

# New Empirical Strategies for the Study of Parliamentary Government Formation

**Garrett Glasgow**

*Department of Political Science, University of California, Santa Barbara, 3719 Ellison Hall,  
Santa Barbara, CA 93106*

*e-mail: glasgow@polsci.ucsb.edu (corresponding author)*

**Matt Golder**

*Department of Political Science, Pennsylvania State University, 306 Pond Laboratory,  
University Park, PA 16801*

*e-mail: mgolder@psu.edu*

**Sona N. Golder**

*Department of Political Science, Pennsylvania State University, 305 Pond Laboratory,  
University Park, PA 16801*

*e-mail: sgolder@psu.edu*

Edited by Vera Troeger

In recent years, a consensus has developed that the conditional logit (CL) model is the most appropriate strategy for modeling government choice. In this paper, we reconsider this approach and make three methodological contributions. First, we employ a mixed logit with random coefficients that allows us to take account of unobserved heterogeneity in the government formation process and relax the independence of irrelevant alternatives (IIA) assumption. Second, we demonstrate that the procedure used in the literature to test the IIA assumption is biased against finding IIA violations. An improved testing procedure reveals clear evidence of IIA violations, indicating that the CL model is inappropriate. Third, we move beyond simply presenting the sign and significance of model coefficients, suggesting various strategies for interpreting the substantive influence of variables in models of government choice.

## 1 Introduction

The theoretical literature dealing with parliamentary government formation is arguably among the largest and most developed areas in all comparative politics (Baron and Ferejohn 1989; Laver and Schofield 1990; Laver and Shepsle 1996; Laver 1998; Golder, Golder, and Siegel 2012). In contrast, empirical work designed to distinguish among different perspectives and predict the party composition of governments has, until recently, lagged behind. One reason for this was that empirical scholars had not found a way to appropriately study data on tens of thousands of potential governments. Traditionally, scholars had used the potential government as the unit of analysis. However, this was problematic because it failed to recognize that the probabilities of the various potential governments taking office in a particular formation opportunity are interrelated—as one potential government becomes more likely to take office, other potential governments must become less likely to take office.

This all changed with the groundbreaking article by Martin and Stevenson (2001), which used the government formation opportunity as the unit of analysis and modeled government choice as a discrete

---

*Authors' note:* We thank Jim Adams, Lanny Martin, Randy Stevenson, Greg Wawro, two anonymous reviewers, and audiences at the 2010 Annual Meeting of the Midwest Political Science Association, the 2010 Summer Political Methodology Meeting, the 2011 Southern California Comparative Political Science conference, and the 2011 European Consortium for Political Research General Conference for helpful comments on this paper. The data and all computer code necessary to replicate the results and figures in this analysis are available at the *Political Analysis* dataverse.

choice problem with each potential cabinet as a choice alternative in a single formation opportunity. Their approach led naturally to McFadden's (1974a, 1974b) conditional logit (CL) model. In the CL model, the probability that government  $j$  is selected out of the set of  $K$  potential governments in formation opportunity  $i$  is

$$P_{ij} = \frac{e^{x_{ij}\beta}}{\sum_{k=1}^K e^{x_{ik}\beta}}, \quad (1)$$

where  $\beta$  represents a vector of coefficients and  $x_{ik}$  represents a vector of independent variables associated with potential government  $k$  in formation opportunity  $i$ . A consensus has quickly developed that the CL model is the most appropriate strategy for modeling government choice, encouraging a rapidly growing empirical literature on government formation (Bäck 2003; Diermeier and Merlo 2004; Druckman, Martin, and Thies 2005; Warwick 2005; Golder 2006; Skjæveland, Serritzlew, and Blom-Hansen 2007; Bäck and Dumont 2008; Indriðason 2008; Kang 2009; Martin and Stevenson 2010).

Although the CL model represents an important advance in the study of government choice, at least two empirical challenges remain. The first involves addressing the role of unobserved heterogeneity. Qualitative scholars have repeatedly shown that idiosyncratic contextual factors, such as personality clashes and ad hoc critical events, that are difficult to quantify can often have a significant impact on the choice of government (von Beyme 1985; Pridham 1986; Laver and Schofield 1990). Substantively, we would like to understand what role these factors play in influencing government formation, and in particular, how these factors induce variation in the influence of theoretically important variables. Methodologically, ignoring unobserved heterogeneity in the effects of our observed variables is likely to violate the CL model's assumption that the errors are independently and identically distributed (IID) across potential governments, leading to inconsistent parameter estimates. Failure to account for unobserved heterogeneity also leads to inaccurate estimates of substitution patterns—predictions of how the probabilities that potential governments will take office change as the characteristics of a government formation opportunity change. Although the CL model recognizes that the probabilities of winning office across potential governments in a formation opportunity are interrelated, it makes the restrictive independence of irrelevant alternatives (IIA) assumption, thereby imposing a particular type of relationship on these probabilities. A model that accounts for unobserved heterogeneity and thus relaxes the IIA assumption will allow us to estimate more accurate substitution patterns.

The second challenge involves extracting and presenting quantities of substantive interest from our empirical models. Although it is well recognized that such practices can add greatly to our understanding of the political world (King, Tomz, and Wittenberg 2000), it is rare for government formation scholars to interpret the substantive importance of the variables in their models. Estimating the substantive significance of variables in models of government choice is difficult for several reasons, as we describe below, and thus, most studies restrict themselves to discussing the sign and statistical significance of their model coefficients.

In this article, we make three methodological contributions that help to address these empirical challenges. First, we introduce the mixed logit (MXL) model to the study of government choice. An MXL with random coefficients allows us to take account of unobserved heterogeneity and improve our substantive understanding of government formation. The MXL also relaxes the IIA assumption, allowing us to estimate more flexible and accurate substitution patterns. Second, while addressing the issue of unobserved heterogeneity, we also demonstrate that the commonly used procedure for testing the IIA assumption in CL models of government choice is biased against finding IIA violations. An improved testing procedure reveals that many potential governments share unobserved attributes, thereby violating the IIA assumption and suggesting that the CL model is inappropriate. Third, we develop several strategies for interpreting the substantive influence of variables in MXL (and CL) models of government choice.

## 2 Unobserved Heterogeneity in the Government Formation Process

A key substantive and methodological problem confronting scholars interested in government formation has to do with the role played by unobserved heterogeneity, or variation in unobserved contextual factors, across formation opportunities. We know from qualitative researchers and country specialists that

unobserved or unmeasured factors can have a significant impact on government formation. The concern is that our observed variables may be crucially important for government choice in some contexts but largely irrelevant or even obstacles in others because of idiosyncratic or case-specific contextual factors.

It seems reasonable, for instance, to think that policy concerns might play a crucial role in government formation in some contexts but be largely inconsequential in others depending on how parties value the trade-off between office and policy (Müller and Strøm 1999; Indriðason 2005). As an example, the theoretical literature suggests that minority governments are more likely to form as the relative weight that parties place on policy increases (Strøm 1990; Golder, Golder, and Siegel 2012). Depending on the value placed on policy by certain parties in a particular formation opportunity, a potential government's minority status could be detrimental, beneficial, or inconsequential. The problem for the analyst, though, is that it is very difficult to measure the value that parties place on policy versus office at particular points in time.

Personality conflicts can also play an important role when it comes to government formation. Following the 2003 elections in the Netherlands, for instance, most party leaders advised the Queen's informateur to put together a center-left cabinet comprising the Christian Democratic (CDA) and Labor (PvdA) parties. This proved impossible, in large part due to the lack of "trust and 'chemistry' between the two partners" (Irwin and Holsteyn 2004, 555–6). Similarly, the unusually long delay in forming a new government after the 2007 Belgian election was exacerbated by the fact that the French-speaking Christian Democratic (cdH) party was "very reluctant" to enter a center-right coalition. This reluctance stemmed in part from "personal reasons" that were causing difficulties between the cdH leader Joëlle Milquet and the Liberal (MR) leader Didier Reynders (Pilet and van Haute 2008, 550). Although media accounts often suggest that personal conflicts between the leading protagonists lead to certain coalitions being ruled out, it is unclear how political scientists can evaluate these claims given the difficulty in systematically measuring the tone and importance of personal relationships between party elites, particularly in a cross-national setting.

If we can observe (and measure) the relevant case-specific contextual factors, then we can capture the conditional relationship between our observed variables and government choice through the use of interaction terms (Brambor, Clark, and Golder 2006; Kam and Franzese 2007, Berry, Golder, and Milton, *forthcoming*). As we have noted, though, it is often the case that this kind of heterogeneity is caused by characteristics of the particular formation opportunity that are either unobserved or difficult to measure in a reliable and systematic fashion. Even if plausible measures are available, including them in empirical models can be a challenge. For example, Martin and Stevenson (2010, 10) include party-specific fixed effects to capture one likely source of unobserved heterogeneity. This strategy requires estimating a large number of additional parameters (73 party-specific dummy variables would be required for the model we present in Table 1), which can create multicollinearity problems for those covariates related to party characteristics. Moreover, it also necessitates some ad hoc restrictions on the coefficients for those party-specific dummy variables related to parties that have never entered government in order to avoid a separation problem (Zorn 2005).<sup>1</sup>

For these reasons, quantitative scholars have typically assumed that the only meaningful variation that exists across cases is captured by the observed variables in their models and, therefore, that idiosyncratic contextual factors, such as personality conflicts, can be safely relegated to a random error term that is IID across observations. Unfortunately, this assumption is problematic when studying government formation on both substantive and methodological grounds.

Substantively, we want to relax the assumption that all formation opportunities are homogeneous, with the only interesting variation across cases captured by our observed variables. Although we cannot observe all the contextual factors influencing government choice, understanding how these unobserved factors introduce variation into the effects of the variables that we can observe is valuable substantive information. Some government alternatives may be more (or less) attractive than others for reasons that are not fully captured by the independent variables in our models. For example, a party with a leader who is known to be difficult may make all government alternatives containing that party less attractive than other potential cabinets with identical values on the independent variables. In general, it is easy to think how features specific to particular countries or elections might make some government alternatives

<sup>1</sup>For example, Martin and Stevenson (2010, 12) constrain these coefficients for their party-specific dummy variables to be  $-3$ .

more or less attractive than others. By allowing this kind of heterogeneity and causal complexity to play a role in our empirical model, we narrow the gap between general theoretical models and case-specific studies, thereby building a more nuanced substantive picture of government choice (Western 1998; Beck and Katz 2007). Unfortunately, the CL model cannot account for this type of unobserved heterogeneity (Train 2009, 42).

