Estimating Legislators' Preferred Points

John Londregan
University of California, Los Angeles

This paper shows that agnostic spatial models that simultaneously attempt to estimate legislators' preferred points and ideological locations for the proposals on which they vote, such as the well-known NOMINATE model of Poole and Rosenthal, are not identified. The problem arises because the agnostic estimators inherit the granularity of the voting data and, so, cannot recapture the underlying continuous parameter space. I propose an alternative estimator that achieves identification by modeling the agenda.

1 Introduction

The spatial model of legislative politics combines an analytically attractive theory with a set of powerful behavioral predictions. The model makes the assumption that each legislator votes for the alternative with an ideological location closest to the legislator's preferred outcome. When this assumption is warranted, an analyst need only know the ideological "locations" of legislative proposals, and the preferred ideological outcomes of legislators, to predict legislative votes with considerable confidence. These attractive features have made the spatial model a cornerstone of the empirical analysis of legislatures.

One aspect of the foundation of empirical spatial models requires especially close attention. An intuitively appealing approach to operationalizing the spatial model of voting in a legislature is to specify a utility function that evaluates outcomes according to their spatial locations and then simultaneously estimate legislators' preferred outcomes and the spatial "locations" of the bills on which they vote. Here I show that this seemingly sensible approach contains a deadly flaw. Essentially the granularity of the data, which consist of votes of "aye" or "nay," imposes an artificial granularity on the parameter space of possible preferred outcomes for legislators and possible locations for bills. This causes a lack of identification that can lead to severely misleading parameter estimates.

These problems have not passed unnoticed by users of scaling methods. Poole and Rosenthal (1991) use Monte Carlo techniques to evaluate the performance of their model. They conclude that their model does a good job of recovering information about legislator ideal points and "cut points" for bills when the number of voters and the number of bills is large, though it is less effective at recovering information about the precise location of the "aye" and "nay" locations for each bill.

Author's note: I am grateful for useful comments from Larry Bartels, Henry Brady, Gary King, Keith Poole, Howard Rosenthal, seminar participants at the 1996 Political Methodology Meetings in Ann Arbor, and from three anonymous referees.

Copyright 1999 by the Society for Political Methodology

The favorable Monte Carlo findings by Poole and Rosenthal notwithstanding, the results here establish that the identification problems persist for any finite number of voters, whether they be the 5 voters in a committee or the 435 voters on the floor of the U.S. House of Representatives. The Monte Carlo results suggest that as the size of the legislature grows the severity of the bias may diminish. However, whatever the reason for the favorable Monte Carlo findings when both the number of voters and the number of bills are large, the negative identification results here heighten the need for users of agnostic scaling methodologies to explain how, why, and when their methods yield acceptable approximations despite their lack of statistical consistency.

Even if large numbers can eventually be shown to render the theoretical biases small, one has to be very careful how one counts! In legislatures where party "whipping" occurs, such as the British Parliament, the number of independent decision makers on "whip" votes is not the number of members in the chamber, but the number of party leaders. It is not enough to count the number of legislators! In particular, we cannot use estimates from agnostic models to test whether "party leadership votes" in the U.S. Congress result from leaders exerting party discipline. If they do, estimated ideal points from agnostic models will be biased and, so, not very useful in testing the hypothesis.

This paper is intended as constructive criticism of the application of empirical spatial models to legislatures, and I go on to offer a solution that avoids the statistical barriers to estimation. This comes at the "expense" of thinking more carefully about the agenda on which legislators vote, which is something we ought to be doing anyway. The estimator recommended here can be applied even in small "legislatures," including legislative committees, judicial panels such as the U.S. Supreme Court, where justices take concurrent or dissenting opinions, and other decision-making bodies that take recorded votes, such as the U.S. International Trade Commission. Scholars interested in these settings should not interpret the flaws of agnostic spatial models as a reason to revert to using factor analytic procedures. Heckman and Snyder (1996) draw a close link between factor analysis and agnostic spatial models, showing that the former coincides exactly with a particular set of assumptions about legislator's preferences and about the random shocks that perturb their vote choices. Those who would avoid agnostic spatial models should, with as much reason, eschew factor analysis. The recommendation here is to move in the other direction, including the formation of the agenda as part of the model, whether this agenda comes in the form of a stream of bills in a legislature or a flow of cases to the Supreme Court.

The genesis of this paper was a research project on Chile's democratic transition. The committees of the Chilean Senate are required to take recorded "roll call" votes on all bills considered at the second reading, creating a gold mine of voting data.[1] It was while writing computer code to implement the NOMINATE model of Poole and Rosenthal in this setting that I discovered the inconsistency argument set forth in the second section of this paper. However, latitude and longitude confer no protection from the lack of identification for agnostic spatial models, especially when the number of independent decision makers is small, as it can be on party leadership votes. Students of the U.S. Congress are by no means immune from the problems identified here, regardless of their having been discovered on the other side of the tropics.

Section 2 of this paper highlights the connection between "self-scoring" educational tests and "agnostic" spatial models that attempt simultaneously to estimate both the ideological

[1] Those interested in seeing the methods developed here at work will find an application to the Chilean Senate given by Londregan (2000).

content of the agenda on which legislators vote and the ideological leanings of the legislators doing the voting. This close connection means that many results can carry over from one literature to the other.

In Section 3 the discussion turns to the biases that plague agnostic spatial models. The essence of the problem is that agnostic models attempt to estimate the locations of legislative proposals located along an ideological continuum from discrete voting data. The attempt to estimate so many parameters granularizes the space of parameters the model attempts to estimate. It is as if analysts were trying to paint a panoramic landscape on a piece of string. Worse, the lack of identification of the parameters describing the legislative agenda "metastasizes" to legislator's preference parameters, which are not identified.

Having outlined the reasons agnostic spatial models fail, in Section 4 I offer a feasible alternative. Instead of attempting to remain almost completely agnostic about the agenda on which the legislature votes, the approach advocated here is to pay something like as much attention to modeling the structure of the agenda as we do to legislators' decision of how to vote on it. Just as we have become sophisticated at modeling legislators' voting decisions, so too we can model the more important activity of making legislative proposals. This enterprise leads to a model that simultaneously encompasses legislators' voting and proposing decisions. Unlike agnostic models of the agenda, this model contains a fixed number of parameters, and it does not suffer the biases and failures of identification that plague its agnostic cousin. The resolution of the technical problems raised by the spatial model has the beneficial effect of forcing us to think more carefully about the agenda being voted and about legislators' motives for casting the votes they do. Section 5 concludes.

2 A Self-Scoring Ideology Measure

Political presure groups take an intense interest in members' voting records, and in the United States these groups often compute summaries of the voting records of members of Congress, marking off points for "incorrect votes" much as teachers might grade a true–false exam and calculating a summary score reporting the percentage of "correct" votes cast by the legislator. This similarity is more than coincidental; in both cases evaluators are attempting to measure a latent characteristic. This section begins with a discussion of a family of educational testing models that is very closely connected with the spatial model and then moves on to a parallel development of "agnostic" spatial models. Both bodies of literature agree that while the agnostic spatial models and their psychometric counterparts are potentially biased, these biases are not important when the number of voters (test takers) and the number of proposals voted (test questions) are large. However both raise warning flags about data sets in which the number of either voters or issues is small.

2.1 *Self-Scoring Educational Tests*

The use of statistical analysis to measure ability and other latent traits has a long history (Thurstone 1925), but a useful point of departure is the Rasch (1961) model. This model emerges as the response to the question, "When is the widespread practice of summarizing a student's performance on a test using the fraction of correct answers justified?" that is, "When is the fraction of correct answers a sufficient statistic for the latent characteristic of ability or subject mastery the test seeks to measure?"

The answer to this question is surprisingly straightforward (Fischer 1995). The fraction of correct answers is a sufficient statistic for latent ability[2] if and only if the probability that the subject v, gives a correct answer to question p, ϕ_{vp} is

$$\phi_{vp} = \frac{e^{\gamma\theta_v - \beta_p}}{1 + e^{\gamma\theta_v - \beta_p}} \tag{1}$$

Many readers will recognize this as the logit probability. It is useful to rewrite this condition in terms of the discriminant function. Subject v will give a correct answer to item p provided that

$$\xi_{vp} < \gamma\theta_v - \beta_p \tag{2}$$

where ξ_{vp} is a random error term that obeys the so-called "extreme value" distribution that gives rise to logit probabilities (Maddala 1986). In this context, ξ_{vp} corresponds to idiosyncratic factors, such as momentary distractions during the test. The parameter γ is an arbitrary positive constant, and by convention it is set equal to 1, an assumption that is relaxed below. This assumption of a constant and positive value for the "item discrimination parameter," γ, means that we know the correct answers to all of the questions and that all of the questions are equally ambiguous.

