Properties of Ideal-Point Estimators[1]

Alexander Tahk

Assistant Professor

Department of Political Science

University of Wisconsin–Madison

April 2015

[1]Paper prepared for the Conference on Ideal Point Models (May 1–2, 2015, at MIT). I thank Simon Jackman, John Ferejohn, Jon Krosnick, Kevin Quinn, and Michael Bailey for valuable comments and suggestions. An earlier version of this paper was presented at the Annual Meeting of the Society for Political Methodology.

Abstract

Ideal-point estimation has become relatively commonplace in political science. However, little is known about the properties of these estimators. Almost all estimators—including Poole and Rosenthal's NOMINATE and Clinton, Jackman, and Rivers's IDEAL—are complicated by the Neyman–Scott problem.

This paper presents several theoretical results regarding ideal-point estimation when the number of bills is allowed to grow but the number of legislators is fixed. I show that consistent estimators of the ideal points are impossible in the most common parametric ideal-point models. I provide counterexamples to consistency of common ideal-point estimators with regard to the rank-order of the ideal points. I also show that maximum likelihood provides a consistent estimate of the rank order when the direction of each bill is known but not otherwise. Finally, I show that consistent estimation of the rank-order of the ideal-points is possible under reasonably weak assumptions even if the most common current estimators do not do so.

1 Introduction

Despite the popularity ideal-point estimation has gained in recent years, relatively little is known about the propoerties of the most common ideal-point estimates. Even basic properties such as consistency have not been established for most models. Further, several basic results regarding the consistency of maximum likelihood estimators and Bayes estimators are not applicable to ideal point models.

One difficulty in establishing the consistency of ideal point estimators is the incidental parameters problem or Neyman–Scott problem (e.g. Heckman and Snyder 1997; Clinton, Jackman, and Rivers 2004). If the number of parameters grows with the number of observations, then the maximum likelihood need not—and, in many cases, does not—provide a consistent estimator even if the parameters of interest are fixed in size and appear in every observation (Neyman and Scott 1948). In ideal point models, each new roll call vote introduces a new set of bill parameters. Thus, even if we are interested in the ideal points of legislators and these are fixed in size (which Neyman and Scott refer to as "structural parameters"), the addition of new bill parameters ("incidental parameters") means that the maximum likelihood estimate of the ideal points may not be consistent.

Frequentists and classical Bayesians, who assume the existence of a true but unknown parameter, generally acknowledge consistency to be an important property.[1] Even subjective Bayesians, who do not believe in an objective, true parameter, have reason to be concerned with consistency. Although Bayes estimates usually converge to the same value under different priors provided the priors are each absolutely continuous with respect to the other (Blackwell and Dubins 1962),[2] we cannot expect

[1]Although some have argued that asymptotic properties are irrelevant because we always encounter finite amounts of data, Diaconis and Freedman (1986b) point out that "if your procedure gives the wrong answer with unlimited data, probably you will not like it so well with a finite sample either."

[2]This means, roughly, that as long as two priors agree on what is possible, we expect the data to eventually swamp our priors and our estimates under the two priors will converge.

this behavior when Bayes estimators are inconsistent Diaconis and Freedman (see, e.g., 1986a).

In the context of ideal-point models, this means that Bayes estimates of ideal points may (and often are) dependent on our choice of prior for the bill parameters. This poses an even greater problem for the common use flat or nearly flat priors on the bill parameters, which are almost always implausible as genuine subjective beliefs.[3] If it does not, then we cannot expect the resulting posterior to approach the correct posterior based on our actual subjective beliefs regardless of the number of bills.

1.1 Incidental parameters problem in ideal-point estimation

The incidental parameters problem is sometimes addressed in the ideal-point literature. One early model that avoided the problem is the linear probability model introduced by Heckman and Snyder (1997). However, this model deviates somewhat substantially from the form of most ideal point models and, in many cases, is implausible.

Other authors have suggested random-effects estimators. Bailey (2001) and Lewis (2001) both use random-effects approaches to handle data involving many legislators and few votes. Londregan (1999) proposes a random effects estimator to data with few legislators and many votes that includes a model of the agenda process. Peress (2009) introduces a somewhat more general random-effects estimator for this same situation. However, as is generally true with random-effects models, their consistency depends upon the correct choice of parametric family for the random effects. When this distribution is misspecified, these estimators may still be inconsistent.

Several other solutions have been applied in the closely-related literature on item-response theory. Andersen (1973) introduces a conditional maximum likelihood ap-

[3]For instance, a flat prior on the bill parameters generally implies that we are certain (in the sense that our prior probability is one) that all votes will be unanimous. If this truly reflected our subjective beliefs, ideal-point estimation would be inappropriate.

proach to estimating the one-parameter Rasch model. However, conditional maximum likelihood can only be applied when a sufficient statistic for the incidental parameters such that the likelihood conditional on this sufficient statistic still depends upon the structural parameters. While this is possible for the one-parameter Rasch model, the use of a discrimination parameter in ideal point models renders it inapplicable in most cases.

However, in the most common ideal point estimators, (e.g. Poole and Rosenthal 1991) the incidental parameters problem is left unaddressed. This includes the standard Bayesian approach (Clinton, Jackman, and Rivers 2004), as Bayes estimators also need not be consistent in the presence of incidental parameters unless the prior distribution of the incidental parameters provides an accurate description of their distribution (Lancaster 2000).[4] This is typically implausible, particularly where a flat or near-flat prior is used.

Poole (2000b) also introduced a non-parametric ideal-point estimator known as Optimal Classification or OC. Instead of estimating a parametric model, optimal classification attempts to find ideal points and cutting planes that maximize the number of correctly classified votes. While intended to provide estimates for ideal-point models with symmetric error, the only investigation seems to be some Monte Carlo results (Poole 2000a; Carroll and Poole 2006) and the observation of a flaw in the OC algorithm (Tahk 2006). Thus, even basic properties of the estimator have yet to be established.

While there have been many variants of the ideal point model, many of which are aimed at avoiding potential problems, there has been little investigation into the property of the most commonly applied ideal point models. Since the incidental parameters problem has eliminated the usual proofs of consistency for the MLE, the properties of the MLE remain uncertain.

[4]In this case, the model essentially becomes equivalent to a random-effects model where the parameters of the random-effects distribution are known.

Regarding the closely-related one-parameter Rasch model, some significant reports have been made. In addition to work using conditional maximum likelihood, several authors have studied the properties of the unrestricted maximum likelihood estimator. For a Rasch model applied to k items with n test-takers, Andersen (1973) establishes inconsistency as $n \to \infty$ for fixed k and an approximation to the bias. Haberman (1977) establishes consistency as $k, n \to \infty$ provided $k^{-1} \log n \to 0$ (see also Lord 1975; Fischer 1981; Leeuw and Verhelst 1986). Others have provided evidence from simulations (e.g. Wright and Douglas 1977; van den Wollenberg, Wierda, and Jansen 1988). Nonetheless, these results have limited implications for ideal-point estimation because of the differences in models and because, as Clinton, Jackman, and Rivers (2004) point out, item-response theory is generally concerned with the estimation of the item parameters, which are equivalent to the bill parameters and generally of little interest in ideal-point estimation.

The primary results regarding the properties of the standard ideal point model come from Londregan (1999). This establishes two theorems regarding the ideal point model for a fixed number of legislators. First, that the MLE of the bill parameters is inconsistent, which follows fairly directly from the fact that they are each observed only a finite number of times. The second uses the coalition probability function, $p^* (x, \beta)$, which defines the vector of probabilities of observing each possible coalition of voters on a given vote with ideal point x and bill parameters β.

> "We say that the coalition probability function p^* is fully agenda manipulable if we can move the vector of probabilities in any direction from $p^* (x, \beta | f)$ by manipulating the agenda. Formally, the coalition probability function 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 $\xi(m,g)$, such that $\Omega(x,\beta)$, defined as

$$\Omega(x,\beta) = \left[\begin{array}{c} \int_{-\infty}^{\infty} \lambda(m,g,x,\beta)\,\xi'(m,g)\,dgdm \\ \int_{-\infty}^{\infty} \xi(m,g)'\,dgdm \end{array} \right]$$

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

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

It follows immediately from Theorem 2 and sufficiency of p^0 for x and β that the voter parameters x are not identified." (Londregan 1999, 47–48)

A number of important questions remain. Most importantly, (Londregan 1999) establishes no conditions under which p^* is fully agenda manipulable. This leaves unclear when this might be expected to hold. While the paper seems to assume that this will always be true, it never establishes that it is for any—let alone all—cases of the general model. In the case of the linear probability model of Heckman and Snyder (1997), p^* is never agenda manipulable. Further, the result is restricted to the case in which the error distribution is symmetric and the ideal points are unidimensional.

There are also some more technical problems. As Rivers (2003) points out, Londregan (1999) somewhat strangely refers to x as being unidentified when this is not, in fact, the case. Similarly, Londregan refers to p^0 as being sufficient when it is not. This points make more sense if we consider a hierarchical model in which the parameters m and g follow a hyperdistribution (without restrictions on its form), in which case both claims are correct, although this is not discussed. Finally, Londregan implicitly assumes a limiting distribution for the bill parameters exists, but it may not.

2 Model

2.1 Utility shape

While other utility models have been proposed, quadratic utility is by far the dominant approach in both formal modeling and empirical research (e.g. Clinton, Jackman, and Rivers 2004; Ladha 1991; Heckman and Snyder 1997), with the notable exception of NOMINATE (Poole and Rosenthal 1991). The use of quadratic utility makes for notable mathematical simplification and leads to a model that mirrors one in item-response theory. Thus, we will restrict our attention to ideal point models with quadratic utility.

In this model, we assume that there is a d-dimensional policy space in which each legislator i has ideal point $x_i \in \mathbb{R}^d$ and each bill j consists of a 'yea' position, $r_j^{(y)} \in \mathbb{R}^d$, and a 'nay' position, $r_j^{(n)} \in \mathbb{R}^d$. The vote of legislator i on bill j will be noted v_{ij}, where $v_{ij} = 1$ indicates a 'yea' vote and $v_{ij} = 0$ indicates a 'nay' vote.

The utility of legislator i receives utility $U_i\left(v_{ij}=1\right) = \left\|x_i - r_j^{(y)}\right\|^2 + \eta_{ij}$ from voting 'yea' and utility $U_i\left(v_{ij}=0\right) = \left\|x_i - r_j^{(n)}\right\|^2 + \nu_{ij}$ from voting 'nay,' where $\|\cdot\|$ indicates the Euclidean norm and η_{ij} and ν_{ij} are random variables such that their differences, $\epsilon_{ij} = \eta_{ij} - \nu_{ij}$,[5] are independently and identically distributed across legislators and bills according to a distribution such that $\Pr\left(\epsilon_{ij} > x\right) = F\left(x\right)$, $F\left(0\right) = \frac{1}{2}$, and F is analytic. This leads to the probability that legislator i votes 'yea' on bill j being

$$\Pr\left(v_{ij}=1\right) = F\left(x_i^\top \beta_j - \alpha_j\right),$$

[5]Note that η_{ij} and ν_{ij} need not be independent of each other or identically distributed.

where $\alpha_j = r_j^{(y)\top} r_j^{(y)} - r_j^{(n)\top} r_j^{(n)}$ and $\beta_j = 2\left(r_j^{(y)} - r_j^{(n)}\right)$. Note that this includes the common cases where ϵ_{ij} follows a normal distribution, as in Clinton, Jackman, and Rivers (2004), or a logistic distribution, as in Ladha (1991).

