A Multinomial Response Model for Varying Choice Sets, with Application to Partially Contested Multiparty Elections*

Teppei Yamamoto†

First Draft: September 15, 2010
This Draft: April 16, 2014

Abstract

This paper proposes a new multinomial choice model which explicitly takes into account variation in choice sets across observations. The proposed *varying choice set logit* (VCL) model relaxes the independence of irrelevant alternatives assumption by allowing the individual random utility function to directly depend on choice set types, and can be applied to a variety of data in which some individuals can only choose from a subset of the theoretically possible responses. Both frequentist and Bayesian simulation-based estimation procedures are developed using the Monte Carlo expectation-maximization algorithm and Markov chain Monte Carlo, respectively. The proposed model can be used to analyze survey data in partially contested multiparty elections in which some political parties do not run their candidates in every district. For illustration, I apply the proposed method to the 1996 Japanese general election, where none of the districts was contested by all of the six major parties.

Word Count: 8,482

Key Words: discrete choice models, independence of irrelevant alternatives, mixed logit, hierarchical Bayes, Markov chain Monte Carlo, Monte Carlo EM algorithm, partially contested multiparty elections

---

*I thank Kosuke Imai, John Londregan, and Nolan McCarty for their guidance and encouragement. I am also grateful to Bernd Beber, Neal Beck, Jeff Gill, Minjung Kyung, Ko Maeda, and seminar participants at Harvard University, Massachusetts Institute of Technology, New York University, Princeton University, and Washington University in St. Louis for their helpful comments and suggestions.

†Assistant Professor, Department of Political Science, 77 Massachusetts Avenue, E53-463, Massachusetts Institute of Technology, Cambridge, MA 02139. Email: teppei@mit.edu, URL: http://web.mit.edu/teppei/www/

1 Introduction

This paper proposes a new multinomial choice model which explicitly takes into account variation in choice sets across observations. The proposed model can be applied to various situations where some individuals in the sample can only choose from subsets of the responses that are theoretically possible. Such situations include, for example, vote choice in multiparty elections where some parties do not run their candidates in every district, transportation mode choice for commuters who may lack access to public transportation, and consumer choice when products may be unavailable in some local markets. The model thus makes it possible to analyze these complex data more accurately than standard models of discrete choice, which often implicitly assume that all choices are available to every observation.

Traditionally, multinomial responses have been analyzed using the multinomial logit model (MNL, also called the conditional logit model), which was popularized by McFadden (1973). Since his seminal work, numerous empirical studies have been conducted using MNL to analyze various aspects of human behavior, such as the choice of transportation mode (McFadden, 1974), consumer brand choice (Guadagni and Little, 1983), coalition government formation (Martin and Stevenson, 2001), public opinion on policy issues (Branton and Jones, 2005), and voting (Whitten and Palmer, 1996). This approach, however, has been criticized because of its restrictive assumption. Namely, the model implicitly makes the independence of irrelevant alternatives (IIA) assumption, which implies that the relative probability of choosing one alternative instead of another does not depend on whether and what other alternatives are also available.

The IIA assumption is especially problematic when the alternative missing from an individual's choice set is a substitute of another alternative for this individual. For example, suppose a voter faces a choice between two candidates: the incumbent from the government party and a challenger from the main opposition party. Assume further that this voter equally likes the two candidates, so that she will vote for either candidate with equal probability. Now, if another candidate — a challenger with the same party affiliation and very similar characteristics to the existing challenger — is added to the race, will the voter still choose between the original two candidates with equal probability? Because the two challengers are close substitutes for this voter, it is natural to expect that the new candidate will take her vote from the challenger with higher probability. However, the IIA assumption implies this probability to be equal, so that the relative chance of winning between the original two candidates is assumed to remain unchanged. Applying the standard MNL to this situation will thus produce biased estimates of the predicted probabilities. In general, many instances of IIA violations similar to this simple example can occur in a wide variety of applications.

In this paper, I propose a solution to this problem which focuses on the actual variation in choice sets that are present in data. As shown formally in Section 4.1, the proposed *varying choice set logit* (VCL) model relaxes IIA by allowing regression coefficients to vary across the groups of individuals defined by the alternatives available in their actual choice sets. This means that, even for two individuals who are identical in terms of predictors, the predicted probabilities can still differ if they face different choice sets.

A key difference between VCL and standard multinomial choice models is that the former explicitly incorporates the information about varying choice sets while other models do not. As described in Section 4.2 and in Supporting Information (SI), VCL is a generalization of the standard MNL, while it is also a particular type of the mixed logit model (MXL; Train, McFadden and Ben-Akiva, 1987; Bhat, 1998; Train, 1998). Unlike the standard MXL, however, VCL produces a set of estimated coefficients for each choice set type as opposed to each individual, implying efficiency gain. At the same time, VCL will still produce consistent estimates of the predicted probabilities because various types of IIA violation can be taken into account by VCL's group-level parameters. Finally, VCL also differs from the multinomial probit model (MNP; Hausman and Wise, 1978), another popular option for relaxing IIA, in that the systematic component of the VCL model is allowed to vary across choice set types unlike MNP. In Section 5, I propose two Monte Carlo-based methods for fitting VCL, Monte Carlo expectation maximization (MCEM) and Markov chain Monte Carlo (MCMC) algorithms, the latter of which is illustrated with Japanese election survey data in Section 6.

The proposed method is applicable to various multinomial response data which involve varying choice sets. In political science, it is particularly useful for analyzing survey data in partially contested multiparty elections. The next two sections discuss why it is substantively important to explicitly deal with varying choice sets generally in such elections (Section 2), and particularly in the context of the Japanese election example (Section 3).

2 Partially Contested Multiparty Elections

Multiparty systems vary in terms of how citizens are organized into constituencies for electing representatives. In countries like the Netherlands and Israel, voters choose one freely from the entire list of political parties. However, in many other systems where the electorate is divided into multiple districts, parties may not run their candidates in some districts because they lack a local organizational basis or they simply want to avoid wasting resources on impossible-to-win races. Voters in such partially contested districts are thus *ex ante* precluded from choosing certain parties at the ballot box regardless of their preferences.

The standard MNL is clearly less than ideal for modeling such elections, because partially contested districts are likely to be systematically different from fully contested districts. Moreover, difference in the set of available parties is likely to affect the way people vote by changing local electoral contexts. For example, consider a district where the main competitors for the office are two candidates from a conservative government party and an equally conservative opposition party. Because the two candidates are not clearly distinguishable in terms of their ideology, we may expect that voters in such a district will base their decisions on how they evaluate the performance of the current government. On the other hand, in districts where no government party candidate is available, people may instead vote based on their ideological proximity to the candidates. Such difference will not only lead to the violation of IIA, but also imply heterogeneity in the predictive powers of explanatory variables across different combinations of locally available parties.

In addition, it has been suggested that people may switch among different modes of voting behavior depending on the situation they are faced with (Iversen, 1994; Lewis and King, 1999; Claassen, 2009). For example, people may choose to vote for their less preferred candidate because their first choice has little chance of winning in the district (Duverger, 1954). Alternatively, people may vote for a candidate representing more extreme ideology than their true preference, in the hope that electing such a candidate will tip the balance of the legislature or coalition government in their preferred direction (Kedar, 2005). A voter mixing these different types of voting behavior may for example cast a proximity vote only when a party sufficiently close to her own position is available in her own choice set. Clearly, such a possibility represents an additional source of IIA violation.

The IIA assumption can be violated not only due to these micro-level mechanisms but also because of macro-level dynamics. For example, coalition politics in multiparty systems often entails dynamism which implies the violation of IIA. For example, parties that are coalition partners often negotiate prior to an election and coordinate on a candidate, with one of the two parties officially endorsing the candidate and the other providing informal support to the candidate within the district.[1] In this case, supporters of the latter party have an additional reason to prefer the first party to *other* running parties, compared to the hypothetical situation where both parties in the coalition had chosen to run. Thus, the IIA assumption is violated, and ignoring the dependence of voters' preferences on choice sets will lead to biased predictions.

In the field of comparative electoral studies, multinomial response models have long been a standard method to analyze voters' choice among multiple political parties. Early analyses typically used MNL and ignored possible violation of IIA (e.g. Whitten and Palmer, 1996). More recently, however, scholars started to pay closer attention to the possibility of IIA violation. In their well-known article, Alvarez and Nagler (1995) used the multinomial probit model (MNP) to analyze the 1992 U.S. presidential election. Unlike MNL, MNP does not assume IIA and thus can be more useful for analyzing substitution patterns, such as how Perot supporters would have voted had only Bush and Clinton run in the election.[2] An alternative approach has been to use MXL, which not only relaxes IIA but also allows the systematic component of the utility function to vary across voters (Glasgow, 2001). However, as I discuss in Section 4.2, neither MNP nor MXL is ideal for the analysis of partially contested multiparty elections, because they do not utilize the information about *how* IIA is violated, which is often a quantity

[1] Evidence of interparty electoral cooperation has been documented in many multiparty systems, including Chile (Rahat and Sznajder, 1998), Estonia (Laitin, 1994), France (Schrijver, 2004), Italy (Cox and Schoppa, 2002), Japan (Christensen, 2000), and Uganda (Bogaards, 2003).

[2] Because of its flexibility (and the increasing availability of needed computational power), many studies have since used MNP to analyze multiparty elections (e.g. Schofield, Martin, Quinn and Whitford, 1998; Alvarez and Nagler, 1998, 2000; Alvarez, Nagler and Bowler, 2000; Alvarez, Boehmke and Nagler, 2006).

of interest, that is contained in the actual variation in choice sets.

The current paper is hardly the first to point out the problem of varying choice sets in the literature of multiparty elections. In their seminal paper, Katz and King (1999) developed a statistical model for *aggregate* election results which explicitly considers the existence of partially contested districts. Their model was subsequently further investigated and modified by other scholars (Jackson, 2002; Tomz, Tucker and Wittenberg, 2002; Honaker, Katz and King, 2002). In this closely related literature, the outcome to be modeled is district party vote shares instead of individual vote choices, so that multinomial response models are unsuitable. The current paper contributes to the study of partially contested elections by offering a statistical model that can be applied to individual-level survey data.

3 Application: The 1996 Japanese General Election

The 1996 Japanese House of Representatives election represented a watershed in the Japanese postwar electoral history in several respects. It was the first national election since the Liberal Democratic Party (LDP) regained its status as a government party after the landslide loss in the 1993 election. The election was also the first since the introduction of the new mixed-member electoral system, which replaced the old system of single non-transferable vote (SNTV). Scholars were particularly interested in whether the new system has made the personal vote (Cain, Ferejohn and Fiorina, 1987) less important in Japan. The old SNTV system provided LDP politicians with incentives to cultivate a personal vote, because the system forced them to compete with other candidates belonging to their own party (Carey and Shugart, 1995; Cox, 1997). This shifted the locus of Japanese political competition to inside of LDP and marginalized opposition parties (Reed, 1994).

Naturally, some argue that the 1996 election were less personalistic because of the new system (Hirano, 2006). In the new single-member districts (SMDs), personal electoral organization (known as *kōenkai*) alone might no longer be sufficient for candidates to pass the threshold for election, providing them with additional incentives to seek support from broader party base. At the same time, party labels might become more meaningful in voters' decision making since parties no longer ran multiple candidates in a single district. Others, however, argue that the level of personal voting was largely unchanged even after the electoral reform because of the persistence of the institutional and cultural legacy developed under the old system (e.g. Krauss and Pekkanen, 2004; McKean and Scheiner, 2000).

The political instability surrounding the 1996 election produced an unusually complex pattern of party competition. Approximately a month prior to the election, 55 members of the National Diet left the two junior coalition member parties (Social Democratic Party [SDP] and New Party *Sakigake* [SKG]) and formed a new party called the Democratic Party of Japan (DPJ). As a result, the election was mainly fought among the three government parties (LDP, SDP and SKG), DPJ, and the two other opposition parties (the New Frontier Party [NFP] and the Japan Communist Party [JCP]).

Strikingly, *all* of the three-hundred SMDs were only partially contested in the 1996 election. As shown in Table 1, no party except JCP was able to run their candidates in every district. For example, even LDP, the largest government party at the time of the election, was missing from four out of the 174 SMDs in the survey data (see Section 6.1). The main opposition party NFP failed to run their candidates in more than 20 percent of these districts (36 out of 174). The third largest party DPJ was missing from as many as about 48 percent of the districts (83 out of 174). As a consequence, only just above 40 percent of the seats (71 out of 174) were contested with all these three parties participating. This is remarkable since LDP, NFP and DPJ were widely regarded as the three main players in the election.

