Measuring Bias and Uncertainty in Ideal Point Estimates via the Parametric Bootstrap

Jeffrey B. Lewis

*Department of Political Science, University of California,
Los Angeles, Los Angeles, CA 90095
e-mail: jblewis@ucla.edu*

Keith T. Poole

*Center for Advanced Study in the Behavioral Sciences and
Department of Political Science, University of Houston,
Houston, TX 77204-3011
e-mail: kpoole@uh.edu*

Over the last 15 years a large amount of scholarship in legislative politics has used NOMINATE or other similar methods to construct measures of legislators' ideological locations. These measures are then used in subsequent analyses. Recent work in political methodology has focused on the pitfalls of using such estimates as variables in subsequent analysis without explicitly accounting for their uncertainty and possible bias (Herron and Shotts 2003, *Political Analysis* 11:44–64). This presents a problem for those employing NOMINATE scores because estimates of their unconditional sampling uncertainty or bias have until now been unavailable. In this paper, we present a method of forming unconditional standard error estimates and bias estimates for NOMINATE scores using the parametric bootstrap. Standard errors are estimated for the 90th U.S. Senate in two dimensions. Standard errors of first–dimension placements are in the 0.03 to 0.08 range. The results are compared with those obtained using the Markov chain Monte Carlo estimator of Clinton et al. (2002, Stanford University Working Paper). We also show how the bootstrap can be used to construct standard errors and confidence intervals for auxiliary quantities of interest such as ranks and the location of the median senator.

1 Introduction

The purpose of this paper is to show a general method for obtaining standard errors, confidence intervals, and other measures of uncertainty for the ideal point estimates obtained from NOMINATE and other similar scaling procedures. The number of parameters estimated by these scaling methods is so large that conventional approaches to obtaining standard errors have proven impractical. Our approach is to use the parametric bootstrap (Efron 1979; Efron and Tibshirani 1993) to obtain standard errors and other measures of estimation uncertainty. Nominal Three-step Estimation (NOMINATE) was originally developed by Poole and Rosenthal (1985, 1991, 1997) to scale U.S. Congressional roll call data. The method is based upon a probabilistic spatial voting model that utilizes a random utility function (McFadden 1976). NOMINATE produces ideal points for the legislators

and two points—one corresponding to the Yea outcome and one corresponding to the Nay outcome—for every roll call along with the parameters of the utility function. If there are 100 legislators and 500 roll calls, NOMINATE estimates 1,101 parameters in one dimension and 2,202 parameters in two dimensions using the 50,000 observed choices. In a classical maximum likelihood framework the standard errors are obtained from either inverting the information matrix or inverting the analytical Hessian matrix directly. Unfortunately, this entails inverting a very large matrix and is computationally difficult even with modern computers.

Because of these difficulties, NOMINATE computes only *conditional* standard errors. For example, the standard errors for a legislator's ideal point parameters are obtained from inverting just the information matrix for those parameters—the roll call parameters are fixed and each legislator's parameters are independent of other legislator parameters. Although computationally easy to compute, the quality of these conditional standard errors is suspect and they probably underestimate the true uncertainty. Indeed, Clinton et al. (2002) show that the standard errors from W-NOMINATE may be smaller than those they derive from a Markov Chain Monte Carlo (MCMC) approach. Unfortunately, a direct comparison is not possible because Clinton et al. use a quadratic utility function while NOMINATE is based upon a normal distribution utility function. Because of the computational intensity of the MCMC approach, it has not yet been applied to the NOMINATE model.

We bridge this gap by applying the parametric bootstrap to W-NOMINATE as well as the Quadratic-Normal (QN) scaling procedure developed by Poole (2000, 2001). The QN procedure is based upon the quadratic utility function so that we can compare the bootstrapped standard errors from both procedures with the standard errors derived from the Clinton et al. MCMC method.

Interestingly, how the scale, location, and rotation of the issue space are defined complicates not only comparisons of the point estimates across models but also comparisons of uncertainty estimates. How the space is identified affects in subtle ways which members' locations are estimated to be more or less uncertain. Thus, while we find similar general levels of uncertainty in the estimates generated by each model, the way in which estimation uncertainty is distributed across members varies considerably. As described below, this should not be surprising given that the definition of the issue space is arbitrary and ultimately what these models recover are the relative locations. Accordingly, much of the apparent difference in the uncertainty estimates evaporates when uncertainties in *rank orders* as opposed to locations are considered. Beyond providing estimates of standard errors, the bootstrap technique can also be used to calculate uncertainty estimates for other other commonly calculated quantities of interest that are constructed from ideal point estimates such as the location or identity of the median senator or the filibuster pivot.