The alternative used by NOMINATE is for utility to be Gaussian, as defined by

$$U_i\left(v_{ij} = 1\right) = \gamma e^{-\frac{1}{2}\sum_{d=1}^{D} w_d^2 \left(x_{id} - r_{jd}^{(y)}\right)^2} + \eta_{ij} \tag{2.1}$$

$$U_i\left(v_{ij} = 0\right) = \gamma e^{-\frac{1}{2}\sum_{d=1}^{D} w_d^2 \left(x_{id} - r_{jd}^{(n)}\right)^2} + \nu_{ij}, \tag{2.2}$$

where w and γ are additional parameters. This leads to voting probabilities

$$\Pr\left(v_{ij} = 1\right) = F\left(\gamma e^{-\frac{1}{2}\sum_{d=1}^{D} w_d^2 \left(x_{id} - r_{jd}^{(y)}\right)^2} - \gamma e^{-\frac{1}{2}\sum_{d=1}^{D} w_d^2 \left(x_{id} - r_{jd}^{(n)}\right)^2}\right), \tag{2.3}$$

where F is the cumulative distribution function of $\epsilon_{ij} = \eta_{ij} - \nu_{ij}$ and ϵ_{ij} follow independent logistic or standard normal distributions. Unlike with quadratic utility, it is not possible to reparameterize this model to replace $r^{(y)}$ and $r^{(n)}$ with α and β.

2.2 Conditions

Without any restrictions, it is easy to see that we run into some immediate difficulties. First, the model is clearly only identified up to an affine transformation of the policy space. Rivers (2003) establishes that identification can be solved by, amongst other things, restricting the ideal points of the first $d+1$ legislators to be $x_i = e_i$ for $i = 1, \ldots, d$ (where e_i is the i^{th} standard basis vector in $\mathbb{R}^d$) and $x_{n+1} = 0$.

Second, even with this restriction, we run into an identification problem when $\beta_j = 0$. Moreover, even when the model is identified, we can run into problems with consistency because we approach an identification problem asymptotically (as, for example, in many cases where $\beta_j \to 0$ or $\alpha_j \to \infty$).

To avoid trivial results, we will therefore assume some basic conditions throughout:

Condition 1. $x_i = e_i$ for $i = 1, \ldots, d$ and $x_{n+1} = 0$.

Condition 2. There exists $a \in \mathbb{R}$ such that $\|\alpha_i\| \leq a$ for all i. This can generally be replaced with the assumption that α_i converges in distribution to G for some distribution G.

Condition 3. There exist $b, c \in \mathbb{R}$ such that $0 < b \leq \|\beta_i\| \leq c$ for all i. This can generally be replaced with the assumption that β_i converges in distribution to H provided H is not a point mass at zero.

These conditions are not only necessary for our theorems establishing the consistency of certain estimators, but also greatly improve counterexamples and proofs of inconsistency by ensuring that they go beyond trivial or degenerate cases.

3 Consistency of estimates of ideal points

Although standard ideal point models do not assume that the bill parameters are drawn from a common distribution, it will be helpful to consider hierarchical ideal point models in which the bill parameters are treated as random variables drawn independently across bills from a common distribution. As in Londregan (1999), it will be easier to represent the model in terms of the probability of observing each of the 2^k possible voting patterns,[6] which we shall denote $p^*(x, \alpha, \beta)$, where k is the number of legislators, rather than a vector of k elements giving the probability of each individual voting 'yea.' Without restrictions, these hierarchical ideal point models are clearly not identified with respect to the bill parameters. The more important question is whether such a model is identified with respect to the ideal points.[7]

[6]Londregan refers to the 2^{k-1} coalition probabilities, ignoring the difference between a vote in which a given set of voters all vote 'yea' with all others voting 'nay' and a vote in which the same coalition votes 'nay' with all others voting 'yea.' This distinction is, in many ways, irrelevant if the error distribution is symmetric, but we use the larger vector which makes this distinction in order to avoid this assumption and other technicalities.

[7]See appendix for the definition of identification with respect to a parameter.

Because it will be easier to prove results about the the hierarchical model, it is useful to establish a relationship between the identification of the hierarchical model and the consistency of the non-hierarchical model. The following lemma establishes a straight-forward connection.

Lemma 3.1. *If the hierarchical ideal point model is not identified with respect to the ideal points, then, given any estimator, $h(X)$, then there does not exist an estimator of the ideal points that is consistent for all possible ideal points, x, and sequence of bill parameters, $\{\alpha_i, \beta_i\}$.*

Proof. See Appendix. □

In essence, if the hierarchical model led to two pairs of ideal points and distributions that imply the same distribution to the data, then this implies that this also holds for the same ideal points and some sample from each of the two distributions.

3.1 Inconsistency results

As Londregan (1999) concluded, there appear to be significant issues with the consistent estimation of the standard ideal-point models in cases where the number of legislators are small. The inconsistency of maximum likelihood estimators can be established in all cases—without the conditions needed by Theorem 2 in Londregan (1999)—by a simple counterexample,[8] leading to the following result:

Theorem 3.1. *For all possible error distributions, the maximum likelihood estimator is not consistent provided either the parameter estimates are not constrained or the parameter estimates are constrained according to Conditions 2 and 3 for sufficiently large values a and b and sufficiently small value b.*

[8] This counterexample is discussed in more detail in the next section, which gives a stronger result establishing the inconsistency of maximum likelihood estimators as estimators of the rank order of the ideal points.

Proof. See Appendix. □

Thus, maximum likelihood never provides a consistent estimator for the ideal points.

The inconsistency of maximum likelihood estimators does not itself imply that consistent estimators might not exist. However, almost all ideal point models[9] use quadratic or Gaussian utility combined with logistic or normal error distributions. Much stronger results establishing the inconsistency of all estimators are possible for these models. Proofs of these results begin with a similar approach to that of Londregan (1999). However, whereas Londregan (1999) did not establish the existence of the necessary bill parameter distributions in any case, these proofs do so.

In the case of the quadratic-logistic model, this leads to the following result:

Theorem 3.2. *If utility is quadratic and the error distribution is logistic (i.e.,* $F(x) = (1 + e^{-x})^{-1}$*), then the hierarchical model is unidentified with respect to the ideal points.*

Proof. See Appendix. □

Thus, the hierarchical version of the quadratic-logistic model will always face an identification problem with respect to the ideal points. A similar result applies for normal errors:

Theorem 3.3. *If utility is quadratic and the error distribution is normal (i.e.,* $F(x) = \Phi(x)$*), then the hierarchical model is unidentified with respect to the ideal points.*

Proof. See Appendix. □

Using Lemma 3.1, this also implies:

Corollary 3.1. If utility is quadratic and the error distribution is logistic or normal, there does not exist an estimator that is consistent with respect to the ideal points.

Proof. This follows directly from Theorems 3.2 and 3.3 and Lemma 3.1. □

[9] The linear probability model of Heckman and Snyder (1997) is the primary exception.

Although these results are applicable only to quadratic utility models, NOMINATE (Poole and Rosenthal 1991) uses a Gaussian utility model. However, the quadratic utility model can be viewed as a limiting case of this Gaussian utility model. If $w_d = \frac{1}{\gamma}$ for all d, then the limit as $\gamma \to \infty$ leads to

$$\lim_{\gamma\to\infty} \Pr(v_{ij}=1) = \lim_{\gamma\to\infty} F\left(\gamma e^{-\frac{1}{\gamma}\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(y)}\right)^2} - \gamma e^{-\frac{1}{\gamma}\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(n)}\right)^2}\right). \tag{3.1}$$

Assuming F is continuous, this becomes

$$\lim_{\gamma\to\infty} \Pr(v_{ij}=1) \tag{3.2}$$
$$= F\left(\lim_{\gamma\to\infty} \gamma e^{-\frac{1}{\gamma}\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(y)}\right)^2} - \gamma e^{-\frac{1}{\gamma}\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(n)}\right)^2}\right). \tag{3.3}$$
$$= F\left(\gamma\left(1-\frac{1}{\gamma}\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(y)}\right)^2\right) - \gamma\left(1-\frac{1}{\gamma}\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(n)}\right)^2\right)\right) \tag{3.4}$$
$$= F\left(\sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(n)}\right)^2 - \sum_{d=1}^{D}\left(x_{id}-r_{jd}^{(y)}\right)^2\right), \tag{3.5}$$

which is a quadratic utility model.

This relationship between the NOMINATE model and quadratic utility models helps to extend the results regarding quadratic utility models to the NOMINATE model.

Theorem 3.4. *If utility is Gaussian and the error distribution is logistic or normal, the hierarchical model is unidentified with respect to the ideal points.*

Proof. See Appendix. □

As before, this has negative implications for the ability to consistently estimate the standard non-hierarchical NOMINATE model.

Corollary 3.2. If utility is Gaussian and the error distribution is logistic or normal, there does not exist an estimator that is consistent with respect to the ideal points.

Proof. This follows directly from Theorem 3.4 and Lemma 3.1. □

4 Consistency of estimates of rank order

For many purposes, precise estimates of the ideal points are unnecessary and only the rank order is needed. Indeed, Ho and Quinn (2010) argue for using only the ordinal values of of ideal-point estimates on the grounds that the cardinal values are "extremely sensitive to trivial changes in modeling assumptions." Although the previous results establish that consistent estimation of the cardinal values of ideal points is not possible, they do not rule out the possibility of consistently estimating their rank order.

Note that rank order is only meaningful when applied to a single dimension. Thus, the discussion of the estimation of rank order of the ideal points in this section will be limited to one-dimensional models throughout.

