifty percent of Quebec voters voting for the party. This strong showing put quite a dent in the Liberal Party's showing within the region and made the BQ a significant player in the Canadian parliament (54 seats, 12.4 percent). Similarly, while not quite on the scale of the BQ, the Green Party was another small party which undoubtedly played a part in reducing the vote share of the Liberal Party. Though support for the party increased in the 2004 elections, its small initial voter base kept it from receiving any seats within parliament. However, it did gain a significant portion of votes in the election (0 seats, 4.3 percent).

To study the 2004 Canadian election we used the survey data for Canada collected by Blais et al. (2006). Table 1 shows vote shares within the sample and the overall vote shares. The similarity between these two sets of shares suggests that the sample is fairly representative of the Canadian electorate. Table 1 also has columns for those voters within Quebec, as Bloc Quebecois only ran within Quebec.

The factor analysis performed on the voters' responses in the survey questions led us to conclude that there were two factors or policy dimensions: one "social," the other "decentralization." The social dimension is a weighted combination of voters' attitudes towards (1) the gap between poor and rich, (2) helping women, (3) gun control, (4) the war in Iraq and (5) their position the left-right scale. We coded the social dimension such that lower values imply higher interest in social programs so as to have a left-right scale along this axis. The decentralization dimension included voters' attitudes towards (1) the welfare state, (2) their standard of living, (3) inter-jurisdictional job mobility, (4) helping Quebec and (5) the influence of

Table 2 Survey items

Inequality	How much to you think should be done to reduce the gap between the rich and the poor in Canada?
	(1) much more—(5) much less
Women	How much do you think should be done for women?
	(1) much more—(5) much less
Gun Only police/military	Only the police and the military should be allowed to have guns.
	(1) strongly agree—(4) strongly disagree
Iraq War	As you may know, Canada decided not to participate in the war against Iraq. Do you think this was a good decision or a bad decision?
	(1) good decision (2) bad decision
Left-Right	In politics, people sometimes talk of left and right. Where would you place yourself on the scale below?
	(0) left—(11) right
Welfare	The welfare state makes people less willing to look after themselves.
	(1) strongly disagree—(4) strongly agree
Standard of Living	The government should see to it that everyone has a decent standard of living.
	(1) leave people behind (2) Don't leave people
Quebec	How much do you think should be done for Quebec?
	(1) much more—(5) much less
Moving Cross Region	If people can't find work in the region where they live, they should move to where the jobs are?
	(1) strongly disagree—(4) strongly agree
Federal-provincial	In general, which government looks after your interests better?
	(1) provincial (2) no difference (3) federal

Federal versus Provincial governments in their lives. A greater desire for decentralization implies higher values on this axis. The questions used in the factor analysis can be found in Table 2.

Using the factor loadings given in Table 3, we computed the value for each voter along the social and decentralization dimensions. The mean and median values of voters' positions along these two dimensions in Canada are at the electoral origin, (0; 0). To illustrate, a voter who thinks that more should be done to reduce the gap between rich and poor would tend to be on the left of the Social axis (x axis), while a voter who believes that the federal government does a better job of looking after peoples' interests would have a negative position on the D axis (y axis), and could be regarded as opposed to decentralization.

The survey asked voters which party they would be voting for, so we estimated party positions as the mean of voters for that party. The party positions in the policy space are given by the vector:

Table 3 Weighting coefficients for Canada

Components	Social	Decentralization
Inequality	0.36	−0.03
Women	0.35	0.07
Gun only police/military	0.20	0.52
Iraq War	0.30	0.20
Left-Right	0.38	−0.06
Welfare	0.37	−0.17
Standard of Living	0.38	−0.05
Quebec	−0.35	0.00
Moving cross region	0.27	−0.48
Federal-provincial	−0.09	−0.65
SD ($\sqrt{var}$)	1.67	1.07
% Var	28	11
Cumulative % Var	28	39

$$z^* = \begin{bmatrix} & Lib. & Con. & NDP & Grn. & BQU \\ S & -0.17 & 1.27 & -0.78 & -0.63 & -1.48 \\ D & -0.38 & 0.32 & 0.05 & -0.13 & 0.23 \end{bmatrix}$$