One possible approach to this complex data structure is to classify the six parties into the government and opposition camps and simplify the problem into a binary outcome model.[3] However, this approach is problematic because the parties that composed each camp were quite different from one another. Both the government and opposition were in fact ideologically split: The government coalition contained the former long-time rivals LDP and SDP, while the opposition camp included conservatives (NFP), the center left (DPJ), and a left-wing protest party (JCP). Moreover, the political roles they were generally expected to play by the public and media were also remarkably different. While NFP was clearly considered the main opposition party, DPJ was widely regarded as the third alternative which could play a pivotal role in the post-election interparty bargaining. Treating these alternatives as one category is thus likely to produce rather misleading inferences.

[3] Steel (2003) used a similar approach of treating all non-LDP parties as one category and applied the binomial probit model.

Choice Set Type	Party: LDP	NFP	DPJ	SDP	SKG	JCP	N_{dist}	N_{obs}
A	✓	✓	✓	✓		✓	8	80
B	✓	✓	✓			✓	63	630
C	✓	✓		✓		✓	10	96
D	✓	✓			✓	✓	2	26
E	✓	✓				✓	50	429
F	✓		✓			✓	19	157
G	✓			✓		✓	4	36
H	✓					✓	13	107
I		✓	✓			✓	1	19
J		✓		✓		✓	1	11
K		✓			✓	✓	3	24
Total							174	1615

Table 1: Variation in the Choice Sets of Candidates in the 1996 Japanese General Election. The table summarizes the number of single member districts in the dataset (N_{dist}) and the number of survey respondents voted in those districts who are included in the analysis in Section 6 (N_{obs}) for each combination of the major parties that actually ran candidates.

A natural alternative is to use a standard multinomial response model, such as MNL, MXL or MNP. However, as I discuss in Section 4.2, these models are less than desirable when the IIA assumption is implausible and there exists actual variation in choice sets in the observed data. The 1996 election clearly represents a case where IIA is problematic, as ample anecdotal evidence suggests the existence of electoral cooperation (*senkyo kyōryoku;* Christensen, 2000) among the three parties comprising the government coalition. In the next section, I propose a new multinomial response model more suitable for the analysis of this difficult but interesting case.

4 The Proposed Model

In this section, I describe the proposed VCL model and its main properties. I also compare it with other standard multinomial response models, including MNL, MXL and MNP.

4.1 The Varying Choice Set Logit Model

Consider a sample of N individual observations, which are randomly drawn from the population of interest. Suppose that individual i faces a choice among a finite number of alternatives. I consider the situation where there exist a total of J alternatives in the population but only a subset of these J alternatives are actually available to each individual. Let R_i denote this choice set, and R the set of the unique choice sets that are present in the sample. For example, if there are three voters, of whom two are choosing between parties 1 and 2 and the other is choosing between parties 1 and 3, then $N = 3$, $J = 3$, $R_1 = R_2 = \{1, 2\}$, $R_3 = \{1, 3\}$, and $R = \{\{1, 2\}, \{1, 3\}\}$.[4]

To simplify the notation, I define the choice set type indicator $m \in \{1, ..., M\}$ where $M = |R|$, such that each integer from 1 to M corresponds to a unique element of R. Because each individual observation belongs to a single choice set type, m can be thought of as a group indicator. I use N_m to denote the size of group m, or the number of observations that has choice set type m. Clearly, $N = \sum_m N_m$. I will use S_m to denote the set of alternatives that corresponds to type m and denote the number of elements in S_m by J_m. This implies that every individual who belongs to type m can only choose from the alternatives in choice set S_m.

Now I define a *random utility model* for individual choice behavior with varying choice sets. Let y^*_{ijm} be the latent utility that individual i (who has choice set type m) receives when choosing alternative j. I assume that y^*_{ijm} can be written as a function of P covariates including an intercept, $x_{ij}^\top = [1, x^1_{ij}, ..., x^{P-1}_{ij}]$, and an error term ε_{ij} which captures the factors that are not included in x_{ij}. That is,

$$y^*_{ijm} = x_{ij}^\top \alpha + z_{ij}^\top \beta_m + \varepsilon_{ij}, \tag{1}$$

where z_{ij} is a Q-dimensional vector composed of the subset of the covariates comprising x_{ij}. Equation (1) says that the utility that voter i obtains from voting for party j can be expressed as the sum of the systematic component, $x_{ij}^\top \alpha + z_{ij}^\top \beta_m$, and stochastic component, ε_{ij}.

The coefficient vector of x_{ij}, α, represents the mean (or "fixed") effect of each element of x_{ij} that is invariant across observations. The coefficient of z_{ij}, β_m, varies across choice set types (hence the subscript m) and represents the deviation from the mean effect of each element of z_{ij}. For now, I as-

[4]Note that this notation also covers the case in which some individuals can choose from the entire choice set $\{1, ..., J\}$ because a set is a subset of itself.

sume that β_m is independently and identically distributed and has the Q-dimensional multivariate normal distribution with mean zero and the covariance matrix $\mathbf{\Sigma}$. That is,

$$\beta_m \mid X \sim \mathcal{N}_Q(\mathbf{0}, \mathbf{\Sigma}), \tag{2}$$

where X is the predictor matrix. Because z_{ij} is composed of a subset of x_{ij}, z_{ij} can be viewed as the set of covariates whose effects on y^*_{ijm} are allowed to vary across choice set types. Because β_m has its own probability distribution, it is conventionally called the "random" effects, and equation (1) can be called the linear utility with random coefficients or "mixed" effects.[5]

An alternative approach for modeling choice set variation is to use dummy variables indicating choice sets and their interaction terms with other covariates instead of random coefficients as in equation (2). Such a model will produce separate estimates of both coefficients and choice probabilities for each choice set type, and it is thus a flexible approach. A similar possible approach is to fit MNL separately for each choice set type, as suggested by Tomz, Tucker and Wittenberg (2002). However, these simple alternatives run into a problem for data like those in Table 1, which contains many "rare" choice set types. Estimators based on partial pooling of information across groups (Gelman and Hill, 2007) like VCL are likely to perform better and therefore preferable in such situations.

In random utility models, the latent utility is mapped to individual choice as follows,

$$y_{ijm} = \mathbf{1}\left\{y^*_{ijm} > y^*_{ijm'} \text{ for all } j' \in S_m \setminus \{j\}\right\}, \tag{3}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and $y_{ijm} \in \{0, 1\}$ is the choice indicator representing whether individual i (in group m) chose alternative j. In words, equation (3) implies that voter i chooses party j when the utility of doing so exceeds the utility she would receive from choosing any other party in the set of parties *from which she can actually choose*. Note that, like other multinomial response models, VCL can accommodate both individual-specific covariates and choice-varying covariates. In the context

[5] As in many random coefficient models, the proposed model can be further extended by assuming a different distribution for β_m or modeling the coefficients with another set of covariates. In the latter case, the model becomes a *multi-level model* which can be estimated with a Bayesian MCMC method (see Section 5.2).

of multiparty elections, the former may include demographic characteristics of survey respondents (e.g., age, gender) and the latter may be candidate characteristics (e.g., incumbency, past office experience). [6]

Now, assume that ε_{ij} is independently and identically distributed with the type-I extreme value distribution. Using the result of McFadden (1973), it can be shown that this assumption implies the following choice probability conditional on the mth choice set,

$$\mathcal{R}_{ijm} \quad = \quad \Pr(y_{ijm} = 1 \mid X, \alpha, \beta_m) \; = \frac{\exp(x_{ij}^{\top}\alpha + z_{ij}^{\top}\beta_m)}{\sum_{k \in S_m} \exp(x_{ik}^{\top}\alpha + z_{ik}^{\top}\beta_m)}. \tag{4}$$

In words, equation (4) shows that, once given the value of random coefficients β_m, the probability that voter i chooses party j from her choice set S_m is equal to the relative magnitude of the systematic component of the latent utility which she receives by voting for party j.

From this, it is straightforward to derive the unconditional choice probability $\mathcal{P}_{ij}$,

$$\begin{aligned} \mathcal{P}_{ij} &= \Pr(y_{ijm} = 1 \mid X, Z, \alpha, \mathbf{\Sigma}) \\ &= \int \frac{\exp(x_{ij}^{\top}\alpha + z_{ij}^{\top}\beta_m)}{\sum_{k \in S_m} \exp(x_{ik}^{\top}\alpha + z_{ik}^{\top}\beta_m)} \phi_Q(\beta_m \mid \mathbf{\Sigma}) d\beta_m, \end{aligned} \tag{5}$$

where $\phi_Q(\cdot \mid \mathbf{\Sigma})$ represents the density function of $\mathcal{N}_Q(\mathbf{0}, \mathbf{\Sigma})$. According to Equation (5), the probability that voter i will choose party j from her actual choice set S_m is equal to the average of equation (4) weighted by the distribution of the random coefficients β_m. Thus, like MXL, VCL can be seen as a mixture of "logits" where the mixing distribution is multivariate normal (see Section 4.2). The likelihood function for the sample of size N can then be written as,

$$\mathcal{L}(y \mid X, Z, \alpha, \mathbf{\Sigma}) \quad = \quad \prod_{m=1}^{M} \prod_{i=1}^{N_m} \prod_{j \in S_m} \{\mathcal{P}_{ij}\}^{y_{ijm}}, \tag{6}$$

which explicitly contains varying choice sets unlike the common representation of the likelihood for a standard multinomial response model.

[6] To simplify the exposition, I implicitly assumed in equations (1) and (2) that the covariates do not include individual-specific covariates. All subsequent discussion in this paper will remain substantively valid even without this simplification. See Section A of SI for full details about the notational framework for VCL.

The primary advantage of VCL lies in the convenient property that it relaxes the IIA assumption in a "minimal" fashion. That is, the model contains complexity just sufficient to allow for a systematic variation of relative choice probabilities across choice sets. This can be seen by showing that under VCL, the (local) odds ratio of choosing alternative j against k between two choice set types is a log-linear function of the random coefficients. That is,

$$OR(j, k, m, m') \equiv \frac{\mathcal{R}_{ijm}/\mathcal{R}_{ikm}}{\mathcal{R}_{ijm'}/\mathcal{R}_{ikm'}} = \exp\{(z_{ij} - z_{ik})^\top(\beta_m - \beta_{m'})\}, \tag{7}$$

which implies that β_m can be interpreted as a parameter capturing the interactive effect between covariates and choice sets. For example, suppose z_{ij} is a binary variable indicating whether party j belongs to the coalition government. The value of β_m then represents how much larger (or smaller) the effect of being a government party on the vote share is in districts with choice set type m than its average effect across all choice sets, on the log odds ratio scale.

Once the mixing distribution ($\mathbf{\Sigma}$) and the fixed coefficients in the logits (α) are estimated, one can obtain predicted choice probabilities by simulating the outcomes based on equation (5). However, an important advantage of VCL is that it can produce other quantities of interest which are directly related to choice set variation. First, one may be interested in how the effects of covariates differ depending on the available choice set. For example, it may be expected that voters' view of government performance plays a larger role when deciding between candidates from the government party and the main opposition party than when the choice is between two opposition parties. The estimates of the random coefficients (β_m) will provide an answer to this kind of empirical question. Second, under the assumption that choice sets are exogenously determined, VCL can be used for counterfactual analyses with respect to the choice sets. In applications to multiparty elections, one potentially interesting question is how different the election outcome would have been if the parties had run in different districts. This type of question can be investigated by setting m to an appropriate value for each observation.

4.2 Comparison with Other Multinomial Response Models

Because VCL is based on a standard random utility model (i.e., equation 1), it is closely related to other multinomial choice models. Here, I briefly discuss three such models: MNL, MXL, and MNP. I argue that VCL is often preferable to these models when there is variation in choice sets in the actual data and one is particularly interested in how choice probabilities depend on the choice sets. More detailed and technical comparison can be found in Section B of SI.

MNL. MNL is a special case of VCL where two simplifying assumptions are typically made. First, the systematic component of the utility function is assumed to be invariant across observations conditional on the values of the covariates. Second, choice sets are also commonly assumed to be identical for every observation, so that every individual can choose any of the J alternatives. The latter assumption, however, can be relaxed without leaving the framework of MNL, although statistical software for fitting MNL does not always have an option for varying choice sets (see Section B of SI).

MNL is often problematic for the analysis of varying choice sets because it relies on the strong IIA assumption. That is, the odds ratio in equation (7) is equal to 1 for any combination of choice sets or alternatives under MNL. This implies that, while both MNL and VCL will consistently estimate choice probabilities when IIA holds, MNL will produce biased estimates when IIA is violated. Additionally, MNL will underestimate statistical uncertainty if observations sharing the same choice set have common unobserved determinants of their utility.