4.1 The Quinn Conjecture

The proposal to use only the rank order of ideal-point estimates raises the obvious question of whether—and which—available estimators produce consistent estimates of the rank order. Ho and Quinn (2010) provide evidence that the rank order of the estimates tends to be similar across models but not that these estimates are consistent for any model. (Poole 2005) refers to the possibility that these estimates are consistent as the Quinn Conjecture:

> "If the voting space is one-dimensional and the noise process is symmetric, then as the number of roll calls goes to infinity, the true rank ordering of the legislators will be recovered.
>
> More technically, suppose the error ϵ is drawn from some continuous symmetric probability distribution with mean 0 and finite variance σ^2

having support on the real line; then the Quinn Conjecture is: If $s = 1$ and $\epsilon \sim f(0, \sigma^2)$ with $-\infty < \epsilon < \infty$ and $\sigma^2 < \infty$, then as $q \to \infty$, $\Pr(X_1 < X_2 < X_3 < \cdots < X_{p-2} < X_{p-1} < X_p | V) \to 1$ where V is the p by q matrix of observed votes and the legislator indices have been permuted so that they correspond to the true ordering of the ideal points." (Poole 2005, 206–207)

This statement of the conjecture leaves some ambiguity. First, it does not explicitly exclude trivial counterexamples (such as when $\beta_j = 0$ for all j or even many cases where $\beta_j \to 0$, $\beta_j \to \pm\infty$, or $\alpha_j \to \pm\infty$) and does not provide a restriction to distinguish between one ordering and its reverse. This is easy to fix by adding the second condition of Subsection 2.2 and combined with the identifying restriction that $x_1 < x_2$.

Second, it is not entirely clear what estimator is being conjectured to be consistent, although it is discussed in the context of Optimal Classification. We will consider three variants of the Quinn Conjecture: that it holds for Optimal Classification,[10] that it holds for maximum likelihood, and that there exists some estimator for which it holds.

4.2 Optimal classification

Optimal Classification (OC) was introduced as a non-parametric estimator of ideal points by Poole (2000b). OC seeks to provide estimated ideal points and cutting points (or, in multiple dimensions, cutting hyperplanes) that minimize the number of misclassifictions, where all ideal points on one side of a cutting point classified as predicted "yea" votes and all on the other side as predicted "nay" votes. The question of the consistency of OC estimates is complicated by its lack of a probabilistic model,

[10]There exists an additional—albeit correctable—complication that Poole's OC algorithm Poole (2000b) does not always lead to a solution that satisfies the optimal classification criterion (see Tahk 2006).

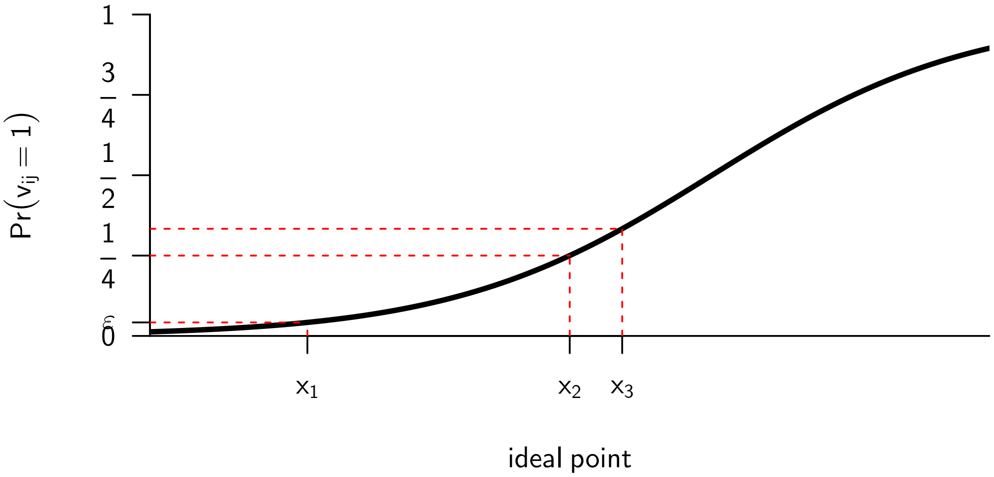

Figure 4.1: A graph of the ideal points and voting parameters described in Section 4.2

parametric or otherwise. We will consider it in the context of models with concave utility and a continuous error distribution.

Other work exploring the conjecture has tended to support it. Carroll and Poole (2006) provide Monte Carlo evidence on the Quinn Conjecture for Optimal Classification which, "though limited, show strong support for the Quinn Conjecture." Poole (2005) also states a belief that the conjecture is true. The following counterexample will demonstrate otherwise.

Consider the case with three ideal points, $x_1 < x_2 < x_3$, and bill parameters such that the probabilities of each legislator voting yea on bill j is given by

$$Pr\left(v_{1j}=1\right)=\varepsilon \qquad Pr\left(v_{2j}=1\right)=\frac{1}{4} \qquad Pr\left(v_{3j}=1\right)=\frac{1}{3}$$

where $0 \leq \varepsilon < \frac{1}{4}$. Note that, as long as the error distribution is continuous, parameters that allow these probabilities must exist. This situation is depicted in Figure 4.1.

Table 4.1 shows the probabilities of different voting patterns for each bill along with the number of misclassifications for the optimal cutting point given the true

Table 4.1: Vote probabilities for the counterexample to the Quinn conjecture for Optimal Classification

Votes				If $x_1 < x_2 < x_3$		If $x_3 < x_1 < x_2$	
v_{1j}	v_{2j}	v_{3j}	Probability	Pattern	Misclass.	Pattern	Misclass.
0	0	0	$(1-\varepsilon)/2$	NNN	0	NNN	0
0	0	1	$(1-\varepsilon)/4$	NNY	0	YNN	0
0	1	0	$(1-\varepsilon)/6$	NYN	1	NNY	0
0	1	1	$(1-\varepsilon)/12$	NYY	0	YNY	1
1	0	0	$\varepsilon/2$	YNN	0	NYN	1
1	0	1	$\varepsilon/4$	YNY	1	YYN	0
1	1	0	$\varepsilon/6$	YYN	0	NYY	0
1	1	1	$\varepsilon/12$	YYY	0	YYY	0
Expected misclassifications				$\frac{2}{12}+\frac{1}{12}\varepsilon$		$\frac{1}{12}+\frac{5}{12}\varepsilon$	

ordering of the ideal points and given the ordering $x_3 < x_1 < x_2$. The expected number of misclassifications is $\frac{2}{12}+\frac{1}{12}\varepsilon$ under the true ordering and $\frac{1}{12}+\frac{5}{12}\varepsilon$ under the alternative ordering. Since $\varepsilon < \frac{1}{4}$, the expectation under the true ordering is greater than under the ordering $x_3 < x_1 < x_2$.

Because votes are independent across bills, the weak law of large numbers implies that the probability that the number of misclassifications under the true ordering is greater under the true ordering than under the alternative is arbitrarily close to one for a sufficiently large number of bills. Thus, not only is OC not a consistent estimator here, but the probability that the OC estimate is incorrect approaches one as the number of bills increases. This counterexample can also be extended to apply to an arbitrary number of legislators by choosing ideal points such that $\Pr(v_{ij} = 1 \forall i \notin \{2,3\}) < \varepsilon$. This disproves the Quinn conjecture for Optimal Classification.

Table 4.2: Vote probabilities for the counterexample to the Quinn conjecture for maximum likelihood

Votes				If $x_1 < x_2 < x_3$		If $x_3 < x_1 < x_2$	
v_{1j}	v_{2j}	v_{3j}	Probability	Pat.	Log-lik.	Pat.	Log-lik.
0	0	0	$(1-\varepsilon)/2$	NNN	0	NNN	0
0	0	1	$(1-\varepsilon)/4$	NNY	0	YNN	0
0	1	0	$(1-\varepsilon)/6$	NYN	$< -2\log 2$	NNY	0
0	1	1	$(1-\varepsilon)/12$	NYY	0	YNY	$\geq -3\log 2$
1	0	0	$\varepsilon/2$	YNN	0	NYN	$\geq -3\log 2$
1	0	1	$\varepsilon/4$	YNY	$< -2\log 2$	YYN	0
1	1	0	$\varepsilon/6$	YYN	0	NYY	0
1	1	1	$\varepsilon/12$	YYY	0	YYY	0
Expected log-likelihood				$< -\left(\frac{1}{3}+\frac{1}{6}\varepsilon\right)\log 2$		$\geq -\left(\frac{1}{4}+\frac{15}{12}\varepsilon\right)\log 2$	

4.3 Maximum likelihood

A similar counterexample can be used with a maximum-likelihood estimator. Again, assume there are three legislators with $x_1 < x_2 < x_3$ and bill parameters such that the probabilities of each legislator voting yea on bill j is given by

$$Pr\left(v_{1j}=1\right)=\varepsilon \qquad Pr\left(v_{2j}=1\right)=\frac{1}{4} \qquad Pr\left(v_{3j}=1\right)=\frac{1}{3}$$

but now assume $0 \leq \varepsilon < \frac{1}{13}$. A table of the probabilities of each vote pattern and bounds on the log-likelihood under that vote pattern are given in Table 4.2.

From Table 4.2 and the assumption that $\varepsilon < \frac{1}{13}$, we see that the upper bound on the expected log-likelihood under the true ordering is less than the lower bound on the expected log-likelihood under the alternative ordering $x_3 < x_1 < x_2$. Thus, the weak law of large numbers implies that the probability that the log-likelihood using the alternative ordering is greater than the log-likelihood using the true ordering approaches one as the number of bills increases. Thus, this Quinn conjecture is also false for maximum likelihood. The maximum likelihood estimator is not consistent and can produce an incorrect ordering with arbitrarily high probability.

4.4 The existence of a consistent estimator

The counterexamples to the consistency of OC and maximum likelihood depend on the fact that the polarity of a bill—i.e., the sign of the discrimination parameter—is unknown and can be misestimated. If, instead, the sign of each discrimination parameter were known, then estimating the rank order would be much simpler. As long as the utility function is quadratic or, at least, concave, the probability of a "yea" vote when the discrimination parameter is positive or "nay" vote when it is negative is an increasing function of the ideal point of a legislator. Thus, under some regularity conditions, simply adding up the number of "yea" votes where the discrimination parameter is positive and "nay" votes where the discrimination parameter is negative produces a consistent estimate of the ideal points. ADA scores can be thought of as estimates of this type.

We can use a similar approach without knowing the sign of the discrimination parameter. Consider the quantity $z_{ij} = v_{ij}(v_{2j} - v_{1j})$. By independence of v_{in}, $E[z_{in}] = (F(\alpha_n + \beta_n x_2) - F(\alpha_n + \beta_n x_1)) F(\alpha_n + \beta_n x_i)$. Since F is increasing and $x_1 < x_2$ by assumption, $(F(\alpha_n + \beta_n x_2) - F(\alpha_n + \beta_n x_1))$ is positive. Therefore, z_{in} is strictly increasing in x_i.