These party positions correspond closely with those estimated by Benoit and Laver (2006), obtained using expert opinions in 2000. As with these estimates, the Liberal Party locates to the left on the social access while the Conservative party lies in the upper right quadrant, as shown in Fig. 1. Figure 1 also shows the distribution of voters in Canada. From this, we see that most voters have a moderately leftist view on social issues and are fairly evenly split on decentralization issues, with most voters lying right in the middle. In Fig. 1, the "Q" represents the electoral mean within Quebec, which is noticeably left of the overall electoral mean. Figure 2 shows the voter distribution for Quebec only. The majority of voters in Quebec advocate more liberal social policies than the average voter in Canada. Similarly, voters in Quebec tend to want more decentralization of government, as Quebec has a strong regional identity and wants to maintain its somewhat independent state. This, along with the differences that are easily seen from the two plots, are evidence that the two regions have strong regional identities.

The survey also collected sociodemographic data. For each respondent, sex, age, and education level were recorded. Age was divided into four categories: 18–29, 30–49, 50–65, 65 and older. Education was divided into three categories: No High School Diploma, High School Diploma but No Bachelors, Bachelors or Higher. Due to the structure of the VCL and the underlying random effects model, sociodemographics are viewed as categorical so that groups can be made. As noted previously, parsimony is very important in the VCL model as the time to convergence and the time necessary to run the Gibbs sampler can be long (each sociodemographic group

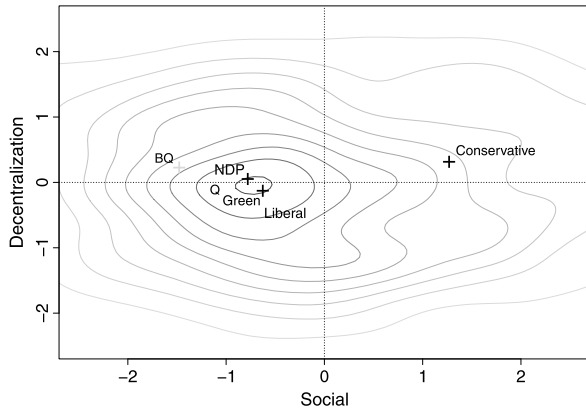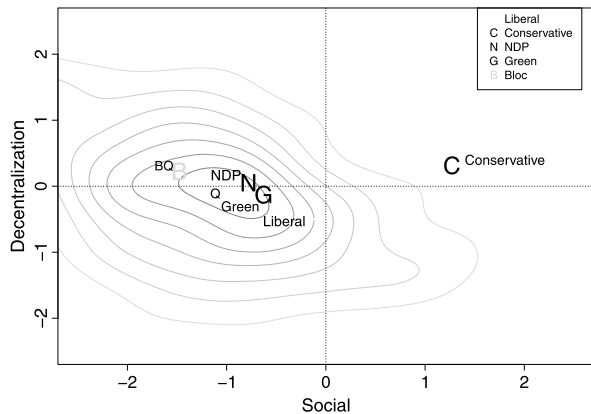**Fig. 1** Distribution of voters and party positions for Canada in 2004

Fig. 2 Distribution of voters and party positions for Quebec in 2004

has a random effect for each region being considered), thus it is always a good idea to examine the relationships between the variables and see if it makes sense to keep them all in the model. In this case, after toying with the model for some time, it seemed that the relationship between sex and vote was yielded spurious by age and education. Thus, to preserve time and allow the Gibbs sampler to run efficiently, our model does not include sex as a variable.

Using the varying choice set logit proposed earlier, we estimate β and the valences for a model with sociodemographics. For the model, given some correlation between the random effects of interest and the independent variable of Euclidian difference, we use the random effects correction procedure proposed earlier. We include the mean difference for each party in each region's respective random effects by setting the mean of the normal priors to the random effects at this value. To assist in convergence of the VCL, we create a diffuse gamma hyperprior for the variance of each prior. As stated before, this model does take a while to converge, so it is necessary to let the Gibbs sampler for this model run a while. We ran each Gibbs sampler for around 100,000 iterations and received nice normal distributions for

each of the parameters of interest. Similarly, allowing the Gibbs sampler to run this long reduces the effects of the inherent autocorrelation that occurs in the sampler.

