Response to Martin and Vanberg: Evaluating a Stochastic Model of Government Formation

Matt Golder   Pennsylvania State University
Sona N. Golder   Pennsylvania State University
David A. Siegel   Duke University

In a 2012 Journal of Politics *article, we presented a zero-intelligence model of government formation. Our intent was to provide a "null" model of government formation, a baseline upon which other models could build. We made two claims regarding aggregate government formation outcomes: first, that our model produces aggregate results on the distributions of government types, cabinet portfolios, and bargaining delays in government formation that compare favorably to those in the real world; and second, that these aggregate distributions vary in theoretically intuitive ways as the model parameters change. In this issue, Martin and Vanberg (MV) criticize our model on theoretical and empirical grounds. Here we not only show how MV's evaluation of our model is flawed, but we also illustrate, using an analogy to common statistical practice, how one might properly attempt to falsify stochastic models such as ours at both the individual and the aggregate level.*

In a recent article in the *Journal of Politics*, we presented a "zero-intelligence" model of government formation (Golder, Golder, and Siegel 2012; henceforth, we refer to this model as the GGS model).[1] Our intent was to provide a "null" model of government formation, a baseline upon which other models could build (428). As such, our non-restrictive institutional assumptions were only two: (1) an incumbent government always exists, and (2) all governments must enjoy the support of a legislative majority. These assumptions are non-restrictive in that they correspond to the two constitutional constraints that bind in any parliamentary government formation process (430). The GGS model is stochastic, as it assumes that government proposals are made and chosen randomly, subject to our two institutional constraints (433). In our article, we used the GGS model to assess government formation in the aggregate. Our claim in this regard was twofold (443). First, our model produces aggregate results on the distributions of government types, cabinet portfolios, and bargaining delays in government formation that compare favorably to those in the real world. And second, these aggregate distributions vary in theoretically intuitive ways as the model parameters change.

Martin and Vanberg (2014, 874) make two criticisms of our model. Theoretically, they claim that our behavioral assumption of random choice is implausible and drives key results. And empirically, they argue that our assessment of our model's performance relies on data poorly designed to test coalition theories. These criticisms are not unexpected given modeling norms in political science. Stochastic models are relatively rare in the discipline, and many scholars are unfamiliar with the appropriate way to assess them. As we demonstrate, much of MV's criticism can be traced to a misunderstanding of how these models should be evaluated. We take this opportunity not only to show how MV's evaluation of our model is flawed but also to illustrate, using an analogy to common statistical practice, how one might properly attempt to falsify stochastic models such as ours at both the individual and the aggregate level.

---

[1]The data and computer code necessary to reproduce the numerical results in this article will be made available at http://homepages.nyu.edu/mrg217/public/ on publication.

Why Assume Random Government Proposals?

Martin and Vanberg do not take issue with our two institutional assumptions. Instead, they focus on the implausibility of our behavioral assumption that government proposals are made and (in the case that more than one is acceptable) selected at random. On one level, MV are correct—our assumption of random choice *is* implausible. We highlighted this ourselves when we wrote that our "model makes no pretense to accurately describe or explain the bargaining behavior of actors involved in the government formation process....We hold no illusions... that party leaders really gather together and simultaneously announce randomly chosen coalitions" (429). MV claim that by abstracting away from the bargaining process in this way we are launching a broad-side "attack" on existing coalition theory (2014, 873, 878). However, this is not true. We explicitly stated that "we [did] not see our model as a direct challenge to existing models that rely on utility maximization and strategic behavior.... [These models have] contributed greatly to our knowledge and [do] particularly well in highlighting the strategic interactions that are possible between parties" (428, 429). Our model is, in fact, complementary to existing approaches. We abstracted away from the bargaining process for the simple reason that examining the strategic aspects of government formation was not our goal.