Now consider two legislators, i and j, such that $x_i < x_j$. Since F is increasing, $E[z_{jn} - z_{in}] = (F(\alpha_n + \beta_n x_2) - F(\alpha_n + \beta_n x_1))(F(\alpha_n + \beta_n x_j) - F(\alpha_n + \beta_n x_i)) \geq \frac{1}{2}$ with equality if and only if $\beta_n = 0$. Thus, as n increases, the probability that $\bar{z}_{jn} > \bar{z}_{in}$ where $\bar{z}_{in} = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$ and $\bar{z}_{jn} = \frac{1}{n}\sum_{k=1}^{n} z_{jk}$ must also increase. If the limit of this probability is one, then $1_{\bar{z}_j > \bar{z}_i}$ is a consistent estimator of $1_{x_j > x_i}$. From this, we can construct a consistent estimator of the order of $x_3, \ldots, x_K$.

This result leaves two problems. First, the limit of this probability might not approach one. Indeed, without any restrictions, this is clearly possible.[11] We can avoid this my imposing relatively mild regularity conditions. The conditions of Subsection 2.2 are sufficient. Note that $\bar{z}_{in}$ and $\bar{z}_{jn}$ need not converge for $\Pr(\bar{z}_{jn} > \bar{z}_{in})$ to converge.

More importantly, we have only estimated the order of $x_3, \ldots x_K$. We do not know where x_1 and x_2 fall in this order aside the restriction that $x_1 < x_2$ nor could we use $\bar{z}_1$ and $\bar{z}_2$ to find this out as we relied on the independence of v_{1n}, v_{2n}, and v_{in} in establishing the properties of $\bar{z}_i$. However, we can perform the same process based around $z_{in}^{(p,q)} = (v_{qn} - v_{pn})\, v_{in}$ for all $i \notin \{p, q\}$ where $q, p \notin \{1, 2\}$ and, importantly, $\bar{z}_{pn}^{(1,2)} < \bar{z}_{qn}^{(1,2)}$. This allows us to asymptotically recover the order of x_i for $i \notin \{p, q\}$. We therefore have a method of positioning the order of x_i relative to x_j and x_k. Doing so, however, requires distinct values for i, j, k, p, q and, therefore, requires that there be at least five legislators, leaving no contradition with Theorem 4.1. This leaves unspecified how to deal with cases in which different choices of p and q return conflicting results for the order of x_i, x_j, and x_k. However, so long as this situation arises asymptotically with probability zero, it is irrelevant how it is resolved—the estimator will still be consistent.

The following theorem formalizes this demonstration that a consistent estimator can be constructed:

Theorem 4.1. *If the number of legislators $K \geq 5$, there exists an estimator of the order of the ideal points which is consistent for any set of ideal points $x_1, \ldots, x_K$ and sequence of bill parameters $\alpha_1, \ldots$ and $\beta_1, \ldots$ satisfying the conditions of Subsection 2.2.*

Proof. See Appendix. □

While the estimator used above is not a particularly efficient use of the data, it does demonstrate that the data contain sufficient information to identify the ideal

[11]For example, when $\beta_n = 0$ for all n or $\beta_n = \frac{1}{n}$.

points up to a monotonic transformation so long as there are at least five voters despite our inability to consistently estimate the ideal points themselves.

4.5 Maximum likelihood with known bill polarity

In some cases, the polarity of each bill may be known and not require estimation from the data. Although simpler consistent estimators are possible in this case, we might still desire to use a maximum-likelihood estimator to increase efficiency. Unlike in the case where the polarity is unknown, maximum likelihood can be shown to be consistent in this case.

Theorem 4.2. *If the error distribution is continuous, the sign of each discrimination parameter, β_j, is known, and the parameter space of each β_j is constrained to be nonpositive or nonnegative according to its sign, then the maximum likelihood estimator is a consistent estimator of the order of the ideal points for any set of ideal points $x_1, \ldots, x_K$ and sequence of bill parameters $\alpha_1, \ldots$ and $\beta_1, \ldots$ satisfying the conditions of Subsection 2.2.*

Proof. See Appendix. □

5 Conclusion

These results suggest that treating ideal-point estimates as cardinal values is problematic for both quadratic utility models like that of Clinton, Jackman, and Rivers (2004) and Gaussian utility models like NOMINATE. Although Ho and Quinn (2010) have already argued in favor of using only ordinal information from ideal-point estimates on the grounds of sensitivity to parametric assumptions, these results strengthen that case. Even if the parametric assumptions are correct and we observe an infinite number of bills, we cannot consistently estimate the cardinal value of the ideal points. The properties of estimates are only likely to be worse in the more common situation in which the parametric assumptions are not precisely correct and the number of bills is finite.

Even when using only the rank order of ideal-point estimates, problems can emerge. Avoiding manual coding of the direction of bills is sometimes regarded as "one main virtue" of ideal-point estimation (Ho and Quinn 2010). However, as shown in this paper, ideal-point estimators can also fail to accurately code the direction of a bill. When this occurs sufficiently often, even estimates of rank order can be inconsistent. Data on the direction of each bill can prevent this problem but often require additional data collection. Further, when the direction of each bill is known, likelihood-based estimators provide consistent estimates of rank order even when their parametric assumptions are incorrect.

This is not to say that previous scholarship using cardinal estimates of ideal points should necessarily be discarded. However, if results depend on the cardinal values of the estimates, then there is cause for suspicion. Problems with the estimates of rank order are also appear more likely in some situations than others. For example, these problems may be unlikely with votes from Congress given the large number of voters and rarity of votes with only one or two voters in the minority. On the other hand, because the Supreme Court has only a small number of voters and minorities of one are not uncommon, these problems may be more likely to emerge with Supreme Court data.

Finally, the development of additional tools for nonparametric ideal-point estimation may be useful. Although existing estimators, including the nonparametric Optimal Classification, may not always provide consistent estimates of ideal points, consistent estimation of the rank order of ideal points is possible without relying on parametric assumptions. Further, if only ordinal information is used, it seems sensible to use estimators and hypothesis tests aimed at ordinal estimation.

References

Andersen, Erling B. 1973. *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygiejnisk Forlag.

Bailey, Michael. 2001. "Ideal Point Estimation with a Small Number of Votes: A Random-Effects Approach." *Political Analysis* 9 (January): 192–210.

Blackwell, David, and Lester Dubins. 1962. "Merging of Opinions with Increasing Information." *The Annals of Mathematical Statistics* 33 (3): 882–886.

Carroll, Royce, and Keith T. Poole. 2006. "Interval Properties of Ideal Point Estimators." Presented at the Conference on Legislative Behavior in Europe, the US, and Beyond, hosted by the Center on Institutions and Governance, Del Mar, CA.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355–370.

Diaconis, Persi, and David Freedman. 1986a. "On the Consistency of Bayes Estimates." *The Annals of Statistics* 14 (Mar): 1–26.

———. 1986b. "Rejoinder: On the Consistency of Bayes Estimates." *The Annals of Statistics* 14 (Mar): 63–67.

Fischer, Gerhard. 1981. "On the Existence and Uniqueness of Maximum-Likelihood Estimates in the Rasch Model." *Psychometrika* 46 (March): 59–77.

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

Heckman, James J., and James S. Snyder, Jr. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *The RAND Journal of Economics* 28 (Special Issue in Honor of Richard E. Quandt): S142–S189.

Ho, Daniel E., and Kevin M. Quinn. 2010. "How Not to Lie with Judicial Votes: Misconceptions, Measurement, and Models." *California Law Review* 98 (3): 813–876.

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

Lancaster, Tony. 2000. "The Incidental Parameter Problem since 1948." *Journal of Econometrics* 95 (2): 391–413.

Leeuw, Jan De, and Norman Verhelst. 1986. "Maximum Likelihood Estimation in Generalized Rasch Models." *Journal of Educational and Behavioral Statistics* , 183 –196.

Lewis, Jeffrey B. 2001. "Estimating Voter Preference Distributions from Individual-Level Voting Data." *Political Analysis* 9 (January): 275–297.

Londregan, John. 1999. "Estimating Legislators' Preferred Points." *Political Analysis* 8 (1): 35–56.

Lord, Frederick. 1975. *Consistent Estimation when Number of Variables and Number of Parameters Increase without Limit.* Princeton: Educational Testing Service.

Neyman, Jerzy, and Elizabeth L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16 (January): 1–32.

Peress, Michael. 2009. "Small Chamber Ideal Point Estimation." *Political Analysis* 17 (July): 276–290.

Poole, Keith T. 2000a. "Appendix to Non-parametric Unfolding of Binary Choice Data." Unpublished manuscript. www.voteview.com/paapp2.pdf

———. 2000b. "Nonparametric Unfolding of Binary Choice Data." *Political Analysis* 8 (3): 211–237.

———. 2005. *Spatial Models of Parliamentary Voting.* Cambridge: Cambridge University Press.

Poole, Keith T., and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35 (1): 228–278.

Rivers, Douglas. 2003. "Identification of Multidimensional Spatial Voting Models." Unpublished manuscript.

Tahk, Alexander. 2006. "A New Approach to Optimal Classification." Paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA.

van den Wollenberg, Arnold L., Folkert W. Wierda, and Paul G. W. Jansen. 1988. "Consistency of Rasch Model Parameter Estimation: A Simulation Study." *Applied Psychological Measurement,* 12: 307–313.

Wright, Benjamin D., and Graham A. Douglas. 1977. "Conditional Versus Unconditional Procedures for Sample-Free Item Analysis." *Educational and Psychological Measurement* 37 (October): 573–586.

Appendix

Definitions

Before discussing identification and consistency of ideal-point models, it is

Definition 1.1. A *discrete statistical model* p_θ consists of a set of probability mass functions $\mathcal{P} = \{p_\theta : \theta \in \Omega\}$ indexed by $\theta \in \Omega$.

As in Rivers (2003), we restrict the definition to cases where $\mathcal{P}$ is dominated by counting measure to avoid measure-theoretic complications which are irrelevant the present inquiry. However, we do not, for the time being, restrict Ω to be a subset of Euclidean space, nor do we restrict it to be finite.

Definition 1.2. If X is distributed according to p_θ, then θ is said to be identified on the basis of X if for all $\theta_1 \neq \theta_2$, $p_{\theta_1} \neq p_{\theta_2}$.

We extend this definition of identification slightly to cover functions of the parameter space. Since we will be primarily concerned with the identification of ideal points without regard to the bill parameters, this is an important feature.

Definition 1.3. If X is distributed according to p_θ, then $g(\theta)$ is said to be identified on the basis of X if, for all θ_1, θ_2, $g(\theta_1) \neq g(\theta_2)$ implies $p_{\theta_1} \neq p_{\theta_2}$.

These definitions lead to analogous definitions for the identification of a model.

Definition 1.4. A class of discrete statistical models $\mathcal{P}$ is said to be identified if for all $\theta_1 \neq \theta_2 \in \Omega$, $p_{\theta_1} \neq p_{\theta_2}$.

Definition 1.5. A class of discrete statistical models $\mathcal{P}$ is said to be identified with respect to g if for all $\theta_1, \theta_2 \in \Omega$, $g(\theta_1) \neq g(\theta_2)$ implies $p_{\theta_1} \neq p_{\theta_2}$.