Methodologically, ignoring unobserved heterogeneity in the effects of our observed variables is likely to violate the CL model's assumption that the errors are IID across potential governments, leading to inconsistent estimates (Cameron and Trivedi 2005, 513).<sup>2</sup> In particular, unobserved heterogeneity is likely to mean that the error terms associated with the potential cabinets in each government formation opportunity are negatively correlated as increases or decreases in the influence of the observed variables will systematically advantage some potential cabinets and disadvantage others.

Failure to address violations of the IID assumption for our error term will also lead us to estimate incorrect substitution patterns between potential governments. In other words, we will incorrectly estimate how the probability that a potential government takes office changes as its characteristics or those of other potential governments in a formation opportunity change. This follows from the fact that the IID error term in the CL model is the basis for the well-known IIA assumption. Due to the types of unobserved heterogeneity described previously, it seems implausible that the IIA assumption will hold in the government formation setting, where there can be tens, hundreds, or even thousands of potential governments that share some unobserved characteristics.

### 3 Testing for IIA Violations in Government Formation Data

Of course, whether the IIA assumption is, in fact, violated in a particular setting is an empirical question, and one that can be evaluated with formal statistical tests such as the Hausman-McFadden (HM) test (Hausman 1978; Hausman and McFadden 1984). Remarkably, a number of recent studies report no evidence of IIA violations in CL models of government choice (Martin and Stevenson 2001; Druckman, Martin, and Thies 2005; Golder 2006; Skjaeveland, Serritzlew, and Blom-Hansen 2007; Bäck and Dumont 2008). As we now demonstrate, though, previous work has reached this conclusion because the commonly used procedure for testing the IIA assumption in CL models of government choice is inappropriate and biased against finding IIA violations.

When considering the ratio of probabilities between government alternatives *A* and *B* in a CL model, the denominators drop out, and with some rearranging, we have

$$P(A)/P(B) = e^{(x_A - x_B)\beta}. \quad (2)$$

Equation (2) illustrates the CL model's IIA assumption—the ratio of choice probabilities for governments *A* and *B* depends only on the characteristics of these two governments and remains constant no matter how the attributes of some other government *C* might change. This IIA assumption will be violated if, for instance, the unobserved attributes of government *A* are correlated with the unobserved attributes of government *C*.

HM tests compare the coefficients estimated by the full CL model under examination ( $\hat{\beta}_F$ ) to the coefficients obtained by estimating a restricted version of this model that omits one or more choice categories ( $\hat{\beta}_R$ ). An intuitive understanding of the HM test for IIA violations can be obtained by examining the ratio of probabilities between two choice alternatives that is presented in equation (2). Since the observed data  $x_A$  and  $x_B$  remain the same no matter which other choice alternatives are included in the model, the only way for this ratio to change in response to a violation of the IIA assumption is for  $\beta$  to change. The test statistic is

$$HM_{IIA} = (\hat{\beta}_R - \hat{\beta}_F)' [\hat{V}(\hat{\beta}_R) - \hat{V}(\hat{\beta}_F)]^{-1} (\hat{\beta}_R - \hat{\beta}_F), \quad (3)$$

where  $\hat{\beta}_F$  is restricted to include only those coefficients also included in  $\hat{\beta}_R$ , and  $\hat{V}(\hat{\beta}_R)$  and  $\hat{V}(\hat{\beta}_F)$  are the estimated variance-covariance matrices from the full and restricted CL models, respectively. This test

<sup>2</sup>Unobserved heterogeneity is less problematic in an ordinary least squares setting, where it leads to inefficient, but consistent, estimates.

statistic is distributed  $\chi^2$  with degrees of freedom equal to the number of coefficients in  $\hat{\beta}_R$ . Significant values of  $HM_{IIA}$  indicate a significant difference in  $\beta$  across the two models, which is evidence that the IIA assumption has been violated.

HM tests have typically been employed in settings where the number of alternatives is relatively small. Generally, researchers test for IIA violations with multiple HM tests, with each test omitting one of the choice alternatives. An HM test with a statistically significant result indicates that the unobserved attributes of the omitted choice alternative are correlated with the unobserved attributes of some other choice alternatives in the model, violating the IIA assumption. It is not immediately obvious, though, how to conduct tests of this type in a setting with thousands of unique choice alternatives such as the government formation process. Dropping a single alternative out of the thousands of choices available is unlikely to produce a statistically significant change in the model coefficients, even if there is an IIA violation.

As a result, an alternative testing procedure has been developed in the government formation literature. This approach begins by dropping a random subset of alternatives from each formation opportunity (anywhere from 10% to 50%, never dropping the government that actually formed) and conducting an HM test for IIA violations. This process is then repeated multiple times, and the average  $p$  value of the HM tests is reported. No study using this procedure has ever reported evidence of a violation of the IIA assumption (Martin and Stevenson 2001; Druckman, Martin, and Thies 2005; Golder 2006; Skjaeveland, Serritzlew, and Blom-Hansen 2007; Bäck and Dumont 2008), and all have therefore proceeded with a CL model.

Unfortunately, this now-standard testing procedure for IIA violations in the government formation literature is flawed and biased against finding IIA violations. First, dropping a certain percentage of potential governments from each formation opportunity at random does not represent a plausible (or, indeed, possible) change to the choice set of potential governments in the real world. For example, such a procedure might lead to the removal of a potential single-party government made up of Party A from the choice set while leaving cabinets that include Party A as possible government alternatives. It is clearly impossible in the real world to have cabinets that include Party A be part of the choice set when Party A has itself been removed from the choice set. As the purpose of dropping choice alternatives in the HM test is to remove some subset of choices that could plausibly serve as a substitute for some of the remaining choices, dropping potential governments at random is a mistake. A better strategy is to drop sets of potential governments that could plausibly have substitutes in the remaining data. For example, we could drop all potential cabinets that include a particular party, examining whether this party has a substitute in the formation opportunity.

Second, an insignificant average  $p$  value from a set of HM tests does not mean that the IIA assumption has not been violated. If any *one* HM test is significant, then this indicates that removing that choice alternative or set of choice alternatives has produced evidence of an IIA violation. For instance, if most potential governments are independent, but a small subset are related for some unobserved reason, then the average  $p$  value on a set of HM tests will likely be insignificant, but the IIA assumption would still be violated. Finally, to take account of the fact that some apparent IIA violations are likely to appear by chance given the large number of HM tests that are typically conducted when studying the government formation process, it is necessary to employ a Bonferroni correction for the test  $p$  values.

To examine whether our corrections to the standard testing procedure employed in the government formation literature reveal evidence of IIA violations, we performed a set of HM tests on the CL model presented in Table 1 below, sequentially correcting each of the three problems described here.<sup>3</sup> We first conducted 100 HM tests, randomly dropping 10% of the potential governments from each formation opportunity (never dropping the government that actually formed). The mean  $p$  value from these tests was .86, which in the previous work would have led scholars to conclude that there were no IIA violations and, therefore, that the CL model was an appropriate empirical strategy for examining government choice. However, rather than averaging the  $p$  values from the tests, we examined each individual  $p$  value to determine if any one test found an IIA violation. We found that the HM test produced a  $p$  value of less than .05 in 5 of the 100 tests, indicating multiple violations of the IIA assumption.

<sup>3</sup>Our upcoming CL model specification is identical to Model 7 from Martin and Stevenson (2001, 45); it is also equivalent to the baseline model in Martin and Stevenson (2010, 513).



We then applied a Bonferroni correction to the  $p$  values of these tests. This correction adjusts the  $p$  value required for statistical significance when multiple tests are applied by dividing the significance level by the number of tests conducted. Specifically, if an analyst conducts  $n$  statistical tests on the same data, then the Bonferroni correction involves evaluating each individual test at a statistical significance level of  $1/n$  times what it would be if only one test were being conducted. Given that we wanted to limit the risk of making at least one Type I error—rejecting the null of IIA when the null is true—to a significance level of  $\alpha = .05$  (“familywise” alpha) for the family of tests, the Bonferroni correction required that the  $p$  value from any individual HM test be less than  $\alpha = .05/100 = .0005$  (“testwise” alpha) for us to conclude that the IIA assumption was violated. Even with this conservative Bonferroni correction, we found that 2 of the 100 HM tests we conducted produced evidence of IIA violations.

Finally, rather than conducting a series of HM tests in each of which we drop a random subset of potential governments, we conducted a series of tests that sequentially dropped each political party and all the potential governments containing this party from a single formation opportunity at a time. For example, one of the HM tests involved dropping all potential governments in the 1949 Austrian government formation opportunity that included the Austrian People’s Party, whereas another test involved dropping all potential governments in the 1987 Irish government formation opportunity that included Fine Gael. Although the average number of potential governments that were dropped in any given test corresponded to less than half of 1% of the entire sample, we still found evidence of IIA violations in 12 tests (4 after the Bonferroni correction). That is, even when we just drop the cabinets related to a *single* party in a *single* formation opportunity from the whole data set (220 formation opportunities and 33,256 potential governments), and even with a Bonferroni correction, we found clear evidence of IIA violations in this CL model.

Overall these (very conservative) tests for IIA violations provide clear evidence that some potential governments are perceived as substitutes for unobserved reasons and, therefore, that the CL model is an inappropriate empirical strategy in the government formation context. Furthermore, it should be noted that even if IIA tests fail to indicate the presence of IIA violations, it is often unwise for scholars to conclude that the IIA assumptions hold. There is growing evidence that IIA tests of all sorts frequently fail to identify IIA violations, even in Monte Carlo studies where the data are explicitly constructed to have IIA violations (Cheng and Long 2007). Perhaps, the best advice on IIA violations is that models that make the IIA assumption, such as the CL, should only be used when the choice categories “can plausibly be assumed to be distinct and weighed independently in the eyes of each decision maker” (McFadden 1974a). This is clearly not the case for potential government alternatives in government formation opportunities.

#### 4 MXL Models for the Government Formation Process

These substantive and methodological concerns with unobserved heterogeneity and IIA violations can be addressed with an MXL specified as a random coefficients model (Train 1998; McFadden and Train 2000; Glasgow 2001).

Substantively, we want to specify a model where the effects of the independent variables are allowed to vary across government formation opportunities. For instance, we might want to allow for the possibility that the effect of minority government status will vary depending on how parties value policy relative to office. As our earlier discussion indicated, minority governments are expected to be more likely when parties place greater value on policy concerns. We can allow the effects of independent variables, such as minority government status, to vary across formation opportunities through a random coefficients setup in our MXL model, where the independent variables have a mean effect  $\beta$  that is adjusted upward or downward by some amount  $\eta_i$  in each formation opportunity  $i$ .