The results of the VCL are shown in Table 4. We show the VCL estimates of the parameter values and the corresponding 95 percent credible intervals. In this example, we use the Liberal Party as the base group, thus their valence is always restricted at 0. For the model, we report β and the aggregate valences first. We then report the regional effect for each party. While the sociodemographic random effect values may be of substantive interest sometimes, they are included simply as controls in this case, thus we do not report these values. We also report the deviance information criterion (DIC), which is a hierarchical model analogue to AIC or BIC. When the posterior distribution is assumed to be multivariate normal (as it is in this case), the DIC functions as a measure of model quality rewarding a model with a small number of parameters, but penalizing a model that does not fit the data well. The DIC can be seen as a measure of the log-likelihood of the posterior density. Lower values of DIC are preferred.

From this model, we can see a number of things. First, as would have been predicted before running the model, the Liberal Party is the highest valence party in Canada outside of Quebec. However, the Conservative Party is almost equivalent in valence level. By simply adding the aggregate valence to the Non-Quebec regional random effect, we can see that the two are almost equivalent in valence outside of Quebec. However, this model shows that the BQ is, in fact, the highest valence party in Canada. This makes sense, given that of the people that could actually vote for the party, nearly 50 percent of them did. This exemplifies one of the strengths of this model, which is that it accurately specifies this party as the highest valence party, even though it is only available to around 25 percent of the electorate. Thus, if we view parties as entities that look down and see a uniform electorate of members without specific regional affiliation or sociodemographic groups, then they would estimate that BQ is the highest valence party.

Outside of Quebec, as mentioned before, the Conservative Party and the Liberal Party are the highest valence parties, with almost equivalent valence. The NDP is of somewhat lower valence as the party simply does not have the same presence as its larger Liberal counterpart. However, its valence and positioning in the preference space of Canada allows it to be a significant competitor outside of Quebec. The lowest valence party outside of Quebec is the Green Party, which makes plenty of sense as it is was (and is still) more of a one-issue dimension party and fails to have mass appeal to the electorate.

Inside Quebec, BQ is the highest valence party, with an even larger valence than that estimated by the aggregate valence measure. The Liberal Party also has a strong presence in Quebec; however, given that BQ and the Liberal Party are in similar areas of the preference space, they compete for many of the same voters and BQ simply has a stronger presence in Quebec. The Conservative Party is of somewhat lower valence within Quebec, as it fails to draw voters that instead choose to vote for BQ. The lowest valence party in Quebec is also the Green Party.

Recall that we are interested in finding where the parties will locate in the policy space in order to maximize their vote share. Because the outcome of the election

Table 4 2004 Canada VCL model given sociodemographics (LPC base)

	NDP		Conservatives		Greens		BQ	
β			0.2598					
			(0.2234, 0.2976)					
λ_j	−0.5883		−0.0864		−1.826		0.4525	
	(−1.678, 0.4625)		(−1.057, 1.107)		(−3.446, 0.2405)		(−0.9229, 2.322)	
	NQ	Q	NQ	Q	NQ	Q	NQ	Q
μ_{rj}	0.0341	−0.6085	0.0655	−0.1827	−0.4038	−0.4915	–	0.1797
	(−1.014, 1.127)	(−1.778, 0.4704)	(−1.135, 1.04)	(−1.465, 0.8110)	(−2.493, 1.217)	(−2.720, 1.062)	–	(−1.67, 1.565)
DIC			2029.291					

depends on these vote shares, we assume that parties use polls and other information at their disposal to form an idea of the anticipated election outcome and then use this information to find their most preferred position taking into account their estimates of where other parties will locate.