So, what was our goal? And if the assumption of random choice is implausible, why make it? Martin and Vanberg (2014, 873–878) give the impression that we are alone in thinking that it might be useful to reconsider the theoretical basis of government formation from the ground up. We were, in fact, responding to a growing number of empirical scholars troubled by the gap between theory and empirics in the government formation literature (428). While game-theoretic models of government formation offer impressive insights into the strategic interactions between parties, they often fail to produce aggregate outcomes in line with empirical observation (431–32).[2] As Warwick and Druckman state with respect to portfolio allocation, for example, we "have an empirical relationship impressive enough to warrant the status of a scientific law [Gamson's Law]...in apparent conflict with a class

of bargaining models that includes the most prominent formal model of legislative bargaining" (2006, 639). It is this type of disconnect that led Laver, de Marchi, and Mutlu to conclude that the topic of government formation "must be reconsidered from the bottom up" (2011, 288). Our stated goal (428) was to take up this challenge by providing a null model of government formation.

Our assumption of random proposals is to be understood in light of this goal (433). The only constitutional constraints in a parliamentary government formation process are that an incumbent government always exists and that a government must enjoy the support of a legislative majority. We make no institutional assumptions other than these. As we indicated, though, "*any* model that requires the production of proposals is necessarily equivalent to a model that delineates some distribution...over possible proposals" (433).[3] Given the necessity of assuming *some* distribution and our desire to create a baseline model, we chose to assume a uniform distribution for proposals. While MV take issue with this choice, they never state what distribution they would prefer us to use. Our assumption of a uniform distribution is similar to the assumption of uninformative priors in Bayesian statistical analysis (433). It also has the same spirit as other simplifying assumptions in game theory such as the use of uniform distributions to represent uncertainty.

As with these similar assumptions, the implausibility of our random proposal mechanism does limit the degree to which the GGS model can describe the particular process underlying any single government formation. However, it does not prohibit the model from producing useful insights.[4] Many scholars have found the assumption of a random proposal mechanism useful (429). While we situated our model primarily in the large and growing literature on zero-intelligence agents in economics, computer science, and other fields (Gode and Sunder 1993), we also pointed to Compte and Jehiel (2004, 2010), who use a random proposal mechanism to examine the effect of institutions on outcomes in the specific context of

---

[2]Our focus on the gap between the predictions of game-theoretic models of government formation and the empirical world is thus closer to the "narrow reading" MV offer of our article (2014, 873).

[3]Martin and Vanberg (2014, 875–876) claim that we mislead readers into thinking that our model only makes minimalist institutional assumptions, but our cited statement here clearly contradicts their claim.

[4]The earth is not a sphere, but it is useful in many cases to assume that it is, much as it can be useful to assume that individuals act as rational maximizers of utility despite experimental evidence to the contrary. The need to assume a mechanism for making government proposals in no way specifies the type of assumption that should be made. Random selection is viewed in many fields as minimal. Adding it to the political science toolkit is likely to be fruitful.

multiparty bargaining—the same context that characterizes the government formation process. A key factor indicating whether a model with an implausible assumption can be useful is the degree to which its results depend on that assumption. Martin and Vanberg claim that the assumption of random proposals "drives [our] key results" (2014, 874). As we now demonstrate, this is incorrect. While random choice is a factor, it is far from the sole factor.

Assessing our Stochastic Model

Martin and Vanberg assess our model in two ways. First, they compare its accuracy in predicting individual government types to that of an empirical model by Martin and Stevenson (2010; henceforth MS). Second, they examine the extent to which the results of our model depend on our random proposal mechanism.