The parameter β_p measures the difficulty of the question, while θ_v measures the test taker's ability or subject mastery. Lower values of β_p and higher values for θ_v correspond to a higher probability of a correct answer.

We could easily reparameterize this model in terms of the number of times the subject answered "true" on a test with binary options, e.g., "true" and "false." In this case instead of using the standard normalization of $\gamma = 1$ for all questions, γ would vary among questions, with $\gamma_p = 1$ for questions for which the correct answer was "true" and $\gamma_p = -1$ for questions whose correct answer was "false". We could then replace the difficulty parameter with a new parameter, β_p^*, equal to β_p for questions whose correct answer was "true" but equal to $-\beta_p$ for questions whose correct answer was "false." This means that examinee v will give an answer of "true" to question p when

$$\xi_{vp} < \gamma_p\theta_v - \beta_p^* \tag{3}$$

This requires that the distribution of ξ_{vp} is symmetrical about 0.

An extension of the basic Rasch model takes this approach a step further. Instead of imposing the value of γ_p a priori, the "two-parameter logit model" treats this as an additional parameter to be estimated (Birnbaum 1968), so that we have the formulation in inequality (3) but without any restriction on the values that γ_p can assume. If we knew the examinees' ability parameters, the θ_v, then estimating the γ_p and β_p parameters would be straightforward. The probability that examinee v with high ability answers "true" to question p will exceed the probability of an answer of "true" from examinee v' with lower

[2] A few technical conditions are necessary: (i) the random variable ξ_{vp} in inequality (2) can take on any value on the real line, albeit with a very low probability; (ii) there is "no guessing," so that at very high levels of mastery the probability that the subject answers correctly converges to 1, while at abysmal levels it converges to 0; and (iii) random disturbances to the subject's response, caused by factors such as momentary distractions during the test, are uncorrelated across questions. Of these three assumptions, the second is the most problematic: we might expect low-ability subjects to guess.

ability if the item discrimination parameter γ_p is positive, so that the correct answer is "true," while if the correct answer is "false," we will have $\gamma_p < 0$, and the high-ability examinee will be less likely to answer "true." If we knew our test takers' ability levels, the θ_v's, we could construct an estimate, $\hat{\gamma}_p$, of the item discrimination parameter. Not only would this tell us the correct answer,[3] something an educational tester probably already knows, but more importantly it will alert us to ambiguous questions by returning a value of $\hat{\gamma}_p$ near 0. In tandem with our estimate of γ_p the estimated value for β_p^* tells us the difficulty of the question, with higher values for $\hat{\beta}_p$ corresponding to more difficult questions when the correct answer is "true" and to easier questions when the appropriate answer is "false."

Of course educators are seldom, if ever, in the position of having absolute knowledge of examinees' abilities. Instead testers often attempt to estimate simultaneously the examinees' ability parameters and the question parameters. This precipitates analysts into what is called the "parameter proliferation problem." As the researcher adds data, for example, by giving the exam to more subjects, she also adds more parameters to estimate (the ability parameters of the newly added test takers). Even if one is willing to bend the definition of consistency somewhat, the conditions needed to obtain consistent estimates in this case are somewhat fanciful (Haberman 1977), namely, that both the the number of examinees and the number of test questions are infinite, as must be the ratio of the number of questions to the number of examinees. The idea behind this is that we want to have so many questions that we can, essentially, estimate the ability of each examinee using a separate, and arbitrarily large, set of questions.

This opens up vexing questions about when N is really large. In a time-series context researchers routinely estimate annual models based on a few dozen years' of data, despite the small sample biases to which such estimates are prey (Marriott and Pope 1954). How severe are the biases that beset psychometric models that act as though all of the relevant N 's identified by Haberman are large? Lord (1983) addresses this question in the context of the so-called "three-parameter model," a further refinement of the two-parameter logit model that allows for guessing by examinees. For the three-parameter model Lord calculates bias correction formulas for item parameters on the Verbal SAT, where there are 96 questions and several million examinees. However, the estimated biases are inversely proportional to the number of questions. When the number of questions becomes small his formulas indicate very large biases.

2.2 *"Agnostic" Spatial Models*

Now let us consider the spatial model of voting, which posits that legislators voting on a proposal choose between alternatives associated with affirmative and negative votes, each of which has an ideological "location." This location might be a point on a continuum from the political "left" to the political "right," or it might require a more complicated representation (if legislators' preferences for income redistribution differed along a left-to-right continuum, while simultaneously their views on social issues such as abortion differed along a "religious" to "secular" continuum). According to the spatial model, legislators have "most-preferred" policy locations and vote in favor of the alternative closest to their most preferred location. The spatial model is most tractable when locations all lie along a single left to right continuum, and it is on this case that I focus here.

[3] If $\hat{\gamma}_p > 0$, we can infer that the correct answer is "true." If $\hat{\gamma}_p < 0$, then our estimate indicates that the correct answer is "false."

The basic ingredients of the model are positions of the proposal and status quo alternatives, and of legislators' most preferred outcomes. Following Stokes (1963), we can add an additional "valence" element of policy, call it q, about which everyone agrees, something like greater efficiency, lower cost, or more lives saved. Both proposals and the status quo have some level of this valence issue. We get the "plain vanilla" spatial model as a special case of this more general model by setting the valence for the status quo and all proposals equal. For proposal p let z_p denote the location of the proposal while $z_{p,sq}$ is the location for the status quo outcome expected to result if proposal p is defeated. Likewise, let q_p and $q_{p,sq}$ denote the valence for the proposal and the corresponding status quo. Note that in a data set with many votes the expected "reversion" to status quo may differ among proposals. This might occur because the status quo changes over the course of a legislative session as new laws are passed and old ones are repealed. If proposal z_p is adopted, then a legislator with a preferred outcome of x_v would enjoy a utility of

$$U(z_p, q_p \mid x_v) + \psi_{vp} \tag{4}$$

Here ψ_{vp} is an idiosyncratic "shock" to preferences specific to a particular legislator, indexed by v and a particular alternative p. This would be sensible if the spatial model captured all of the considerations that entered legislators' voting decisions save for the most idiosyncratic and personal ones, such as whether the legislator has some preexisting expertise in the area, and then only if the author does not "call in" all outstanding favors at the same time! The utility function U has the property that $U(z_p, q_p \mid x_v)$ rises as q_p rises and as $|z_p - x_v|$ falls.

Analysts typically remain agnostic about the locations of the proposal z_p and status quo $z_{p,sq}$, and also about legislators' most preferred outcomes, the x_v, much as educational testers attempt to calibrate the item parameters and test takers' ability levels simultaneously. Many analysts assume that $q_p = q_{p,sq}$ for every proposal, though we shall see that this has less practical effect than one might think. In the context of legislative voting, the basic strategy is to derive the link between the location parameters and legislators' votes of "aye" or "nay" and then to "read backward," using the observed votes to infer locations for proposals and legislators' most preferred outcomes.

Following this development, the model tells us that legislator v will vote in favor of alternative z_p if

$$U(z_p, q_p \mid x_v) + \psi_{vp} > U(z_{p,sq}, q_p \mid x_v) + \psi_{vp,sq} \tag{5}$$

Letting $\epsilon_{vp} = \psi_{vp,sq} - \psi_{vp}$ denote the net "idiosyncratic shock" in favor of the status quo, the probability that legislator v favors proposal p equals the probability that

$$\epsilon_{vp} < U(z_p, q_p \mid x_v) - U(z_{p,sq}, q_{p,sq} \mid x_v) \tag{6}$$

To operationalize the model one needs to specify a probability distribution for ϵ_{vp} and a functional form for U. Two utility functions have received special attention in the literature, the quadratic utility function and the Gaussian utility function. The quadratic utility function,

$$U(z, q \mid x_v) = -\frac{1}{2}(z - x_v)^2 + \alpha q \tag{7}$$

has the appealing property that more distant departures from the preferred outcome yield lower utility, while it leads to relatively tractable estimation formulas (Hinich and Ordeshook 1978; Brady 1991). It has has been used by several authors (Ladha 1991; Heckman and Snyder 1996) to analyze legislative voting data. The α term in Eq. (7) measures the weight placed on valence.