Proofs

Lemma 3.1. *If the hierarchical ideal point model is not identified with respect to the ideal points, then, given any estimator,* $h(X)$*, then there does not exist an estimator of the ideal points that is consistent for all possible ideal points,* x*, and sequence of bill parameters,* $\{\alpha_i, \beta_i\}$*.*

Proof. Let $\mathcal{P}\left(p^*_{x,\nu} \in \Omega\right)$ be a hierarchical ideal-point model with k legislators, where x indexes the ideal points, ν indexes the distribution of the bill parameters with density function ν defined with respect to a common dominating measure, and $p^*_{x,\nu}$ is defined in terms of the non-hierarchical ideal-point model as

$$p_{x,\nu}(V) = \int p^*_{x,\alpha,\beta}(V)\, \nu(\alpha,\beta)\, d\mu_{\alpha,\beta}.$$

Assume $\mathcal{P}$ is not identified with respect to the ideal points. Therefore, there exist ideal points x and x' and distributions ν and ν' such that $x \neq x'$ but $p^*_{x,\nu} = p^*_{x',\nu'}$. Take any such $x \neq x'$, ν, and ν'. So,

$$\int p^*_{x,\alpha,\beta}(V)\, \nu(\alpha,\beta)\, d\mu_{\alpha,\beta} = \int p^*_{x',\alpha,\beta}(V)\, \nu'(\alpha,\beta)\, d\mu_{\alpha,\beta}.$$

Assume h is a consistent estimator of the ideal points under all possible ideal points, x, and sequence of bill parameters, $\{\alpha_i, \beta_i\}$. Thus, if $V_i \overset{iid}{\sim} p^*_{x,\alpha_i,\beta_i}$, then $h(V_1, \ldots, V_n) \xrightarrow{p} x$.

Let $\epsilon = \frac{\|x-x'\|}{2}$. Note that since $x \neq x'$, $\epsilon > 0$. Let

$$A_n = \{(V_1, \ldots, V_n) : \|h(V_1, \ldots, V_n) - x\| < \epsilon\}$$

and

$$B_n = \{(V_1, \ldots, V_n) : \|h(V_1, \ldots, V_n) - x'\| < \epsilon\}.$$

If $\|h(V_1,\ldots,V_n)-x\|<\epsilon=\frac{\|x-x'\|}{2}$ and $\|h(V_1,\ldots,V_n)-x'\|<\epsilon=\frac{\|x-x'\|}{2}$, then

$$\|h(V_1,\ldots,V_n)-x\|+\|h(V_1,\ldots,V_n)-x'\|<\|x-x'\|.$$

But by the triangle inequality, $\|h(V_1,\ldots,V_n)-x\|+\|h(V_1,\ldots,V_n)-x'\|>\|x-x'\|$, so there can be no such $h(V_1,\ldots,V_n)$. Therefore A_n and B_n are disjoint.

Since h is a consistent estimator, $\lim_{n\to\infty} p^*_{x,\alpha,\beta}(A_n)=1$ and $\lim_{n\to\infty} p^*_{x',\alpha',\beta'}(B_n)=1$. Thus, there exists an N such that, for all $n>N$, $\Pr(A_n)>\frac{2}{3}$. Similarly, there exists an M such that, for all $m>M$, $\Pr(B_m)>\frac{2}{3}$. Take any such N and M. Let $n=\max(M,N)+1$. So, $p^*_{x,\alpha,\beta}(\|h(V_1,\ldots,V_n)-x\|<\epsilon)>\frac{2}{3}$ and $p^*_{x',\alpha',\beta'}(\|h(V_1,\ldots,V_n)-x'\|<\epsilon)>\frac{2}{3}$.

Consider $p^*_{x,\nu}(A_n)$. By definition,

$$p^*_{x,\nu}(A_n)=\int p^*_{x,\alpha,\beta}(A_n)\,\nu(\alpha,\beta)\,d\mu_{\alpha,\beta}\geq\int\frac{2}{3}\nu(\alpha,\beta)\,d\mu_{\alpha,\beta}=\frac{2}{3}.$$

Likewise,

$$p^*_{x',\nu'}(B_n)=\int p^*_{x',\alpha,\beta}(B_n)\,\nu'(\alpha,\beta)\,d\mu_{\alpha,\beta}\geq\int\frac{2}{3}\nu'(\alpha,\beta)\,d\mu_{\alpha,\beta}=\frac{2}{3}.$$

But since $p^*_{x,\nu}=p^*_{x',\nu'}$,

$$\int p^*_{x,\alpha,\beta}(B_n)\,\nu(\alpha,\beta)\,d\mu_{\alpha,\beta}=\int p^*_{x',\alpha,\beta}(B_n)\,\nu'(\alpha,\beta)\,d\mu_{\alpha,\beta}.$$

Thus, $\int p^*_{x,\alpha,\beta}(B_n)\,\nu(\alpha,\beta)\,d\mu_{\alpha,\beta}\geq\frac{2}{3}$.

Since A_n and B_n are disjoint,

$$\begin{aligned}\int p^*_{x,\alpha,\beta}(A_n\cup B_n)\,\nu(\alpha,\beta)\,d\mu_{\alpha,\beta} &= \int p^*_{x,\alpha,\beta}(B_n)\,\nu(\alpha,\beta)\,d\mu_{\alpha,\beta}+\int p^*_{x,\alpha,\beta}(B_n)\,\nu(\alpha,\beta)\,d\mu_{\alpha,\beta}\\ &\geq \frac{4}{3}.\end{aligned}$$

But since $p^*_{x,\nu}$ is defined to be a probability measure,

$$p^*_{x,\nu}\left(A_n \cup B_n\right) = \int p^*_{x,\alpha,\beta}\left(A_n \cup B_n\right) \nu\left(\alpha,\beta\right) d\mu_{\alpha,\beta} > 1$$

is a contradition. Therefore, does not exist an estimator of the ideal points that is a consistent for all possible ideal points, x, and sequence of bill parameters, $\{\alpha_i, \beta_i\}$. $\square$

Theorem 3.1. *For all possible error distributions, the maximum likelihood estimator is not consistent provided either the parameter estimates are not constrained or the parameter estimates are constrained according to Conditions 2 and 3 for sufficiently large values a and b and sufficiently small value b.*

Proof. Consider the case where the number of dimensions $d = 1$ and $k \geq 3$ legislators. Let v_{ij} denote the vote of legislator i on bill j and equal one for a 'yea' vote (with probability $F\left(\beta_j x_i - \alpha_j\right)$) and zero for a 'nay' vote. Let ℓ_{ij} denote the log-likelihood function for observation i, j, $\ell_{ij}\left(x'_i, \alpha'_j, \beta'_j\right) = v_{ij} \log F\left(x'_i \beta'_j - \alpha'_j\right) + \left(1 - v_{ij}\right) \log F\left(x'_i \beta'_j - \alpha'_j\right)$, ℓ_j denote the log-likelihood for the observations from vote j, $\ell_j\left(x', \alpha'_j, \beta'_j\right) = \sum_i \ell_{ij}\left(x'_i, \alpha'_j, \beta'_j\right)$, and ℓ denote the total log-likelihood function, $\ell\left(x', \alpha', \beta'\right) = \sum_j \ell_j\left(x', \alpha', \beta'\right)$.

Because of the identifying constraints, $x_1 = 1$ and $x_2 = 0$. Let

$$\alpha_j = F^{-1}\left(1 - \left(1 - \frac{1}{6k^2 - 1}\right)^{\frac{1}{k-2}}\right)$$

and $\beta_j = F^{-1}\left(\frac{1}{2k}\right) + \alpha_j$ for all j. Let $x_3 = \left(F^{-1}\left(\frac{1}{3k}\right) + \alpha_j\right)\beta_j^{-1}$. Finally, take any $x_i < 0$ for all $i > 3$. Note that this implies that

$$F\left(\beta_j x_3 - \alpha_j\right) = \frac{1}{3k} < \frac{1}{2k} = F\left(\beta_j x_1 - \alpha_j\right).$$

Further, $F(\beta_j x_i - \alpha_j) < 1 - \left(1 - \frac{1}{6k^2-1}\right)^{\frac{1}{k-2}} = F(\beta_j x_1 - \alpha_j)$, so

$$\Pr(v_{1j} = 0 \wedge v_{ij} = 0 \forall i > 3) = \frac{6k^2-2}{6k^2-1} > \Pr(v_{3j} = 0) = \frac{3k-1}{3k} > \Pr(v_{3j} = 0) = \frac{2k-1}{2k}.$$

First, note that in all cases, $\ell_j\left(x'\alpha'_j, 0\right) = k \log F(0) = -k \log 2$ as the log-likelihood when $\beta'_j = 0$ does not depend on v_{ij}. Further, the likelihood of an observation cannot exceed one, so the log-likelihood cannot exceed zero. Thus, in all cases, $-k \log 2 \leq \sup_{\alpha',\beta'} \ell_j(x', \alpha', \beta') \leq 0$.

Consider the case $\hat{x}_1 = 1$, $\hat{x}_2 = 0$, $\hat{x}_3 = -1$, $-1 < \hat{x}_i < 0 \forall i > 3$. Take any such values, $\hat{x}$. If $v_{1j} = 1$ and $v_{ij} = 0$ for all $i \neq 1$, then the data are perfectly separating and so $\sup_{\hat{\alpha},\hat{\beta}} \ell_j\left(\hat{x}, \hat{\alpha}, \hat{\beta}\right) = 0$. If $v_{3j} = 1$ and $v_{ij} = 0$ for all $i \neq 3$, then the data are again perfectly separating and so $\sup_{\hat{\alpha},\hat{\beta}} \ell_j\left(\hat{x}, \hat{\alpha}, \hat{\beta}\right) = 0$. And if $v_{ij} = 0$ for all i, then in the limit as $z \to \infty$, $\ell_j(\hat{x}, z, z^2) \to 0$ and so $\sup_{\hat{\alpha},\hat{\beta}} \ell_j\left(\hat{x}, \hat{\alpha}, \hat{\beta}\right) = 0$. These three outcomes occur with probability

$$\begin{aligned} &\Pr(v_{ij} = 0 \forall i) + \Pr(v_{1j} = 1, v_{ij} = 0 \forall i \neq 1) + \Pr(v_{3j} = 1, v_{ij} = 0 \forall i \neq 3) > \\ &\frac{2k-1}{2k} \cdot \frac{3k-1}{3k} \cdot \frac{6k^2-2}{6k^2-1} + \frac{1}{2k} \cdot \frac{3k-1}{3k} \cdot \frac{6k^2-2}{6k^2-1} + \frac{2k-1}{2k} \cdot \frac{1}{3k} \cdot \frac{6k^2-2}{6k^2-1} = \frac{3k^2-1}{3k^2} \end{aligned}$$