In the next section we briefly describe the bootstrap. In Section 3 we explain how the parametric bootstrap is applied to W-NOMINATE and QN. In Section 4 we present the results of applying the bootstrap to roll call data from the 90th Senate in two dimensions and we compare the bootstrap results for NOMINATE and QN with Clinton et al.'s. IDEAL model applied to the same data. The 90th Senate is selected because it is known from previous research to reveal two dominant dimensions and because the 90th Senate (1967–68) includes a senator (Goodell of New York) with an abbreviated voting record whose ideal point we expect to be measured with considerable uncertainty.[1] In Section 5 we use the bootstrap to calculate the uncertainty in additional quantities of interest arising from ideal

---

[1]Charles Goodell replaced Robert Kennedy in September 1968 and cast votes on only 39 of the 596 roll calls taken in the 90th Senate.

point estimation. In this section, we turn to the 105th Senate. The 105th Senate (1997–98) reveals a single dominant dimension where quantities such as the median and filibuster pivots are more salient and where the identities of the senators are more likely to be familiar to the reader. In Section 6, we conclude.

2 The Parametric Bootstrap

Excellent discussions of the bootstrap are provided in Efron and Tibshirani (1993), Hall (1985), Mooney (1996), and Young (1994). We will only briefly discuss the bootstrap here and we will focus on the less common parametric form of the bootstrap that we employ. Typically the bootstrap is used to provide estimates of the standard errors and confidence intervals that do not rely on asymptotic normality. When the nonparametric bootstrap is employed, the resulting confidence intervals and standard errors are nonparametric in the sense that they do not rely on the correctness of the likelihood function for the data. This can be particularly useful in cases in which the robustness to distributional assumptions is of great concern, the estimation is itself nonparametric, or the samples are too small to rely on asymptotic approximations. In our case, the reason to apply the bootstrap is mainly computational convenience, though, as we will show in Section 5, the bootstrap also allows us to estimate the uncertainty of auxiliary quantities of interest such as the location of the median legislator. Recovering the variance-covariance matrix of parameter estimates by forming and inverting the full (estimated) information matrix for roll call voting models such as NOMINATE or QN is sufficiently difficult that the bootstrap is an attractive and tractable alternative.

Following Efron (1979), let $\boldsymbol{\theta}$ be a vector of parameters to be estimated and $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$. The sampling distribution of $\hat{\boldsymbol{\theta}}$ is dependent on the joint distribution of the data. Let F be the joint cumulative distribution of the data. We can then write $\hat{\boldsymbol{\theta}}(F)$. If F were known, the sampling distribution of $\hat{\boldsymbol{\theta}}$ could be ascertained directly by analytic or simulation methods. Using simulation methods, repeated samples would be drawn from F, $\hat{\boldsymbol{\theta}}$ calculated for each sample, and features of the sampling distribution of $\hat{\boldsymbol{\theta}}$ approximated with arbitrary precision (as the number of pseudosamples grows large).[2] Efron (1979) shows that $\hat{\boldsymbol{\theta}}(\hat{F})$ can provide a good approximation of $\hat{\boldsymbol{\theta}}(F)$ where $\hat{F}$ is an estimate of F based on sample data. Even in small samples, approximating F by some $\hat{F}$ will in many situations provide excellent estimates of the sampling distribution of $\hat{\boldsymbol{\theta}}(F)$. Some asymptotic properties of $\hat{\boldsymbol{\theta}}(\hat{F})$ in fairly general (usually univariate) settings are given in Hall (1994) and references therein.

In simple settings where the data are independently and identically distributed (i.i.d.), the nonparametric maximum likelihood (ML) estimate of the marginal distribution of each observation is simply the empirical distribution of the observations. In this case, an approximate draw from the joint distribution F can be made by sampling with replacement n draws from the observed data, where n is the number of observations. Consider the case in which the data are observations on a single variable Y. Letting $\vec{y} = (y_1, y_2, \ldots, y_n)$ be a vector of observations on Y, drawing a sample from $\hat{F}$ is a matter of sampling with replacement n values in turn from $\vec{y}$ where each element of $\vec{y}$ is selected with probability $1/n$ at each turn.