Incorporating the valence element of policy is less straightforward for the Gaussian utility function, and analysts using Gaussian utility typically ignore the valence aspect of policy (Poole and Rosenthal 1991) while they add an additonal parameter, β, that calibrates how well individuals' voting behavior is described by the model (Poole and Rosenthal 1991):

$$U(z \mid x_v, \beta) = \beta e^{-\frac{1}{8}(z-x_v)^2} \tag{8}$$

Instead of describing the "yes" and "no" alternatives in terms of their locations, z_p and $z_{p,sq}$, it is often convenient to represent the choice in terms of the midpoint,[4] m_p, and the gap, g_p, between the "yes" and the "no" alternatives associated with proposal p, $m_p = (z_p + z_{p,sq})/2$ and $g_p = z_p - z_{p,sq}$. This representation contains the same information as the initial locations, which can readily be recovered from m_p and g_p,

$$z_p = m_p + g_p/2 \qquad \text{and} \qquad z_{p,sq} = m_p - g_p/2 \tag{9}$$

One can substitute from Eq. (9) to reexpress the condition for a legislator with quadratic utility voting in favor of a proposal given in Eq. (7) in terms of the g_p and m_p parameters

$$\epsilon_{vp} < g_p(x_v - m_p) + \alpha(q_p - q_{p,sq}) \tag{10}$$

A similar condition can be derived using Gaussian utility.[5] Collecting terms in Eq. (10) we can rewrite the condition for voting in favor of proposal p as

$$\epsilon_{vp} < g_p x_v - b_p^* \tag{11}$$

This condition is identical to that given in (3), although the interpretation of the variables is quite different. The item discrimination parameter, γ_p, from the two-parameter logit model of educational tests is replaced here by g_p, the gap between the location of the proposal and that of the status quo. Similarly the ability parameter, θ_v, is replaced here by x_v, the legislator's most preferred policy outcome. Finally, the difficulty parameter, β_p^*, is replaced by b_p^*, which is an amalgam of the proposal midpoint, m_p, the gap between the proposal and the status quo alternatives, and the valence advantage, $\alpha(q_p - q_{p,sq})$, of the proposal over the status quo. Formally

$$b_p^* \equiv g_p m_p - \alpha(q_p - q_{p,sq}) \tag{12}$$

Eq. (12) tells us that we cannot separately identify the midpoint, m_p, and a proposal's "valence advantage," $\alpha(q_p - q_{p,sq})$. A proposal that moves policy moderately rightward

[4] In applications in which there are no valence considerations this is often referred to as the "cut point" because, absent valence considerations and preference shocks, all legislators with ideal points to the left of m_p will vote one way, while those to the right will vote the other.

[5] This is $\epsilon_{vp} < \beta(1 - e^{-\frac{g_p}{4}(x_v - m_p)})e^{-\frac{1}{8}(x_v - m_p - \frac{g_p}{2})^2}$.

may receive wide support because it has a very high valence or because the status quo against which it is being compared is far to the left, leading to a midpoint m_p that puts most voters on the same side as the proposal. Some analysts solve this problem by setting $q_p = q_{p,sq}$ for every proposal. This has the effect of placing all of the weight on the midpoint. While this is an analytically tractable solution, it leads to the implication that an unsuccessful proposal from a political moderate failed because it was ideologically extreme, rather than because of a low valence. In most legislatures bills sponsored by political moderates enjoy a success rate well below 100%. The hypothesis that this imperfect success rate is due to moderates occasionally making ideologically extreme proposals seems less plausible than the alternative that at least some of the variation is attributable to success or failure at crafting high valence proposals.

Similarly to the educational testing models, when both the number of voters and the number of proposals is large but finite, the parameter estimates exhibit little bias. Poole and Rosenthal find that for their model, which is based on Gaussian utility, and in legislatures with 100 or more voters, the parameter estimates from their estimator exhibit little appreciable bias, save for "extreme" legislators at the edges of the ideological spectrum. Likewise, both the analytical results of Lord and the Monte Carlo analysis of Poole and Rosenthal indicate that more severe biases result if either the number of voters or the number of proposals voted remains small as the sample size grows. Our attention now turns to the sources of these biases and to the potential remedies.

3 Issues in Model Identification

Spatial voting models, and their psychometric cousins, the Rasch models, are not identified without further restrictions. This is the foundation of the parameter bias results from the legislative voting (Poole and Rosenthal 1991) and psychometrics (Lord 1983) literatures. The problem arises because each proposal is put to a vote only one time. If two proposals each receive the support of the same group of "aye" voters, and meet with the opposition of the same coalition of "nay" voters, then their estimated locations will be identical. This is a classic case of overfitting, and in the current context it has mischievous consequences, preventing identification of any of the parameters.[6] Our attention now turns to the details of why this failure of identification occurs.

3.1 *Operationalizing the Spatial Model*

The agnostic spatial voting models all take some important steps in common. Each of them posits a utility function for voters and a probability distribution for the ϵ_{vp} term in Eq. (6). Three probability distributions have been suggested for this purpose: the normal, the uniform, and the extreme-value distributions. Ladha (1991) assumes that ψ_{vp} and $\psi_{vp,sq}$ are both normally distributed. This implies that ϵ_{vp} is normally distributed as well and leads to a probit model of voting. Heckman and Snyder (1996) posit a uniform distribution for ϵ_{vp}, which leads them to a linear-probability model (LPM), though in this case the ψ_{vp} and $\psi_{vp,sq}$ terms must obey distinct probability distributions! Poole and Rosenthal posit that ϵ_{vp} follows an extreme-value distribution, leading them to a nonlinear version of the

[6] In addition, there is the trivial need to normalize the scale, for example, by placing one legislator on the "left," at -1, and another on the "right," at 1. This resembles the arbitrary normalizations on temperatures imposed by the Fahrenheit and Celsius scales.

logit model. They have subsequently obtained very similar results assuming a normally distributed error instead of a logit (Poole and Rosenthal 1996).

The assumptions made by these various authors about the probability distribution of ϵ_{vp} and the form of U imply voting probabilities. Because they ignore the valence aspects of policy, Heckman and Snyder, and Ladha, are able to identify m_p from b_p^*. For the quadratic preferences analyzed by these authors, ignoring valence amounts to setting $\alpha = 0$ in Eq. (12), so that b_p^* can then be expressed in terms of m_p,

$$b_p^* = m_p g_p \tag{13}$$

For Ladha's model the probability legislator v, with preferred policy x_v, votes in favor of proposal p, with characteristics $\{m_p, g_p\}$, is

$$p_L(m_p, g_p \mid x_v) = \Phi(x_v g_p - m_p g_p) \tag{14}$$

where $\Phi(u)$ is the standard normal cumulative distribution function.

For the LPM the probability of a vote for the affirmative is

$$p_{LPM}(m_p, g_p \mid x_v) = \begin{cases} 0 & \text{for} \quad x_v g_p - m_p g_p \leq 0 \\ x_v g_p - m_p g_p & \text{for} \quad 0 < x_v g_p - m_p g_p < 1 \\ 1 & \text{for} \quad x_v g_p - m_p g_p \geq 1 \end{cases} \tag{15}$$

For the model of Poole and Rosenthal the probability legislator i votes for the "yes" alternative is

$$p_{PR}(m_p, g_p \mid x_v, \beta) = \frac{1}{1 + e^{\beta e^{-\frac{1}{8}\left(x_v - m_p - \frac{g_p}{2}\right)^2 \left(1 - e^{-\frac{g_p}{4}(x_v - m_p)}\right)}}} \tag{16}$$

In what follows we can represent any of Eqs. (14)–(16) by a generic probability that voter v votes "yes" on proposal p,[7] $p(m_p, g_p \mid x_v, \beta)$.

3.2 *"First-Order" Identifying Restrictions*

A little algebra establishes that the probability of a "yes" vote in the Poole–Rosenthal model depends only on the values of $g_p(x_v - m_p)$, $(x_v - m_p)^2 + g_p^2/4$ and β. While this is enough to pin down β, to identify the remaining parameters we need to impose the normalizations $x_v = 0$ and $x_{v'} > x_v$. The location of legislator v is normalized to equal zero, while legislator v' is normalized as being to the right of legislator v. This is the well-known arbitrariness of whether to place liberals on the "left" of the spectrum and conservatives on the "right," or vice versa.