If we could observe this adjustment for each formation opportunity, then the probability that potential government  $j$  is selected in formation opportunity  $i$  is simply the CL probability from equation (1), with some coefficients adjusted by adding the constant  $\eta_i$  to  $\beta$ :

$$P_{ij} = \frac{e^{x_{ij}\beta + x_{ij}\eta_i}}{\sum_{k=1}^K e^{x_{ik}\beta + x_{ik}\eta_i}} \quad (4)$$

Of course,  $\eta_i$  is not actually observed because it is a function of the unobserved or unmeasured factors that affect government choice. We solve this problem by specifying a joint probability distribution  $g(\eta|\Omega)$

for  $\eta$ , where  $\Omega$  are the fixed parameters of the distribution  $g$ —this specifies the distributions of our random coefficients. We can then obtain the unconditional probability that a government is selected by integrating the term in equation (4) over all possible values of  $\eta$  weighted by the density function of  $\eta$  as given by  $g$ :

$$P_{ij} = \int \left[ \frac{e^{x_{ij}\beta + x_{ij}\eta_i}}{\sum_{k=1}^K e^{x_{ik}\beta + x_{ik}\eta_i}} \right] g(\eta|\Omega) d\eta. \quad (5)$$

This is the MXL model, so named because the choice probabilities are a mixture of CL probabilities, each with different values for the  $\eta$ s and a weight determined by the mixing distribution  $g$ . Estimating this MXL gives us  $\beta$  and  $\Omega$ —the means and covariance matrix of our random coefficients.

MXL models cannot be estimated through standard maximum likelihood techniques because the integral for the choice probabilities in equation (5) has no closed-form solution. Instead, they are estimated through maximum simulated likelihood. For each formation opportunity, a value for  $\eta_i$  is drawn from  $g(\eta|\Omega)$  and used to calculate  $\hat{P}_{ij}$ , the conditional choice probability in equation (4). This process is repeated  $R$  times and the integration over  $g(\eta|\Omega)$  is approximated by averaging over the  $R$  conditional choice probabilities for each formation opportunity. A simulated log-likelihood function is then created from these simulated probabilities and is maximized with conventional maximum likelihood techniques. In our estimates below, we use 200 Halton draws to calculate the conditional choice probabilities. The use of quasi-random sequences such as Halton draws allows for a better coverage of the interval over which the integration is to be performed (for a given number of draws) than if we were to draw randomly, greatly increasing the speed of estimation (Train 2009, 228).<sup>4</sup>

Note that the estimation of MXLs is, in general, a more demanding process than the estimation of CLs. First, it takes significantly longer to estimate an MXL, even when Halton draws are used. For instance, the estimation time for the MXL (Model 2) shown in Table 1 was 52 min compared to an estimation time of just 4 seconds for the CL model.<sup>5</sup> Second, unlike the log-likelihood function for the CL, the simulated log-likelihood function for the MXL may not be globally concave in all cases. This does not appear to be a serious issue when scholars specify normally distributed random coefficients but can pose a problem if they specify random coefficients with bounded distributions, such as lognormal distributions (Huber and Train 2001). Thus, using several different sets of starting values to ensure that the model has not converged to a local maximum is advisable.<sup>6</sup> Finally, comparing the results obtained from estimating the same MXL with different numbers of Halton draws to simulate the integrals will help to verify that simulation error is not unduly affecting the model results. If the model results change significantly as the number of draws changes, the number of draws used in the final model should be increased until they fall into a range where the coefficient estimates are stable (Hensher and Greene 2003).

#### 4.1 Specification Issues

A number of specification issues arise with MXLs. The first is determining which coefficients should have random distributions. One approach is to simply allow all the MXL coefficients to be random. Although theoretically possible, this can sometimes lead to identification problems in practice (Revelt and Train 1998; Train 2009, 141). The most common problem that we have encountered is in specifying random coefficients on variables that create a separation problem, that is, those variables that almost

<sup>4</sup>In theory, a random coefficients multinomial probit (MNP) model could also address our substantive and methodological concerns with unobserved heterogeneity and IIA violations (Hausman and Wise 1978; Haaijer et al. 1998). Indeed, MNP models have been employed in a similar capacity in the political science literature in the past (Alvarez and Nagler 1998a, 1998b). However, estimating an MNP model involves solving a  $K - 1$ -dimensional integral, where  $K$  is the number of choice alternatives in the data set. This means that estimating an MNP model in the government formation setting, where the number of potential governments in a single formation opportunity can number in the thousands, is a practical impossibility with standard maximum simulated likelihood techniques. The MXL model does not face the same estimation hurdle as the dimension of integration is defined by the number of random coefficients in the model, not the number of choice alternatives.

<sup>5</sup>These models were estimated with *Stata* 10.0 on a machine with a 3.8-GHz processor and 1.5 GB of RAM, running Windows XP. The MXL was estimated using the *mixlogit* package (Hole 2007).

<sup>6</sup>In the upcoming empirical analysis, we first estimated our MXL using the CL estimates as the starting values for the mean coefficients and 0.1 as the starting values for the SDs. We then estimated each model five more times, perturbing each coefficient obtained from the first estimation with a uniform draw on the interval  $[-0.5, 0.5]$  multiplied by a constant we incremented from 1 to 5. All model results were similar. We report the models with the highest simulated log-likelihood in Table 1.

perfectly predict whether a potential government will form or not (Zorn 2005). For example, our upcoming empirical application based on a model from Martin and Stevenson (2001) includes a dichotomous variable identifying potential governments that contain parties that had publicly refused to form a government together. As only one of the 220 governments that formed in our sample were based on such a preelection “anti-pact,” nonzero values on this variable almost perfectly predict that such a coalition will not enter office. Estimating a random coefficient on such a tenuously identified variable will usually result in scaling problems and unstable parameter estimates.

A more parsimonious approach to specifying MXLs was developed by McFadden and Train (2000), who present a Lagrange multiplier test to determine if an MXL with a given set of random coefficients would fit the data better than a CL. This useful specification test is based solely on the CL results and does not depend on any assumptions about the distribution of the possible random coefficients. For each independent variable  $x$ , the test first creates an “artificial variable”  $z_{ij} = \frac{1}{2}(x_{ij} - \sum_{k=1}^K x_{ik} P_{ik})^2$ , where  $x_{ij}$  is the value of independent variable  $x$  for potential government  $j$  in formation opportunity  $i$  and  $P_{ik}$  is the CL probability for potential government  $k$  in formation opportunity  $i$ . After adding a set of these artificial variables to the CL, one then examines their joint significance by conducting a Wald or likelihood ratio test with degrees of freedom equal to the number of artificial variables being tested. Note that each artificial variable is essentially the squared deviation of the independent variable from its weighted mean, with the weights given by the CL probabilities. If we find that this “leftover heterogeneity” is related to the choice outcome, then this is evidence that there was unobserved heterogeneity that was not being captured by the CL model. McFadden and Train (2000) show that this test is equivalent to a Lagrange multiplier test comparing the CL to an MXL with random coefficients on the set of variables being tested. Although this test does not explicitly identify which of the independent variables should have random coefficients, McFadden and Train suggest using random coefficients for any variables that have an accompanying artificial variable with an asymptotic  $t$ -statistic greater than one.

A second specification issue concerns the distribution(s) to be used for the random coefficients. MXL models are extremely flexible, and researchers can specify any distribution they wish for the  $\eta$ s.<sup>7</sup> By far the most common choice in empirical applications is normally distributed random coefficients. An appealing characteristic of normally distributed random coefficients is that they can be justified in terms of the central limit theorem—when a large number of random shocks, some positive and some negative, change the effect of a particular independent variable in an additive manner, then the distribution of the effect of this independent variable becomes normal. In some cases, randomly distributed coefficients with bounded support, such as triangular and uniform distributions, may be theoretically appropriate, although in most empirical applications they appear to produce similar substantive results as models with normally distributed random coefficients (Hensher and Greene 2003). For those random coefficients that are theoretically expected to be of only one sign, the lognormal distribution is the standard choice.<sup>8</sup> A coefficient  $\beta_k$  follows a lognormal distribution if the log of  $\beta_k$  is normally distributed. Typically, a lognormal distribution is parameterized in terms of the underlying normal, estimating parameters  $\mu$  and  $\sigma^2$  that represent the mean and variance of the log of the coefficient:  $\ln \beta_k \sim N(\mu, \sigma^2)$ . The lognormal assumes that the distribution of coefficients on an independent variable is nonnegative, but nonpositive distributions can be estimated by simply entering the negative of the independent variable into the model. In our upcoming models, we use normally distributed random coefficients (or lognormally distributed coefficients for those variables predicted to have only nonnegative or nonpositive effects).

A third specification issue concerns whether the random coefficients should be correlated or not. Many of the first discrete choice models in political science to relax the IIA assumption specified correlated random coefficients on a set of alternative-specific constant terms in order to capture the unobserved factors that led decision makers to view choice alternatives as similar (Alvarez and Nagler 1995; Glasgow 2001). These models examined cases with a relatively small number of choice alternatives that were identical across all observations. For instance, Alvarez and Nagler (1995) considered correlations

<sup>7</sup>In most applications, the mixing distribution  $g$  is assumed to be continuous. However, when the mixing distribution is discrete and the random coefficients can only take on a finite set of distinct values, the MXL becomes what is more commonly called a latent class or finite mixture model.

<sup>8</sup>Note that in some cases the unbounded tail of the lognormal can lead to difficulties in both empirical estimation and substantive interpretation (Huber and Train 2001; Hensher and Greene 2003).



that might arise due to the unobserved attributes of the three major candidates in the 1992 U.S. Presidential election. This particular type of specification, which involves including a set of alternative-specific constants, obviously cannot be replicated in studies of government formation as each choice alternative in each formation opportunity is unique. Although it is possible to examine correlations in the effects of the random coefficients, the theoretical and empirical justification for doing this is unclear. Furthermore, it should be noted that allowing for correlated random coefficients raises a set of fairly complicated identification issues that do not arise with uncorrelated random coefficients (Walker, Ben-Akiva, and Bolduc 2007). Finally, note that although  $\eta$  in equation (5) varies across government formation opportunities, it does not vary across potential governments in a single formation opportunity. This introduces correlation across potential governments into the factors that affect the probability that a cabinet will enter office, thereby relaxing the IIA assumption even when the random coefficients are uncorrelated. For these reasons, we estimate uncorrelated random coefficients in our models below.