Predicting Individual Government Types

Martin and Vanberg's first way of assessing our model is both misplaced and flawed, exhibiting a fundamental misunderstanding of stochastic models. It is misplaced because the "goal of our model [was] simply to discover whether the two constitutional constraints that exist in all parliamentary systems are sufficient to generate empirically realistic predictions about the types of governments that form, portfolio allocation, and delays to government formation, independent of context-specific factors, norms of behavior, or bargaining protocols" (429). Maximizing the predictive accuracy of our model was never a goal. We explicitly recognized that these unmodeled factors, many of which have been identified in the empirical literature to which we ourselves have contributed (Golder, Golder, and Glasgow 2011, 2012; Golder 2010; Golder and Glasgow, n.d.; Golder and Thomas 2014), play "a significant role...in predicting government outcomes" (429). Martin and Stevenson's empirical analysis includes many of these unmodeled factors, and, as a result, it would be unsurprising if it outperformed ours in predicting government types. Recall that our motivation for developing the GGS model had to do with the gap between the predictions of game-theoretic models and the real world. Importantly, the MS analysis is not based on a unified theoretical model, game-theoretic or otherwise. Instead, it represents a survey of the disparate claims about government formation

made by scholars over the past 50 years.[5] Notably, MV never present evidence that an existing game-theoretic model is able to outperform ours when it comes to predicting real-world distributions on government types, bargaining delays, and portfolio allocation.

Not only is MV's approach to assessing our model misplaced, but it is also flawed in its implementation. Standard game-theoretic models produce point predictions on behavior, and scholars are typically interested in these equilibrium outcomes or the manner in which they vary with model parameters. While one can also derive comparative statics from computational models (Bendor et al. 2011; Siegel 2009), stochastic models like ours generally do not produce point predictions. Instead, they produce *distributions* over potential outcomes. Each iteration of a stochastic model produces a single realization of a distribution. Multiple simulations map out the distribution, and summary statistics can be drawn from this distribution. This procedure is no different from Markov Chain Monte Carlo simulations commonly used in statistics. Populating a posterior distribution or its equivalent is the primary goal of this type of simulation analysis.

If one wants to understand the behavior of a system such as the government formation process, one must simulate the system enough times for each set of parameter values to produce a reasonable approximation of the distribution of outcomes. In other words, if one wants to explain what one expects to happen in a *particular* government formation process, one should simulate that particular process—with its associated constant set of parameter values—many times, until the distribution of outcomes arising from it can be discerned. Importantly, a single draw from this distribution is typically meaningless. Focusing on a single draw would be like rolling a die and claiming that the number shown perfectly captures all of the numbers that could be shown on other rolls of the die. This is clearly problematic. However, this is precisely what MV do when evaluating the predictive accuracy of our model—they take a single draw from the distribution of government types and compare this against the real-world government that formed in the equivalent party system. In effect, they make the reverse of an ecological inference error—they attempt to derive

---

[5]The 2010 MS analysis is based on their 2001 analysis, which, as they note, was designed to evaluate "a good sampling of the received wisdom on the systematic determinants of government formation" (2001, 38). The main difference with the 2010 analysis, other than sample size, is that it includes 70% more covariates.

information about the distribution of government types from a single point on the distribution.

In contrast, we performed no analysis at the level of *individual* governments or made any related claims. As a result, we could not, by definition, have engaged in any of the ecological fallacies that MV repeatedly ascribe to us (2014, 874, 876, 879). This does not mean that it is not possible to evaluate individual-level claims from our model. To falsify our model at the individual level, one would have to construct a posterior distribution for a given set of parameter values and compare the real-world outcome to this distribution. If the real-world value were sufficiently far from the mean of the posterior distribution, we might conclude that it is unlikely to have come from the posterior distribution. This practice, central to statistical inference, is how one must assess stochastic models such as ours at the individual level. We did not follow this particular path because such a test is unlikely to be informative here due to the fact that our variance, as we acknowledge, is likely to be large given our minimal assumptions (429). Unlike us, MV *do* attempt to make individual-level claims (2014, 878). That they fail to produce distributions over outcomes, though, means that their claims are effectively meaningless. Note that we are not criticizing the method by which MV derive their results. Other than their use of 317 real-world party systems rather than our 50,000 random systems (weighted by their likelihood to appear in the real world), they use the same procedure for generating output as we do. Rather, our critique is of the inferences they attempt to draw from their analysis at the individual level.[6]

The Random Proposal Mechanism Does *Not* Drive Our Results

Although we made no individual-level inferences about government formation, we did make aggregate-level ones. Martin and Vanberg state that we should have not only compared the aggregate results from our model to those from the real world as we did, but that we should also have compared them to those obtained from a purely random model that does not impose our two institutional assumptions. They claim (2014, 874, 877, 878) that there is little difference between the purely random model and the GGS model and, hence, that our results are driven by our random proposal mechanism. As we now demonstrate, this claim is false.