The predictions of the Ladha and linear probability models each hinge on the value of $x_v g_p - b_p^*$. Straightforward calculations reveal that we can pin down the remaining coefficients if we impose the restrictions, $x_v = 0$ and $x_{v'} = 1$, for some distinct pair (v, v'). The first normalization is identical to the Poole–Rosenthal model, while the second places

[7] The β parameter is included to allow the notation to encompass the Poole and Rosenthal model. For the Heckman and Snyder, and the Ladha, models, think of β as a known constant.

the second voter to the right of the first, as with the model of Poole and Rosenthal, and uses the distance between these voters as a yardstick against which the distances among the other legislators are measured.

3.3 *Insufficiency of the "First-Order" Restrictions for the Proposal Parameters*

The preceding identifying restrictions do not solve a more profound set of statistical problems plaguing this class of models. Before turning to the details of this problem it is useful to establish the following lemma.

Lemma 1. Provided that the probability density function for ϵ_{vp} is symmetrical about 0, the voting probability function stated in terms of the m_p and g_p parameters will satisfy the condition

$$p(m_p, g_p \mid x_v, \beta) = 1 - p(m_p, -g_p \mid x_v, \beta) \tag{17}$$

Proof. See the Appendix.

This result makes formulating the log of the likelihood function more straightforward. Consider a legislature with V voters that considers P proposals. To keep matters simple I focus on the case in which there are no abstentions, so that every legislator votes either "yes" or "no" on every proposal. The log of the likelihood function is

$$l^* = \sum_{p=1}^{P}\sum_{v=1}^{V}\{\delta_{vp}\ln[p(m_p, g_p \mid x_v, \beta)] + (1-\delta_{vp})\ln[p(m_p, -g_p \mid x_v, \beta)]\} \tag{18}$$

Here δ_{vp} is 0 if legislator v votes against proposal p and 1 if she votes for it.

Now attention turns to a key feature in the analysis of the entire class of "agnostic" voting models considered here: if every legislator votes the same way on each of two proposals, then the maximum-likelihood estimates of the parameters corresponding to the two proposals will be identical. Lemma 2 makes this precise.

Lemma 2. If $\delta_{vp} = \delta_{vp'}$ for each $v \in \{1, 2, \ldots, V\}$ and the log of the likelihood function is twice continuously differentiable with a unique maximum, then the maximum-likelihood estimates will satisfy $m_p^{\mathrm{ML}} = m_{p'}^{\mathrm{ML}}$ and $g_p^{\mathrm{ML}} = g_{p'}^{\mathrm{ML}}$.

Proof. See the Appendix.

This happens because the legislators' votes are the only proposal-specific information used by the maximum-likelihood algorithm. When this voting information is the same for both proposals, the maximum-likelihood algorithm has identical information about the two proposals and, so, arrives at identical estimates for the corresponding proposal parameters.

There are 2^V possible voting profiles. We can create an arbitrary index

$$c^*(\delta_1, \ldots, \delta_V) = \sum_{v=1}^{V} \delta_v \cdot 2^{v-1} \tag{19}$$

which labels each possible voting profile with a different value between 0 and $2^V - 1$, and let $P^*(c)$ denote the set of all proposals that give rise to voting profile c. For any pair $(p, p') \in P^*(c)$, Lemma 2 implies that $m_p^{\mathrm{ML}} = m_{p'}^{\mathrm{ML}}$. Let m_c^{ML} represent this common

value for all of the proposals in $P^*(c)$. Likewise Lemma 2 implies that $g_p^{\mathrm{ML}} = g_{p'}^{\mathrm{ML}}$; call this common value g_c^{ML}.

Letting N_c denote the number of proposals in $P^*(c)$, one can rewrite the log of the likelihood from Eq. (18) as

$$l^* = \sum_{c=0}^{2^V-1} N_c \left\{ \sum_{v \in \mathrm{Yes}(c)} \ln p(m_c, g_c \mid x_v, \beta) + \sum_{v \in \mathrm{No}(c)} \ln p(m_c, -g_c \mid x_v, \beta) \right\} \qquad (20)$$

The parameter estimates corresponding to the type c votes chosen to maximize this likelihood function, m_c^{ML} and g_c^{ML}, are functions of the remaining parameters of the model, β and $\mathbf{x} = (x_1, \ldots, x_V)'$, $m_c^{\mathrm{ML}} = m_c^*(\mathbf{x}, \beta)$, and $g_c^{\mathrm{ML}} = g_c^*(\mathbf{x}, \beta)$.

Now consider whether the maximum-likelihood parameter estimates are consistent for the proposal parameters m_p and g_p.[8] Essentially consistency means that as the sample grows arbitrarily large, the probability distribution over the proposal parameter estimates comes arbitrarily close to placing all of the probability on the true parameter value. Theorem 1 establishes that this will not occur for the proposal parameters:

Theorem 1. The proposal parameters, m_j and g_j, are not consistently estimated.

Proof. See the Appendix.

While the probability distribution for a consistent estimator for the proposal parameters, g_p, and m_p, would converge toward a probability mass at the true parameter values, the probability distribution for the agnostic parameter estimates for each proposal places discrete lumps of probability at a dispersed set of locations. For example, suppose that there is a three-person legislature and that the voters have Gaussian utility with logistic voting errors, as in the Poole and Rosenthal model. Suppose that the true midpoint for a proposal is $m_p = 0.1$, with a gap between the proposal and the status quo of $g_p = 0.5$, while the preferred outcomes for the three voters are $x_1 = -0.3$, $x_2 = 0$ and $x_3 = 0.5$. Even when calculated under the overly optimistic assumption that the true values for the x's are known, the probability density is not centered at the true value of m_p. Instead there is a neighborhood of the true midpoint that contains no probability of a parameter estimate! The probability distribution function for this estimator places probability masses of 0.130, 0.265, and 0.279 at -5.010, -0.150, and 0.254, while it has a probability of 0.13 of exploding toward ∞. This is a classic case of "overfitting"; each proposal is observed only once, and there are only eight observable voting outcomes for our three voters.

3.4 *Lack of Identification for the Legislators' Preference Parameters*

Now that it is established that the proposal parameters are not consistently estimated, our attention turns to legislator's preference parameters. As we shall see in Theorem 2 the preference parameters inherit the lack of identification of the proposal parameters. Before I state and prove this important result, it is useful to develop a few more preliminaries.

Returning to the likelihood function in Eq. (20), a close examination reveals a collinearity problem that arises because the pair of alternatives most likely to divide any two sets of

[8] There are several definitions of statistical "consistency" in common use: weak consistency, strong consistency, and Fisher consistency (Cox and Hinkely 1974). The notion used here corresponds to weak consistency: "T_n is *weakly consistent* for θ if $T_n = \theta + o_p(1)$," where "$o_p(1)$" signifies "a random variable Z_n such that for any $\epsilon > 0$, $lim_{n\to\infty} pr(|Z_n| > \epsilon) = 0$" (Cox and Hinkely 1974).

voters will be the same, whether the first group all votes "aye" and the second group "nay," or vice versa. The following lemma establishes that proposal parameters associated with these two voting profiles are perfectly collinear.

Lemma 3. Consider two voting profiles, c and c'. If the probability density function for ϵ_{vp} is symmetrical about 0 and $\text{Yes}(c') = \text{No}(c)$, while $\text{No}(c') = \text{Yes}(c)$, then $m_c(\mathbf{x}, \beta) = m_{c'}(\mathbf{x}, \beta)$ and $g_c(\mathbf{x}, \beta) = -g_{c'}(\mathbf{x}, \beta)$.

Proof. See the Appendix.

Lemma 3 tells us that behind the 2^V voting profiles there are 2^{V-1} coalition profiles, each consisting of a different division of the V voters in the legislature. We can index each coalition by the voting profile that belongs to it in which voter V casts a negative vote: $a \in \{0, 1, 2, \ldots, 2^{V-1} - 1\}$. Let the set $A(a)$ denote the set of all proposals exhibiting coalition profile a. Proposal p is an element of $A(a)$ if it exhibits either voting profile a or its mirror image.[9] We can say that two profiles, c and c', are "mirror images" of one another if $\text{Yes}(c) = \text{No}(c')$ and $\text{No}(c) = \text{Yes}(c')$. Coalition a emerges a total of $N_a^* = N_a + N_{c'}$ times, where voting profile c' is the mirror image of the voting profile a.

As a corollary of Lemmas 1 and 3, the contribution to the log of the likelihood function by any pair of proposals with the same coalition profile is the same.

Corollary. Any proposal from either $P^*(c)$ or $P^*(c')$ makes an identical contribution to the likelihood function (20) provided that $\text{Yes}(c') = \text{No}(c')$ and $\text{No}(c') = \text{Yes}(c)$.