Since $\sup_{\hat{\alpha},\hat{\beta}} \ell_j\left(\hat{x}, \hat{\alpha}, \hat{\beta}\right) \geq -k \log 2$ for all other cases and these occur with probability less than $\frac{1}{3k^2}$, we have

$$E\left[\sup_{\hat{\alpha}_j,\hat{\beta}_j} \ell_j\left(\hat{x}, \hat{\alpha}_j, \hat{\beta}_j\right)\right] > -\frac{1}{3k} \log 2.$$

Now, consider the case where $\tilde{x}_1 = 1$, $\tilde{x}_2 = 0$, $\tilde{x}_3 > 0$, and $\tilde{x}_i < 0$ for all $i > 3$. If $v_{3j} = 1$ and $v_{ij} = 0$ for all $i \neq 3$, then since $\tilde{x}_2 < \tilde{x}_3 < \tilde{x}_1$, the likelihood cannot exceed $\frac{1}{4}$ and, thus, $\sup_{\tilde{\alpha}_j,\tilde{\beta}_j} \ell_j\left(\tilde{x}, \tilde{\alpha}_i, \tilde{\beta}_j\right) \leq -2 \log 2$. This voting outcomes occur with probability

$$\Pr\left(v_{3j}=1, v_{ij}=0 \forall i\neq 3\right)=\frac{2k-1}{2k}\cdot\frac{1}{3k}\cdot\frac{6k^2-2}{6k^2-1}>\frac{1}{4k}$$

since $k \geq 3$. Since, in all other cases, $\sup_{\tilde{\alpha}_j,\tilde{\beta}_j} \ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right) \leq 0$, we have

$$E\left[\sup_{\tilde{\alpha}_j,\tilde{\beta}_j} \ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right)\right] < -\frac{1}{2k}\log 2.$$

Thus,

$$E\left[\sup_{\hat{\alpha}_j,\hat{\beta}_j} \ell_j\left(\hat{x},\hat{\alpha}_j,\hat{\beta}_j\right) - \sup_{\tilde{\alpha}_j,\tilde{\beta}_j} \ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right)\right] > \frac{1}{6k}\log 2 > 0$$

and, so, there exist $\hat{\alpha}_j$ and $\hat{\beta}_j$ such that

$$E\left[\ell_j\left(\hat{x},\hat{\alpha}_j,\hat{\beta}_j\right) - \sup_{\tilde{\alpha}_j,\tilde{\beta}_j} \ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right)\right] > 0. \tag{3.1}$$

This has, so far, assumed that the parameter estimates are unbounded. However, even if the parameter space is restricted to values where $|\hat{\alpha}_j| < a$ and $0 < b < \left|\hat{\beta}_j\right| < c$, it must be the case that there are sufficiently large values of a and c and sufficiently small values of b such that this relationship holds, as any such $\hat{\alpha}_j$ satisfies $|\hat{\alpha}_j| < a$ for some sufficiently large a, $\hat{\beta}_j$ satisfies $\left|\hat{\beta}_j\right| < c$ for sufficiently large c, and, even if $\hat{\beta}_j = 0$, there must be another value, $\hat{\beta}'_j \neq 0$ that satisfies Equation 3.1 by continuity of the likelihood function, so $a < \left|\hat{\beta}'_j\right|$ is satisfied for some sufficiently small $a > 0$.

By the strong law of large numbers applied to Equation 3.1, there exist vectors $\hat{\alpha}$ and $\hat{\beta}$ such that

$$\Pr\left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\left\{\ell_j\left(\hat{x},\hat{\alpha}_j,\hat{\beta}_j\right) - \sup_{\tilde{\alpha}_j,\tilde{\beta}_j} \ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right)\right\} > 0\right) = 1.$$

Thus,

$$\begin{aligned}
\Pr\left(\lim_{n\to\infty}\frac{1}{n}\left\{\sum_{j=1}^{n}\sup_{\hat{\alpha}_j,\hat{\beta}_j}\ell_j\left(\hat{x},\hat{\alpha}_j,\hat{\beta}_j\right)-\sum_{j=1}^{n}\sup_{\tilde{\alpha}_j,\tilde{\beta}_j}\ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right)\right\}>0\right) &= 1\\
\Pr\left(\lim_{n\to\infty}\frac{1}{n}\left\{\sup_{\hat{\alpha},\hat{\beta}}\sum_{j=1}^{n}\ell_j\left(\hat{x},\hat{\alpha}_j,\hat{\beta}_j\right)-\sup_{\tilde{\alpha},\tilde{\beta}}\sum_{j=1}^{n}\ell_j\left(\tilde{x},\tilde{\alpha}_i,\tilde{\beta}_j\right)\right\}>0\right) &= 1\\
\Pr\left(\lim_{n\to\infty}\frac{1}{n}\left\{\sup_{\hat{\alpha},\hat{\beta}}\ell\left(\hat{x},\hat{\alpha},\hat{\beta}\right)-\sup_{\tilde{\alpha},\tilde{\beta}}\ell\left(\tilde{x},\tilde{\alpha},\tilde{\beta}\right)\right\}>0\right) &= 1.
\end{aligned}$$

Therefore, given a sufficient number of observations, there are almost surely (i.e., with probability one) vectors of parameter estimates $\hat{x}$, $\hat{\alpha}$, and $\hat{\beta}$ satisfying $\hat{x}_3 < \hat{x}_1 < \hat{x}_2$ with a likelihood greater than the likelihood for any parameter vectors $\tilde{x}$, $\tilde{\alpha}$, and $\tilde{\beta}$ satisfying $\tilde{x}_1 < \tilde{x}_3 < \tilde{x}_2$. Therefore, the maximum likelihood estimate of the ideal points almost surely does not converge to any values satisfying $x_1 < x_3 < x_2$. Since, by assumption, the true values do satisfy these two inequalities, the maximum likelihood estimator cannot be a consistent estimator of the ideal points, x. □

Theorem 3.2. *If utility is quadratic and the error distribution is logistic (i.e.,* $F(x) = (1+e^{-x})^{-1}$*), then the hierarchical model is unidentified with respect to the ideal points.*

Proof. Let π_k $(k = 1, \ldots, 2^K)$ be the distinct elements of $\mathcal{P}(\{1, \ldots, K\})$. Assume $\sum_{i\in\pi_k} x_i^\top v$ are unique for some d-dimensional unit vector, v, which is true for almost all x (i.e., except on a set of Lebesgue measure zero). Take any such v.

The vote pattern probabilities are given by

$$p_k^*(x,\alpha,\beta) = \frac{\prod_{i\in\pi_k}\exp\left(x_i^\top\beta-\alpha\right)}{\prod_{i=1}^{K}\left(1+\exp\left(ideal_i^\top\beta-\alpha\right)\right)} \tag{3.2}$$

$$= \frac{\exp\left(\sum_{i\in\pi_k}\left(x_i^\top\beta-\alpha\right)\right)}{\prod_{i=1}^{K}\left(1+\exp\left(x_i^\top\beta-\alpha\right)\right)}. \tag{3.3}$$

Assume $p_k^*(x,\alpha,\beta)$ are not linearly independent functions of α and β. Thus, by assumption, there exists a nontrivial solution to

$$\sum c_k p_k^* (x, \alpha, \beta) = 0. \tag{3.4}$$

Let $k' = \arg\max_{k:c_k \neq 0} \sum_{i\in\pi_k} x_i^\top v$ (indexing the set with the largest summed ideal points subject with $c_k \neq 0$). Observe that

$$\sum c_k \frac{\exp\left(\sum_{i\in\pi_k}\left(x_i^\top\beta - \alpha\right)\right)}{\prod_{i=1}^K\left(1+\exp\left(\left(x_i^\top\beta-\alpha\right)\right)\right)} = 0 \tag{3.5}$$

$$\sum c_k \exp\left(\sum_{i\in\pi_k}\left(x_i^\top\beta - \alpha\right)\right) = 0 \tag{3.6}$$

$$\sum c_k \exp\left(\sum_{i\in\pi_k} x_i^\top\beta\right) = 0 \tag{3.7}$$

$$\sum c_k \frac{\exp\left(\sum_{i\in\pi_k} x_i^\top\beta\right)}{\exp\left(\sum_{i\in\pi_{k'}} x_i^\top\beta\right)} = 0 \tag{3.8}$$

Let $\beta = v\beta_0$ with β_0 a scalar. As $\beta_0 \to \infty$, all terms with $k \neq k'$ drop out as $\sum_{i\in\pi_k} x_i^\top v < \sum_{i\in\pi_{k'}} x_i^\top v$. Thus, we have $c_{k'} = 0$, which contradicts the definition of k' ($c_{k'} \neq 0$). Thus, $p_k^*(x, \alpha, \beta)$ are linearly independent functions.

Take 2^K pairs (α_k, β_k) such that the vectors $p^*(x, \alpha_k, \beta_k)$ span $\mathbb{R}^{2^K}$. For any w such that $\sum_{k=1}^{2^K} w_k = 1$, let $H(w)$ be the distribution of (α, β) such that $\Pr((\alpha_k, \beta_k) = (\alpha_k, \beta_k)) = w_k$ ($k = 1, \ldots, 2^K$). Take any w' such that $w'_k > 0 \forall k$. Take any sequence of ideal points, x_n, satisfying the identifiability constraint where $x_n \to x$ and $x_n \neq x \forall n$. By continuity of p^*, $E_{H(w')} p^*(x_n, \alpha, \beta) \to E_{H(w')} p^*(x, \alpha, \beta)$. For each n, there exists a $w^{(n)}$ such that $E_{H(w')} p^*(x_n, \alpha, \beta) = E_{H\left(w^{(n)}\right)} p^*(x, \alpha, \beta)$ since $p^*(x, \alpha_k, \beta_k)$ span $\mathbb{R}^{2^K}$ and $\sum_{k=1}^{2^K} w_k^{(n)} = 1$. Since $w' \in (0,1)^{2^K}$ and $w^{(n)} \to w'$, there exists an n such that $w^{(n)} \in (0,1)^{2^K}$. Thus, there exists an n such that $x_n \neq x$ and $E_{H(w')} p^*(x_n, \alpha, \beta) = E_{H\left(w^{(n)}\right)} p^*(x, \alpha, \beta)$. □

Theorem 3.3. *If utility is quadratic and the error distribution is normal (i.e., $F(x) =$*

$\Phi(x)$), *then the hierarchical model is unidentified with respect to the ideal points.*

Proof. Let π_k $(k = 1, \ldots, 2^K)$ be the distinct elements of $\mathcal{P}(\{1, \ldots, K\})$. Assume $\sum_{i \in \pi_k} x_i^\top v$ are unique for some d-dimensional unit vector, v, which is true for almost all x (i.e., except on a set of Lebesgue measure zero). Take any such v.