MXL. Instead of allowing coefficients to vary across choice set types (m), MXL lets each individual observation (i) have its own unique coefficients. That is, the random coefficients β in equation (2) are subscripted by i instead of m under MXL. This implies that VCL is a special case of MXL.

Given that MXL is more flexible than VCL, should we always use the former? The answer is no for several reasons. First, the additional assumption made for VCL is inconsequential for the purpose of analyzing how choice probabilities may depend on variation in choice sets. This is because VCL still produces a unique estimate of the relative choice probability for each choice set type even after the restriction. Second, if the additional assumption is correct, VCL is likely to produce estimates with less statistical uncertainty. VCL, therefore, can be seen as an efficient middle ground between MNL and MXL that is flexible enough to fully capture choice-set dependence.

MNP. MNP is a popular alternative to MNL which does not assume IIA. As Alvarez and Nagler (1998) point out, however, analyses using MNP "only relax the IIA assumption through the specification of the stochastic (random) component of the model" (p.85), treating the IIA violation simply as nuisance even when it may be theoretically interesting. In particular, the effects of covariates in MNP are assumed to be fixed across individuals because the systematic component of the utility function is invariant by assumption. In contrast, VCL allows the coefficients to vary across choice sets and thus can be used to analyze how differently covariates affect choice probabilities depending on which alternatives are actually available.

5 Estimation Strategies

In this section, I present two methods for estimating VCL. Because VCL can be seen as a special case of MXL (see Section 4.2), many existing estimation procedures for the latter model (e.g. Train, 2009) can be used for VCL with minor modifications. I first summarize the *Monte Carlo expectation maximization* (MCEM) algorithm and then a Bayesian procedure based on *Markov chain Monte Carlo* (MCMC). Throughout this section, I assume for the sake of simplicity that the coefficient on every covariate in the model is allowed to vary randomly across choice sets, i.e., $P = Q$. This implies that the utility function (equation 1) can be written as $y^*_{ijm} = x_{ij}^\top \eta_m + \varepsilon_{ij}$ where $\eta_m \sim \mathcal{N}_Q(\alpha, \mathbf{\Sigma})$.

5.1 Monte Carlo Expectation Maximization Algorithm

The first procedure uses the MCEM algorithm (Wei and Tanner, 1990). In general, the EM algorithm is suited for situations in which the actual (or "observed-data") likelihood function is complex and difficult to directly maximize but can be simplified by augmenting the data by some auxiliary information. The maximum likelihood estimator (MLE) can then be obtained by first computing the expected value of the augmented (or "complete-data") likelihood with respect to the auxiliary data given the current value of the parameters, maximizing this expected value with respect to the parameters to update the parameter values, and then repeating these steps until convergence (Dempster, Laird and Rubin, 1977). In the

MCEM algorithm, a Monte Carlo approximation of the expected complete-data likelihood is used in order to avoid intensive computation of the exact complete-data likelihood.

The proposed MCEM algorithm for VCL is based on the procedure developed for MXL by Train (2008). The key idea is to treat the random coefficients, η_m, as the auxiliary information for augmentation, or "missing data," and repeat the following two steps until convergence.

- *E step:* Calculate the simulated expectation of the complete-data log-likelihood using D Monte Carlo draws of η_m^d from $\mathcal{N}_Q(\alpha^t, \mathbf{\Sigma}^t)$ as,

$$\check{\mathcal{Q}}_i^*(\alpha, \mathbf{\Sigma} \mid \alpha^t, \mathbf{\Sigma}^t, X, y) \quad = \quad \frac{1}{D}\sum_{d=1}^{D} \frac{\left\{\prod_{j\in S_m} L_{ij}(\eta_m^d)^{y_{ijm}}\right\} \log \phi_Q(\eta_m^d \mid \alpha, \mathbf{\Sigma})}{\frac{1}{D}\sum_{d'=1}^{D}\left\{\prod_{j\in S_m} L_{ij}(\eta_m^{d'})^{y_{ijm}}\right\}}, \tag{8}$$

  where $(\alpha^t, \mathbf{\Sigma}^t)$ denotes the current (tth) value of $(\alpha, \mathbf{\Sigma})$ and L_{ij} is defined in Section C.1 of SI.

- *M step:* Find the value of $(\alpha, \mathbf{\Sigma})$ that maximizes the sum of equation (8) over N observations, and set the result as $(\alpha^{t+1}, \mathbf{\Sigma}^{t+1})$.

After the convergence of the algorithm, an estimate of the random effects can be obtained as their posterior mean evaluated at the MLE of $(\alpha, \mathbf{\Sigma})$. More details of this algorithm are discussed in Section C.1 of SI.

5.2 Bayesian Markov Chain Monte Carlo

The second procedure is based on MCMC, a common strategy for the Bayesian inference of multinomial choice models (e.g. Allenby and Lenk, 1994; Allenby and Rossi, 1998). Specifically, I propose the following adaptive Metropolis-within-Gibbs sampler, which is derived and justified in Section C.2 of SI.

1. Draw η^{t+1} from its conditional posterior given $(\alpha^t, \mathbf{\Sigma}^t)$ using a random-walk Metropolis sampler:

$$f(\eta \mid Y, \alpha^t, \mathbf{\Sigma}^t, X) \quad \propto \quad \prod_{m=1}^{M} \exp\left(-\frac{1}{2}(\eta_m - \alpha^t)^\top \left(\mathbf{\Sigma}^t\right)^{-1}(\eta_m - \alpha^t)\right) \prod_{i=1}^{N_m} \frac{\sum_{k\in S_m} y_{ikm} \exp(x_{ik}^\top \eta_m)}{\sum_{k\in S_m} \exp(x_{ik}^\top \eta_m)},$$

   where the variance of the jumping distribution is adjusted once every 50 iterations, as proposed by Roberts and Rosenthal (2009).

2. Draw α^{t+1} from its conditional posterior given $(\mathbf{\Sigma}^t, \eta^{t+1})$, which is proportional to the density of $\mathcal{N}(\sum_{m=1}^{M} \eta_m^{t+1}/M, \mathbf{\Sigma}^t/M)$ with an improper flat prior on α.

3. Draw $\mathbf{\Sigma}^{t+1}$ from its conditional posterior given $(\alpha^{t+1}, \eta^{t+1})$, which is proportional to the density of $\text{IW}(Q + M, (Q\boldsymbol{I}_Q + M\boldsymbol{S}^{t+1})/(Q + M))$ when the vague conjugate prior $\text{IW}(Q, Q\boldsymbol{I}_Q)$ is used on $\mathbf{\Sigma}$. Here, $\text{IW}(a, A)$ denotes the inverse-Wishart distribution with a degrees of freedom and scale A and $\boldsymbol{S}$ is defined in Section C.2 of SI.

These three steps are repeated until a likely convergence to the target joint posterior distribution is obtained. As described in Section 6.2, the proposed MCMC algorithm has been implemented in R. As it turns out, the use of the adaptive Metropolis sampler (as opposed to a naïve Metropolis or Metropolis-Hastings sampler) plays a key role for achieving convergence within reasonable computation time, along with standard computational techniques such as standardizing predictors and overparameterizing the model (Gelman et al., 2004). The R code is made publicly available on the author's website.

6 Empirical Analysis

Now I apply the proposed method to the 1996 Japanese general election. As discussed in Section 3, this election involved a complex pattern of variation across districts in the set of the parties that actually ran their local candidates. VCL can be used for such a situation in order to analyze heterogeneity across choice set types, as well as voters' hypothetical behavior when their choice sets had been different.

6.1 Data and Model Specification

To analyze voting behavior at the individual level, I use survey data from the Japan Election Study II. The survey covers a nationally representative sample of 181 out of the 300 SMDs in the 1996 election. I use the total of 1862 cases from the last two waves of this seven-wave panel survey, which yields the final study sample of 1615 respondents from 174 SMDs after excluding cases with missing information, non-voters, and those coming from districts with unusually strong candidates from minor parties or without party affiliation.

The outcome variable is the party membership of the candidate for whom the respondent cast his or her SMD vote. There are two sets of key predictors. First, as discussed in Section 3, scholars were interested in whether the new electoral system had changed the importance of personal voting, and particularly to what extent incumbent legislators were advantaged because of their personal support organizations. To answer these questions, I include *past office experience* and *incumbency status* as candidate-varying predictors. The past experience of a candidate is likely to be strongly correlated with the strength of personal support organizations and can be seen as a proxy for the candidate's personal vote (but see Section 6.2 for possible problems). Then, one may hypothesize that the predictive power of incumbency status will vanish after conditioning on past office experience if the incumbency advantage is predominantly due to candidates' personal support networks. On the other hand, if Japanese incumbent legislators gain advantage from other sources, such as their higher public recognition and greater campaign resource (Stokes and Miller, 1962; Erikson, 1971; Cox and Katz, 1996), then incumbent candidates will have higher probability of being chosen even after their past office experience is taken into account.

Second, many scholars have been interested in the extent to which electoral campaigns affect voters' decision making (e.g. Jacobson, 1990; Levitt, 1994; Hillygus and Jackman, 2003). In the Japanese context, some have wondered whether the introduction of the mixed-member system changed the significance of electoral campaigns in legislative elections (e.g. Christensen, 1998; Reed, 2003). To investigate this question, I include two candidate-varying predictors, *received mail* and *asked to vote*, which indicate whether the respondent received a postcard from the candidate and whether she was personally asked by others to vote for the candidate, respectively. In addition, I include interaction terms between these two variables and incumbency status in order to examine whether campaigns are more effective for challengers than incumbents, a focus of the recent scholarly debate in the field of campaign politics (e.g., Jacobson, 1990; Gerber, 1998, 2004).

Along with the above two sets of predictors, I also include demographic variables for both respondents (gender, age and education) and candidates (gender and age). The summary statistics of these variables are given in Table 2 in SI. Further description of the data can also be found in Section D of SI. All together, the model contains $4 \times (6 - 1) = 20$ fixed effects for the voter characteristics (including intercepts) and eight for the candidate-varying covariates, along with eleven random effects and their variances and covariances for each of these parameters.

6.2 Heterogeneity of Effects across Choice Set Types

Based on the model specification described above, I fit VCL using the MCMC procedure described in Section 5.2. To ensure approximate convergence to the stationary distribution, five Markov chains are run in parallel in R with different starting values for each parameter. After one million iterations with the first 500,000 discarded as burn-ins, the chains seem to be fully mixed based on standard diagnostics (with Gelman-Rubin scores less than 1.1 for all parameters) and the visual inspection of the trace plots. The total computation time was approximately 40 hours, with the five chains parallelized to individual CPU cores (Intel Xeon X5670 running at 2.93GHz). All predictors are standardized to have mean 0 and variance 1 prior to the computation in order to accelerate convergence.

The first question one might ask is how the key predictors on average affected Japanese voters' party choice. This question can be answered by examining the VCL estimates of the fixed effects coefficients (α). In Figure 1, these estimates are presented along with their 95% Bayesian intervals (blue solid squares with horizontal bars). First, the coefficient on the incumbent indicator is positive and significantly different from zero. The VCL estimate of the coefficient is equal to 0.315 with the 95% interval of $[0.057, 0.573]$ even after the candidates' past office experience is taken into account. In fact, the coefficient on the experience variable is statistically indistinguishable from zero (0.089, $[-0.162, 0.336]$) based on VCL. This may imply that under the new electoral system, Japanese incumbent candidates have significant electoral advantage over challengers even if they do not have long political careers and thus are only supported by weak personal organizations. An alternative interpretation, however, is that past office experience is only a poor proxy for the strength of personal support organizations (e.g., due to the existence of second-generation politicians who simply inherit the personal organizations of their parents) and the incumbency variable still captures much of personal voting.