Proof. See the Appendix.

This means that we can rewrite the log of the likelihood function in concentrated form as

$$l^c(\mathbf{x}, \beta \mid \mathbf{N}^*) = \sum_a N_a^* \sum_{v \in \text{Yes}(a)} \ln\{p[m_a^*(\mathbf{x}, \beta), g_a^*(\mathbf{x}, \beta) \mid x_v, \beta]\} + \sum_a N_a^* \sum_{v \in \text{No}(a)} \ln\{p[m_a^*(\mathbf{x}, \beta), -g_a^*(\mathbf{x}, \beta) \mid x_v, \beta]\} \tag{21}$$

This concentrated likelihood recognizes that the proposal parameter estimates depend entirely on the voter parameters, $\mathbf{x}$ and β, and on the observed frequencies for the various coalition profiles. This is what allows us to rewrite the likelihood function in terms of the voter parameters and the frequencies at which the coalitions are observed. Put another way, the vector of coalition frequencies, $\mathbf{N}^*$, is sufficient for the voter parameters, $\mathbf{x}$ and β.

It is useful to characterize the "limiting" population of proposals that emerges as the sample size grows with a density function. Define $f(m, g)$ as the relative density of proposals with midpoint m and policy gap g. I normalize this density so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(m, g) dg dm = 1 \tag{22}$$

This means that $f(m, g)$ has all of the properties of a probability density function. Here I treat $f(m, g)$ as continuously differentiable, though this requirement can be relaxed somewhat.

[9] That is, $A(a) = P(a) \cup P(2^V - a - 1)$, where the labeling convention takes advantage of the fact that the indices for any voting profile and its mirror image must sum to $2^V - 1$.

The voting coalition probabilities depend on the legislators' preference parameters, the agenda, as represented by the density over proposals, $f(m, g)$, and on the probability that coalition a emerges, $\lambda(a \mid m, g, \mathbf{x}, \beta)$, which is itself a function of the true proposal and voting parameters,

$$\lambda(a \mid m, g, \mathbf{x}, \beta) = \prod_{v \in \text{Yes}(a)} p(m, g \mid x_v, \beta) \cdot \prod_{v \in \text{No}(a)} p(m, -g \mid x_v, \beta) + \prod_{v \in \text{No}(a)} p(m, g \mid x_v, \beta) \cdot \prod_{v \in \text{Yes}(a)} p(m, -g \mid x_v, \beta). \quad (23)$$

Combining Eqs. (22) and (23), we obtain a compact expression for the probability that we will observe the coalition profile a,

$$p_a^*(\mathbf{x}, \beta \mid f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lambda(a \mid m, g, \mathbf{x}, \beta) f(m, g) dg dm \quad (24)$$

We can stack these terms to obtain a summary of all of the coalition probabilities conditional on a particular proposal, with true parameters (m, g): $\boldsymbol{\lambda}(m, g, \mathbf{x}, \beta)$, and the corresponding coalition probabilities when the agenda can be represented by a density function $f(m, g)$ over the proposal parameters: $\mathbf{p}^*(\mathbf{x}, \beta \mid f)$.

As the number of proposals tends to infinity the maximum-likelihood parameter estimates $\{\mathbf{x}^{\text{ML}}, \beta^{\text{ML}}\}$ will be those that maximize $l^x[\mathbf{x}, \beta \mid \mathbf{p}^*(\mathbf{x}, \beta \mid f)]$. Hence $\mathbf{p} = \mathbf{p}^*(\mathbf{x}, \beta \mid f)$ is asymptotically sufficient for $\mathbf{x}$ and β (Cox and Hinkely 1974). Everything that can be learned about the legislators' preference parameters must come via the coalition probabilities.

The problem for identification is that the sufficient statistic, $\mathbf{p}$, is dependent on $f(m, g)$ as well as on $\mathbf{x}$ and β. Since we cannot directly observe $f(m, g)$, we will not, in general, be able to distinguish among any of an infinite number of combinations of ideal points and agendas.

Before turning to a formal statement of the result we need to develop one bit of terminology for describing the relationship between $\mathbf{p}$ and the underlying agenda. We say that the coalition probability function $\mathbf{p}^*$ is *fully agenda manipulable* if we can move the vector of probabilities in any direction from $\mathbf{p}^*(\mathbf{x}, \beta \mid f)$ by manipulating the agenda. Formally, the coalition probability function $\mathbf{p}^*$ is *fully agenda manipulable* if there exist $2^{V-1} + 1$ distinct probability densities $\xi_0(m, g), \ldots, \xi_{2^{V-1}}(m, g)$, which stack to make the vector $\boldsymbol{\xi}(m, g)$, such that $\boldsymbol{\Omega}(\mathbf{x}, \beta)$, defined as

$$\boldsymbol{\Omega}(\mathbf{x}, \beta) = \begin{bmatrix} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \boldsymbol{\lambda}(m, g, \mathbf{x}, \beta) \boldsymbol{\xi}'(m, g) dg dm \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \boldsymbol{\xi}(m, g)' dg dm \end{bmatrix} \quad (25)$$

has full rank. This definition allows us to state the main result for the preference parameters.

Theorem 2. If $\mathbf{p}^0 = \mathbf{p}^*(\mathbf{x}, \beta \mid f)$ is fully agenda manipulable, then for any voter parameters $\mathbf{x}$ and β, and an agenda represented by the probability density function $f(m, g)$, there is another, distinct set of voter preferred points $\mathbf{x}'$ that differs in every element from $\mathbf{x}$ and another agenda represented by the probability density function $h(m, g)$ that leads to the same coalition probabilities: $\mathbf{p}^0 = \mathbf{p}^*(\mathbf{x}', \beta \mid h)$.

Proof. See the Appendix.

It follows immediately from Theorem 2 and sufficiency of $\mathbf{p}^0$ for $\mathbf{x}$ and β that the voter parameters $\mathbf{x}$ are not identified.

The results summarized by Theorems 1 and 2 tell us that none of the parameters of the agnostic spatial model are identified! No matter how many votes we observe, neither the proposal parameters nor those for the voters can be recovered from the data. The inability to capture the proposal parameters is not, by itself, a surprise. With very few exceptions of bills that are reconsidered with no changes in their wording, we observe each proposal only once, and only a finite number of legislators cast votes on that occasion. Lemma 2 makes clear why this is a problem. The finiteness of the legislature means that there is only a fixed set of possible voting coalitions, and the number of possible values that the proposal parameters can take on is bounded above by the number of observable voting coalitions. It is Theorem 2 that delivers the real punch: the lack of identification for the proposal parameters is inherited by the estimates for the voter parameters. This is because the same set of coalition probabilities can arise from various combinations of an agenda and a set of legislators voting on that agenda. Even with infinitely many votes we can estimate only the coalition probabilities, and these are not enough, by themselves, to tell us which combination of proposal and voter parameters produced the data.

4 Modeling the Agenda

A remedy for the identification problems that beset the agnostic model is to construct a structural model of the agenda. Instead of estimating individual proposal parameters, the analyst can construct a structural model with a finite number of agenda parameters. This results in a likelihood function that is then maximized with respect to the voters' unidentified preference parameters and the parameters characterizing the agenda. In contrast to the agnostic models, as the number of observations in the structural model grows without bound, the number of parameters to be estimated remains fixed. This means that the estimated m and g parameters are no longer confined to a finite lattice of points. With enough data, and given suitable regularity conditions, one can identify all of the parameters of the model. The model here identifies voters' preference parameters using the structure placed on the agenda. This identification comes at a cost: the interpretation of the preference parameters is contingent on the distribution of proposals posited by the structural model.

Building on the simple spatial voting model, we found that the condition for legislator v to vote in favor of proposal p was given by Eq. (11), which treats every proposal as entirely unique and unrelated to any other proposal. But sometimes we see strings of amendments in the same issue area sponsored by the same group of legislators. It would be surprising indeed if the ideological displacements for these amendments (the g_p) were in opposite directions and of very different magnitudes. Instead we might expect that the sponsors pursue a coherent agenda, much as we assume that voters have stable ideal points and vote a coherent set of issue preferences. More generally, we could imagine the content of proposals depending on other explanatory variables as well. For example, when legislators are voting on bills related to their districts, the types of policies they propose might shift. If we were interested in the impact of farm interests in a legislator's district on how he or she proposes, we could allow g_p to depend on the percentage of the labor force in the member's district who are employed in agricultural industries. Instead of estimating a different g_p for each proposal from a given author, we might expect that each proposal sharing a given set of observable characteristics, including the identity of its author, was made in pursuit of a

stable set of ideological objectives with

$$g_p = \mathbf{g}'\mathbf{w}_{gp} + \zeta_p \tag{26}$$

where ζ_p is idiosyncratic to proposal p, while $\mathbf{w}_{gp}$ is a vector of explanatory variables. It might include such features as the author's identity, or district variables, such as the fraction of the legislator's district employed in agriculture. The estimator might also differentiate among proposals by issue area.