In the first three columns of Table 1, we compare real-world data on government type from Western Europe to the weighted simulation results from (1) the GGS model and (2) a purely random model where no account is taken of whether proposed governments have majority support or are preferred to the status quo government.[7] As we can see, the GGS results are not only similar in many respects to the real-world data, but they are also substantially different from the purely random results in all but the case of minority coalitions. In all cases, the GGS results are closer to the real-world data than the purely random results.

In our original article, we recognized that this similarity to real-world data "in and of itself, is not a sufficient test of our model's predictive power" (436).[8] This is why we presented additional evidence in support of our model. Before turning to this additional evidence, consider the last two columns of Table 1. These columns present results from models where we do not weight our party systems to better reflect their likelihood of occurring in the

---

[6]This brings up a subtle point. The method that MV and we employ to examine *aggregate* behavior involves drawing a single realization from the distribution of outcomes for each set of parameter values. Why is drawing a single realization appropriate in this setting? Here we rely on the fact that some draws from outcome distributions produce values higher than the mean, and some lower. With many draws over many distributions, these differences average out. We verified this by drawing multiple instances of 50,000 simulated party systems and obtaining nearly identical results. It is not clear that the 317 party systems MV use are sufficient for this purpose. A comparison of two sets of 317 random party systems yields insufficient consilience to inspire confidence. An additional advantage of our using so many simulated party systems is that we are able to produce more general inferences—we can detail how outcomes vary with parameters related to the full range of variation in possible party systems. MV's focus on just 317 party systems makes this difficult. This is a minor point, but it does explain why some of our results differ from theirs.

[7]The values from the GGS model differ ever so slightly from our original analysis because we replicated all earlier results, and some variation remains even after averaging over 50,000 simulation runs. Our results for the purely random model differ from MV's for three reasons. First, our model is based on 50,000 rather than just 317 party systems. Our random party systems exhibit more heterogeneity than MV's, resulting, all else equal, in more majority and surplus majority governments than they get. Second, we include single-party majority governments as we believe these governments can also be explained by our model. Third, we don't include measures of statistical significance because they are not interesting with 50,000 data points.

[8]This directly contradicts MV's claim that the "key piece of evidence [we] adduce in support of [our] model [is that] the aggregate distribution of government types in postwar Europe corresponds reasonably closely to the aggregate distribution that emerges out of [our] simulations" (2014, 874).

TABLE 1   Types of Parliamentary Governments

Government Type	Real-World Data Western Europe	Weighted		Unweighted	
		GGS	Random	GGS	Random
Single-Party Majority	.130	.156	.006	.086	.006
Minimal Winning Coalition	.307	.255	.156	.195	.155
Minority Coalition	.118	.255	.295	.462	.297
Single-Party Minority	.229	.226	.059	.149	.058
Surplus Majority	.216	.108	.484	.109	.485

Note: The numbers indicate the proportion of each government type. The real-world data are for 17 countries. "GGS" refers to the GGS model, while "Random" refers to a purely random version of the GGS model that does not impose either of our two institutional constraints. Our "weighted" results are designed to reflect the types of party systems that we observe in the real world. For a discussion and justification of our weighting scheme, see Golder, Golder, and Siegel (2012, 435).

real world. As we might expect, the unweighted GGS results differ from the weighted GGS results. For example, minority coalitions are more frequent in the unweighted GGS analysis than in the weighted GGS analysis, reflecting the fact that party systems conducive to the emergence of minority coalitions (many effective parties, strong weight on policy relative to office, extremist parties) are less commonly observed in the real world than they are in our 50,000 random party systems. Note, however, that there is *no difference* at all between the weighted and unweighted results from the purely random model.