Second, the main effects of both campaign variables have positive signs and significantly different from zero, indicating that electoral campaigns like sending postcards and employing grassroots organizations are effective in mobilizing votes for non-incumbents. The estimated values of these coefficients in

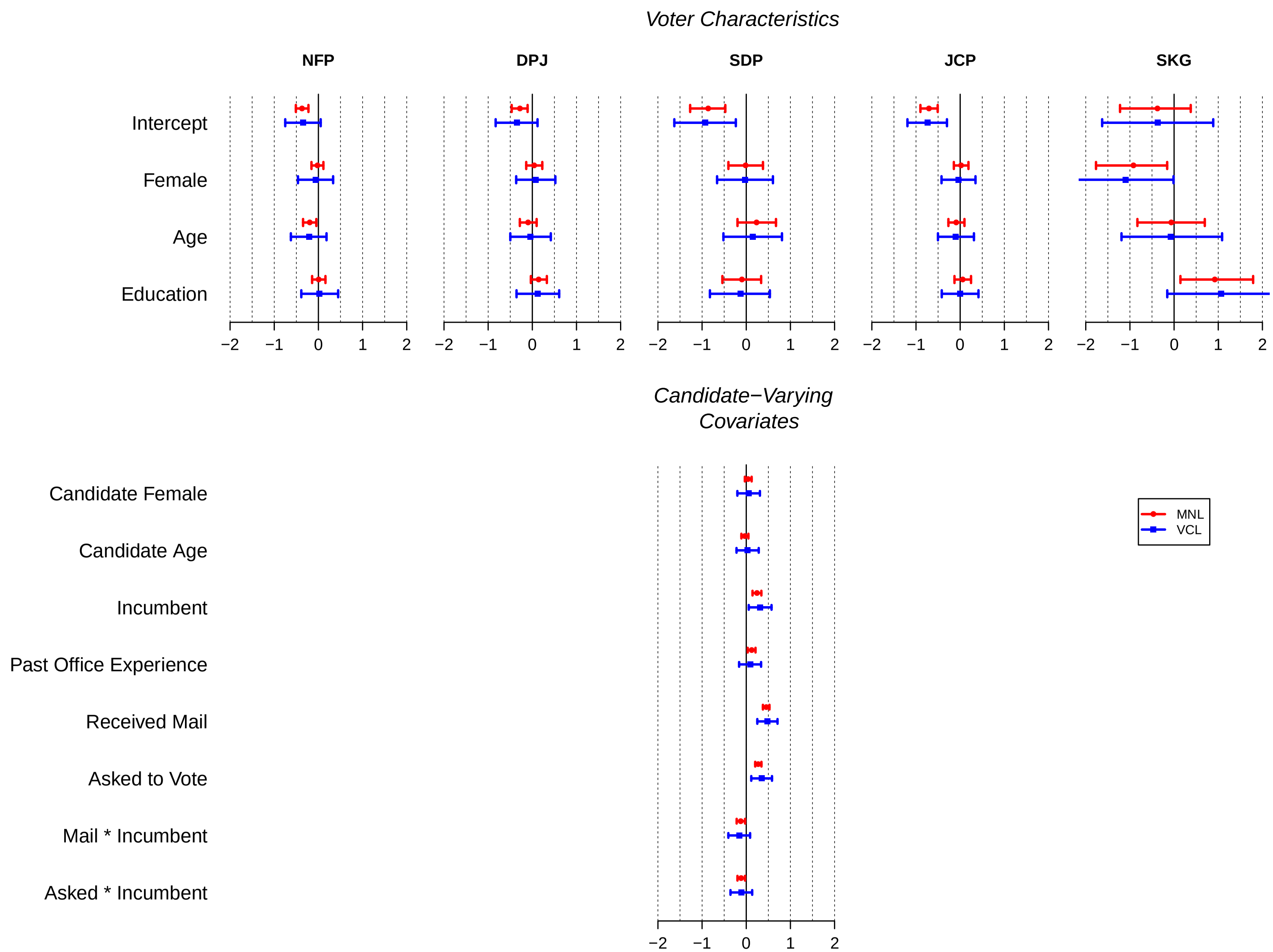

Figure 1: Estimated Model Coefficients for MNL and VCL. For each predictor, the red solid circle (top) and blue solid square (bottom) indicate the point estimates of the corresponding coefficient under MNL and VCL, respectively, along with their 95% Bayesian intervals. The VCL estimates are obtained via the MCMC procedure discussed in Section 5.2. All variables are standardized.

VCL are 0.481 and 0.346 with the 95% intervals of $[0.250, 0.708]$ and $[0.114, 0.583]$, respectively. Moreover, there is evidence that these campaign variables may interact with incumbency status, with both interaction terms having negative coefficients (-0.158 and -0.111, $[-0.405, 0.087]$ and $[-0.357, 0.135]$ for the mail and "asked" interactions, respectively). I return to this possibility shortly below.

For the sake of comparison, Figure 1 also shows the estimates based on MNL obtained via `MCMCmnl` (Martin, Quinn and Park, 2011) (red solid circles with horizontal bars). The MNL estimates of the coefficients are roughly similar to the VCL estimates in terms of posterior means. However, the two sets of estimates are strikingly different in terms of statistical uncertainty, with MNL producing much narrower Bayesian intervals. For example, the estimated coefficient on the experience variable (0.121) is now statistically significant under MNL, with the 95% interval not covering zero ($[0.033, 0.210]$). Indeed, the 95% intervals for MNL are on average less than half as wide as those for VCL. This difference can be attributed to the strong assumption underlying MNL that each voter can be viewed as an independent observation regardless of their districts or choice set types. In contrast, VCL allows for clustering at the choice set level via the inclusion of random coefficients. Because voters who face the same set of available parties are likely to share some unobserved characteristics of local electoral competition (see Section 2), it is important for a statistical model to accommodate the increased uncertainty due to such clustering.

VCL also allows for the direct estimation of effect heterogeneity across choice set types. In Figure 2, I present the estimated coefficients on the key candidate-varying covariates for each of the eleven choice set types (A to K, as labeled in Table 1) along with their 95% intervals (green horizontal bars). The results show that these covariates indeed have moderately heterogeneous effects on voters' decision making, although the effects are estimated with rather high degree of uncertainty for some choice set types. For example, the main effect of incumbency appears to be relatively large for the elections fought between an LDP candidate and a candidate from one of the opposition parties (types E, F and H) and small in races where competition occurred within either the government or opposition camp (B, C and G). One possible interpretation is that incumbency becomes an important advantage only when voters choose between a clear government candidate and a viable opposition candidate. Such interpretation

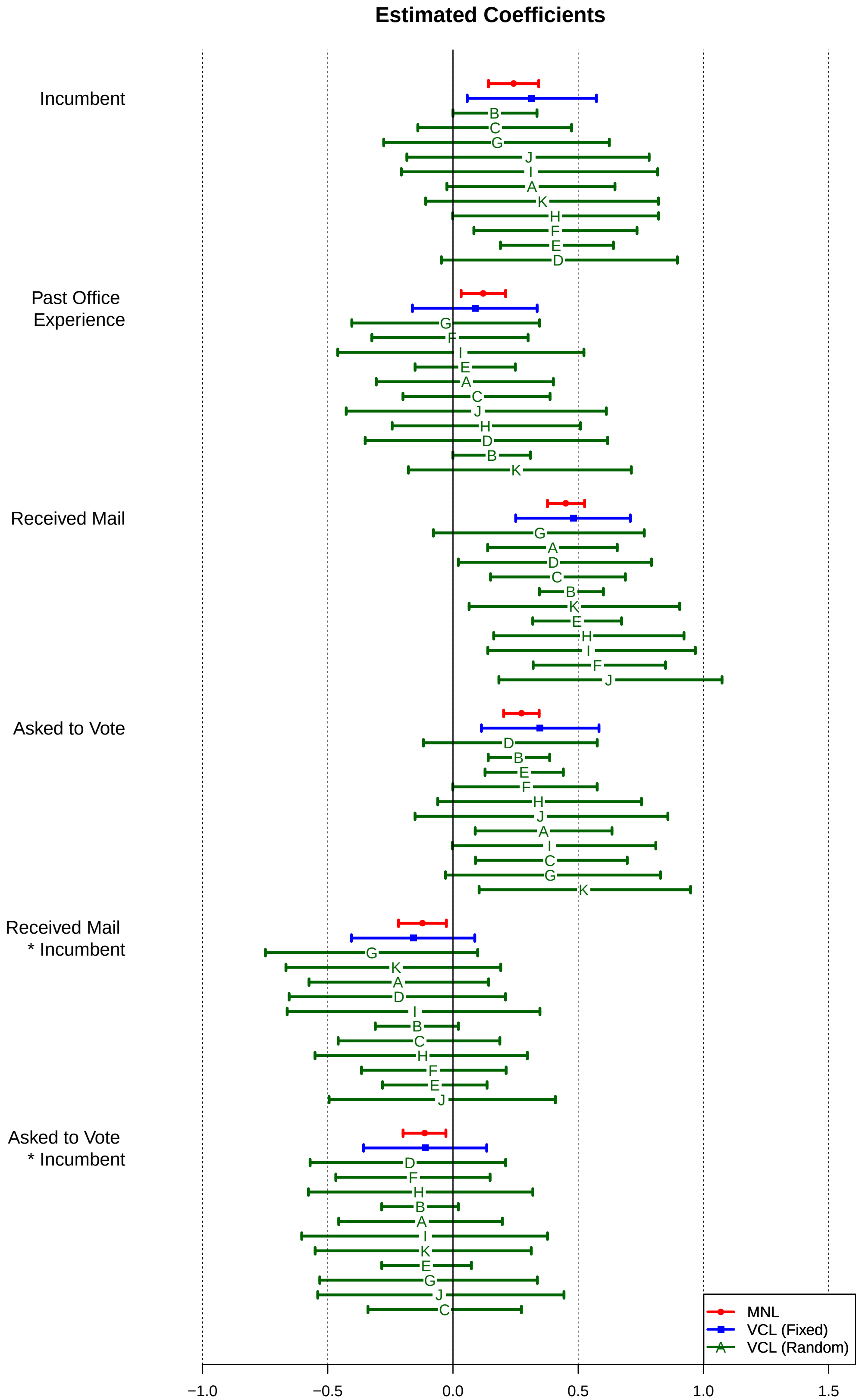

Figure 2: Effect Heterogeneity across Choice Set Types. The panel shows estimated VCL fixed (blue squares) and random effects (green letters) for key candidate-varying covariates and interaction terms along with their 95% Bayesian intervals. The letters A–K indicate choice set types as defined in Table 1. The MNL results are also shown for comparison (red circles).

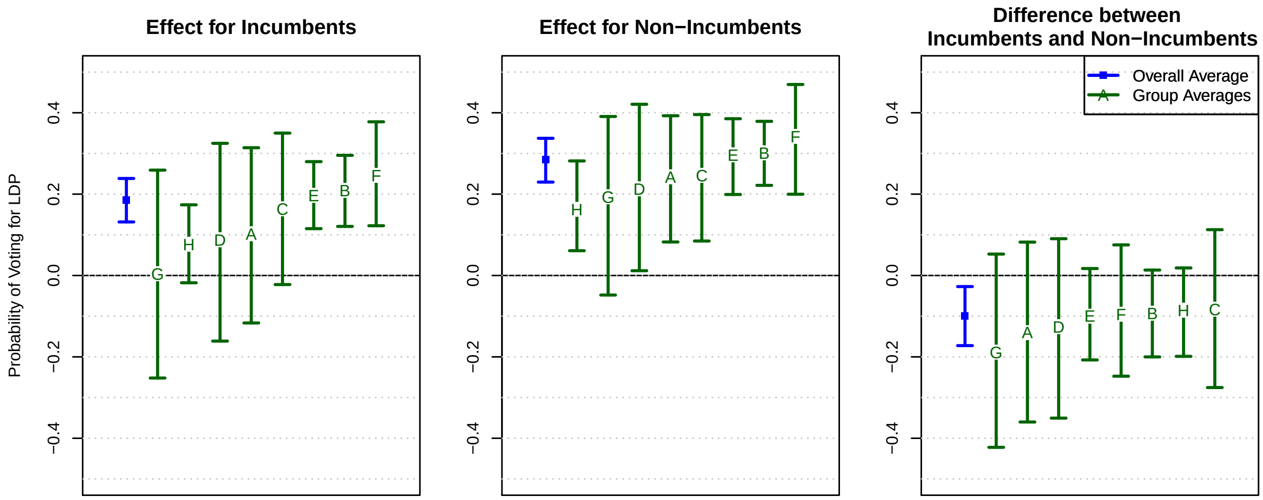

Figure 3: Effect of Receiving Mail on the Probability of Voting for LDP Incumbents and Non-Incumbents. See the caption of Figure 2 for the explanation of graph elements. The choice sets without an LDP candidate (I, J and K) are excluded.

should only be made cautiously given the wide confidence intervals, but it is nonetheless stimulating for further investigation.