## 5 An Empirical Application

To demonstrate how an MXL can add to our substantive understanding of government choice, we begin by replicating a well-known CL analysis originally conducted by Martin and Stevenson (2001, 45, Model 7).<sup>9</sup> The data come from 14 democratic countries—Austria, Belgium, Canada, Denmark, West Germany, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Sweden, and the United Kingdom—from the mid-1940s to the mid-1980s. Given their focus on *coalition* governments, Martin and Stevenson exclude formation opportunities in which a single party controls a legislative majority. This yields a total of 220 formation opportunities and 33,256 potential governments. The particular model specification employed here is designed to test a variety of hypotheses about how institutions and government characteristics relating to size and ideology affect the likelihood that a particular cabinet forms. One set of hypotheses focuses on how many parties are in a potential government and whether the potential cabinet contains the largest party, the median ideological party, or the party of the previous prime minister. Other hypotheses focus on whether the potential government is a minority coalition, a surplus majority coalition, or a minimal winning coalition; whether the potential cabinet is the incumbent government; whether there are large ideological divisions in the potential government and in the opposition; whether the potential cabinet contains an anti-system party; and whether the potential government faces an investiture vote. Finally, two hypotheses address whether the parties in a potential government have reached a pre-electoral coalition agreement and whether at least one party publicly refuses to go into office with another party from the same potential cabinet.

We compare this CL model to two different MXL specifications. The first specification (Model 1) allows all the MXL coefficients to be random, with the exception of the *Anti-System Presence in the Coalition*, *Pre-Electoral Pact associated with the Coalition*, and *Anti-Pact associated with the Coalition* variables. These three variables create a separation problem. Nonzero values on the anti-system and anti-pact variables almost perfectly predict that a potential government will not enter office—only 6 of the 220 governments that actually formed contained an anti-system party and only one contained parties with an anti-pact. Only 30 of the 33,256 potential governments had a pre-electoral pact, 7 of which went on to form a government. Attempting to estimate random coefficients on these variables led to scaling problems, so they were held fixed in our specification. The second specification (Model 2) relies on the McFadden-Train specification test to identify the variables that should have random coefficients. The results from the Lagrange multiplier test revealed statistically significant coefficients and asymptotic *t*-ratios greater than one on seven of the “artificial variables.” Random coefficients were specified for all these variables except for the anti-system variable, which was held fixed due to the separation problem just mentioned.

The random coefficients in both MXLs were all specified as uncorrelated normal distributions, with the exception of those on *Ideological Divisions in the Coalition*, *Ideological Divisions within Majority Opposition*, and *Median Party in the Coalition*. Ideological divisions within a potential government are not theoretically expected to increase the likelihood that it will enter government, but they may have a negative effect, with the severity of the negative effect varying across government formation

<sup>9</sup>We should note that we were able to replicate their analysis perfectly.

opportunities depending on how parties value the trade-off between office and policy. Similarly, theory indicates that minority governments are never less likely to form when potential majority coalitions are ideologically divided, and governments are never less likely to form when they contain the median party. The three random coefficients on these particular variables were specified as uncorrelated lognormal distributions. Since the lognormal assumes that the distribution of coefficients on an independent variable is nonnegative, we used the negative of the *Ideological Divisions in the Coalition* variable in our model (Model 1).

The results from estimating the three models described above are presented in Table 1. The CL model is presented in the first column, whereas the MXL results are presented in the following four columns. For each MXL model, the fixed coefficients and the means of the random coefficients are presented in the first column for that model, whereas the standard deviations (SDs) of the random coefficients are presented in the second column. Random coefficients that follow a lognormal distribution are shaded in gray as a reminder that these numbers are actually logged coefficients.

The CL results presented in Table 1 indicate that minimal winning coalitions, governments containing the largest party, minority governments faced with an ideologically divided opposition, incumbent governments, and governments with pre-electoral pacts are all more likely to form than governments without these characteristics. Conversely, additional parties, ideological divisions within the government, minority governments in systems where investiture votes are required, an anti-system party in the government, and anti-pacts associated with the coalition make potential cabinets less likely to enter office.

The fixed and mean coefficients in the MXLs lead to similar substantive insights as the corresponding coefficients in the CL, with the exception of the *Minimal Winning Coalition* variable, which is not statistically significant at conventional levels in the MXLs. When comparing the CL and MXL results, it is useful to remember three things. First, the log of the coefficients is presented for the lognormally distributed coefficients (shaded in gray). Second, the negative of *Ideological Divisions in the Coalition* was entered into Model 1 in order to estimate the lognormally distributed coefficient for that variable. As a result, readers should reverse the sign of the coefficient on this variable when interpreting its effect on government choice. And third, the unobserved influences captured in the random coefficients in the MXLs are assumed to be part of the IID error term in the CL. This results in different scales being used for the coefficients, with MXL coefficients tending to be larger than CL coefficients.

Although the effects of the independent variables in the CL model are similar to the mean effects in the MXLs, the MXLs make an additional contribution to our substantive understanding of government choice by revealing unobserved heterogeneity in the effects of several of these variables. Four of the normally distributed random coefficients had statistically significant SDs, those on *Minority Coalition*, *Number of Parties in the Coalition*, *Largest Party in the Coalition*, and *Incumbent Coalition*.

For all four of these variables, the substantive conclusions that we would draw from the CL model differ quite markedly from those that we would draw from the MXLs. Consider the results related to *Minority Coalition*. The negative and statistically significant mean coefficient in both MXLs indicates that minority governments are less likely to enter office *on average* than other potential governments. As we noted previously, though, theory suggests that minority governments may be more likely to form in some formation opportunities than others depending on the extent to which parties value policy relative to office. The large and statistically significant SD on *Minority Coalition* is consistent with this. The fact that the mean coefficient on *Minority Coalition* in Model 2 is  $-1.69$  with an SD of 3.37 indicates that 69.2% of the distribution is below zero and 30.8% is above zero. Practically speaking, this means that minority coalitions are disadvantaged a little over two-thirds of the time but advantaged a little under a third of the time. These MXL results directly contradict the substantive inference from the CL model that minority governments are *always* disadvantaged when it comes to government formation.

Now consider the results related to *Incumbent Coalition*. The positive and statistically significant mean coefficient in both MXLs indicates that incumbent governments are more likely to enter office *on average* than other potential governments. However, idiosyncratic case-specific events are known to have influenced the likelihood that incumbents are returned to office. Prior to the 1997 legislative elections in Norway, for example, Thorbjørn Jagland made an ill-advised declaration that his Labor government would resign if it received less than 36.9% of the national vote (the percentage Labor had won in the previous election). Although the elections resulted in the Labor Party being the plurality winner and likely PM party, the fact that it won only 35% of the vote effectively ruled it out as a governmental party

**Table 1** CL and MXL models of government choice

Regressor	Dependent variable: choice of government (1, 0)				
	CL	MXL		MXL	
	(M & S Model 7)	(Model 1)	(Model 2)		
	Coefficients	Fixed/mean coefficients	SD of coefficients	Fixed/mean coefficients	SD of coefficients
Minority Coalition	-0.85* (0.48)	-1.66* (0.89)	2.87*** (1.12)	-1.69** (0.87)	3.37*** (1.32)
Minimal Winning Coalition	0.55** (0.26)	0.22 (0.49)	2.03* (1.17)	0.49 (0.33)	
Number of Parties in the Coalition	-0.31** (0.14)	-0.47** (0.21)	0.79*** (0.24)	-0.44** (0.20)	0.69*** (0.24)
Largest Party in the Coalition	1.40*** (0.27)	2.18*** (0.76)	2.30** (1.14)	2.36*** (0.92)	2.52** (1.27)
Median Party in the Coalition	0.32 (0.20)	-0.59 (0.52)	0.05 (1.03)	-0.66 (0.56)	0.10 (1.45)
Ideological Divisions in the Coalition	-2.92*** (0.85)	1.13*** (0.36)	0.03 (0.87)	-3.10*** (1.08)	
Ideological Divisions within Majority Opposition	2.64*** (0.83)	1.60*** (0.32)	0.06 (0.40)	1.48*** (0.34)	0.01 (0.81)
Previous Prime Minister in the Coalition	-0.13 (0.25)	0.33 (0.41)	0.18 (0.76)	0.33 (0.42)	
Incumbent Coalition	1.89*** (0.20)	1.43** (0.64)	5.11*** (1.35)	1.47** (0.65)	4.50*** (1.21)
Minority Coalition where Investiture Vote Required	-0.91*** (0.34)	-2.54* (1.33)	3.27 (2.33)	-1.89* (1.01)	
Anti-System Presence in the Coalition	-19.13*** (3.71)	-22.08*** (4.59)		-21.75*** (4.59)	
Pre-Electoral Pact associated with the Coalition	2.72*** (0.62)	5.04*** (1.38)		4.25*** (1.03)	
Anti-Pact associated with the Coalition	-4.10*** (1.04)	-7.90*** (2.14)		-7.83*** (2.01)	
Formation opportunities	220	220		220	
Potential governments	33,256	33,256		33,256	
Log-likelihood/simulated log-likelihood	-558.74	-532.99		-533.75	
Likelihood-ratio $\chi^2$ versus CL		51.50***		49.98***	

*Note.* The first column presents the results from a CL model. This particular model replicates Model 7 from Martin and Stevenson (2001, 45) and is equivalent to the baseline model in Martin and Stevenson (2010, 513). The next four columns present the results from two MXL models. For each MXL, the fixed coefficients and the means of the random coefficients are presented in the first column for that model, whereas the SDs of the random coefficients are presented in the second column. All random coefficients are normally distributed except for those shaded in gray, which are lognormally distributed; the lognormal coefficients are presented as logged coefficients. Note that it was necessary to enter the negative of *Ideological Divisions in the Coalition* into Model 1. As a result, readers should reverse the sign of the coefficient on this variable in Model 1 when interpreting its effect on government choice. Standard errors are shown in parentheses.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$  (two-tailed).

(Strøm 2000). The large and statistically significant SD on *Incumbent Coalition* in Model 2 indicates that incumbent governments are advantaged about 62.8% of the time but disadvantaged 37.2% of the time. This result again contradicts the substantive inference from the CL model that incumbent governments are *always* advantaged when it comes to government formation.<sup>10</sup>

<sup>10</sup>The substantive conclusions that one would draw from the CL and MXLs also differ in a similar way with respect to the *Number of Parties in the Coalition* and *Largest Party in the Coalition* variables.