One possibility is that all parties will locate at their respective electoral means, meaning that $\mathbf{z}^*$ is as follows:

$$
\mathbf{z}^* = \begin{bmatrix} & \textit{Lib.} & \textit{Con.} & \textit{NDP} & \textit{Grn.} & \textit{BQU} \\ S & 0 & 0 & 0 & 0 & -1.11 \\ D & 0 & 0 & 0 & 0 & -0.08 \end{bmatrix}
$$

Notice that this means that BQ will not locate at the same position as the other parties as it only runs in Quebec, so its regional mean is at the mean of voters in Quebec. Given this vector of party positions and the information about the voter ideal points, we can calculate the Hessian of the vote function for each party as well as the convergence coefficient, $c(\mathbf{z}^*)$ for each party. For the Hessians, we are interested in the eigenvalues associated with the Hessians for each party; if they are both negative, then the Hessian is negative definite and the party location is at a local maximum. Given $\mathbf{z}^*$, if any of the Hessians are not negative definite, then one of the parties will not choose to locate at this position in equilibrium. Similarly, we can check the convergence coefficients to see if they meet the necessary condition for convergence. Given that any of these conditions fail, the party for which they fail will choose to move elsewhere in the policy space at equilibrium and. Given that the Green Party is the lowest valence party in both regions, as well as at the aggregate level, we can assume that if a party is going to move, it will be the Green Party. We now examine the Hessians and $c(\mathbf{z}^*)$ for each party.

$$
\mathcal{H}_{Lib} = \begin{bmatrix} -0.0365 & -0.0004 \\ -0.0004 & -0.0705 \end{bmatrix}; \quad \mathcal{H}_{NDP} = \begin{bmatrix} 0.0021 & 0.0012 \\ 0.0012 & -0.0362 \end{bmatrix}
$$

$$
\mathcal{H}_{Con} = \begin{bmatrix} -0.0326 & -0.0002 \\ -0.0002 & -0.0676 \end{bmatrix}; \quad \mathcal{H}_{GPC} = \begin{bmatrix} 0.0085 & 0.0085 \\ 0.0085 & -0.0091 \end{bmatrix}
$$

$$
\mathcal{H}_{BQ} = \begin{bmatrix} -0.1194 & 0.0034 \\ 0.0034 & -0.1286 \end{bmatrix}
$$

$$
eigen(\mathcal{H}|z^*) = \begin{bmatrix} & \textit{Lib.} & \textit{NDP} & \textit{Con.} & \textit{Grn.} & \textit{BQ} \\ \textit{Eigen}1 & -0.0365 & 0.0021 & -0.0326 & 0.0085 & -0.1183 \\ \textit{Eigen}2 & -0.0705 & -0.0361 & -0.0676 & -0.0092 & -0.1297 \end{bmatrix}
$$

$$
c_j(\mathbf{z}^*) = \begin{bmatrix} & \textit{Lib.} & \textit{NDP} & \textit{Con.} & \textit{Grn.} & \textit{BQ} \\ c(\mathbf{z}^*) & 1.031 & 1.518 & 1.071 & 1.945 & -0.5921 \end{bmatrix}
$$

From the Hessian's and their corresponding eigenvalues, we can see that two parties will diverge from the vector of electoral means. The NDP and the Green Party both have positive eigenvalues, meaning that $\mathbf{z}^*$ is not a vote maximizing position for them and, thus, not a LNE. It is interesting to note that both of these parties $\mathbf{z}^*$ is

a saddle point. Thus, when they choose a better position, it will still be on the mean of the decentralization axis as the second eigenvalue represents that axis.

We can also utilize the test of convergence coefficients to assess convergence to the vector of interest. Here, we see that all of the convergence coefficients, except for BQ's, are greater than one but less than w (which in this case is 2),[4] thus we need to check the largest one to see if it indicates convergence to the mean vector. The largest convergence coefficient belongs to the Green Party and examination of the constituent portions of its $c(\mathbf{z}^*)$ shows:

$$c_{GPC}(\mathbf{z}^*) = 1.379 + 0.5657$$

where 1.379 corresponds to the social axis. This means that the Green Party is not maximizing its vote share at the mean social position. These values indicate that the Green Party is also located at a saddle point when given the mean vector, just as the Hessian test did.