In multinomial choice models, the coefficients themselves can be misleading because they are not directly informative about quantities of interest such as changes in predicted choice probabilities. Translating estimated coefficients into more interpretable quantities is therefore essential (King, Tomz and Wittenberg, 2000). In Figure 3, I illustrate one such translation by calculating the interactive effects of receiving mail and incumbency for voters with different choice sets. The first two panels show the estimated average effect of receiving mail from an LDP incumbent (left) and non-incumbent (center) on the probability of voting for the candidate, assuming that there was no other incumbent and no such mail was received from any other candidate. The overall average effect is estimated to be 18.5 and 28.4 percentage points for incumbent and non-incumbent LDP candidates with the 95% Bayesian intervals ranging $[13.2, 23.8]$ and $[23.0, 33.7]$, respectively. The difference, -9.9 percentage points, is statistically different from zero ($[-17.2, -2.7]$) and substantively large (right panel). Interestingly, these effects appear to vary in size depending on choice set types, with some having mostly null effects (D and G) and others substantially large effects (B, E and F). The interaction effects, however, are estimated to be less variable, with all choice set types (perhaps except G) showing magnitudes similar to the overall average. Overall, the result suggests that electoral campaigns are more effective for challengers than incumbent politicians, similarly to the American context (Jacobson, 1990; Gerber, 2004).

6.3 Counterfactual Analysis

As discussed in Section 4, another important advantage of VCL over other models is that one can explicitly conduct counterfactual analyses with respect to choice set types. Because in VCL each choice set type is associated with a unique set of coefficients, it is possible to analyze how choice probabilities would have changed when voters had faced different sets of parties in their districts simply by setting the voters' choice set types to appropriate values and recomputing the predicted probabilities.

This type of analysis is particularly relevant for the case of the 1996 Japanese election. It is widely believed among Japanese political observers that the leaders of DPJ originally planned to merge SDP and SKG in their entirety into a new center-left party (e.g. Ishikawa, 2004, p.188). In reality, only 35 out of the 114 SDP legislators and 15 out of 27 SKG legislators defected from their original parties and joined DPJ. A natural question then is what would have happened if the DPJ leaders had adhered to their original plan and all SDP and SKG legislators had become members of DPJ. Namely, how could the choices of the Japanese voters have been affected by such change in choice sets?

Figure 4 presents the result of a counterfactual analysis. For each choice set group which included either SDP or SKG, I first calculated the predicted vote shares for all the parties that ran their candidates in the *actual* 1996 election. These estimates are represented by the black square on the left side of each column in the seven panels of Figure 4 with the 95% intervals (vertical bars). Next, I estimated the predicted vote shares in the *counterfactual* 1996 election, in which the same SDP and SKG candidates ran except that they changed their party affiliations to DPJ, thereby also changing the choice set type. This can be achieved by setting the random coefficients to the values corresponding to the new choice set type while using the original covariate values, assuming that these same candidates still ran the same electoral campaigns. These estimates are shown as red solid circles, again with the 95% intervals.

The results indicate interesting heterogeneity across choice set types. The top left panel of Figure 4 shows the average effect of the hypothetical SDP–SKG merger across all voters facing choice including these two parties. On average, DPJ would have gained roughly the same number of votes as SDP and SKG combined; the predicted DPJ vote share jumps by about 17 percentage points, which is approxi-

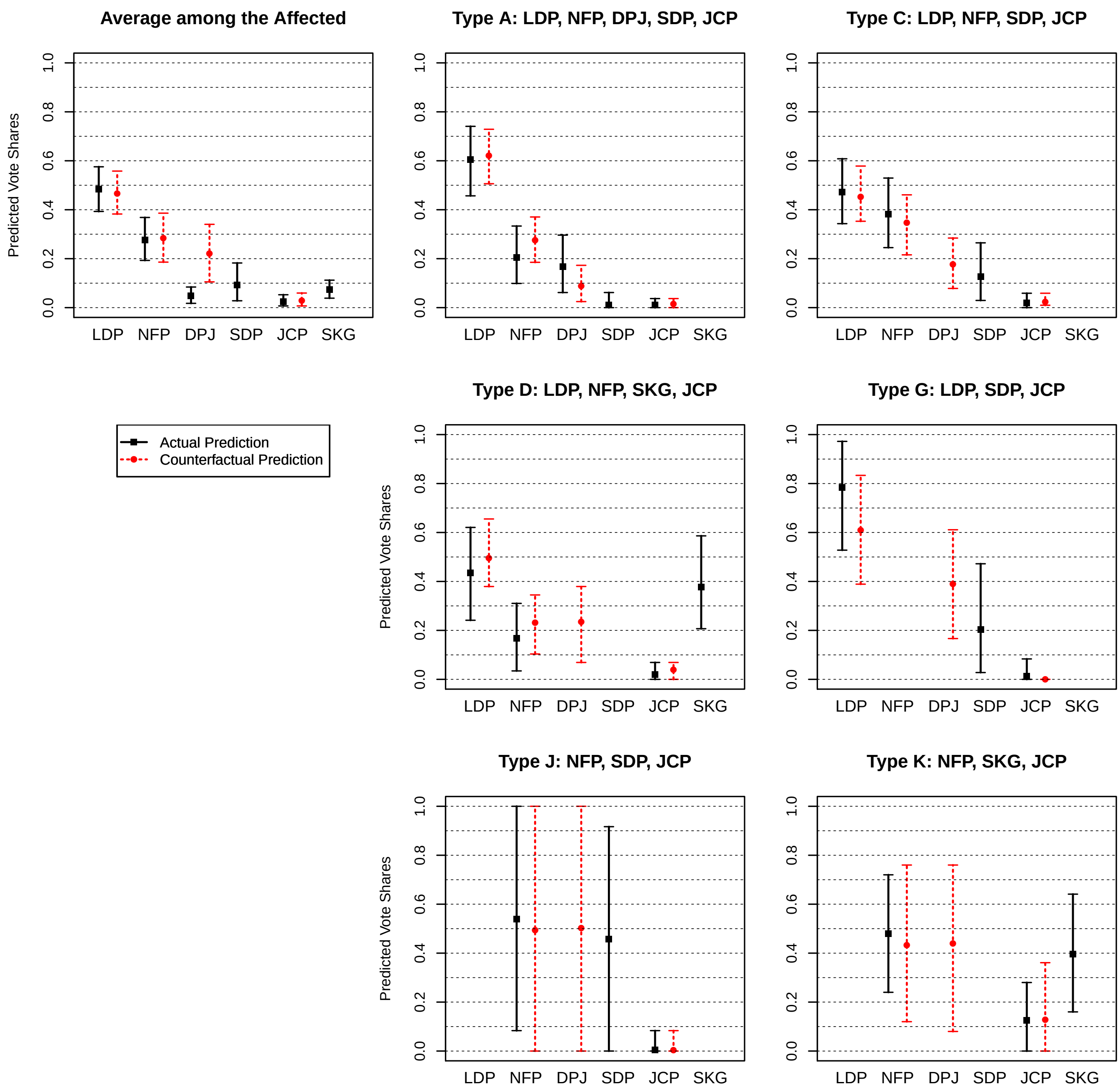

Figure 4: What If SDP and SKG Had Been Merged into DPJ? In each panel, predicted vote shares under the actual choice set (black squares) and the hypothetical choice set that would have realized had SDP and SKG both become merged to DPJ (red circles) are shown for each party, along with their 95% confidence intervals. The top left panel presents the average predicted vote shares for all voters whose choice sets would have affected by the hypothetical full merger, while the other six panels show the results for each of the relevant choice set types.

mately equal to the sum of the actual SDP and SKG vote shares.

However, it would be hasty to conclude that the SDP and SKG supporters would simply have migrated to DPJ. The other six panels in Figure 4 suggest that there is substantial difference between the SDP supporters and SKG supporters. For all choice set types involving an SDP candidate (A, C, G and J), DPJ would have gained at least as many votes as implied by the actual SDP vote shares. On the contrary, there is evidence that the actual SKG supporters would not have simply migrated to DPJ and instead a significant fraction of them are likely to have voted for LDP or NFP rather than DPJ (D). One possible explanation for this difference is that SKG candidates are on average ideologically closer to LDP and NFP candidates than DPJ candidates on many policy dimensions.

The above result must be interpreted with caution, since it rests on the strong assumption that the choice set types are conditionally exogenous after controlling for the covariates in the model. The violation of this assumption is a serious concern in this context because parties may have chosen districts in which they ran their own candidates strategically based on some information we do not observe. With this caveat in mind, it is still reasonable to conclude that VCL produces interesting insights on how changes in choice set types might affect choice behavior.

7 Conclusion

Multinomial response models are widely used across scientific disciplines for the purpose of modeling choice behavior involving more than two alternatives. Standard models, however, typically neglect the fact that individual choices must often be made between a subset of the alternatives that are theoretically available. This is especially problematic if the IIA assumption is violated, because the relative probability of choosing an alternative will be different depending on which particular choice set the alternative is being selected from. Such a situation happens when an alternative represents a substitute of another alternative for some individuals in the analysis.

In this paper, I develop a new multinomial response model which explicitly takes into account the variation in choice sets actually observed in data. The proposed VCL model relaxes the IIA assumption by allowing parameters in the individual random utility function to vary across choice set types, thereby generalizing MNL in the way specifically tailored for the analysis of choice set dependence. To show the advantages of the proposed model, I apply it to a partially contested multiparty election in which parties did not run their candidates in every district and voters thus had to choose only from subsets of the nationally contesting parties. Through the empirical analysis of this Japanese general election, I show that VCL is useful for analyzing the extent to which the predictors included in the model had heterogeneous effects on voting behavior across choice set types. The analysis also indicates that the counterfactual behavior of the supporters of the two junior coalition member parties would have been different depending on choice sets.

Several improvements and extensions are left for future research. First, alternative specifications of the random coefficients, including nonparametric distributions, may be desirable in some situations where the normality assumption in equation (2) is insufficient. Second, the conditional exogeneity of choice set types is a strong assumption and relaxing this assumption (by, for example, explicitly modeling the selection process) will be a key extension. Finally, the current model can be naturally extended to a hierarchical model incorporating characteristics of choice set types. These extensions are especially important for the analysis of multiparty elections, as choice sets themselves are often strategically determined by parties and multi-level data sources are becoming increasingly available. The existence of these diverse future possibilities makes the study of varying choice set models an exciting field for political methodologists.

References

Allenby, Greg M. and Peter E. Rossi. 1998. "Marketing Models of Consumer Heterogeneity." *Journal of Econometrics* 89(1–2):57–78.

Allenby, Greg M. and Peter J. Lenk. 1994. "Modeling Household Purchase Behavior with Logistic Normal Regression." *Journal of the American Statistical Association* 89(428):1218–1231.

Alvarez, R. Michael, Frederick J. Boehmke and Jonathan Nagler. 2006. "Strategic Voting in British Elections." *Electoral Studies* 25:1–19.

Alvarez, R. Michael and Jonathan Nagler. 1995. "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39(3):714–744.

Alvarez, R. Michael and Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42(1):55–96.

Alvarez, R. Michael and Jonathan Nagler. 2000. "A New Approach for Modeling Strategic Voting in Multiparty Elections." *British Journal of Political Science* 30:57–75.

Alvarez, R. Michael, Jonathan Nagler and Shaun Bowler. 2000. "Issues, Economics and the Dynamics of Multi-party Elections: the British 1987 General Election." *American Political Science Review* 94:131–149.

Bhat, Chandra R. 1998. "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling." *Transportation Research A: Policy and Practice* 32(7):495–507.

Bogaards, Matthijs. 2003. "Electoral Choices for Divided Societies: Multi-Ethnic Parties and Constituency Pooling in Africa." *Commonwealth & Comparative Politics* 41(3):59–80.

Branton, Regina P. and Bradford S. Jones. 2005. "Reexamining Racial Attitudes: The Conditional Relationship Between Diversity and Socioeconomic Environment." *American Journal of Political Science* 49(2):359–372.

Cain, Bruce, John Ferejohn and Morris Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*. Cambridge, MA: Harvard University Press.

Carey, John M. and Matthew S. Shugart. 1995. "Incentives to Cultivate a Personal Vote: A Rank Ordering of Electoral Formulas." *Electoral Studies* 14:417–440.

Christensen, Ray. 1998. "Putting New Wine into Old Bottles: The Effect of Electoral Reforms on Campaign Practices in Japan." *Asian Survey* 38:986–1004.

Christensen, Ray. 2000. *Ending the LDP Hegemony: Party Cooperation in Japan*. Honolulu, HI: University of Hawaii Press.

Claassen, Ryan L. 2009. "Direction Versus Proximity: Amassing Experimental Evidence." *American Politics Research* 37(2):227–253.

Cox, Gary. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge University Press.