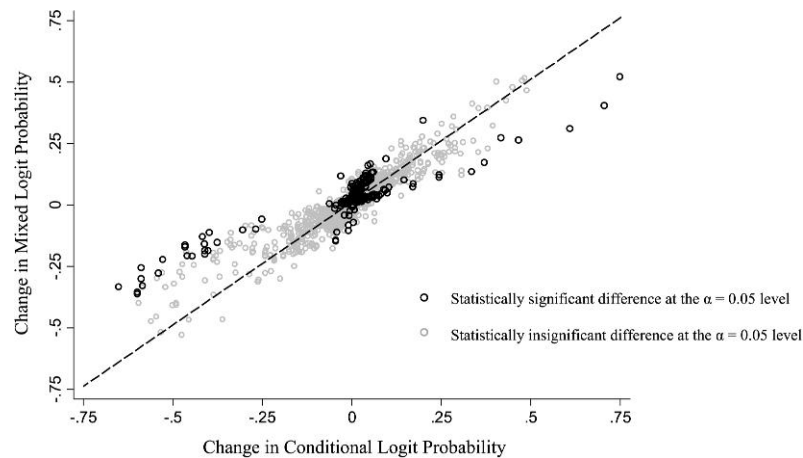
In some cases, the discovery of unobserved heterogeneity in the effect of an independent variable in an MXL model can be a precursor to further theory development, leading scholars to think about exactly what unobserved or unmeasured factors might be causing it. These factors can then be included as conditioning variables in future analyses. Indeed, two recent studies suggest that at least some of the unobserved heterogeneity in the effect of *Incumbent Coalition* has to do with the performance of incumbent parties while in office and with whether the incumbent government ended in public conflict or not (Martin and Stevenson 2010; Glasgow, Golder, and Golder 2011). It is worth noting, though, that many of the factors that qualitative scholars believe are being ignored in large- $N$  quantitative analyses, and which lead to possible unobserved heterogeneity, are case specific, and it will be impossible to construct a general theory to explain them. For example, it is hard to come up with a general theory to explain why political actors can overcome personal animosity to form coalitions in some situations but not in others.

Methodologically, the statistically significant SDs on the random coefficients in Table 1 clearly indicate the presence of unobserved heterogeneity, demonstrating that the CL model's assumption that the error term is IID across potential governments has been violated. Likelihood ratio tests confirm that both MXL models are a better fit to the data than the CL model. It is a widespread misconception that since parameter estimates from a CL model are often similar to those from models that relax the IIA assumption, such as the MXL or MNP, IIA violations are substantively inconsequential. This is incorrect. Failure to address violations of the assumption of an IID error term will lead to inconsistent parameter estimates in the CL model. IIA violations will also lead us to estimate incorrect substitution patterns between potential governments. In other words, we will incorrectly estimate how the probability that a potential government takes office changes as its characteristics or those of other potential governments change. Previous research suggesting that IIA violations are "rarely relevant" (Dow and Endersby 2004) has only considered differences in parameter estimates, neglecting to study how IIA violations influence predicted substitution patterns.

Figure 1 demonstrates how IIA violations can lead to incorrect substantive inferences in the present context. This figure plots the change in predicted probability for each potential government in our data set under a hypothetical scenario where we switch largest party status between the largest and second largest parties in each government formation opportunity. The changes in probability estimated by the CL model are plotted on the horizontal axis, whereas the changes for the MXL model (Model 2) are plotted on the vertical axis. The dashed 45-degree line indicates our expectation if the CL and MXL models produced the same estimated changes in probabilities in this hypothetical scenario. Statistically significant differences (at the  $\alpha = .05$  level) between the CL and MXL probability changes are indicated with black circles, whereas statistically insignificant changes are indicated with gray circles.<sup>11</sup>

It should be immediately clear that the inconsistent CL model tends to overestimate the changes in the predicted probability that a potential government enters office when the identity of the largest party changes. Indeed, regressing the changes in predicted probability from the MXL model on the changes in predicted probability from the CL model produces a slope coefficient of .77. The difference between the CL and MXL predictions under our counterfactual scenario is statistically significant for 2356 of the 33,256 potential governments. Although the magnitude of the changes in the CL and MXL predicted probabilities are quite similar for the mass of cases where the change in the identity of the largest party has little impact, they are quite marked for those cases where the change in the identity of the largest party matters most. In fact, the difference between the changes in predicted probability from the CL and MXL models can be as large as .32. Although partially influenced by the cases where the change in the identity of the largest party has little effect, it is also worth noting that the changes in predicted probability from the CL model are off by over 50% in 67% of the potential governments in our sample. If we restrict ourselves to just those cases where there was a statistically significant difference between the CL and MXL results, the changes in predicted probability from the CL model are still off by over 50% more than half the time.

<sup>11</sup>Statistical significance was calculated by taking 1000 draws from the multivariate normal distribution defined by the coefficient vector and variance-covariance matrix for both the CL and the MXL models and calculating 1000 different estimates of the probability differences under the counterfactual scenario for each model. As these probability differences are independent normal distributions, we can simply calculate the mean and SD of the difference between these differences. We report the difference between the CL and MXL probability changes as statistically significant if it was 1.96 SDs from zero.



**Fig. 1** Comparing MXL and CL substitution patterns. This figure plots the predicted change in probability for each potential government in our data set as we switch largest party status from the largest to the second largest party in each government formation opportunity. The changes in probability estimated by the CL model are plotted on the horizontal axis, whereas the changes for the MXL model are plotted on the vertical axis. The probability changes would line up along the dashed 45-degree line if the CL and MXL models produced the same estimated changes in probabilities for this hypothetical scenario. Statistically significant differences (at the  $\alpha = .05$  level) between the CL and MXL probability changes are indicated with black circles, whereas statistically insignificant changes are indicated with gray circles.

Whether we are interested in statistical or substantive significance, it seems clear that any counterfactual analysis we might undertake with the CL model is likely to be misleading.

## 6 Substantive Interpretation of Models of Government Formation

Although the substantive interpretation of one's model coefficients, whether from a CL or an MXL model, can add greatly to one's understanding of government formation, almost all studies of government choice have restricted themselves to reporting the sign and statistical significance of their model coefficients. One reason for this is that the substantive interpretation of government formation models poses challenges that are not normally encountered in other applications of multinomial choice models in political science.

Multinomial choice models in political science typically employ data with a common and relatively small set of choice alternatives across decision-making units. For instance, Alvarez and Nagler's (1998b) study of voter choice in the 1987 British election involved a sample of 2131 voters and 3 choice alternatives. The now-standard approach to substantive interpretation in this type of setting is to create a hypothetical observation by setting the values of the independent variables to create a hypothetical baseline scenario, and then examining how the predicted probabilities for the choice alternatives change as we alter the values of the independent variables to create different hypothetical scenarios (King, Tomz, and Wittenberg 2000).

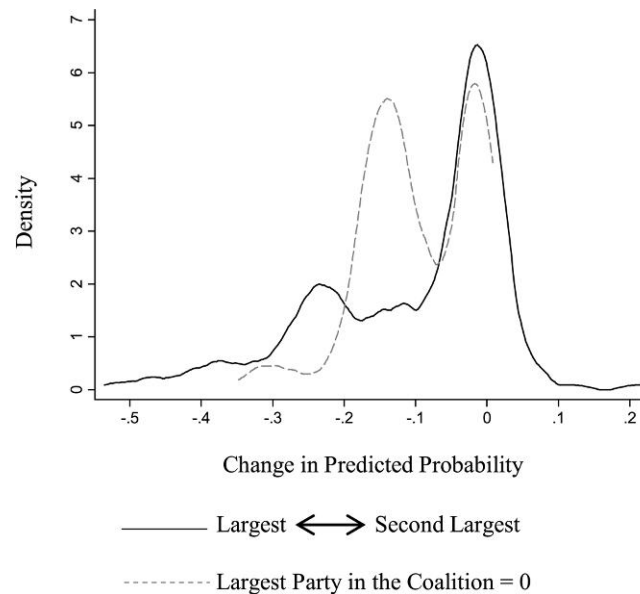
Most data on government choice differs from this standard setup in at least three ways that have important implications for the substantive interpretation of government formation models. First, the choice alternatives (potential governments) are not common across decision-making units (formation opportunities). Each potential government is unique, and this limits the generalizability of any single hypothetical observation one might create.

Second, the large number of potential governments compared to formation opportunities results in many near-zero predicted probabilities. Indeed, the median predicted probability for a potential government based on the MXL or CL results in Table 1 is less than .001. Although changing the values of the independent variables to create different hypothetical scenarios may have a large effect on some potential governments, care must be taken to ensure that the presentation of these results is not swamped by the thousands of near-zero probabilities.



Third, many of the independent variables in models of government choice are logically linked across potential governments in a formation opportunity. For instance, creating a counterfactual case in which a particular potential government no longer contains the largest party typically implies that some other governments in the formation opportunity now *do* contain the largest party. Furthermore, this change implies changes in other things, such as the set of minimal winning and minority coalitions. The failure to account for this type of relationship between independent variables will, in many cases, produce logically inconsistent counterfactuals—for example, scenarios with “minimal winning coalitions” that do not control a majority of seats. This is obviously problematic if one’s goal is to study real-world substantive effects. The issue here is analogous to calculating the substantive effect of a variable that is also part of an interaction term—without accounting for the interaction term, statements about the substantive effect of the variable will be misleading. An implication of all this is that there is no single effect associated with a counterfactual analysis—the effect of changing some independent variable on the probability of entering office will necessarily vary across both potential governments and formation opportunities.

In Fig. 2, we demonstrate the difference between constructing a logically coherent counterfactual scenario and simply setting a variable of interest to a counterfactual value. This figure presents kernel density plots over changes in the predicted probabilities derived from the MXL (Model 2) for the actual government that formed in each of the 220 formation opportunities in our sample using two different approaches to counterfactual scenarios. The first counterfactual approach examines what happens when the largest party status is transferred from the largest party in the formation opportunity to the second largest. Transferring the largest party status from one party to another can result in other changes such as whether the potential government is minimal winning or minority—all these changes are taken into account. The kernel density plot for this logically coherent counterfactual, which is, in fact, the same



**Fig. 2** Kernel density plots of changes in the predicted probability of entering office for actual governments when changing largest party status. This figure presents kernel density plots over changes in the predicted probability derived from the MXL (Model 2) for the actual government that formed in each of the 220 formation opportunities in our sample using two different approaches to counterfactual scenarios. The solid black line presents a logically coherent counterfactual created by transferring largest party status to the second largest party and accounting for any resulting changes in the identity of minority or minimal winning coalitions. The dashed gray line presents a counterfactual that disregards logical coherence and simply sets the dichotomous variable indicating whether the largest party is in the coalition to 0 for all potential governments. Both density plots use an Epanechnikov kernel with a bandwidth of 0.02.

counterfactual that we examined in Fig. 1, is captured by the solid black line in Fig. 2. The second counterfactual approach ignores our advice and simply sets the dichotomous variable *Largest Party in the Coalition* to 0 for all potential governments. In effect, it assumes that the government containing the largest party no longer contains this party and makes no effort to construct a logically coherent counterfactual. The kernel density plot for this counterfactual is captured by the dashed gray line in Fig. 2.