We can treat the "consensus parameter," b_p^*, similarly: $b_p^* = -\mathbf{b}'\mathbf{w}_{bp} - \omega_p^*$. All voters are more attracted to proposals with low values for b_p^*, which may be low either because of a high valence advantage or because the proposals confront extreme alternatives.

We might expect variation in the valence of proposals with the same characteristics, such as authorship. In part, this is because authors sometimes make mistakes on technical grounds.[10] Likewise, we can allow for the possibility that the mean value for ϵ_{vp} differs among voters.[11] Let $E\{\epsilon_{vp}\} = -c_v$ and let η_{vp} denote the deviation of the idiosyncratic shock from its mean: $\eta_{vp} = \epsilon_{vp} - E\{\epsilon_{vp}\}$. As with the proposer parameters, so the voter parameters can also depend on explanatory variables: $x_v = \mathbf{x}'\mathbf{w}_{xv}$, and $c_v = \mathbf{c}'\mathbf{w}_{cv}$. This framework encompasses the widespread practice of attempting to scale each voter's preferred outcome separately.[12]

Substituting into Eq. (11) and rearranging terms, legislator v will vote in favor of proposal p if

$$\eta_{vp} < (\mathbf{g}'\mathbf{w}_{gp})(\mathbf{x}'\mathbf{w}_{xv}) + (\mathbf{b}'\mathbf{w}_{bp}) + (\mathbf{c}'\mathbf{w}_{cv}) + \zeta_p(\mathbf{x}'\mathbf{w}_{xv}) + \omega_p^* \tag{27}$$

While many analysts use scaled locations of legislators as the dependent variables for regressions that test competing explanations about the foundations of their ideologies, this framework, which allows for joint estimation of the preferred outcomes and the impact of the explanatory variables, offers an attractive alternative.

To complete the model we need to specify distributions for the random variables η_{vp}, ζ_p, and ω_p^*. By allowing all three to be normally distributed we are led to a multivariate probit model. As it is formulated, this is a multivariate dichotomous choice model with heteroscedastic errors caused by the $\zeta_p(\mathbf{x}'\mathbf{w}_{xv})$ term on the right-hand side of inequality (27). Heteroscedasticity makes estimation more complex. The ζ_p parameter captures variation in the ideological content of proposals with the same observed characteristics. While we expect the technical quality of proposals to vary, even when they come from the same legislator and pertain to the same policy area, their ideological content is another matter: most legislators are professional ideologues with a keenly developed sense of the ideological content of a proposal. Provided that the status quo is fairly static, we might expect a given legislator consistently to propose moving the status quo to the same location, implying little variation in the g_p.

[10] See Londregan (2000) for some examples in the Chilean context.

[11] This would happen if there were voters who were biased on procedural grounds against amendments per se. While this would not be consistent with rational behavior, it is straightforward to incorporate into the model, and we are then free to test whether we could have restrained all of the means to equal 0.

[12] To scale each voter separately, our explanatory variables, $\mathbf{w}_{xv}$ and $\mathbf{w}_{cv}$, would consist solely of legislator-specific dummies.

Thus we might start with a tractable model with the property that $\zeta_p = 0$ for all p. We also need to calibrate the relative importance of the η_{vp} and ω_p^* terms. This can be done by first making η_p normal with mean 0 and variance 1. This is just the standard normalization for the error term in probit models. Then replace ω_p^* with $\alpha\omega_p$, where ω_p is normally distributed with mean 0 and variance 1. This formulation allows us to capture the variance of ω_p^* as $\mathrm{Var}(\omega_p^*) = \alpha^2$. Under these conditions, legislator v will back proposal p from proposer $a(p)$ if

$$\eta_{vp} < a_{vp} + \alpha\omega_p \tag{28}$$

where

$$a_{vp} = (\mathbf{g}'\mathbf{w}_{gp})(\mathbf{x}'\mathbf{w}_{xv}) + (\mathbf{b}'\mathbf{w}_{bp}) + (\mathbf{c}'\mathbf{w}_{cv}) \tag{29}$$

This leads to a voting probability for voter v on proposal p of $\Phi(a_{vp} + \alpha\omega_p)$.

Let $\mathsf{V}(p)$ denote the set of legislators who voted on proposal p. The function $\mathrm{Vote}(v)$ is equal to 1 if legislator v voted in favor of the proposal and is equal to -1 if, instead, she voted against it. Using this notation we can write the log of the likelihood for this model as

$$l(\mathbf{b}, \mathbf{c}, \mathbf{g}, \mathbf{x}, \alpha) = \sum_{p=1}^{P} \ln\left(\int_{-\infty}^{\infty} \prod_{v\in\mathsf{V}(p)} \Phi[\mathrm{Vote}(v)a_{vp} + \alpha\omega_p]\phi(\omega_p)d\omega_p\right) \tag{30}$$

Unlike its agnostic cousins, this model can be consistently estimated using maximum likelihood methods. At some price in the form of additional complexity, the model can be extended to include random variation in the ideological content of proposals from a particular individual, as well as variation in their valence. Because the integration over the common random term, the ω_p, in Eq. (30) can be done using numerical methods, it is easy to alter the probability distribution that generates it.

The lack of correlation between voters' idiosyncratic shocks and the proposal-specific random term, the ω_p built into the model is most plausible when making proposals is easy and the voting agenda is difficult to manipulate. This is the case for the Chilean legislative data that initially led to the formulation of this model (Londregan 2000). In other settings the agenda is manipulated by those who control it, as it is often alleged occurs in the U.S. House of Representatives. Consider Fenno's assertion that Ways and Means Chairman Wilbur Mills was successful at gaining passage for his bills by knowing just how far he could get the floor to go, that is, by tailoring their content to the values of members' idiosyncratic shocks. In such a highly manipulated setting the model could be extended, at some cost in complexity,[13] to allow for correlation between the random variation among proposals from the same, strategic, author and the voters' idiosyncratic random preference shocks.

The approach advocated here for modeling the agenda is more difficult than applying an agnostic model, or some related factor analytic technique, but there is no ready substitute. Moreover, the proposal side of the model can be extended and adapted to reflect our understanding of the context, such as the degree of agenda control exercised over proposals.

[13]This amounts to allowing the α parameter to differ among voters.

5 Conclusions

While the spatial model has been widely applied to empirical analysis of legislative voting, the statistical foundations of this model have received little attention. The key result of this paper is that statistical procedures that take the intuitively appealing route of jointly estimating separate spatial "locations" for each proposal, and separate preferred outcomes for each legislator, are not identified. The results here establish the reason these models fail: the granularity of the voting data imposes a false granularity on the set of parameters the model is capable of estimating. This failure is severe when the number of voters is small.

I propose an alternative estimator which overcomes the identification failure of the agnostic models by incorporating a model of the agenda along with the model of voting. The approach advocated here does not come without consequences. This structural model of legislation incorporates the proposal process, along with voting. In one sense this is disappointing: if we misspecify the proposal process, this specification error will, in general, contaminate our estimates of the voting parameters as well as those pertaining to proposal making. But the agnostic approach is not an option. It results in biased estimates of *all* of the parameters in question. Viewed from a different vantage, the proposal process is an important part of legislation. How many laws are named for lawmakers who were sponsors, and how many are named for legislators who merely joined the majority in voting for them? The genesis of legislative proposals is part of the process political scientists seek to explain. Sooner or later we were going to have to go beyond agnostic models of the bill proposal process anyway. The statistical issues raised here merely add an additional reason to move in this direction. The model of proposal making set forth here is not meant to be the last word on how to model proposal making. Instead it is offered as a stepping-off point on the important journey of incorporating the agenda into empirical models of voting.