Cox, Gary W. and Jonathan N. Katz. 1996. "Why Did the Incumbency Advantage in US House Elections Grow?" *American Journal of Political Science* 40(2):478–497.

Cox, Karen E. and Leonard J. Schoppa. 2002. "Interaction Effects in Mixed-Member Electoral Systems: Theory and Evidence from Germany, Japan, and Italy." *Comparative Political Studies* 35(9):1027–1053.

Dempster, A.P., N.M. Laird and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B* 39(1):1–38.

Duverger, Maurice. 1954. *Political Parties: Their Organization and Activity in the Modern State*. London: Methuen.

Erikson, Robert S. 1971. "The Advantage of Incumbency in Congressional Elections." *Polity* 3(3):395–405.

Gelman, Andrew, Gareth O. Roberts and W.R. Gilks. 1996. "Efficient metropolis jumping hules." *Bayesian statistics* 5:599–608.

Gelman, Andrew and Jeniffer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC.

Gerber, Alan. 1998. "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables." *American Political Science Review* 92(2):401–411.

Gerber, Alan S. 2004. "Does Campaign Spending Work?: Field Experiments Provide Evidence and Suggest New Theory." *American Behavioral Scientist* 47(5):541–574.

Glasgow, Garrett. 2001. "Mixed Logit Models for Multiparty Elections." *Political Analysis* 9(2):116.

Guadagni, Peter M. and John D.C. Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science* 2(3):203–238.

Hausman, Jerry A. and David A. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46(2):403–426.

Hillygus, D. Sunshine and Simon Jackman. 2003. "Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy." *American Journal of Political Science* 47(4):583–596.

Hirano, Shigeo. 2006. "Electoral Institutions, Hometowns, and Favored Minorities: Evidence from Japan's Electoral Reforms." *World Politics* 59(1):51–82.

Honaker, James, Jonathan N. Katz and Gary King. 2002. "An Fast, Easy, and Efficient Estimator for Multiparty Electoral Data." *Political Analysis* 10(1):84–100.

Ishikawa, Masumi. 2004. *Sengo Seiji-shi, Shin-ban [Postwar Political History, New Edition]*. Tokyo Japan: Iwanami Shoten.

Iversen, Torben. 1994. "Political Leadership and Representation in West European Democracies: A Test of Three Models of Voting." *American Journal of Political Science* 38(1):45–74.

Jackson, John E. 2002. "A Seemingly Unrelated Regression Model for Analyzing Multiparty Elections." *Political Analysis* 10(1):49–65.

Jacobson, Gary C. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34(2):334–362.

Katz, Jonathan N. and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review* 93(1):15–32.

Kedar, Orit. 2005. "When Moderate Voters Prefer Extreme Parties: Policy Balancingin Parliamentary Elections." *American Political Science Review* 99(2):185–199.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44:341–355.

Krauss, Eliss S. and Robert Pekkanen. 2004. "Explaining Party Adaptation to Electoral Reform: The Discreet Charm of the LDP?" *The Journal of Japanese Studies* 30(1):1–34.

Laitin, David D. 1994. "The Russian-Speaking Nationality in Estonia: Two Quasi-Constitutional Elections." *East European Constitutional Review* 3:23–27.

Levitt, Steven D. 1994. "Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the US House." *Journal of Political Economy* pp. 777–798.

Lewis, Jeffrey B. and Gary King. 1999. "No Evidence on Directional vs. Proximity Voting." *Political Analysis* 8(1):21–33.

Martin, Andrew D., Kevin M. Quinn and Jong Hee Park. 2011. "MCMCpack: Markov Chain Monte Carlo in R." *Journal of Statistical Software* 42(9):22.
URL: *http://www.jstatsoft.org/v42/i09/*

Martin, Lanny W. and Randolph T. Stevenson. 2001. "Government Formation in Parliamentary Democracies." *American Journal of Political Science* 45(1):33–50.

McFadden, Daniel. 1973. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka. New York: Academic Press pp. 105–142.

McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3:303–328.

McKean, Margaret and Ethan Scheiner. 2000. "Japan's New Electoral System: la plus ça change...." *Electoral Studies* 19(4):447–477.

Rahat, Gideon and Mario Sznajder. 1998. "Electoral Engineering in Chile: The Electoral System and Limited Democracy." *Electoral Studies* 17(4):429–442.

Reed, Steven R. 1994. "Democracy and the Personal Vote: A Cautionary Tale from Japan." *Electoral Studies* 13(1):17–28.

Reed, Steven R. 2003. *Japanese Electoral Politics: Creating a New Party System*. New York, NY: Routledge.

Roberts, Gareth O. and Jeffrey S. Rosenthal. 2009. "Examples of Adaptive MCMC." *Journal of Computational and Graphical Statistics* 18(2):349–367.

Schofield, Normal, Andrew D. Martin, Kevin M. Quinn and Andrew B. Whitford. 1998. "Multiparty Electoral Competition in the Netherlands and Germany: A Model Based on Multinomial Probit." *Public Choice* 97(3):257–293.

Schrijver, Frans J. 2004. "Electoral Performance of Regionalist Parties and Perspectives on Regional Identity in France." *Regional & Federal Studies* 14(2):187–210.

Steel, Gill. 2003. "Gender and Voting Preferences in Japanese Lower House Elections." *Japanese Journal of Political Science* 4(1):1–39.

Stokes, Donald E. and Warren E. Miller. 1962. "Party Government and the Saliency of Congress." *Public Opinion Quarterly* 26(4):531–546.

Tomz, Michael, Joshua A. Tucker and Jason Wittenberg. 2002. "An Easy and Accurate Regression Model for Multiparty Electoral Data." *Political Analysis* 10(1):66–83.

Train, Kenneth E. 1998. "Recreation Demand Models with Taste Differences over People." *Land economics* 74(2):230–239.

Train, Kenneth E. 2008. "EM Algorithms for Nonparametric Estimation of Mixing Distributions." *Journal of Choice Modelling* 1(1):40–69.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Second ed. New York, NY: Cambridge University Press.

Train, Kenneth E., Daniel L. McFadden and Moshe Ben-Akiva. 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *The Rand Journal of Economics* 18(1):109–123.

Wei, Greg C.G. and Martin A. Tanner. 1990. "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms." *Journal of the American Statistical Association* 85(411):699–704.

Whitten, Guy D. and Harvey D. Palmer. 1996. "Heightening Comparativists' Concern for Model Choice: Voting Behavior in Great Britain and the Netherlands." *American Journal of Political Science* 40(1):231–260.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT press.

Supporting Information

A Details of the VCL Notational Framework

The exposition of the VCL model in Section 4.1 is made simple by the assumption that the model only contains choice-varying covariates. Here, I provide a more general representation of the VCL model.

First, suppose that the predictors of the model are composed of K individual characteristics and L choice-varying covariates. Because I estimate separate coefficients on individual characteristics for each alternative j, the fixed-effect coefficients for the K individual characteristics are actually represented by a vector of length $J \times K$, i.e., $\alpha_w^\top = [\alpha_w^{1\top} \ldots \alpha_w^{J\top}]$ where $\alpha_w^{j\top} = [\alpha_w^{j1} \ldots \alpha_w^{jK}]$ for every $j \in \{1, ..., J\}$. On the other hand, the vector of coefficients on the choice-varying covariates are of the same length as the number of the covariates L, such that $\alpha_v^\top = [\alpha_v^1 \ldots \alpha_v^L]$. Combining these with the vector of choice-specific intercepts $\alpha_0^\top \equiv [\alpha_0^1 \ldots \alpha_0^J]$, the vector of the fixed-effect effect coefficients has $J \times (K + 1) + L$ elements and can generally be defined as $\alpha^\top = [\alpha_0^\top \; \alpha_w^\top \; \alpha_v^\top]$.

Next, consider the linear latent utility function for individual i who belongs to the choice set type m. Because this individual can choose only from a subset S_m of the entire set of J theoretically available alternatives, only a subvector α_m of the coefficient vector α enters into this individual's utility function. This subvector can be written as $\alpha_m = [\alpha_{0m}^\top \; \alpha_{wm}^\top \; \alpha_v^\top]$, where α_{0m} and α_{wm} contain the elements of α_0 and α_w which correspond to the alternatives in S_m only and thus have the length of J_m and $J_m \times K$, respectively. This implies that α_m is now of length $J_m \times (K + 1) + L$. The vector of the random-effect coefficients β_m for this individual can be defined similarly, except that it only contains the coefficients corresponding to the predictors included in z_{ij}. Denoting the numbers of these predictors K^* and L^* for the individual characteristics and choice-varying covariates, respectively, β_m can be written as $\beta_m^\top = [\beta_{0m}^\top \; \beta_{wm}^\top \; \beta_v^\top]$ and is of length $J_m \times (K^* + 1) + L^*$. Using these, the latent utility of choosing alternative j for individual i who belongs to choice set type m can be written as

$$y_{ijm}^* = x_{ij}^\top \alpha_m + z_{ij}^\top \beta_m + \varepsilon_{ij},$$

where $x_{ij}^\top = [0 \ldots 1 \ldots 0 \; \mathbf{0}^\top \ldots w_i^\top \ldots \mathbf{0}^\top \; v_{ij}]$ so that only the elements of α_m corresponding to alternative

j will enter the utility function. The other predictor vector z_{ij} is defined likewise. This represents the general version of equation (1) in Section 4.1.

The specification of the distribution of β_m must also be modified to reflect the fact that the length of β_m now varies across m. First, define the full random-effect coefficients vector β in the same manner as α, i.e., $\beta^\top = [\beta_0^\top \ \beta_w^\top \ \beta_v^\top]$ where $\beta_0^\top = [\beta_0^1 \ ... \ \beta_0^J]$ and $\beta_w^\top = [\beta_w^{1\top} \ ... \ \beta_w^{J\top}] = [\beta_w^{11} \ ... \ \beta_w^{jk} \ ... \ \beta_w^{JK}]$. Then, VCL assumes that β is independently and identically distributed multivariate normal with mean zero and variance $\mathbf{\Sigma}$ even after conditioning on the predictors, i.e., $\beta \mid X, Z \ \sim \ \mathcal{N}_{Q^*}(\mathbf{0}, \mathbf{\Sigma})$ where Q^* is equal to the length of β, $J \times (K^* + 1) + L^*$. Because a marginal distribution of a multivariate normal random variable is also multivariate normal with the mean and variance simply equal to the corresponding elements of the original mean vector and variance matrix, the distribution of β_m can be written as,

$$\beta_m \mid X, Z \ \sim \ \mathcal{N}_{Q_m^*}(\mathbf{0}, \mathbf{\Sigma}_m),$$

where $Q_m^* = J_m \times (K^* + 1) + L^*$ and $\mathbf{\Sigma}_m$ is the $Q_m^* \times Q_m^*$ submatrix of $\mathbf{\Sigma}$ which is composed of the rows and columns corresponding to S_m. This expression is the general version of equation (2).

Finally, the vector of latent utilities for individual i can be written in a stacked form as

$$y_{im}^* \ = \ X_{im}\alpha_m + Z_{im}\beta_m + \varepsilon_{im},$$

where

$$X_{im} \ = \ \left[\begin{array}{ccc|ccc|c} 1 & & \mathbf{0} & w_i^\top & & \mathbf{0} & v_{i1}^\top \\ & \ddots & & & \ddots & & \vdots \\ \mathbf{0} & & 1 & \mathbf{0} & & w_i^\top & v_{ij_m}^\top \end{array}\right]$$

$y_{im}^* = [y_{i1m}^* \ ... \ y_{iJ_m m}^*]$, $\varepsilon_{im}^* = [\varepsilon_{i1}^* \ ... \ \varepsilon_{iJ_m}^*]$, and Z_{im} is defined similarly to X_{im}. This can be further stacked with respect to individuals for each choice set type m and written in the following simple expression,

$$y_m^* \;=\; X_m\alpha_m + Z_m\beta_m + \varepsilon_m,$$

where $y_m^{*\top} = [y_{1m}^{*\top} \,...\, y_{N_m m}^{*\top}]$, $X_m^\top = [X_{1m}^\top \,...\, X_{N_m m}^\top]$, $Z_m^\top = [Z_{1m}^\top \,...\, Z_{N_m m}^\top]$, and $\varepsilon_m^\top = [\varepsilon_{1m}^\top ... \varepsilon_{N_m m}^\top]$. These simplified forms are convenient for implementation in computer software.

B Comparison with Other Multinomial Response Models

In this section, I compare the proposed VCL model to the three multinomial response models commonly used in the literature (MNL, MXL and MNP). A condensed version of this discussion appears in the main paper as Section 4.2.