Figure 2 reveals a marked difference in the estimated changes in probability across the two different counterfactuals. Although setting *Largest Party in the Coalition* to zero for all governments always results in a decrease in the predicted probability that a government will form, the picture is considerably more complex if we construct a more logically coherent counterfactual. Once we account for which party is the new largest party, and the changes this has on other coalition characteristics, we see a much wider range of expected changes in probability, and in some cases, these changes can even increase the probability that the government forms.<sup>12</sup> In short, logically coherent counterfactuals can produce markedly different results from “naïve” counterfactuals that simply set a covariate to a hypothetical value.

Only a few studies of government choice have made an effort to provide some substantive interpretation of their model coefficients. Those that do typically report odds ratios for changes in particular independent variables (Druckman, Martin, and Thies 2005; Martin and Stevenson 2010). For example, we might think to use the results from the CL model presented in Table 1 to calculate the odds that a potential government containing the largest party would form versus some alternate government as  $e^{(1.40 \times 1) - (1.40 \times 0)} = 4.06$ . This reveals that a potential government containing the largest party was 4.06 times as likely to form as a potential government without the largest party, holding all other variables constant. However, drawing substantive inferences from odds ratios may be misleading due to our third point above. Moreover, this method of calculating odds ratios requires the IIA assumption, something that, as we have seen, is unlikely to hold in the government formation context.

In what follows, we present three strategies for interpreting the substantive effects of independent variables in models of government choice that address the difficulties in interpretation described above. All these methods can be applied to both CL and MXL models.

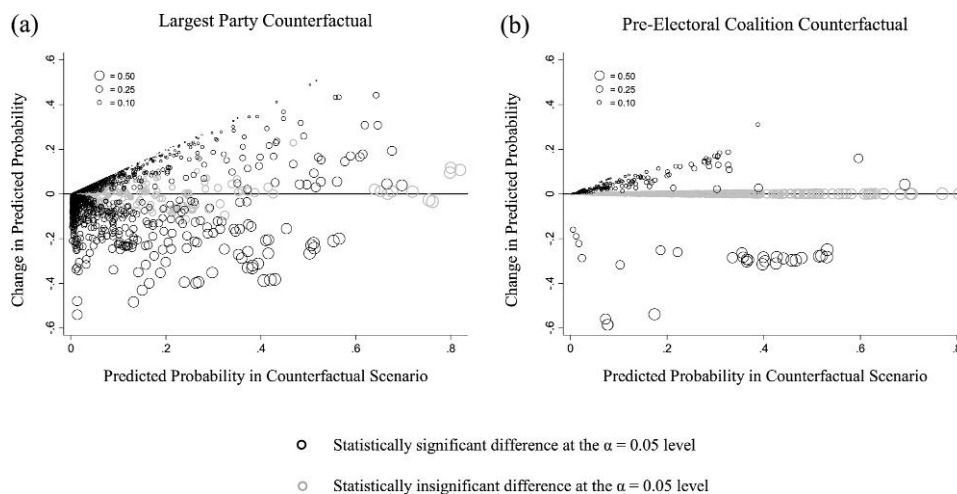
First, bubble plots based on probability changes under some counterfactual scenario can be a useful way to examine the substantive influence of a variable across all potential governments. In Fig. 3, we present the changes in the predicted probabilities derived from the MXL (Model 2) for all potential governments for two different counterfactual scenarios. The first counterfactual, shown in Fig. 3a, is the logically coherent largest party counterfactual described earlier where we switch the identity of the largest party in each formation opportunity to the second largest party. The second counterfactual, shown in Fig. 3b, considers the changes in predicted probability if potential governments did not form pre-electoral pacts.

The predicted probability that each potential government enters office in the counterfactual scenario is plotted on the horizontal axis, whereas the changes in predicted probability as we shift from the baseline to the counterfactual scenario are plotted on the vertical axis. The size of each circle represents the predicted probability for each potential government under the baseline scenario. Large circles near the bottom left corner of the graph indicate potential governments that were likely to enter office under the baseline scenario but that lost probability and are now unlikely under the counterfactual scenario. Small circles near the upper right corner of the graph indicate potential governments that were unlikely under the baseline scenario but that gained probability and are now likely under the counterfactual scenario. The upward sloping “border” in the top portion of each bubble plot is created by potential governments that had essentially zero probability of entering office in the baseline scenario but that have seen this probability increase under the counterfactual. Statistically significant differences (at the  $\alpha = .05$  level) are indicated with black circles, whereas statistically insignificant changes are indicated with gray circles.<sup>13</sup>

A comparison of the two counterfactuals suggests that largest party status had a much more obvious influence on the probabilities that different potential governments would form than pre-electoral pacts—

<sup>12</sup>Note that the governments will still contain the largest party if they included both the largest and the second largest parties.

<sup>13</sup>Statistical significance was calculated by taking 1000 draws from the multivariate normal distribution defined by the coefficient vector and variance-covariance matrix and calculating 1000 different estimates of the probability differences under the counterfactual scenario. A change was regarded as statistically significant if the 2.5th and 97.5th percentile of the 1000 probability differences were both of the same sign. The actual probabilities presented in this graph are the medians of these 1000 probability differences.

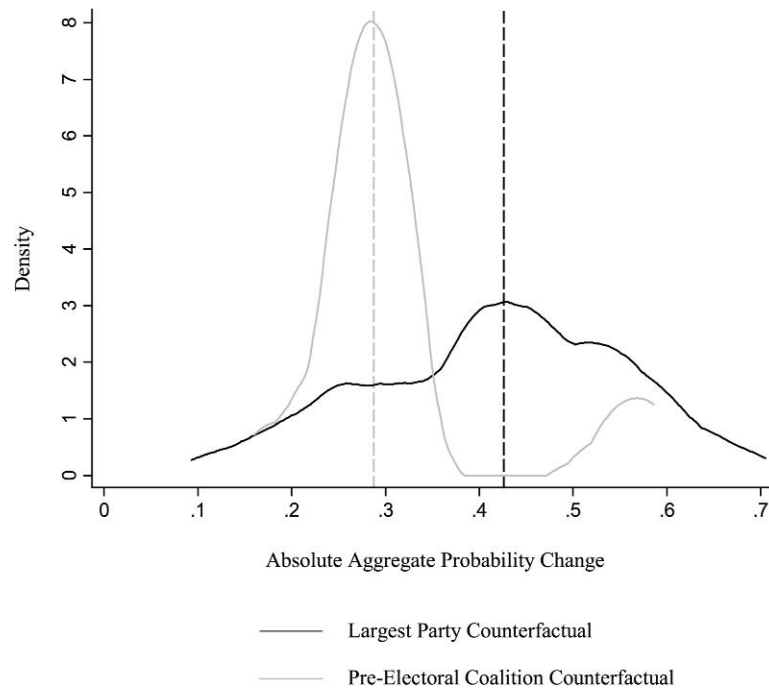


**Fig. 3** A bubble plot based on probability changes under two counterfactual scenarios. This figure presents changes in the predicted probabilities derived from the MXL (Model 2) for all potential governments for two different counterfactual scenarios. Panel (a) presents the changes resulting from a counterfactual that transfers largest party status to the second largest party, whereas panel (b) presents the changes resulting from a counterfactual that eliminates all pre-electoral pacts. The counterfactual probability estimated for each potential government by the MXL model is plotted on the horizontal axis, whereas the changes in probability estimated under each counterfactual are plotted on the vertical axis. The size of each circle represents the probability estimated for each potential government under the baseline scenario. Statistically significant changes (at the  $\alpha = .05$  level) are indicated with black circles, whereas statistically insignificant changes are indicated with gray circles.

many more of the circles have moved from the zero line in Fig. 3a than in Fig. 3b. Indeed, 48.6% of the probability changes under the largest party counterfactual were statistically significant, whereas only 1.8% of the changes under the pre-electoral pact counterfactual were significant. However, these numbers are slightly misleading. It is important to remember that only 30 of the 220 formation opportunities in our sample actually had potential governments based on pre-electoral pacts. As a result, our pre-electoral pact counterfactual could not influence the choice probabilities in many cases. It turns out that if we had limited our analysis to only the 30 formation opportunities in which there were pre-electoral pacts, then we would have found that *all* the probability changes under the pre-electoral counterfactual would have been statistically significant. To some extent, this is because the pre-electoral pacts we observed were in formation opportunities with a small number of potential governments (primarily in Germany and Austria), which tend to produce larger probability shifts for each potential government under the counterfactual change. The sorts of comparisons that can be made with the aid of bubble plots like this are useful in producing an overall picture of how important different independent variables are in the government formation process, either across all formation opportunities as is the case here or across some subset of interest.

A second approach is to consider the influence of these types of counterfactuals within formation opportunities. One way to do this is to calculate a statistic that we term the *absolute aggregate probability change*. This statistic measures how much probability shifts between the potential governments within a formation opportunity.<sup>14</sup> The absolute aggregate probability change for formation opportunity  $i$  is calculated as  $\Delta_i = (\sum_j |\Delta P_{ij}|)/2$ , where  $\Delta P_{ij}$  is the change in probability for potential government  $j$  under the counterfactual. The absolute value prevents the sum of changes from equaling zero, and division by two corrects for counting gains and losses as separate shifts. This statistic will range between zero and one, with larger numbers indicating greater shifts in the relative probabilities of the potential governments within the formation opportunity.

<sup>14</sup>This statistic is similar to the absolute average probability change statistic proposed by Long (1997, 166).



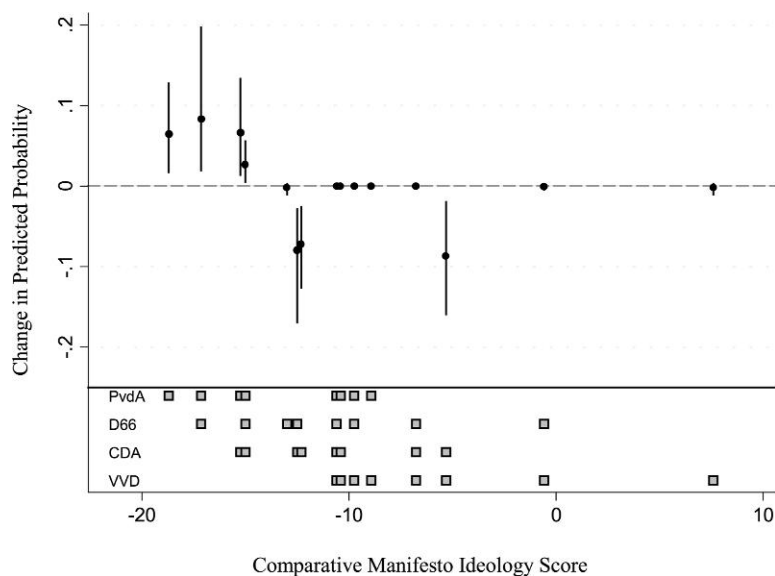
**Fig. 4** Kernel density plots of the absolute aggregate probability change across formation opportunities under the largest party and pre-electoral pact counterfactuals. This figure presents kernel density plots for the absolute aggregate probability change under the largest party (black line) and pre-electoral pact (gray line) counterfactuals. The vertical dashed lines represent the median value of the absolute aggregate probability change across the relevant formation opportunities (220 for the largest party counterfactual, 30 for the pre-electoral pact counterfactual). Both density plots use an Epanechnikov kernel with a bandwidth of 0.03.