The vote pattern probabilities are given by

$$p_k^*(x, \alpha, \beta) = \prod_{i \in \pi_k} \Phi\left(x_i^\top \beta - \alpha\right) \prod_{i \notin \pi_k} \left(1 - \Phi\left(x_i^\top \beta - \alpha\right)\right) \tag{3.9}$$

Assume $p_k^*(x, \alpha, \beta)$ are not linearly independent functions of α and β. Thus, by assumption, there exists a nontrivial solution to

$$\sum c_k p_k^*(x, \alpha, \beta) = 0. \tag{3.10}$$

Take such a nontrivial solution and let $k' = \arg\max_{k: c_k \neq 0} \sum_{i \in \pi_k} \left(x_i^\top v\right)^2$ (indexing the set with the largest summed of squared ideal points along vector v with $c_k \neq 0$).

Recall the limiting behavior of the upper tail of the standard normal CDF implies that $\lim_{z \to \infty} \frac{1 - \Phi(z)}{\exp -\frac{1}{2} z^2 / \left(z\sqrt{2}\pi\right)}$. Since $\lim_{z \to \infty} \Phi(z) = 1$, we also have $\lim_{z \to \infty} \frac{(1 - \Phi(z))/\Phi(z)}{\exp -\frac{1}{2} z^2 / \left(z\sqrt{2}\pi\right)} =$

1 and $\lim_{z\to\infty} \frac{\Phi(z)/(1-\Phi(z))}{\exp \frac{1}{2} z^2 \left(z\sqrt{2}\pi\right)} = 1$. Observe that

$$\sum c_k \left\{ \prod_{i\in\pi_k} \Phi\left(x_i^\top \beta - \alpha\right) \right\} \left\{ \prod_{i\notin\pi_k} \left(1 - \Phi\left(x_i^\top \beta - \alpha\right)\right) \right\} = 0 \tag{3.11}$$

$$\sum c_k \left\{ \prod_{i\in\pi_k} \Phi\left(x_i^\top \beta - \alpha\right) \right\} \left\{ \prod_{i\notin\pi_k} \left(1 - \Phi\left(x_i^\top \beta - \alpha\right)\right) \right\} \left\{ \prod_{i=1}^{K} \left(1 - \Phi\left(x_i^\top \beta - \alpha\right)\right)^{-1} \right\} = 0 \tag{3.12}$$

$$\sum c_k \prod_{i\in\pi_k} \frac{\Phi\left(x_i^\top \beta - \alpha\right)}{1 - \Phi\left(x_i^\top \beta - \alpha\right)} = 0 \tag{3.13}$$

$$\sum c_k \frac{\prod_{i\in\pi_k} \frac{\Phi\left(x_i^\top \beta - \alpha\right)}{/} \left\{1 - \Phi\left(x_i^\top \beta - \alpha\right)\right\}}{\prod_{i\in\pi_{k'}} \frac{\Phi\left(x_i^\top \beta - \alpha\right)}{/} \left\{1 - \Phi\left(x_i^\top \beta - \alpha\right)\right\}} = 0 \tag{3.14}$$

Let $\beta = v\beta_0$ and $\alpha = \min_i \left(x_i^\top v\beta_0\right) - \beta_0$ with β_0 a scalar. As $\beta_0 \to \infty$, $x_i^\top \beta - \alpha \to \infty$

for all i. Using the limiting behavior of $\Phi(z)/(1-\Phi(z))$, the limit as $\beta_0 \to \infty$ gives

$$\lim_{\beta_0\to\infty} \sum c_k \frac{\prod_{i\in\pi_k} \frac{\Phi\left(x_i^\top\beta-\alpha\right)}{/} \left\{1-\Phi\left(x_i^\top\beta-\alpha\right)\right\}}{\prod_{i\in\pi_{k'}} \frac{\Phi\left(x_i^\top\beta-\alpha\right)}{/} \left\{1-\Phi\left(x_i^\top\beta-\alpha\right)\right\}} = 0 \tag{3.15}$$

$$\sum c_k \lim_{\beta_0\to\infty} \frac{\prod_{i\in\pi_k} \frac{\Phi\left(x_i^\top\beta-\alpha\right)}{/} \left\{1-\Phi\left(x_i^\top\beta-\alpha\right)\right\}}{\prod_{i\in\pi_{k'}} \frac{\Phi\left(x_i^\top\beta-\alpha\right)}{/} \left\{1-\Phi\left(x_i^\top\beta-\alpha\right)\right\}} = 0 \tag{3.16}$$

$$\sum c_k \lim_{\beta_0\to\infty} \frac{\prod_{i\in\pi_k} \exp\left(\frac{1}{2}\left(x_i^\top\beta-\alpha\right)^2 + \log\left(x_i^\top\beta-\alpha\right) + \log\left(\sqrt{2}\pi\right)\right)}{\prod_{i\in\pi_{k'}} \exp\left(\frac{1}{2}\left(x_i^\top\beta-\alpha\right)^2 + \log\left(x_i^\top\beta-\alpha\right) + \log\left(\sqrt{2}\pi\right)\right)} = 0 \tag{3.17}$$

$$\sum c_k \lim_{\beta_0\to\infty} \frac{\exp\left(\sum_{i\in\pi_k} \left\{\frac{1}{2}\left(x_i^\top\beta-\alpha\right)^2 + \log\left(x_i^\top\beta-\alpha\right) + \log\left(\sqrt{2}\pi\right)\right\}\right)}{\exp\left(\sum_{i\in\pi_{k'}} \left\{\frac{1}{2}\left(x_i^\top\beta-\alpha\right)^2 + \log\left(x_i^\top\beta-\alpha\right) + \log\left(\sqrt{2}\pi\right)\right\}\right)} = 0 \tag{3.18}$$

$$\sum c_k \lim_{\beta_0\to\infty} \frac{\exp\left(\sum_{i\in\pi_k} \left\{\beta_0^2\frac{1}{2}\left(x_i^\top v - \min\left(x_i^\top v\right) - 1\right)^2 + \log\beta_0 + \log\left(x_i^\top v - \min\left(x_i^\top v\right) - 1\right)\right\}\right)}{\exp\left(\sum_{i\in\pi_{k'}} \left\{\beta_0^2\frac{1}{2}\left(x_i^\top v - \min\left(x_i^\top v\right) - 1\right)^2 + \log\beta_0 + \log\left(x_i^\top v - \min\left(x_i^\top v\right) - 1\right)\right\}\right)} = 0. \tag{3.19}$$

All terms with $k \neq k'$ drop out as the β_0^2 terms dominate as $\beta_0 \to \infty$ and $\sum_{i\in\pi_k}\left(x_i^\top v\right)^2 < \sum_{i\in\pi_{k'}}\left(x_i^\top v\right)^2$. Thus, we have $c_{k'} = 0$, which contradicts the definition of k' ($c_{k'} \neq 0$). Thus, $p_k^*(x,\alpha,\beta)$ are linearly independent functions.

Take 2^K pairs (α_k,β_k) such that the vectors $p^*(x,\alpha_k,\beta_k)$ span $\mathbb{R}^{2^K}$. For any w such that $\sum_{k=1}^{2^K} w_k = 1$, let $H(w)$ be the distribution of (α,β) such that $\Pr((\alpha_k,\beta_k) = (\alpha_k,\beta_k)) = w_k$ $(k = 1,\ldots,2^K)$. Take any w' such that $w'_k > 0 \forall k$. Take any sequence of ideal points, x_n, satisfying the identifiability constraint where $x_n \to x$ and $x_n \neq x \forall n$. By continuity of p^*, $E_{H(w')}p^*(x_n,\alpha,\beta) \to E_{H(w')}p^*(x,\alpha,\beta)$. For each n, there exists a $w^{(n)}$ such that $E_{H(w')}p^*(x_n,\alpha,\beta) = E_{H\left(w^{(n)}\right)}p^*(x,\alpha,\beta)$ since $p^*(x,\alpha_k,\beta_k)$

span $\mathbb{R}^{2^K}$ and $\sum_{k=1}^{2^K} w_k^{(n)} = 1$. Since $w' \in (0,1)^{2^K}$ and $w^{(n)} \to w'$, there exists an n such that $w^{(n)} \in (0,1)^{2^K}$. Thus, there exists an n such that $x_n \neq x$ and $E_{H(w')} p^* (x_n, \alpha, \beta) = E_{H\left(w^{(n)}\right)} p^* (x, \alpha, \beta)$. □

Theorem 3.4. *If utility is Gaussian and the error distribution is logistic or normal, the hierarchical model is unidentified with respect to the ideal points.*

Proof. Let $\Pr_{w,\gamma}$ and $p^*_{w,\gamma}$ denote probability and vote pattern probabilities, respectively, under a Gaussian utility model with parameters w and γ. Let $\Pr_Q$ and p^*_Q denote probability and vote pattern probabilities, respectively, under a quadratic utility model. Under quadratic utility and parameterizing the bill parameters model in terms of $r^{(y)}$ and $r^{(n)}$, the previous two theorems imply that there exist ideal points, x, and 2^K pairs of bill parameters, $\left(r_j^{(y)}, r_j^{(n)}\right)$, such that $p^*_Q \left(x, r_j^{(y)}, r_j^{(n)}\right)$ spans $\mathbb{R}^{2^K}$. Take any such x and $\left(r_j^{(y)}, r_j^{(n)}\right)$ $(j = 1, \ldots, 2^K)$.

Since, as shown in the text,

$$\lim_{\gamma \to \infty} \Pr_{1/\beta,\beta} (v_{ij} = 1) = F \left(\sum_{d=1}^{D} \left(x_{id} - r_{jd}^{(n)} \right)^2 - \sum_{d=1}^{D} \left(x_{id} - r_{jd}^{(y)} \right)^2 \right) \tag{3.20}$$

$$= \Pr_Q (v_{ij} = 1), \tag{3.21}$$

it must also be true that

$$\lim_{\gamma \to \infty} p^*_{1/\gamma,\gamma} \left(x, r_j^{(y)}, r_j^{(n)}\right) = p^*_Q \left(x, r_j^{(y)}, r_j^{(n)}\right). \tag{3.22}$$

Since the set of singular $2^K \times 2^K$ matrices is closed and the matrix

$$\left[p^*_Q \left(x, r_1^{(y)}, r_1^{(n)}\right) \quad \ldots \quad p^*_Q \left(x, r_{2^K}^{(y)}, r_{2^K}^{(n)}\right) \right] \tag{3.23}$$

is, by assumption, nonsingular, there must be some sufficiently large γ' such that

$$\left[p^*_{1/\gamma',\gamma'}\left(x, r^{(y)}_1, r^{(n)}_1\right) \quad \ldots \quad p^*_{1/\gamma',\gamma'}\left(x, r^{(y)}_{2^K}, r^{(n)}_{2^K}\right)\right] \tag{3.24}$$

is also nonsingular as, otherwise, its limit would also be singular.