In cases in which the data are not i.i.d., simple resampling from the data does not generally yield draws from a joint distribution, $\hat{F}$, that is a good approximation of draws from F. If the dependence in the data is temporal or spatial, "block" resampling schemes that draw randomly groups of adjacent observations from the data have been suggested (see Hall

---

[2]Monte Carlo experiments, for example, posit a given F and then recover the sampling distribution of $\hat{\boldsymbol{\theta}}(F)$.

1985, 1994, and references therein). In the case of roll call voting data both the rows and columns of the data matrix are dependent. Indeed, it is these dependencies that are exploited in recovering the ideal point and vote parameters. However, unlike time series or spatial data, we have no *ex ante* expectations about which elements of the vote matrix are "close" to which others.[3] Given this lack of *ex ante* information about how to structure a block resampling scheme, there is no obvious way (at least to us) of implementing the nonparametric bootstrap in this case.[4]

On the other hand, the parametric bootstrap is easy to apply to maximum likelihood estimators such as NOMINATE or QN. In the parametric bootstrap, $\hat{F}$ is estimated directly from the likelihood itself. That is, the joint distribution of the data is approximated by the likelihood evaluated at $\hat{\theta}$. In either QN or NOMINATE individual vote choices are independent conditional on the value of the roll call and ideal point parameters. Thus, conditional on the estimated roll call parameters and ideal points, draws from the joint distribution of the data matrix can be made by drawing from each element of the data matrix independently. Because the estimated parameters are not equivalent to the true parameters, the estimated joint distribution of the data, $\hat{F}$, differs from F, as in the nonparametric case. Note that estimating F based on the parameter estimates is similar to substituting the information matrix evaluated at the estimated parameter values (as opposed to the true values) when approximating the variance–covariance matrix of ML estimators in the usual way (see Efron 1982).

By the Slutsky theorem, the parametric bootstrap estimate of F will be consistent if $\hat{\theta}$ is consistent for θ.[5] Precise conditions under which the models described above are consistent have yet to be established. The models are known not to be consistent as the number of roll calls or the number of legislators goes to infinity, though it may be that sending the number of members, the number of votes, and the ratio of votes to members to infinity is sufficient (Londregan 2000). Thus we cannot appeal to standard asymptotic results to establish the admissibility of the parametric bootstrap estimator in this case. However, extensive Monte Carlo experiments on both these models and similar models in psychometrics suggest that accurate and reliable estimates are obtained if the data matrix has a rank of 100 (Lord 1983; Poole and Rosenthal 1997)—the rank of the Senate roll data considered below.[6]

Finally, it should be noted that unlike the nonparameteric bootstrap, the parametric bootstrap is not robust to violations of distributional assumptions. However, the quality of the parameter estimates themselves depends on these same assumptions and, if these assumptions are correct, the parametric bootstrap will provide a more efficient estimator of F than the nonparametric bootstrap. As mentioned above, the parametric bootstrap turns on the same logic as the standard inverse of the sample information estimator that is typically used to infer the uncertainty of ML estimates.[7]

---

[3]Obviously, elements in the same column or row will be expected to be particularly dependent.

[4]In structural equations modeling, which also involves dependent data, a nonparametric bootstrap is possible because these models operate on the variance-covariance matrix of the data, which can be simulated by simple resampling techniques (Bollen and Stein 1992).

[5]To appeal to the Slutsky theorem, certain conditions must hold. In particular, F must be a continuous function of θ.

[6]Our own Monte Carlo experiments confirm the effectiveness of the parametric bootstrap technique in this setting.

[7]Efron (1982) shows that the typical inverse of the sample information estimator for the variance matrix of ML estimates is a second-order approximation to the variance matrix that would be recovered by the parametric bootstrap.

3 Application of the Parametric Bootstrap to the 90th Senate

The parametric bootstrap is very simple conceptually. In a maximum likelihood or probabilistic framework, the first step is to compute the likelihood function of the sample. The second step is to draw, for example, 1000 samples from the likelihood density and compute for each sample the maximum likelihood estimates of the parameters of interest. Finally, the sample variances computed from these 1000 values are the estimators of the variances of the parameters (Efron and Tibshirani 1993, ch. 6).

When applied to a scaling method such as W-NOMINATE, the first step is to run the program to convergence and then calculate the probabilities for the observed choices. This produces a legislator by roll call matrix containing the estimated probabilities for the corresponding actual roll call choices of the legislators.