However, taken as they are, we do not know if these two tests actually match the vote maximizing tendencies of the parties. Thus, in order to give validity to the proposed tests, we need to use optimization methods to show that the vote maximizing positions for parties are not located on the mean vector. In a Gibbs sampling style of optimizer, we create an optimization method in which each party optimizes its vote share given the positions of the other parties. If we do this for each party in rotation beginning at some arbitrary starting values, the parties should eventually converge on the equilibrium set of positions where no party can do any better by moving given the positions of the other party. This method is necessary given that each party can potentially be optimizing over a different portion of the electorate. In this case, while the other four parties are attempting to optimize their respective vote shares over all of Canada, BQ is only trying to optimize its vote share among those voters in Quebec. Thus, this style of optimizer is necessary for finding the optimizing positions in Canada.

Figure 3 shows the vote optimizing positions for each party in Canada, which are as follows:

$$z^*_{opt} = \begin{bmatrix} & \textit{Lib.} & \textit{Con.} & \textit{NDP} & \textit{Grn.} & \textit{BQ} \\ S & 0.0524 & 0.0649 & 1.099 & 2.337 & -1.069 \\ D & -0.0259 & -0.0264 & 0.0266 & 0.2281 & -0.1290 \end{bmatrix}$$

Fortunately for our measures, the vote optimizing positions echo what we were told by the convergence coefficients: the NDP and the Green Party have incentive to move away from the electoral mean while the other parties want to stay there. Given that these two parties are of relatively low valence, their relocation has little effect on the maximizing positions for the largest three parties. However, in accordance with

---

[4]It is interesting to note that the convergence coefficient need not be positive, as is the case with $c_{BQ}(z^*)$. This simple indicates a particularly strong desire to stay in the given position. A negative convergence coefficient indicates a quickly changing local maximum, meaning that a small departure from this position would result in a large decrease in vote share.
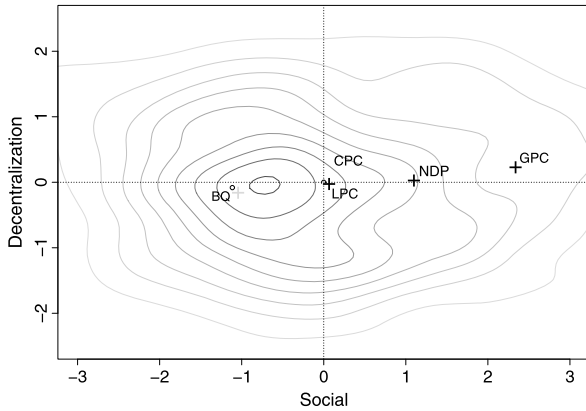
Fig. 3 Vote maximizing positions in Canada 2004

Table 5 Vote shares given various z^*s

	Current	Mean	Optimal
LPC	36.71	33.42	33.43
CPC	29.66	33.34	33.29
NDP	15.65	17.89	16.96
GPC	4.29	3.55	3.80
BQ	12.42	11.79	12.52

the equilibrium theory of proposed by Schofield (2007), the parties locate along the same axis, with distances away from their electoral means proportional to their respective perceived valence differences.

This begs the question, though, how much better can the parties do at these positions than they did at their current positions? Table 5 shows the vote shares in the sample for each party at their current positions, at the electoral mean, and at the vote maximizing positions determined by the optimization routine. These vote shares are predicted using the actual valences from each region (i.e. the aggregate valences plus the regional random effects).

This table strengthens our notion that the vector of means is not a LNE as the Green Party, the BQ, and the Liberals all do better when the Green Party and the NDP locate away from the mean. As the Green Party is one of the parties that is dissatisfied with the electoral mean, it can choose to move to a more extreme position and do better. The NDP is forced to adapt and do worse than it would if the parties all located at their respective electoral means.