MNL. MNL is the simplest of the four alternative multinomial response models considered in this paper. Compared to VCL, the "textbook" version of MNL can be characterized by two simplifying assumptions. First, the systematic component of the individual utility function is invariant across observations after the values of the covariates are taken into account. This implies $\boldsymbol{\Sigma} = 0$ so that the random coefficients have a degenerate distribution which has all its probability mass on $\beta_m = 0$. Second, choice sets are also identical for every observation, so that everyone can choose any of the J alternatives. This can be written as $R_i = \{1, ..., J\}$ for all i.

Under these two assumptions, the individual utility function (equation 1) now becomes

$$y_{ijm}^* \;=\; x_{ij}^\top\alpha + \varepsilon_{ij}, \tag{9}$$

where $j \in \{1, ...J\}$, and the textbook MNL model can be defined by the following choice probability,

$$\begin{aligned}\mathcal{P}_{ij} &= \Pr(y_{ijm} = 1 \mid X, Z, \alpha, \boldsymbol{\Sigma}) \\ &= \frac{\exp(x_{ij}^\top\alpha)}{\sum_{k=1}^J \exp(x_{ik}^\top\alpha)}. \end{aligned}\tag{10}$$

This expression can be found in many standard references (e.g. Wooldridge, 2002, Ch.15). This basic model, however, cannot be used if the data contain any observation for which some alternatives are unavailable, because the summation in the denominator is undefined for such observations. Ignoring this problem and applying a preprogrammed function in standard statistical software (e.g., `mlogit` in Stata) will result in either dropped observations or biased estimates.[7]

However, the assumption of invariant choice sets can be relaxed even within the framework of MNL. In fact, the original MNL (or conditional logit) model proposed by McFadden (1973) allows for varying choice sets and is defined by the following choice probability,

$$\mathcal{P}_{ij} = \frac{\exp(x_{ij}^{\top}\alpha)}{\sum_{k \in S_m} \exp(x_{ik}^{\top}\alpha)}, \tag{11}$$

which differs from equation (10) only in the range of the summation operator. It is, however, important to keep in mind that this model still relies on the IIA assumption. To see this, note that as can be seen from equation (9) the utility individual i obtains from choosing alternative j is distributed independently of her utility about any other alternative. In fact, the relative choice probability between any two alternatives stays constant regardless of choice set types, i.e., $OR(j, k, m, m') = 1$ for all $m, m' \in \{1, ..., M\}$ and $j, k \in S_m \times S_{m'}$ under the choice model defined by equations (9) and (11).

In sum, both the textbook and original versions of MNL are special cases of VCL and rely on the strong IIA assumption. This implies that, while both MNL and VCL will consistently estimate choice probabilities and lead to substantively identical conclusions when the IIA assumption holds, the former will produce biased predictions when IIA is violated.

MXL. VCL can be seen as a special case of MXL. To see this, suppose that each observation, instead of each choice set group, has its own unique coefficient on z_{ij}. Then, individual utility function can be written as

$$y_{ijm}^{*} = x_{ij}^{\top}\alpha + z_{ij}^{\top}\gamma_i + \varepsilon_{ij}, \tag{12}$$

[7] The latter case would occur if unavailable alternatives were treated as if they had been available but not chosen by those individuals. Analysts must be careful so that their software is not automatically doing this.

and the choice probability is

$$\mathcal{P}_{ij} \;=\; \int \frac{\exp(x_{ij}^{\top}\alpha + z_{ij}^{\top}\gamma_i)}{\sum_{k\in S_m}\exp(x_{ik}^{\top}\alpha + z_{ik}^{\top}\gamma_i)} f_Q(\gamma_i \mid \theta) d\gamma_i, \tag{13}$$

where $f_Q(\cdot \mid \theta)$ is the Q-dimensional density function which is fully specified by parameters θ. This is the mixed logit model (see e.g., Train, 2009, Ch.6),[8] and further assumptions are typically made about the form of $f_Q(\cdot \mid \theta)$ (for example, $f_Q(\cdot \mid \theta) = \phi_Q(\cdot \mid \mathbf{\Sigma})$). Now, suppose that one is willing to make an additional assumption that choice probabilities are equal as long as individuals choose from an identical choice set, i.e., $\gamma_i = \gamma_{i'}$ if $R_i = R_{i'}$. This means that there are as many unique values of γ_i as the number of choice set groups, which allows us to write $\beta_m = \gamma_i$ if $S_m = R_i$, $m = 1, ...M$. Equation (13) then becomes identical to equation (4) by further assuming these Q group-level random coefficients are jointly normally distributed with covariance $\mathbf{\Sigma}$.

MXL is therefore a more general model than VCL. Does that mean that we should always use MXL instead of VCL? The answer is no for several reasons. First, the additional assumption made for VCL is inconsequential for the purpose of analyzing how choice probabilities may depend on the choice sets that are actually available. This is because VCL still produces a unique estimate of the relative choice probability, $\mathcal{R}_{ijm}/\mathcal{R}_{ikm}$, for each choice set type m. Second, if the additional assumption is correct, VCL is more efficient than MXL and will produce more accurate estimates of choice set dependence. Typically, when MXL is used, researchers do not attempt to estimate individual values of γ_i because there is not enough information in data.[9] They instead only report the estimate of θ. In contrast, β_m in VCL can be more precisely estimated because the number of choice set groups (M) is usually much smaller than sample size (N) and each group often contains enough observations to produce useful estimates of β_m. As I illustrate in Section 6 using the Japanese election data, these estimates can then be directly used to examine how the effects of the covariates vary depending on choice set types.

[8]Note, however, that typical representations of MXL assume the choice sets to be invariant like the textbook version of MNL. In that case, the summation in the denominator of equation (13) becomes over $k = 1, ..., J$.

[9]An exception is when the data contain repeated observations for each i. In such a case, there may be sufficient information within individual to obtain a meaningful estimate of γ_i.

MNP. MNP is an alternative multinomial model which does not assume IIA. MNP is based on the same utility function as MNL (i.e., equation 9) except that the error term ε_{ij} is allowed to be correlated across alternatives within each observation and that the J_m-dimensional vector of individual error terms has the multivariate normal distribution, i.e., $\varepsilon_{im} \sim \mathcal{N}_{J_m}(\mathbf{0}, \mathbf{\Omega})$ where $\mathbf{\Omega}$ is a J_m-dimensional symmetric positive-definite matrix representing the covariances.[10] Unlike MNL, MNP does not rest on the IIA assumption and thus, for each individual, relative choice probability for any pair of alternatives depends on the utilities of all the alternatives in her choice set (unless $\mathbf{\Omega}$ is assumed to be the identity matrix; see Hausman and Wise, 1978). This means that the predicted choice probabilities estimated by MNP may accurately reflect their dependence on choice sets and the estimates of α will not be biased because of IIA violation.

This may appear to imply that MNP can be used in place of VCL even when there is variation in choice sets in the actual data. However, MNP has a major limitation when the pattern of choice set dependence itself is of interest. As Alvarez and Nagler (1998) correctly point out, analyses using MNP "only relax the IIA assumption through the specification of the stochastic (random) component of the model" (p.85). That is, MNP treats the violation of IIA as nuisance even when it is theoretically interesting. The drawback of this in particular is that the effects of covariates are assumed to be fixed across individuals, because the systematic component of the utility function is invariant by assumption. In contrast, VCL allows the coefficients to vary across choice sets because of the inclusion of the random effects (β_m) and thus can be used to analyze how the effects of covariates may differ depending on which alternatives are actually available.

C Description of the Proposed Estimation Procedures

C.1 Monte Carlo Expectation Maximization Algorithm

The following procedure is based on the MCEM algorithm originally developed for MXL by Train (2008). The key idea here is to view the random coefficients as the auxiliary information for augmenta-

[10] Again, most treatments of MNP additionally assume the choice set to be invariant across observations. This can be easily relaxed by setting the interval of integration from negative infinity to positive infinity (instead of greater than the utility for the chosen alternative) for the missing alternatives in the expression for choice probability.

tion, or "missing data." That is, the complete-data likelihood at iteration t is constructed by augmenting the observed-data likelihood by the vector of random coefficients, η_m, which by assumption follows $\mathcal{N}_Q(\alpha^t, \mathbf{\Sigma}^t)$. Thus, for each observation i, the expectation of the augmented log likelihood with respect to η_m conditional on the current parameter values is

$$\begin{aligned}\mathcal{Q}_i(\alpha, \mathbf{\Sigma} \mid \alpha^t, \mathbf{\Sigma}^t, X, y) &= \int \log f(y_{im}, \eta_m \mid \alpha, \mathbf{\Sigma}, X_{im}) dF(\eta_m \mid \alpha^t, \mathbf{\Sigma}^t, y_{im}) \\ &= \int \left\{ \log f(y_{im} \mid \eta_m, X_{im}) + \log \phi_Q(\eta_m \mid \alpha, \mathbf{\Sigma}) \right\} dF(\eta_m \mid \alpha^t, \mathbf{\Sigma}^t, y_{im}).\end{aligned}$$

However, because the distribution of y_{im} does not directly depend on either α or $\mathbf{\Sigma}$, maximizing this function with respect to $(\alpha, \mathbf{\Sigma})$ is equivalent to maximizing the following simplified function,

$$\begin{aligned}\mathcal{Q}_i^*(\alpha, \mathbf{\Sigma} \mid \alpha^t, \mathbf{\Sigma}^t, X, y) &= \int \log \phi_Q(\eta_m \mid \alpha, \mathbf{\Sigma}) dF(\eta_m \mid \alpha^t, \mathbf{\Sigma}^t, y_{im}) \\ &= \int \log \phi_Q(\eta_m \mid \alpha, \mathbf{\Sigma}) \frac{f(y_{im} \mid \eta_m, \alpha^t, \mathbf{\Sigma}^t) f(\eta_m \mid \alpha^t, \mathbf{\Sigma}^t)}{f(y_{im} \mid \alpha^t, \mathbf{\Sigma}^t)} d\eta_m \\ &= \int \frac{\left\{ \prod_{j \in S_m} L_{ij}(\eta_m)^{y_{ijm}} \right\} \log \phi_Q(\eta_m \mid \alpha, \mathbf{\Sigma})}{\int \left\{ \prod_{j \in S_m} L_{ij}(\eta'_m)^{y_{ijm}} \right\} dF(\eta'_m \mid \alpha^t, \mathbf{\Sigma}^t)} dF(\eta_m \mid \alpha^t, \mathbf{\Sigma}^t), \quad (14)\end{aligned}$$

where $L_{ij}(\eta_m) = \exp(x_{ij}^\top \eta_m) \big/ \sum_{k \in S_m} \exp(x_{ik}^\top \eta_m)$. Calculating this expectation exactly would be computationally difficult because of the Q-dimensional integrals which do not have a closed form. However, we can instead use the following *simulated* expectation,

$$\check{\mathcal{Q}}_i^*(\alpha, \mathbf{\Sigma} \mid \alpha^t, \mathbf{\Sigma}^t, X, y) = \frac{1}{D} \sum_{d=1}^{D} \frac{\left\{ \prod_{j \in S_m} L_{ij}(\eta_m^d)^{y_{ijm}} \right\} \log \phi_Q(\eta_m^d \mid \alpha, \mathbf{\Sigma})}{\frac{1}{D} \sum_{d'=1}^{D} \left\{ \prod_{j \in S_m} L_{ij}(\eta_m^{d'})^{y_{ijm}} \right\}}, \quad (15)$$

where η_m^d is the dth of D Monte Carlo draws from $\mathcal{N}_Q(\alpha^t, \mathbf{\Sigma}^t)$. Note that these draws can be obtained at each iteration using the same draws from the standard multivariate normal, $\mathcal{N}_Q(\mathbf{0}, \boldsymbol{I}_Q)$, and transforming them with the current values of the mean and variance parameters, $(\alpha^t, \mathbf{\Sigma}^t)$.