Figure 4 presents kernel density plots of the absolute aggregate probability change under our largest party and pre-electoral pact counterfactuals across the relevant formation opportunities in our sample (220 for the largest party counterfactual, 30 for the pre-electoral pact counterfactual). The median value of this statistic for both the largest party counterfactual (0.43) and the pre-electoral pact counterfactual (0.29) are represented with vertical dashed lines. Figure 4 reveals that changes in largest party status tended to produce larger shifts in probability between the potential governments within formation opportunities as compared to changes in the status of pre-electoral pacts, even when we limit our analysis of pre-electoral pacts to just those formation opportunities where they were observed.

Finally, scholars may wish to supplement these types of “aggregate” substantive interpretations with one or more in-depth examinations of how changes in their independent variables alter the relative probabilities of entering office for the potential governments within a single formation opportunity. This can easily be done by creating one or more hypothetical formation opportunities (which can be based on real-world cases) and then examining how their counterfactual scenarios influence the likelihood that each of the potential governments in these hypothetical formation opportunities enter office.<sup>15</sup> Using this approach in conjunction with the aggregate measures described above can produce insight into the substantive effects of an independent variable both across and within formation opportunities.

As an example of this kind of within-case substantive analysis, we start with a hypothetical case based on a government formation opportunity resembling the one that took place in the Netherlands following the 1981 election. We examine a Dutch formation opportunity because it is widely recognized that Dutch elections are representative of the types of elections that occur in multi-party parliamentary democracies and because we wish to follow in a long line of studies that have used Dutch elections to demonstrate methodological advances in the use of discrete choice models (Alvarez and Nagler 1998b;

<sup>15</sup>Däubler and Debus (2009) adopt a similar approach in their study of government formation in German *Länder* elections.



**Fig. 5** Changes in the predicted probability of entering office for various potential governments when changing largest party (CDA→PvdA, Netherlands 1981). The results presented here are for a government formation opportunity resembling the one that took place in the Netherlands in 1981. The four parties in this hypothetical scenario are shown at the bottom, with gray squares indicating which parties belong to each potential government. The potential governments are presented in ideological order from left to right, based on the seat-share weighted average of the CMP ideological scores for each party: PvdA  $-18.7$ , D66  $-13.1$ , CDA  $-12.3$ , and VVD  $7.6$ . The ideological positions of some potential governments have been shifted slightly for clarity of presentation. Medians and 95% confidence intervals for the changes in predicted probability that result from a counterfactual case that changes the largest party from the CDA to the PvdA are presented. Probability estimates are based on simulations using 1000 draws from the estimated coefficient vector and variance-covariance matrix.

Schofield et al. 1998; Quinn, Martin, and Whitford 1999; Glasgow and Alvarez 2005). Although there were 10 parties in the Dutch legislature in 1981, we focus on potential governments involving the 4 largest political parties: CDA (Christian Democrats), PvdA (Labor), VVD (Conservative-Liberal), and the D66 (Progressive-Liberal).

In Fig. 5, we use the results from the MXL in Table 1 (Model 2) to calculate the changes in predicted probability when we shift from a baseline scenario where the independent variables take on the actual values observed in the real-world formation opportunity to a counterfactual scenario where the PvdA has replaced the CDA as the largest party. Each of the 15 governments that could be formed from these four parties is presented in ideological order from left to right, based on the seat-share weighted average of the Comparative Manifesto Project (CMP) ideological scores for each party (Budge et al. 2001). According to the CMP scores, the CDA, D66, and the PvdA were all centrist parties, with the CDA slightly to the right and the PvdA and D66 slightly to the left. The VVD was more ideologically distinct and clearly on the right. The four parties are listed at the bottom of the graph, with gray squares indicating which parties are members of each potential government. To calculate the probability differences between the two scenarios, we first took 1000 draws from the estimated coefficient vector and variance-covariance matrix, calculated the predicted probability for each potential government for each draw under both scenarios, and finally calculated the difference between these two vectors of probabilities. The medians and 95% confidence intervals (based on the 2.5th and 97.5th percentile of the 1000 probability differences) are presented here. Note that the 95% confidence intervals will not necessarily be symmetrical around the median.

An examination of Fig. 5 reveals the influence of both largest party status and ideological divisions within a coalition in our model of government choice. As expected, coalitions containing the PvdA tend to benefit, and coalitions containing the CDA tend to suffer, in the counterfactual scenario when we transfer largest party status from the CDA to the PvdA. Note, however, that some coalitions containing the CDA actually have a higher likelihood of entering office in the counterfactual scenario; these particular



**Table 2** The substantive influence of largest party status on government choice (Netherlands 1981)

Coalition	CDA		CDA → PvdA		CDA → VVD		CDA → D66	
	Predicted probability	Change in probability	Predicted probability	Change in probability	Predicted probability	Change in probability	Predicted probability	Change in probability
CDA-VVD	.26 (.05)	-.08** (.03)	.18 (.05)	.52 (.07)	.15 (.05)	.26*** (.08)	.15 (.05)	-.11*** (.03)
CDA-PvdA	.15 (.04)	.07*** (.03)	.22 (.05)	.01 (.004)	.01 (.004)	-.14*** (.04)	.01 (.004)	-.15*** (.04)
CDA-D66	.13 (.05)	-.07** (.03)	.07 (.03)	.08 (.04)	.34 (.09)	-.06** (.02)	.34 (.09)	.20** (.08)
CDA-PvdA-D66	.06 (.02)	.02*** (.01)	.09 (.02)	.05 (.02)	.03 (.01)	-.02 (.02)	.03 (.01)	-.03 (.02)

*Note.* The results presented here are for a government formation opportunity resembling the one that took place in the Netherlands in 1981. The first column indicates the MXL (Model 2) predicted probabilities for four selected potential governments in the baseline scenario when the independent variables take on the actual values observed in the real-world selection opportunity. The remaining columns indicate either predicted probabilities or changes in predicted probabilities as we transfer largest party status (and party size) from the CDA to one of the other three major parties. Standard errors are shown in parentheses. Statistically significant changes in predicted probability are indicated: \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$  (two-tailed). Estimates are based on simulations using 1000 draws from the estimated coefficient vector and variance-covariance matrix.

coalitions happen to contain the PvdA (CDA-PvdA and CDA-PvdA-D66). Also note that ideologically divided coalitions have a very low probability of forming, even if one of their members is the largest party. For instance, the predicted probability for the PvdA-VVD coalition does not measurably increase when the PvdA becomes the largest party.

Extending this type of analysis to consider multiple counterfactual scenarios is straightforward. In many applications, scholars will be interested in how the probabilities of substantively important potential governments change under different counterfactual scenarios. In our 1981 Dutch formation opportunity, the incumbent government was a CDA-VVD coalition. This government had formed in 1977 following a tortuous government formation process that had lasted 208 days. Many Dutch voters considered it to be illegitimate because it did not include the social-democratic PvdA, a party that was widely considered to have won a “landslide victory” at the 1977 polls. The CDA-VVD government had made matters worse by implementing a more right-wing economic policy than many Dutch wanted. Whereas the PvdA played a purely obstructionist role during the legislative term, the D66 acted as a much more “reasonable opposition” and had seen its seat share rise as a result in the 1981 elections (van Mierlo 1981, 300). Much of the speculation during the 1981 government formation process centered on whether the largest political party, the CDA, would form a coalition government with the VVD, the PvdA, the D66, or both the PvdA and the D66. In Table 2, we present the predicted probabilities that each of these potential governments would enter office based on the MXL (Model 2) results presented in Table 1. In the first column, we indicate the predicted probabilities that they would come to power in the baseline scenario where the independent variables take on the actual values observed in the real-world formation opportunity. In the subsequent columns, we indicate the predicted probabilities that they come to power when we transfer largest party status from the CDA to the PvdA, the VVD, or the D66. Finally, we also indicate the change in predicted probability from the baseline scenario to the counterfactual one and whether the change is statistically significant.

Table 2 reveals that the incumbent CDA-VVD government was the most likely outcome in the baseline scenario. Whereas transferring largest party status from the CDA to the PvdA increases the likelihood of a CDA-PvdA or CDA-PvdA-D66 coalition and reduces the probability of a CDA-VVD coalition, transferring largest party status from the CDA to the VVD has the opposite effect—the probability of a CDA-PvdA coalition decreases but that of a CDA-VVD coalition increases. These two counterfactuals reveal that as a large, centrist party, the CDA was likely to be a member of the governing coalition whether the largest party was on the left or the right. The government that actually formed in 1981 was the center-left, three-party coalition comprising the PvdA, CDA, and D66. That the incumbent CDA-VVD government was not returned to power was likely due to some of the contextual factors specific to the Dutch formation opportunity that we mentioned earlier. This fits with our MXL results in Table 1 showing that there is considerable unobserved heterogeneity in the effect of the *Incumbent Coalition* variable on government choice and that incumbent governments will find themselves disadvantaged about 37.2% of the time. Note that the predicted probability of a CDA-PvdA-D66 government is the lowest of the four potential governments in Table 2. It turns out that the PvdA-CDA-D66 government was extremely short-lived and was replaced by a caretaker government after just 8 months. The next duly-mandated government that formed following new elections in 1982 was a CDA-VVD coalition.

## 7 Conclusion

Understanding why certain parties get into government but others do not is crucial to our understanding of democratic politics. The reason for this is that the government formation process represents a key step in the delegation chain linking citizen preferences to policy outcomes (Strøm 2003).

The current “state of the art” in the voluminous literature comprising empirical studies of government formation is to use the government formation opportunity as the unit of analysis and employ a CL to model the choice of a single government out of the set of all potential governments. Although the adoption of the CL model represents a significant advance in the study of government choice, it is not an ideal empirical strategy because it cannot deal with unobserved heterogeneity and it makes the implausible assumption of the IIA.

The MXL model that we present in this paper represents both a substantive and a methodological advance over the CL model and is something that should appeal to a wide variety of comparative politics scholars. Qualitative scholars often question the utility of constructing general theories of government formation and testing them with cross-national statistical models because they believe that these theories and tests inevitably omit idiosyncratic case-specific factors that can have an important impact on the government formation process. Although quantitative scholars generally recognize that these types of case-specific factors can be important for things like government choice, they have essentially ignored them in their analyses under the presumption that doing so does not affect their results or that there are no empirical strategies to handle them. As we have demonstrated, though, both of these presumptions are wrong.