Appendix: Proofs

Proof of Lemma 1. Substituting from Eq. (9) into Eq. (6) and ignoring valence considerations, which is tantamount to setting both q_p and q_{sq} equal to zero, the condition for legislator v to vote in favor of proposal p is

$$\epsilon_{vp} < U(m_p + g_p/2, 0 \mid x_v, \beta) - U(m_p - g_p/2, 0 \mid x_v, \beta) \tag{31}$$

while legislator v will vote against the proposal if, instead,

$$\epsilon_{vp} > U(m_p + g_p/2, 0 \mid x_v, \beta) - U(m_p - g_p/2, 0 \mid x_v, \beta) \tag{32}$$

which happens with a probability of $1 - p(m_p, g_p \mid x_v, \beta)$. However, multiplying both sides of Eq. (32) by -1 yields

$$-\epsilon_{vp} < U(m_p - g_p/2, 0 \mid x_v, \beta) - U(m_p + g_p/2, 0 \mid x_v, \beta) \tag{33}$$

Symmetry about 0 of the probability density function for ϵ_{vp} means that the probability of drawing a value of $-\epsilon_{vp}$ that satisfies Eq. (33) is the same as the probability of drawing a value of ϵ_{vp} that satisfies

$$\epsilon_{vp} < U(m_p - g_p/2, 0 \mid x_v, \beta) - U(m_p + g_p/2, 0 \mid x_v, \beta) \tag{34}$$

But this is just Eq. (31) evaluated at a proposal with characteristics $(-g_p, m_p)$: the probability of voting "nay" on a proposal with characteristics (g_p, m_p) is the probability of voting "aye" on a proposal with characteristics $(-g_p, m_p)$, namely, $p(m_p, -g_p \mid x_v, \beta)$. Q.E.D.

Proof of Lemma 2. The first-order conditions for a maximum of the likelihood function in Eq. (18) with respect to m_p and g_p are

$$\frac{\partial l}{\partial m_p} = 0$$
$$\frac{\partial l}{\partial g_p} = 0$$

Letting $\text{Yes}(p)$ denote the set of legislators who voted in favor of proposal p, while $\text{No}(p)$ indicates the voters who opposed it, one can rewrite the first-order conditions as:

$$\sum_{v \subset \text{Yes}(p)} \frac{p_{m_p}\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} + \sum_{v \subset \text{No}(p)} \frac{p_{m_p}\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} = 0 \quad (35)$$

$$\sum_{v \subset \text{Yes}(p)} \frac{p_{g_p}\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} - \sum_{v \subset \text{No}(p)} \frac{p_{g_p}\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} = 0 \quad (36)$$

Similarly, the first-order conditions for a maximum with respect to $m_{p'}^{\text{ML}}$ and $g_{p'}^{\text{ML}}$ are

$$\sum_{v \subset \text{Yes}(p')} \frac{p_{m_p}\left(m_{p'}^{\text{ML}}, g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_{p'}^{\text{ML}}, g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} + \sum_{v \subset \text{No}(p')} \frac{p_{m_p}\left(m_{p'}^{\text{ML}}, -g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_{p'}^{\text{ML}}, -g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} = 0 \quad (37)$$

$$\sum_{v \subset \text{Yes}(p')} \frac{p_{g_p}\left(m_{p'}^{\text{ML}}, g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_{p'}^{\text{ML}}, g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} - \sum_{v \subset \text{No}(p')} \frac{p_{g_p}\left(m_{p'}^{\text{ML}}, -g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_{p'}^{\text{ML}}, -g_{p'}^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} = 0 \quad (38)$$

By assumption, $\text{Yes}(p') = \text{Yes}(p)$ and $\text{No}(p') = \text{No}(p)$, so Eqs. (37) and (38) can be rewritten as:

$$\sum_{v \subset \text{Yes}(p)} \frac{p_{m_p}\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} + \sum_{v \subset \text{No}(p)} \frac{p_{m_p}\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} = 0 \quad (39)$$

$$\sum_{v \subset \text{Yes}(p)} \frac{p_{g_p}\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} - \sum_{v \subset \text{No}(p)} \frac{p_{g_p}\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)}{p\left(m_p^{\text{ML}}, -g_p^{\text{ML}} \mid x_v^{\text{ML}}, \beta^{\text{ML}}\right)} = 0 \quad (40)$$

But these equations, (39) and (40), are identical to Eqs. (35) and (36), and so any solution to Eqs. (35) and (36) is also a solution to Eqs. (39) and (40), and vice versa. The desired result then follows from uniqueness of the maximum-likelihood parameter estimates $(g_p^{\text{ML}}, m_p^{\text{ML}})$. Q.E.D.

Proof of Theorem 1. If the maximum of the likelihood function is not uniquely valued, then the result is established. If the likelihood does have a unique maximum, then to show inconsistency for m_p^{ML} there are two cases to consider.

(i) $m_c^{\mathrm{ML}} \to m^*$ for all c as the number of proposals goes to infinity. In this case, consider two proposals with two distinct cutpoints, $m_p \neq m'_p$. At most one of these cut points will actually equal m^*; without loss of generality, proposal p' does not, so that

$$m_p^{\mathrm{ML}} \to m^* = m_{p'} + (m_p - m_{p'}) \tag{41}$$

where $(m_p - m_{p'})$ is a constant.

(ii) There is at least one pair of voting patterns, $\hat{c}$ and c', such that $m_{\hat{c}}^* \neq m_{c'}^*$, where $m_{\hat{c}}^{\mathrm{ML}} \to m_{\hat{c}}^*$ and $m_{c'}^{\mathrm{ML}} \to m_{c'}^*$. Next define π_p using

$$\ln(\pi_p) = min_c \left\{ \sum_{i \subset \mathrm{Yes}(c)} \ln p(m_p, g_p \mid x_v, \beta) + \sum_{i \subset \mathrm{No}(c)} \ln p(m_p, -g_p \mid x_v, \beta) \right\} \tag{42}$$

This means that every observable voting pattern has a probability of at least π_p of being observed as a response to proposal p. Without loss of generality, suppose that $m_p \neq m_{\hat{c}}^*$. The probability of observing $\hat{c}$ is at least $\pi(p)$, and leading to a probability of $\pi(p) > 0$ so that as the number of proposals goes to infinity, we observe

$$m_p^{\mathrm{ML}} = m_{\hat{c}}^* = m_p + (m_{\hat{c}}^* - m_p) \tag{43}$$

where $(m_{\hat{c}}^* - m_p)$ is a nonzero constant and π_p is not o_p.

To show inconsistency for g_p^{ML} there are likewise two cases to consider.

(i) $g_c^{\mathrm{ML}} \to g^*$ for all c as the number of proposals goes to infinity. In this case, consider two proposals with distinct gaps, $g_p \neq g_{p'}$. At most one of these gaps will actually equal g^*; without loss of generality, let it be proposal p'. For the other proposal, p,

$$g_p^{\mathrm{ML}} \to g_p + (g^* - g_p) \neq 0 \tag{44}$$

where $g^* - g_p$ is a positive constant and π_p is not o_p.

(ii) There is at least one pair of voting patterns, $\hat{c}$ and c', such that $g_{\hat{c}}^* \neq g_{c'}^*$, where $g_{\hat{c}}^{\mathrm{ML}} \to g_{\hat{c}}^*$ and $g_{c'}^{\mathrm{ML}} \to g_{c'}^*$, while either $g_p \neq g_{\hat{c}}^*$ or $g_p \neq g_{c'}^*$. Suppose, without loss of generality, that $g_p \neq g_{\hat{c}}^*$. There is a probability of at least π_p that as the number of proposals goes to infinity, one will observe

$$g_p^{\mathrm{ML}} = g_{\hat{c}}^* = g_p + (g_{\hat{c}}^* - g_p) \tag{45}$$

where $(g_{\hat{c}}^* - g_p)$ is a positive constant and π_p is not o_p. Q.E.D.