The proposed algorithm then proceeds by finding the values of $(\alpha, \mathbf{\Sigma})$ that maximize the sum of equation (15) over N observations, i.e., $\check{\mathcal{Q}}^* = \sum_{i=1}^{N} \check{\mathcal{Q}}_i^*$, and setting these values as $(\alpha^{t+1}, \mathbf{\Sigma}^{t+1})$. How can we find this maximum? As can be seen from the form of equation (15), $\check{\mathcal{Q}}^*$ is identical to the log likelihood of a weighted random sample of size ND from $\mathcal{N}(\alpha, \mathbf{\Sigma})$ with the weight for each observation equal to $w_{id}^t \equiv \left\{\prod_{j\in S_m} L_{ij}(\eta_m^d)^{y_{ijm}}\right\} \Big/ \left(\frac{1}{D}\right) \sum_{d'=1}^{D} \left\{\prod_{j\in S_m} L_{ij}(\eta_m^{d'})^{y_{ijm}}\right\}$. Thus, $\check{\mathcal{Q}}^*$ takes its maximum value when

$$\alpha^{t+1} = \frac{1}{ND}\sum_{m=1}^{M}\sum_{d=1}^{D} w_{id}^t N_m \eta_m^d, \tag{16}$$

$$\mathbf{\Sigma}^{t+1} = \frac{1}{ND}\sum_{m=1}^{M}\sum_{d=1}^{D} w_{id}^t N_m^2 (\eta_m^d - \alpha^{t+1})(\eta_m^d - \alpha^{t+1})^\top. \tag{17}$$

The MCEM procedure consists of repeating the above steps until the updated values of $(\alpha^t, \mathbf{\Sigma}^t)$ converge. The resulting values of these parameters, $(\hat{\alpha}, \hat{\mathbf{\Sigma}})$, are consistent estimates of the true parameter values. Note that D must be sufficiently large compared to sample size N because it can be shown that the estimator has a limiting distribution only when D goes to infinity faster than $\sqrt{N}$ (Train, 2009).

Finally, once the algorithm converges the estimate of the random effects can be obtained as their posterior mean evaluated at $(\hat{\alpha}, \hat{\mathbf{\Sigma}})$. That is,

$$\hat{\eta}_m = \int \eta_m f(\eta_m \mid y_m, \hat{\alpha}, \hat{\mathbf{\Sigma}}) d\eta_m, \tag{18}$$

where y_m is the stacked vector of choice indicators for all observations with choice set m. Applying Bayes' rule, the posterior distribution can be expressed as,

$$f(\eta_m \mid y_m, \hat{\alpha}, \hat{\mathbf{\Sigma}}) = \frac{\left\{\prod_{i=1}^{N_m}\prod_{j\in S_m} L_{ij}(\eta_m)^{y_{ijm}}\right\} \phi_Q(\eta_m \mid \hat{\alpha}, \hat{\mathbf{\Sigma}})}{\int \left\{\prod_{i=1}^{N_m}\prod_{j\in S_m} L_{ij}(\eta'_m)^{y_{ijm}}\right\} \phi_Q(\eta'_m \mid \hat{\alpha}, \hat{\mathbf{\Sigma}}) d\eta'}, \tag{19}$$

which can be simulated using the draws in the last iteration of the algorithm. The estimate of η_m in equation (18) can thus be obtained via simulation and written as

$$\hat{\eta}^* = \frac{1}{D}\sum_{d=1}^{D} \frac{\eta_m^d \left\{\prod_{i=1}^{N_m}\prod_{j\in S_m} L_{ij}(\eta_m^d)^{y_{ijm}}\right\}}{\frac{1}{D}\sum_{d'=1}^{D}\left\{\prod_{i=1}^{N_m}\prod_{j\in S_m} L_{ij}(\eta_m^{d'})^{y_{ijm}}\right\}}, \tag{20}$$

where η_m^d here is the dth draw in the last iteration.

C.2 Bayesian Markov Chain Monte Carlo

The second procedure differs from the MCEM algorithm in that it is a Bayesian approach and all parameters $(\alpha, \mathbf{\Sigma}, \eta)$ are regarded as random quantities. The basic idea here is that the evaluation of high dimensional integrals, which constitute a major problem for likelihood-based procedures, can be avoided in the Bayesian analysis by successively drawing from the conditional distribution of each parameter given all other parameters and data. MCMC has been widely used for the Bayesian inference of multinomial choice models. For example, Allenby and Lenk (1994) applied an MCMC algorithm for a panel logistic-normal regression model; Allenby and Rossi (1998) used a similar procedure for the random coefficients MNP. Here, I develop a simple algorithm for VCL.

First, using the conditional independences implied by the model, the joint distribution of the parameters and responses can be decomposed as follows,

$$f(Y, \alpha, \eta, \mathbf{\Sigma} \mid X) \quad = \quad f(Y \mid X, \eta) f(\eta \mid \alpha, \mathbf{\Sigma}) f(\alpha) f(\mathbf{\Sigma}). \tag{21}$$

Based on this relationship, the proposed MCMC procedure can now be derived. Denote the starting values of the parameters at iteration t by $(\alpha^t, \mathbf{\Sigma}^t, \eta^t)$. Then, the stationary distribution generated by the following three-step adaptive Metropolis-within-Gibbs sampling algorithm is the target joint distribution.

1. Generate η^{t+1} conditional on $(\alpha^t, \mathbf{\Sigma}^t)$. The conditional distribution of η can be expressed as,

$$\begin{aligned} f(\eta \mid Y, \alpha^t, \mathbf{\Sigma}^t, X) &\propto \prod_{m=1}^{M} f(\eta_m \mid \alpha^t, \mathbf{\Sigma}^t) f(Y_m \mid X, \eta_m) \\ &\propto \prod_{m=1}^{M} \exp\left(-\frac{1}{2}(\eta_m - \alpha^t)^\top \left(\mathbf{\Sigma}^t\right)^{-1} (\eta_m - \alpha^t)\right) \prod_{i=1}^{N_m} \frac{\sum_{k \in S_m} y_{ikm} \exp(x_{ik}^\top \eta_m)}{\sum_{k \in S_m} \exp(x_{ik}^\top \eta_m)}, \end{aligned}$$

where the first line follows from equation (21) and the second line holds because $\eta_m \sim \mathcal{N}_Q(\alpha, \mathbf{\Sigma})$, $f(Y_m \mid X, \eta_m) = \prod_{i=1}^{N_m} \prod_{j \in S_m} \Pr(y_{ijm} = 1 \mid X, \eta_m)$, and equation (4). Since this posterior does not correspond to any common distribution, an alternative sampling technique, such as the Metropolis-

Hastings algorithm, must be employed. I use the adaptive Metropolis sampler, as proposed by Roberts and Rosenthal (2009), where I adjust the variance of the jumping distribution at every 50th iteration so that the acceptance rate becomes closer to the theoretically optimal value of 0.23 (Gelman, Roberts and Gilks, 1996).

2. Generate α^{t+1} conditional on $(\boldsymbol{\Sigma}^t, \eta^{t+1})$. The next step is to generate draws of the fixed effects, α^{t+1}. The conditional distribution is

$$f(\alpha \mid Y, \eta^{t+1}, \boldsymbol{\Sigma}^t, X) \quad \propto \quad f(\alpha) f(\eta^{t+1} \mid \alpha, \Sigma^t), \tag{22}$$

which follows from equation (21). Since the distribution of η_m given α and $\boldsymbol{\Sigma}$ is Q-dimensional multivariate normal, we can use either a conjugate normal prior or an improper flat prior for α so that the posterior distribution of α is also normal. In the application, I use a flat prior $f(\alpha) \propto 1$ and draw from the resulting multivariate normal distribution, which is $\mathcal{N}(\sum_{m=1}^{M} \eta_m^{t+1}/M, \boldsymbol{\Sigma}^t/M)$.

3. Generate $\boldsymbol{\Sigma}^{t+1}$ conditional on $(\alpha^{t+1}, \eta^{t+1})$. Finally we generate draws of the covariance matrix of the mixing distribution. The conditional posterior is

$$f(\boldsymbol{\Sigma} \mid Y, \alpha^{t+1}, \eta^{t+1}, X) \quad \propto \quad f(\boldsymbol{\Sigma}) f(\eta^{t+1} \mid \alpha^{t+1}, \boldsymbol{\Sigma}), \tag{23}$$

which follows directly from equation (21). Since the distribution of η_m given α and $\boldsymbol{\Sigma}$ is Q-dimensional multivariate normal, the conjugate prior for $\boldsymbol{\Sigma}$ is inverse-Wishart with a $Q \times Q$ scale matrix. In the application, I use $\boldsymbol{\Sigma} \sim \text{IW}(Q, Q\boldsymbol{I}_Q)$, a diffuse prior suggested by Train in a working paper for a closely-related MCMC algorithm for MXL (Train, 2009). This prior implies that equation (23) is $\text{IW}(Q + M, (Q\boldsymbol{I}_Q + M\boldsymbol{S}^{t+1})/(Q+M))$, where $\boldsymbol{S}^{t+1} = \sum_{m=1}^{M} (\eta_m^{t+1} - \alpha^{t+1})(\eta_m^{t+1} - \alpha^{t+1})^\top$.

D The Japanese Election Survey Data

All variables used in the empirical analysis in Section 6 are taken from the last two waves of the Japan Election Study II (JES II) panel survey data and auxiliary candidate information contained in the same dataset. The JES II study was originally conducted and made publicly available by Ikuo Kabashima, Joji

Voter Characteristics

Female	Age	Education
0.450	55.4	2.12
0.248	*18.2*	*0.922*

	Candidate-Varying Covariates						Outcome	
	Female	Age	Incumbent	Experience	Mail	Asked	Overall	Available
LDP	0.019	54.8	0.549	2.77	0.429	0.192	0.466	0.482
	0.018	*14.2*	*0.248*	*9.48*	*0.245*	*0.155*	*0.249*	*0.250*
NFP	0.032	49.0	0.480	1.56	0.341	0.275	0.255	0.313
	0.031	*12.3*	*0.250*	*5.83*	*0.225*	*0.199*	*0.190*	*0.215*
DPJ	0.084	46.8	0.318	1.06	0.201	0.103	0.129	0.236
	0.077	*13.8*	*0.217*	*3.23*	*0.161*	*0.092*	*0.113*	*0.180*
SDP	0.130	55.3	0.498	2.09	0.166	0.054	0.023	0.166
	0.114	*11.4*	*0.251*	*7.07*	*0.139*	*0.051*	*0.022*	*0.139*
JCP	0.207	50.0	0.015	0.18	0.062	0.062	0.116	0.116
	0.165	*14.4*	*0.015*	*0.96*	*0.058*	*0.058*	*0.102*	*0.102*
SKG	0.000	57.0	1.000	2.50	0.360	0.160	0.011	0.360
	0.000	*4.1*	*0.000*	*0.00*	*0.235*	*0.137*	*0.011*	*0.235*

Table 2: Summary Statistics. The table shows summary statistics for the variable used in the empirical analysis in Section 6. In each cell, the sample mean (top) and variance (bottom, italic) are shown for each variable. For the outcome variable, the statistics are computed both for the entire sample (i.e., ignoring unavailability; left) and excluding voters for whom the choice was not available in their local districts (right). See text (Section D) for how these variables are constructed from the original survey data.

Watanuki, Ichiro Miyake, Yoshiaki Kobayashi, and Ken'ichi Ikeda. It is a comprehensive panel election study conducted over the period of 1993–1996 and based on two-stage stratified sampling.

Female is a binary indicator variable and equal to one when the respondent is female and zero if male. *Age* is the age of the respondent. *Education* is a four-point scale variable, where $1 =$ less than high school education, $2 =$ high school graduate, $3 =$ some college, and $4 =$ bachelor degree or above.

As for the candidate-varying covariates, $female$ is an indicator variable which equals to one if the candidate is female. Age is the candidate's age; because this variable was only recorded in five intervals (between 25 and 39, 40s, 50s, 60s, and older than 69), I used the midpoints for the first four intervals and 75 for the oldest category. $Incumbency$ is an indicator variable which equals to one for incumbent candidates. $Experience$ indicates how many times the candidate was elected to the national legislature in the past; this variable is also measured in intervals (0, 1, 2 to 3, 4 to 5, 6 to 9, and more than 9) and thus I use midpoints for the first four categories and the value of 10 for the last category. Finally, $mail$ and $asked$ are indicator variables representing whether the respondent was targeted by the candidate's electoral campaign of these types. The summary statistics (mean and variance) of these variables are given in Table 2 along with the statistics for the outcome variable.

As mentioned in Section 6.1, the final analysis sample further excludes seven districts for various reasons. The districts included Tochigi 1 (with Hajime Funada being the strong non-major party candidate), Saitama 13 (Shinako Tsuchiya), Tokyo 4 (Shokei Arai), Shizuoka 4 (Yoshio Mochidsuki), Osaka 11 (Hirofumi Hirano), and Tottori 1 (Shigeru Ishiba). Five of these six candidates joined LDP not long after the election; the remaining candidate joined DPJ. The Kanagawa 14 district also dropped out of the study sample because of non-voting and missing covariates.