To understand the importance of this result, one must remember that the GGS model produces a distribution over outcomes for each party system. Weighting our aggregate data amounts to counting some of these distributions less and some more in the average. If each distribution varies according to our model parameters (434), then weighting will alter the value of the average. This is what we see with the GGS model. If each distribution does not vary with the model parameters, then weighting will have no effect. This is what we see with the purely random model. In other words, if MV were correct that our results were driven by random choice and not variation in model parameters, then we would expect our weighting scheme to have no effect. Since we see a large effect, we can infer that they are incorrect in their assessment.

We now briefly discuss the additional evidence in support of the GGS model that we provided in our original article, evidence that MV chose to ignore. In Figure 2 (437), we indicated not only that the distribution of government types produced by the GGS model is similar to that observed in the real world, but also that it varied in theoretically intuitive ways as we altered the weight that parties place on policy. For example, we showed that both types of

minority government become relatively more likely as parties place more weight on policy as opposed to office. We repeated this analysis for our purely random model and found, unsurprisingly, that the distribution of government types does not vary with policy weight.

In addition to examining government types, we looked at portfolio allocation and bargaining delays. Martin and Vanberg do not address these analyses even though they had equal weight in our article. Gamson's Law states that there should be a strong positive relationship between the percentage of seats a party contributes to the government's majority and its share of ministerial portfolios. The results from the GGS model exhibit such a relationship and are in line with those from empirical studies of portfolio allocation (438). Equivalent results from a purely random model are substantially weaker. Our analysis also indicated that a party's share of cabinet portfolios varied in an intuitive way with the parameters in the GGS model (438–41). Repeating our analysis with a purely random model reveals that this variation is, well, random. The results with respect to bargaining delays are the most striking of all. There is, of course, *no* delay with purely random choice, and so the purely random model can describe *none* of the variation in bargaining delays that the GGS model does. The GGS model, as we indicated, comes close to "replicating the distribution of delays observed in the real world" (441, Figure 3). And this does not even take into account our analysis of the manner in which bargaining delays vary in understandable ways with the parameters in the GGS model (441–43).

In sum, we believe there is substantial evidence indicating that the GGS model has interesting things to say about aggregate government formation behavior. Moreover, these interesting things are not driven entirely by the stochastic elements of

our model as MV claim, but rather in important and intuitive ways by our institutional assumptions and model parameters. While the GGS model does not say much about individual cases, the fact that outcomes vary in a predictable manner with model parameters in a way that is simply not true with a purely random model indicates that behavior at the individual level, if not sufficient for prediction (the standard MV use to assess it), is still responding appropriately to the incentives instantiated in the institutional assumptions of our model. As such, the GGS model provides a complementary method to game-theoretic models that has the benefit of being more closely tied to aggregate empirical analysis.

But what of MV's claim that aggregate data "are ultimately not the most relevant for assessing the empirical success of models of government formation" (2014, 877)? We find this claim unjustified. Even if an individual game-theoretic model is intended to apply only to a single government formation opportunity, logic dictates that it should apply to every such formation opportunity fitting its assumptions. To claim that one cannot aggregate the predictions of game-theoretic models in order to compare them to the sorts of aggregate data that empirical scholars believe are important implies that either (1) the substantive domain of these theories, as dictated by their assumptions, is insufficient to do so, or (2) there is an unmodeled interaction between government formation opportunities that makes simple aggregation impossible. While the latter would be an interesting conjecture, it is not one that MV—or anyone else to our knowledge—has made. The former, of course, if true, would provide additional support to the need for alternative modeling assumptions to game theory in order to widen the domain of applicability.