Proof of Lemma 3. The contribution to the log of the likelihood function in Eq. (30) from proposals with voting profile c is

$$l_c = N_c \left\{ \sum_{v \subset \mathrm{Yes}(c)} \ln p(m_c, g_c \mid x_v, \beta) + \sum_{v \subset \mathrm{No}(c)} \ln p(m_c, -g_c \mid x_v, \beta) \right\} \tag{46}$$

while the contribution to the log of the likelihood from proposals with voting profile c' is

$$l_{c'} = N_{c'}\left\{ \sum_{v\subset \mathrm{Yes}(c')} \ln p(m_{c'}, g_{c'} \mid x_v, \beta) + \sum_{v\subset \mathrm{No}(c')} \ln p(m_{c'}, -g_{c'} \mid x_v, \beta)\right\} \tag{47}$$

However, by hypothesis, $\mathrm{Yes}(c') = \mathrm{No}(c)$ and $\mathrm{No}(c') = \mathrm{Yes}(c)$, so Eq. (47) becomes

$$l_{c'} = N_{c'}\left\{ \sum_{v\subset \mathrm{No}(c)} \ln p(m_{c'}, g_{c'} \mid x_v, \beta) + \sum_{v\subset \mathrm{Yes}(c)} ln\, p(m_{c'}, -g_{c'} \mid x_v, \beta)\right\} \tag{48}$$

By setting $m_c = m_{c'}$ and $g_c = -g'_c$, Eqs. (46) and (48) coincide up to a factor of proportionality. Thus, if m_c^* and g_c^* maximize the expression in Eq. (46), then $m_{c'} = m_c^*$ and $g'_c = -g_c^*$ must maximize Eq. (48). Q.E.D.

Proof of Corollary to Lemmas 1 and 3. The contribution to the likelihood function from proposal p in $P^*(c)$ is

$$l_p^* = \sum_{v\in \mathrm{Yes}(c)} \ln p(m_c, g_c \mid x_v, \beta) + \sum_{v\in \mathrm{No}(c)} \ln p(m_c, -g_c \mid x_v, \beta) \tag{49}$$

while the contribution from a proposal p' in $P^*(c')$ is

$$l_p^* = \sum_{v\in \mathrm{Yes}(c')} \ln p(m_{c'}, g_{c'} \mid x_v, \beta) + \sum_{v\in \mathrm{No}(c')} \ln p(m_{c'}, -g_{c'} \mid x_v, \beta) \tag{50}$$

If voting profile c' is a "mirror image" of c, then we know that $\mathrm{Yes}(c') = \mathrm{No}(c)$ and $\mathrm{No}(c') = \mathrm{Yes}(c)$, while it follows from Lemma 2 that $m_{c'} = m_c$ and $g_{c'} = -g_c$. Substituting into the contribution to the likelihood from the proposal in $P^*(c')$, we have

$$l_{p'}^* = \sum_{v\in \mathrm{Yes}(c)} \ln p(m_c, g_c \mid x_v, \beta) + \sum_{v\in \mathrm{No}(c)} \ln p(m_c, -g_c \mid x_v, \beta) \tag{51}$$

which is identical to the contribution from a vote in $P^*(c)$. Q.E.D.

Proof of Theorem 2. To prove the theorem it is sufficient to show that if one were to change X by some small amount **dx**, one could change the density over proposal characteristics to maintain the vector of coalition probabilities unchanged. Define the density $h_{\mathbf{w}}$ as

$$h_{\mathbf{w}}(m, g) = f(m, g) + \mathbf{w}'\boldsymbol{\xi}(m, g) \tag{52}$$

where **w** is a vector of weights, which can take on either positive or negative values. For $h(m, g \mid \mathbf{w})$ to be a probability density over proposals, it needs to integrate to 1,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(m, g \mid \mathbf{w}) dg dm = 1 \tag{53}$$

Substituting from Eq. (52) and noting that $f(m, g)$ is a probability density [see Eq. (22)] yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{\xi}(m, g)\mathbf{w}\, dgdm = 0 \tag{54}$$

Suppose that at $\mathbf{w} = 0$ the vector of coalition probabilities is

$$\mathbf{p} = \mathbf{p}^*(\mathbf{x}, \beta \mid f) = \mathbf{p}^*(\mathbf{x}, \beta \mid h_{\mathbf{w}}) \tag{55}$$

For a given local change in $\mathbf{x}$ by $\mathbf{dx}$, one can change $\mathbf{w}$ by $\mathbf{dw}$ so that both condition (54) and condition (55) continue to be satisfied, provided that $\mathbf{dw}$ simultaneously satisfies

$$\nabla_{\mathbf{x}'}\mathbf{p}^*(\mathbf{x}, \beta \mid h_{\mathbf{w}})\mathbf{dx} + \nabla_{\mathbf{w}'}\mathbf{p}^*(\mathbf{x}, \beta \mid h_{\mathbf{w}})\mathbf{dw} = \mathbf{0} \tag{56}$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{\xi}(m, g)' dgdm\mathbf{dw} = 0 \tag{57}$$

Define the matrix $\boldsymbol{\Psi}$ as

$$\boldsymbol{\Psi} = \begin{bmatrix} \nabla_{\mathbf{x}'}\mathbf{p}^*(\mathbf{x}, \beta \mid f) \\ \mathbf{0} \end{bmatrix} \tag{58}$$

Substituting $\boldsymbol{\Psi}$ from Eq. (58) and $\boldsymbol{\Omega}$ as defined in Eq. (25), conditions (56) and (57) become

$$\boldsymbol{\Psi}\begin{bmatrix} \mathbf{dx} \\ \mathbf{0} \end{bmatrix} + \Theta\mathbf{dw} = \mathbf{0} \tag{59}$$

Provided that Θ is of full rank, one can choose $\mathbf{dw}$ so that

$$\mathbf{dw} = -\Theta^{-1}\Psi\begin{bmatrix} \mathbf{dx} \\ 0 \end{bmatrix} \tag{60}$$

This means that, provided $\mathbf{dw}$ satisfies the condition in Eq. (60), we can make an arbitrary local change in the $\boldsymbol{Vx1}$ vector of legislators' preferred outcomes $\mathbf{x}$ by $\mathbf{dx}$ and still adjust the $2^{V-1} + 1$ weights $\mathbf{w}$ to preserve all of the coalition profile probabilities intact. All local perturbations of $\mathbf{x}$ are consistent with the observed frequencies for the voting coalitions for *some* choice of agenda.

There is an additional feasibility condition, namely, that $h(m, g) > 0$ for all $\{m, g\}$, but $f(m, g) > 0$ for all $\{m, g\}$, guarantees that this will continue to hold provided $\mathbf{dw}$ is small enough, which it will be for a sufficiently small $\mathbf{dx}$. Thus, for *any* choice of $\mathbf{x}$ and its associated set of coalition probabilities $\mathbf{p}$, one can always find another nearby set of legislator parameters $\mathbf{x} + \mathbf{dx}$ and a probability density over proposals that will also give rise to the same $\mathbf{p}$ as the asymptotic coalition probabilities. Q.E.D.

References

Birnbaum, A. 1968. "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability." In *Statistical Theories of Mental Test Scores*, eds. F. M. Lord and M. R. Novick. Reading, MA: Addison Wesley.

Brady, Henry. 1991. "Traits Versus Issues: Factor Versus Ideal-Point Analysis of Candidate Thermometer Ratings." *Political Analysis* 2:97–129.

Cox, D. R., and D. V. Hinkely. 1974. *Theoretical Statistics*. New York: Chapman and Hall.

Fischer, Gerhard H. 1995. "Derivations of the Rasch Model." In *Rasch Models: Foundations, Recent Developments and Applications*, ed. Gerhard H. Fischer. New York: Springer Verlag.

Haberman, S. J. 1977. "Maximum Likelihood Estimates in Exponential Response Models." *Annals of Statistics* 5:815–841.

Heckman, James J., and James Snyder. 1996. "Linear Factor Models of the Demand for Attributes with an Application to Estimating the Preferences of Legislators," Department of Economics, University of Chicago. Unpublished paper.

Hinich, Melvin J., Lawerence Cahoon, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting." In *Graphical Representation of Multivariate Data*, ed. P. C. Wang. New York: Harwood, pp. 243–279.

Ladha, Krishna. 1991. "A Spatial Model of Legislative Voting with Perceptual Error." *Public Choice* 68:151–174.

Londregan, John B. 2000. *Legislative Institutions and Ideology in Chile's Democratic Transition*. New York: Cambridge University Press.

Lord, Frederic M. 1983. "Unbiased Estimates of Ability Parameters, of Their Variance, and of Their Parallel Forms Reliability." *Psychometrika* 48:477–482.

Maddala, G. S. 1986. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge.

Marriott, F. H. C., and J. A. Pope. 1954. "Bias in the Estimation of Autocorrelations." *Biometrika* 41:390–402.

Poole, Keith T., and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35:228–278.

Poole, Keith T., and Howard Rosenthal. 1996. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Rasch, Georg. 1961. "On General Laws and the Meaning of Measurement in Psychology." *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability* 4:321–333.

Stokes, Donald. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57:368–377.

Thurstone, L. L. 1925. "A Method of Scaling Psychological and Educational Tests." *Journal of Educational Psychology* 16:433–451.