5 Conclusion

In this paper, we proposed a method for examining the vote maximizing positions of parties in electoral systems with parties that do not run in every region. When par-

ties do not run in every region, different voters have different party bundles at the polls and existing theories of valence and empirical methods for estimating valence are no longer appropriate. We proposed a more generalized notion of the convergence coefficient which is able to handle any generalized vector of party positions and tell us whether or not these positions are a local Nash equilibrium for the given electoral system. We also proposed a new method for estimating the parameters necessary to utilize the convergence coefficient that does not rely on the IIA assumption. Though methods of doing so already exist, the sheer amount of information gained from the Varying Choice Set Logit makes it the ideal model to run when examining voting tendencies within complex electorates that have clear hierarchical structures.

Using these methods, we examined the 2004 Canadian elections. Using the new empirical methods, we found that even though it only ran in Quebec, a region that makes up around 25 percent of Canada's population, the Bloc Quebecois was the highest valence party in Canada in the 2004 elections. Using these empirical findings, we found that parties were not able to maximize their respective vote shares by locating at the joint electoral mean, which included BQ locating at the mean of voters in Quebec and not at the join electoral mean. Rather, the lower valence parties were able to maximize vote shares by taking more extreme positions within the policy space. This finding is in direct contrast of widely accepted theories that political actors can always maximize their vote shares by taking positions at the electoral center.

Given the accurate outcomes of these methods, there are a number of more complex situations in which these methods can be used. First, this type of model is not limited to the two region case and can be applied to cases where there are numerous "party bundles" which arise in a nation's electorate. A region, in this case, is equivalent to a party bundle; thus, a region can be a combination of many regions (the case when a party runs in two out of three regions, for example). Similarly, in further uses of this model, it is possible to examine equilibria where parties have perfect information about each of the voters, meaning that parties know each voter's region, sociodemographic group, and ideal point. Given this information, new equilibria can be computed and differences can be examined. This further demonstrates the general nature of the new definition of the convergence coefficient and its ability to handle an even wider variety of electorate types than previously.

Appendix

This appendix gives the algorithm for the Gibbs sampling.

```
model{

for(i in 1:N) {
for(k in 1:K) {
v[i,k] <- alpha[k] + beta[1]*(d[(N*(k-1))+i]-d[i]) +
```

```
m[region[i],k] + ed[region[i], education[i], k] +
ag[region[i],education[i],age[i],k]

expv[i,k] <- exp(v[i,k])
pv[i,k] <- expv[i,k]/sum(expv[i,1:K])
vote[i]  ~ dcat(pv[i, 1:K])
}}

beta[1] ~ dnorm(0,taub[1])I(-5,5)

alpha[1] <- 0
alpha[2] ~ dnorm(0,taua[2])
alpha[3] ~ dnorm(0,taua[3])
alpha[4] ~ dnorm(0,taua[4])
alpha[5] ~ dnorm(0,taua[5])

m[1,1] <- 0
m[1,2] ~ dnorm(0,taum[1,2])
m[1,3] ~ dnorm(0,taum[1,3])
m[1,4] ~ dnorm(0,taum[1,4])
m[1,5] <- -100000
m[2,1] <- 0
m[2,2] ~ dnorm(0,taum[2,2])
m[2,3] ~ dnorm(0,taum[2,3])
m[2,4] ~ dnorm(0,taum[2,4])
m[2,5] ~ dnorm(0,taum[2,5])

taub[1] ~ dgamma(.1,.1)I(.1,10)
taua[2] ~ dgamma(.1,.1)I(.1,10)
taua[3] ~ dgamma(.1,.1)I(.1,10)
taua[4] ~ dgamma(.1,.1)I(.1,10)
taua[5] ~ dgamma(.1,.1)I(.1,10)
taum[1,2]~dgamma(.1,.1)I(.1,10)
taum[1,3]~dgamma(.1,.1)I(.1,10)
taum[1,4]~dgamma(.1,.1)I(.1,10)
taum[2,2]~dgamma(.1,.1)I(.1,10)
taum[2,3]~dgamma(.1,.1)I(.1,10)
taum[2,4]~dgamma(.1,.1)I(.1,10)
taum[2,5]~dgamma(.1,.1)I(.1,10)

for(f in 1:e){
ed[1,f,5] <- -10000
}

for(f in 1:e){
for(z in 1:4){
ed[1,f,z] ~ dnorm(0,taued[1,f,z])
taued[1,f,z] ~ dgamma(.01,.01)I(.01,10)
}}
```

```
for(f in 1:e){
for(z in 1:5){
ed[2,f,z] ~ dnorm(0,taued[2,f,z])
taued[2,f,z] ~ dgamma(.01,.01)I(.01,10)
}}

for(f in 1:e){
for(w in 1:a){
ag[1,f,w,5] <- -10000
}}

for(f in 1:e){
for(z in 1:4){
for(w in 1:a){
ag[1,f,w,z] ~ dnorm(0,tauag[1,f,w,z])
tauag[1,f,w,z] ~ dgamma(.01,.01)I(.01,10)
}}}

for(f in 1:e){
for(z in 1:5){
for(w in 1:a){
ag[2,f,w,z] ~ dnorm(0,tauag[2,f,w,z])
tauag[2,f,w,z] ~ dgamma(.01,.01)I(.01,10)
}}}

for(f in 1:e){
for(z in 1:4){
for(w in 1:a){
tot[1,f,w,z] <- alpha[z] + m[1,z] + ed[1,f,z] + ag[1,f,w,z]
}}}

for(f in 1:e){
for(z in 1:5){
for(w in 1:a){
tot[2,f,w,z] <- alpha[z] + m[2,z] + ed[2,f,z] + ag[2,f,w,z]
}}}
}
```