The MXL provides valuable substantive information about how unobserved contextual factors introduce variation into the effects of our observed variables that is simply not available from more traditional estimation strategies. In some sense, the MXL narrows the gap between qualitative and quantitative scholars by providing a way for political scientists to test cross-national theories of government choice while also taking account of case-specific and idiosyncratic contextual factors.

Methodologically, ignoring unobserved heterogeneity in discrete choice settings such as that of government formation results in inconsistent estimates and incorrect substitution patterns that leave counterfactual analyses and substantive inferences suspect. As we have indicated, the IIA testing procedure currently employed in the government formation literature to justify the use of a CL model is inappropriate and biased against finding IIA violations. Our own IIA tests, although extremely conservative, reveal clear evidence of unobserved heterogeneity and IIA violations, thereby indicating that the CL model is almost certainly an inappropriate empirical strategy in the government formation context. Furthermore, we have demonstrated that the substantive inferences drawn from MXLs can differ from those one would draw from a CL in several important ways.

We have also proposed several strategies for interpreting the substantive influence of variables in models of government choice. To date, most empirical studies of government choice have reported the sign and statistical significance of their model coefficients, without interpreting their substantive importance. As we have demonstrated, though, extracting and presenting quantities of substantive interest from these models is not overly difficult and can add greatly to our understanding of the government formation process.

Finally, we recognize that the methodological techniques that we have discussed here are somewhat more difficult to employ than the more familiar CL, with the primary drawback coming from increased estimation time and effort in avoiding problems such as local maxima. However, recent advances in maximum simulated likelihood and the introduction of easy-to-use code place all the techniques we describe well within the grasp of the applied researcher. Even in cases where the substantive differences between the MXL and CL turn out to be slight, the relatively low additional cost of the MXL means that there is little reason to continue using a model known to have inferior substantive and methodological properties—and as we have seen here, the substantive differences between MXL and CL models can be quite substantial. The techniques that we have presented for improving the substantive interpretation of model coefficients are also straightforward.

In sum, the new empirical strategies to model parliamentary government formation developed here offer substantive and methodological advantages over previous approaches and are not much more difficult than the current “state of the art” approach. It is our hope, therefore, that these strategies will become part of the “empirical toolkit” for researchers in this area and will help to continue the rapid advance in our empirical understanding of the government formation process that we have observed over the last decade.

## References

- 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* 39:714–44.
- . 1998a. Economics, entitlements, and social issues: Voter choice in the 1996 presidential election. *American Journal of Political Science* 42:1349–63.

- . 1998b. When politics and models collide: Estimating models of multiparty elections. *American Journal of Political Science* 42:55–96.
- Bäck, Hanna. 2003. Explaining and predicting coalition outcomes: Conclusions from studying data on local conditions. *European Journal of Political Research* 42:441–72.
- Bäck, Hanna, and Patrick Dumont. 2008. Making the first move: A two-stage analysis of the role of formateurs in parliamentary government formation. *Public Choice* 135:353–73.
- Baron, David P., and John Ferejohn. 1989. Bargaining in legislatures. *American Political Science Review* 83:1181–206.
- Beck, Nathaniel, and Jonathan N. Katz. 2007. Random coefficient models for time-series-cross-section data: Monte Carlo experiments. *Political Analysis* 15:182–95.
- Berry, William, Matt Golder, and Daniel Milton. Forthcoming. Improving tests of theories positing interaction. *Journal of Politics*.
- Brambor, Thomas, William Clark, and Matt Golder. 2006. Understanding interaction models: Improving empirical analyses. *Political Analysis* 14:63–82.
- Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tanenbaum. 2001. *Mapping policy preferences: Estimates for parties, electors, and governments 1945–1998*. New York: Oxford University Press.
- Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*. New York: Cambridge University Press.
- Cheng, Simon, and J. Scott Long. 2007. Testing for IIA in the multinomial logit model. *Sociological Methods Research* 35:583–600.
- Däubler, Thomas, and Marc Debus. 2009. Government formation and policy formulation in the German states. *Regional and Federal Studies* 19:73–95.
- Diermeier, Daniel, and Antonio Merlo. 2004. An empirical investigation of coalitional bargaining procedures. *Journal of Public Economics* 88:783–97.
- Dow, Jay, and James W. Endersby. 2004. Multinomial probit and multinomial logit: A comparison of choice models for voting research. *Electoral Studies* 23:107–22.
- Druckman, James N., Lanny W. Martin, and Michael F. Thies. 2005. Influence without confidence: Upper chambers and government formation. *Legislative Studies Quarterly* 30:529–48.
- Glasgow, Garrett. 2001. Mixed logit models for multiparty elections. *Political Analysis* 9:116–36.
- Glasgow, Garrett, Matt Golder, and Sona N. Golder. 2011. Who ‘wins’? Determining the party of the prime minister. *American Journal of Political Science* 55:937–54.
- Glasgow, Garrett, and R. Michael Alvarez. 2005. Voting behavior and the electoral context of government formation. *Electoral Studies* 24:245–64.
- Golder, Matt, Sona N. Golder, and David A. Siegel. Forthcoming 2012. Modeling the institutional foundations of parliamentary government formation. *Journal of Politics*.
- Golder, Sona N. 2006. *The logic of pre-electoral coalition formation*. Columbus, OH: Ohio State University Press.
- Haaijer, M., M. Wedel, M. Vriens, and T. Wansbeek. 1998. Utility covariances and context effects in conjoint MNP models. *Marketing Science* 17:236–52.
- Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46:1251–72.
- Hausman, Jerry A., and D. Wise. 1978. A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica* 48:403–29.
- Hausman, Jerry A., and Daniel McFadden. 1984. A specification test for the multinomial logit model. *Econometrica* 52:1219–40.
- Hensher, David A., and William H. Greene. 2003. The mixed logit model: The state of practice. *Transportation* 30:133–76.
- Hole, Arne Risa. 2007. Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal* 7:388–401.
- Huber, Joel, and Kenneth Train. 2001. On the similarity of classical and Bayesian estimates of individual mean Partworths. *Marketing Letters* 12:259–69.
- Indriðason, Indriði. 2005. A theory of coalitions and clientelism: Coalition politics in Iceland, 1945–2000. *European Journal of Political Research* 44:439–64.
- . 2008. Does terrorism influence domestic politics? Coalition formation and terrorist incidents. *Journal of Peace Research* 45:241–60.
- Irwin, Galen A., and Joop J. M. Van Holsteyn. 2004. The 2002 and 2003 parliamentary elections in the Netherlands. *Electoral Studies* 23:551–7.
- Kam, Cindy D., and Robert J. Franzese, Jr. 2007. *Modeling and interpreting interactive hypotheses in regression analysis*. Ann Arbor, MI: University of Michigan Press.
- Kang, Shin-Goo. 2009. The influence of presidential heads of state on government formation in European democracies: Empirical evidence. *European Journal of Political Research* 48:543–72.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science* 44:341–55.
- Laver, Michael. 1998. Models of government formation. *Annual Review of Political Science* 1:1–25.
- Laver, Michael, and Kenneth A. Shepsle. 1996. *Making and breaking governments: Cabinets and legislatures in parliamentary democracies*. New York: Cambridge University Press.
- Laver, Michael, and Norman Schofield. 1990. *Multiparty government*. New York: Oxford University Press.
- Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables*. London: Sage Publications.
- Martin, Lanny W., and Randolph T. Stevenson. 2001. Government formation in parliamentary democracies. *American Journal of Political Science* 45:33–50.
- . 2010. Incumbency, context, and government formation in multiparty parliamentary democracies. *American Political Science Review* 104:503–18.

- McFadden, Daniel. 1974a. Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, ed. P. Zarembka, 105–42. New York: Academic Press.
- . 1974b. The measurement of urban travel demand. *Journal of Public Economics* 3:303–28.
- McFadden, Daniel, and Kenneth Train. 2000. Mixed MNL models of discrete response. *Journal of Applied Econometrics* 15:447–70.
- Müller, Wolfgang C., and Kaare Strøm, eds. 1999. *Policy, office, or votes? How political parties in Western Europe make hard decisions*. New York: Cambridge University Press.
- Pilet, Jean-Benoit, and Emilie van Haute. 2008. The federal elections in Belgium, June 2007. *Electoral Studies* 27:547–50.
- Pridham, Geoffrey, ed. 1986. *Coalitional behavior in theory and practice: An inductive model for Western Europe*. New York: Cambridge University Press.
- Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. 1999. Voter choice in multi-party democracies: A test of competing theories and models. *American Journal of Political Science* 43:1231–47.
- Revelt, D., and Kenneth Train. 1998. Mixed logit with repeated choices. *Review of Economics and Statistics* 80:647–57.
- Schofield, Norman, Andrew D. Martin, Kevin M. Quinn, and Andrew B. Whitford. 1998. Multiparty competition in the Netherlands and Germany: A model based on multinomial probit. *Public Choice* 97:257–93.
- Skjaeveland, Asbjørn, Søren Serritzlew, and Jens Blom-Hansen. 2007. Theories of coalition formation: An empirical test using data from Danish local government. *European Journal of Political Research* 46:721–45.
- Strøm, Kaare. 1990. *Minority governments and majority rule*. Cambridge: Cambridge University Press.
- . 2000. Delegation and accountability in parliamentary democracies. *European Journal of Political Research* 37:261–89.
- . 2003. Parliamentary democracy and delegation. In *Delegation and accountability in parliamentary democracies*, eds. Kaare Strøm, Wolfgang C. Müller, and Torbjörn Bergman, 55–106. New York: Oxford University Press.
- Train, Kenneth E. 1998. Recreation demand models with taste differences over people. *Land Economics* 74:230–9.
- . 2009. *Discrete choice methods with simulation*. 2nd ed. New York: Cambridge University Press.
- van Mierlo, Hans J. G. A. 1981. The 1981 Netherlands election. *West European Politics* 4:297–301.
- von Beyme, Klaus. 1985. *Political parties in Western democracies*. London: Gower.
- Walker, Joan L., Moshe Ben-Akiva, and Denis Bolduc. 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. *Journal of Applied Econometrics* 22:1095–125.
- Warwick, Paul V. 2005. Do policy horizons structure the formation of parliamentary governments? The evidence from an expert survey. *American Journal of Political Science* 49:373–87.
- Western, Bruce. 1998. Causal heterogeneity in comparative research: A Bayesian hierarchical modelling approach. *American Journal of Political Science* 42:1233–59.
- Zorn, Christopher. 2005. A solution to separation in binary response models. *Political Analysis* 13:157–70.