Conclusion

While we appreciate the engagement of MV, as well as their spur to compare the results of the GGS model to what would occur under a purely random model, we believe their attempt to defend the status quo is misguided. We are not attacking the utility of game-theoretic models in illustrating the incentives that lead to any particular individual government. On the contrary, we explicitly noted the great contribution of such models to our understanding. Rather, the intent of introducing our model was to tackle a different question: what factors affect the distribution of governments formed and other important aggregate properties of government formations? We chose to tackle this question in response to calls from empiricists who are measuring facets of this population of government formations and despairing of any appropriate theory that gets close to addressing their work. We claim explicitly that ours is only the first step in getting there, but we believe it is a step worth taking, and one complementary to existing analyses.

Acknowledgments

References

Bendor, Jonathan, Daniel Diermeier, David A. Siegel, and Michael Ting. 2011. *A Behavioral Theory of Elections*. Princeton, NJ: Princeton University Press.

Compte, Olivier, and Philippe Jehiel. 2004. "Bargaining Over Randomly Generated Offers: A New Perspective on Multi-Party Bargaining." Unpublished manuscript, C.N.R.S.

Compte, Olivier, and Philippe Jehiel. 2010. "Bargaining and Majority Rules: A Collective Search Perspective." *Journal of Political Economy* 118 (2): 189–221.

Gode, Dhananjay K., and Shyam Sunder. 1993. "Allocative Efficiency of Markets with Zero Intelligence Traders: Market as a Partial Substitute for Individual Rationality." *Journal of Political Economy* 101 (1): 119–37.

Golder, Matt, Sona Golder, and David A. Siegel. 2012. "Modeling the Institutional Foundations of Parliamentary Government Formation." *Journal of Politics* 74 (2): 427–45.

Golder, Matt, Sona Golder, and Garrett Glasgow. 2011. "Who 'Wins'? Determining the Party of the Prime Minister." *American Journal of Political Science* 55 (4): 937–54.

Golder, Matt, Sona Golder, and Garrett Glasgow. 2012. "New Empirical Strategies for the Study of Parliamentary Government Formation." *Political Analysis* 20 (2): 248–70.

Golder, Sona. 2010. "Bargaining Delays in the Government Formation Process." *Comparative Political Studies* 43 (1): 3–32.

Golder, Sona, and Garrett Glasgow. N. d. "A New Approach to the Study of Parties Entering Government." *British Journal of Political Science*. Forthcoming.

Golder, Sona, and Jacquelyn A. Thomas. 2014. "Portfolio Allocation and the Vote of No Confidence." *British Journal of Political Science* 44 (1): 29–39.

Laver, Michael, Scott de Marchi, and Hande Mutlu. 2011. "Negotiation in Legislatures over Government Formation." *Public Choice* 147 (3/4): 285–304.

Martin, Lanny W., and Georg Vanberg. 2014. "A Step in the Wrong Direction: An Appraisal of the Zero-Intelligence Model of Government Formation." *Journal of Politics* 76 (4): 873–879.

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

Martin, Lanny W., and Randolph T. Stevenson. 2010. "The Conditional Impact of Incumbency on Government Formation." *American Political Science Review* 104 (3): 503–18.

Siegel, David A. 2009. "Social Networks and Collective Action." *American Journal of Political Science* 53 (1): 122–38.

Warwick, Paul V., and James N. Druckman. 2006. "The Portfolio Allocation Paradox: An Investigation into the Nature of a Very Strong but Puzzling Relationship." *European Journal of Political Research* 45 (4): 635–65.

Matt Golder is an Associate Professor in the Department of Political Science at Pennsylvania State University, University Park, PA 16802.

Sona N. Golder is an Associate Professor in the Department of Political Science at Pennsylvania State University, University Park, PA 16802.

David A. Siegel is an Associate Professor in the Department of Political Science at Duke University, Durham, NC 27708.