Thus, the vectors $p^*_{1/\gamma',\gamma'}\left(x, r^{(y)}_j, r^{(n)}_j\right)$ span $\mathbb{R}^{2^K}$. It follows from the same arguments as in the previous two theorems that there exist ideal points, $x' \neq x$, and bill parameter distributions ν and ν' such that the vote pattern probabilities marginalized over the bill parameters are equal. Therefore, the hierarchical model with Gaussian utility and logistic or normal error distribution is not identified with respect to the ideal points. □

Theorem 4.1. *For any fixed number of legislators $K \geq 3$ and any error distribution, there is a set of ideal points $x_1, \ldots, x_K$ and sequence of bill parameters $\alpha_1, \ldots$ and $\beta_1, \ldots$ such that Optimal Classification is not a consistent estimate of the order of the ideal points.*

Proof. Let v_{ij} denote the vote of legislator i on bill j and equal one for a 'yea' vote (with probability $F(\beta_j x_i - \alpha_j)$) and zero for a 'nay' vote. Let $m_j(x')$ denote the minimum number of misclassifications possible on vote j given ideal points x' and let $m(x') = \sum_{j=1}^{n} m_j(x')$ denote the minimum number of misclassifications over all votes. Note that in all cases, $0 \leq m_j \leq \frac{k}{2}$, as it is always possible to classify over half the votes correctly by classifying all votes together.

Because of the identifying constraints, $x_1 = 1$ and $x_2 = 0$. Let

$$\alpha_j = F^{-1}\left(1 - \left(1 - \frac{1}{6k^2 - 1}\right)^{\frac{1}{k-2}}\right)$$

and $\beta_j = F^{-1}\left(\frac{1}{2k}\right) + \alpha_j$ for all j. Let $x_3 = \left(F^{-1}\left(\frac{1}{3k}\right) + \alpha_j\right)\beta_j^{-1}$. Finally, take any

$x_i < 0$ for all $i > 3$. Note that this implies that

$$F(\beta_j x_3 - \alpha_j) = \frac{1}{3k} < \frac{1}{2k} = F(\beta_j x_1 - \alpha_j).$$

Further, $F(\beta_j x_i - \alpha_j) < 1 - \left(1 - \frac{1}{6k^2-1}\right)^{\frac{1}{k-2}} = F(\beta_j x_1 - \alpha_j)$, so

$$\Pr(v_{1j} = 0 \wedge v_{ij} = 0 \forall i > 3) = \frac{6k^2-2}{6k^2-1} > \Pr(v_{3j} = 0) = \frac{3k-1}{3k} > \Pr(v_{3j} = 0) = \frac{2k-1}{2k}.$$

Let $m_j(x')$ denote the minimum number of misclassifications possible on vote j given ideal points x' and let $m(x') = \sum_{j=1}^{n} m_j(x')$ denote the minimum number of misclassifications over all votes. Note that in all cases, $0 \leq m_j \leq \frac{k}{2}$, as it is always possible to classify over half the votes correctly by classifying all votes together.

Consider the case where the estimated ordering is $\hat{x}_3 < \hat{x}_4 < \ldots < \hat{x}_k < \hat{x}_1 < \hat{x}_2$. If $v_{1j} = 1$ and $v_{ij} = 0$ for all $i \neq 1$, then the data are perfectly separating and so $m_j(\hat{x}) = 0$. If $v_{3j} = 1$ and $v_{ij} = 0$ for all $i \neq 3$, then the data are again perfectly separating and so $m_j(\hat{x}) = 0$. And if $v_{ij} = 0$ for all i, $m_j(\hat{x}) = 0$. These three outcomes occur with probability

$$\begin{aligned} \Pr(v_{ij} = 0 \forall i) + \Pr(v_{1j} = 1, v_{ij} = 0 \forall i \neq 1) + \Pr(v_{3j} = 1, v_{ij} = 0 \forall i \neq 3) &> \\ \frac{2k-1}{2k} \cdot \frac{3k-1}{3k} \cdot \frac{6k^2-2}{6k^2-1} + \frac{1}{2k} \cdot \frac{3k-1}{3k} \cdot \frac{6k^2-2}{6k^2-1} + \frac{2k-1}{2k} \cdot \frac{1}{3k} \cdot \frac{6k^2-2}{6k^2-1} &= \frac{3k^2-1}{3k^2} \end{aligned}$$

Since $m_j(\hat{x}) \leq \frac{k}{2}$ for all other cases and these occur with probability less than $\frac{1}{3k^2}$, we have

$$E[m_j(\hat{x})] < \frac{1}{6k}.$$

Now, consider the case where $\tilde{x}_4 < \ldots < \tilde{x}_k < \tilde{x}_1 < \tilde{x}_3 < \tilde{x}_2$, which matches the true ordering. If $v_{3j} = 1$ and $v_{ij} = 0$ for all $i \neq 3$, then perfect classification is not possible, so $m_j(\tilde{x}) \geq 1$. This voting outcomes occur with probability

$$\Pr(v_{3j} = 1, v_{ij} = 0 \forall i \neq 3) = \frac{2k-1}{2k} \cdot \frac{1}{3k} \cdot \frac{6k^2-2}{6k^2-1} > \frac{1}{4k}$$

since $k \geq 3$. Since, in all other cases, $m_j \geq 0$, we have

$$E[m_j(\tilde{x})] > \frac{1}{4k}.$$

Thus,

$$E[m_j(\tilde{x}) - m_j(\hat{x})] > \frac{1}{12k}.$$

By the strong law of large numbers,

$$\Pr\left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \{m_j(\tilde{x}) - m_j(\hat{x})\} > 0\right) = 1.$$

Thus,

$$\begin{aligned}
\Pr\left(\lim_{n\to\infty} \frac{1}{n} \left\{\sum_{j=1}^{n} m_j(\tilde{x}) - \sum_{j=1}^{n} m_j(\hat{x})\right\} > 0\right) &= 1 \\
\Pr\left(\lim_{n\to\infty} \frac{1}{n} \{m(\tilde{x}) - m(\hat{x})\} > 0\right) &= 1 \\
\Pr\left(\exists N : \forall n > N, \sum_{j=1}^{n} m_j(\hat{x}) < \sum_{j=1}^{n} m_j(\tilde{x})\right) &= 1.
\end{aligned}$$

So, given sufficiently many votes, the maximum number of correct classifications under $\hat{x}$ will almost surely exceed the maximum number under $\tilde{x}$, which is the correct ordering. Since optimal classification produces an estimated order which maximizes the number of correct classifications, optimal classification is not a consistent estimator of the order of the ideal points. □

Theorem 4.2. *For any fixed number of legislators $K \geq 3$ and any error distribution, there is a set of ideal points $x_1, \ldots, x_K$ and sequence of bill parameters $\alpha_1, \ldots$ and*

$\beta_1, \ldots$ *such that maximum likelihood is not a consistent estimate of the order of the ideal points.*

Proof. This follows directly from the counterexample given in Theorem 3.1. $\square$

Theorem 4.3. *If the number of legislators* $K \geq 5$*, there exists an estimator of the order of the ideal points which is consistent for any set of ideal points* $x_1, \ldots, x_K$ *and sequence of bill parameters* $\alpha_1, \ldots$ *and* $\beta_1, \ldots$ *satisfying the conditions of Subsection 2.2.*

Proof. Take any $a \neq b \in \{1, \ldots, K\}$. Without loss of generality, assume $x_a < x_b$. Consider the quantity $z_{in} = (v_{bn} - v_{an})\, v_{in}$ for all $i \neq a, b$. By independence of v_{in}, $E[z_{in}] = (F(\alpha_n + \beta_n x_b) - F(\alpha_n + \beta_n x_a))\, F(\alpha_n + \beta_n x_i)$. Since F is increasing and $x_a < x_b$ by assumption, $(F(\alpha_n + \beta_n x_b) - F(\alpha_n + \beta_n x_a))$ is positive. Therefore, z_{in} is strictly increasing in x_i.

Now consider two legislators, i and j, such that $x_i < x_j$ and $i, j \neq a, b$. Since F is increasing, $E[z_{jn} - z_{in}] = (F(\alpha_n + \beta_n x_b) - F(\alpha_n + \beta_n x_a))\, (F(\alpha_n + \beta_n x_j) - F(\alpha_n + \beta_n x_i)) \geq 0$. Moreover, treating x as fixed, there exists a $\zeta > 0$ such that $E[z_{jn} - z_{in}] > \zeta$ because of Conditions 2 and 3. Thus, as n increases, the probability that $\bar{z}_{jn} > \bar{z}_{in}$ where $\bar{z}_{in} = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$ and $\bar{z}_{jn} = \frac{1}{n}\sum_{k=1}^{n} z_{jk}$ must also increase. Further, since $E[z_{jn} - z_{in}] > \zeta$ and $-1 \leq \bar{z}_{jn} - \bar{z}_{in} \leq 1$, the bounded weak law of large numbers implies that $\lim_{n\to\infty} \Pr(1 \geq \bar{z}_{jn} - \bar{z}_{in} > \zeta > 0) = 1$. Thus, using $\mathbf{1}_x$ to denote the indicator function, $\mathbf{1}_{\bar{z}_j > \bar{z}_i} \xrightarrow{p} \mathbf{1}_{x_j > x_i}$ for all $i, j \neq a, b$ such that $x_i \neq x_j$ which, by assumption, is true for all i, j.

Note that the rank of x_j excluding x_a and x_b is $\operatorname{rank}(x_j, \{x_i : i \neq a, b\}) = \sum_{i \neq a,b} \mathbf{1}_{x_j > x_i}$. Thus, since the sum is over a finite number of terms, $\sum_{i \neq a,b} \mathbf{1}_{\bar{z}_j > \bar{z}_i} \xrightarrow{p} \sum_{i \neq a,b} \mathbf{1}_{x_j > x_i}$ and, so, $\sum_{i \neq a,b} \mathbf{1}_{\bar{z}_j > \bar{z}_i} \xrightarrow{p} \operatorname{rank}(x_j, \{x_i : i \neq a, b\})$. This gives a consistent estimator of the rank of all ideal points other than x_a and x_b assuming $x_a < x_b$ and gives it up to a monotone transformation so long as $x_a \neq x_b$, which, by assumption, is true for all a, b. If we begin with $a = 2$ and and $b = 1$, where our identifying constraint establishes that $x_a < x_b$, we asymptotically recover the subrank of the remaining ideal points. Given their subrank (excluding x_1 and x_2), we can recover the position of x_1 and x_2 within this subrank by setting a and b to other values until we have recovered their position. This yields an estimator of the rank of all ideal points. Since there are only finitely-many, all these subrankings converge in probability to the true subrankings. Therefore, this estimator of the rank converges in probability to the true rank and is consistent. Of course, this does not define behavior of the estimator when the subrankings have inconsistencies, but as the probability of this happening goes to zero as $n \to \infty$, this choice is irrelevant. $\square$