References

Adams J (2001) Party competition and responsible party government. University of Michigan Press, Ann Arbor

Adams J, Merrill S III (1999) Modeling party strategies and policy representation in multiparty elections: why are strategies so extreme? Am J Polit Sci 43:765–781

Benoit K, Laver M (2006) Party policy in modern democracies. Routledge, London

Blais A, Fournier P, Gidengil E, Nevitte N, Everitt J (2006) Election 2006: how big were the changes... really? Working paper, Universite de Montreal

Caplin A, Nalebuff B (1991) Aggregation and social choice: a mean voter theorem. Econometrica 59:1–23

Dow JK, Endersby J (2004) Multinomial logit and multinomial probit: a comparison of choice models for voting research. Elect Stud 23:107–122

Downs A (1957) An economic theory of democracy. Harper and Row, New York

Gelman A, Park D, Shor B, Bafumi J, Cortina J (2008) Red state, blue state, rich state, poor state: why Americans vote the way they do. Princeton University Press, Princeton

Hotelling H (1929) Stability in competition. Econ J 39:41–57

Kramer G (1978) Existence of electoral equilibrium. In: Ordeshook P (ed) Game theory and political science. New York University Press, New York, pp 375–389

Merrill S, Grofman B (1999) A united theory of voting. Cambridge University Press, Cambridge

Patty JW (2005) Local equilibrium equivalence in probabilistic voting models. Games Econ Behav 51:523–536

Patty JW (2007) Generic difference of expected vote share and probability of victory maximization in simple plurality elections with probabilistic voters. Soc Choice Welf 29:149–173

Penn E (2003) A model of far-sighted voting. Unpublished paper, Harvard University

Poole K, Rosenthal H (1984) US presidential elections 1968–1980. Am J Polit Sci 28:283–312

Quinn K, Martin A, Whitford A (1999) Voter choice in multiparty democracies. Am J Polit Sci 43:1231–1247

Riker WH, Ordeshook PC (1973) An introduction to positive political theory. Prentice-Hall, Englewood Cliffs

Schofield N (2007) The mean voter theorem: necessary and sufficient conditions for convergent equilibrium. Rev Econ Stud 74:965–980

Schofield N, Martin A, Quinn K, Whitford A (1998) Multiparty electoral competition in the Netherlands and Germany: a model based on multinomial probit. Public Choice 97:257–293

Train K (2003) Discrete choice methods for simulation. Cambridge University Press, Cambridge

Yamamoto T (2011) A multinomial response model for varying choice sets, with application to partially contested multiparty elections. Working paper