To draw a random sample we simply treat each probability as a weighted coin and we "flip" the coin. We do this by drawing from a uniform distribution over zero to one— $U(0,1)$—and if the random draw is less than or equal to the estimated probability then our sampled value is the observed choice. If the random draw is greater than the estimated probability, then our sampled value is the opposite of the observed choice; that is, if the observed choice is Yea then our sampled value is Nay. We then apply W-NOMINATE to this sample roll call matrix. This process is repeated 1000 times and the variances of the legislator ideal points are calculated using the 1000 estimated bootstrap configurations.

Technically, let c_{ij} be the observed choice for the ith legislator ($i = 1, \ldots, p$) on the jth roll call ($j = 1, \ldots, q$) where the possible choices are Yea or Nay. In the U.S. Congress there is very little policy–related abstention (Poole and Rosenthal 1997) so we treat nonvoting as missing data. Let $\hat{P}_{ijc}$ be the estimated probability for the observed choice and let ϕ be a random draw from $U(0, 1)$. Let $\hat{c}_{ij}$ be the sampled choice. The sample rule is

$$\hat{c}_{ij} = \begin{cases} c_{ij}, & \text{if } \phi \leq \hat{P}_{ijc} \\ \sim c_{ij}, & \text{if } \phi > \hat{P}_{ijc}, \end{cases} \tag{1}$$

where $\sim c_{ij}$ represents the opposite choice to c_{ij}. This technique allows the underlying uncertainty to propagate through to all the estimated parameters. To see this, note that as $\hat{P}_{ijc} \to 1$, then $\hat{c}_{ij} \to c_{ij}$, that is, sample choices become the observed choices so that the bootstrapped variances for the parameters of the model go to zero. If the fit of the model is poor—for example, if the $\hat{P}_{ijc}$ are between 0.5 and 0.7, then the bootstrapped variances for the parameters will be large.

Although we obtain bootstrap estimates of the means and standard deviations of all the parameters, we focus our analysis on the legislator ideal points because they are used in a wide variety of secondary analyses by many researchers. Let $\hat{\mathbf{X}}$ be the $p \times s$ matrix of legislator coordinates estimated by either W-NOMINATE or QN. Let $h = 1, \ldots, m$ be the number of bootstrap trials and let $\mathbf{X}_h$ be the $p \times s$ matrix of legislator coordinates estimated on the hth bootstrap trial. The legislator and roll call coordinates are identified only up to an arbitrary rotation in the s-dimensional space. This arbitrary rotation must be removed to ensure that the bootstrap process produces accurate estimates of the standard deviations of the parameters. In particular, we assume that

$$\hat{\mathbf{X}} = \mathbf{X}_h \mathbf{V} + \mathbf{E}. \tag{2}$$

$\mathbf{V}$ is an $s \times s$ matrix such that $\mathbf{V'V} = \mathbf{VV'} = \mathbf{I}_s$ where $\mathbf{I}_s$ is an $s \times s$ identity matrix and $\mathbf{E}$ is a $p \times s$ matrix of errors. In psychometrics, Eq. (2) is known as the *orthogonal*

procrustes problem. We use Schonemann's (1966) solution to remove the arbitrary rotation, **V**. Note that we are *rigidly rotating* $\mathbf{X}_h$, we are not altering the estimated points vis-à-vis one another in any way. Consequently, in our discussion below we will simply denote the hth bootstrap trial matrix as $\mathbf{X}_h$ to avoid notational clutter.

For our example we apply the parametric bootstrap to W-NOMINATE and QN for the 90th Senate (1967–68) in two dimensions. We performed 1000 bootstrap trials as described above and computed the means and standard deviations of all the estimated parameters. For example, for the ith legislator on the kth dimension the mean of the bootstrap trials is

$$\bar{x}_{ik} = \frac{\sum_{h=1}^{m} x_{hik}}{m}, \tag{3}$$

where $m = 1000$ is the number of trials and x_{hik} is the estimated coordinate on the hth trial. The corresponding standard deviation is

$$\text{SE}(x_{ik}) = \sqrt{\frac{\sum_{h=1}^{m} (x_{hik} - \hat{x}_{ik})^2}{m-1}}, \tag{4}$$

where $\hat{x}_{ik}$ is the coordinate estimated by W-NOMINATE or QN.

We take a conservative approach and use the estimated coordinate, $\hat{x}_{ik}$, rather than the mean of the bootstrap trials, $\bar{x}_{ik}$, as our "sample mean" in our calculation of the standard deviation. This inflates the standard deviations somewhat but we feel it is better to err on the safe side and not *underreport* the standard deviations.

Figure 1 shows the estimated ideal points for the 90th Senate from W-NOMINATE along with the bootstrapped standard errors. The crosshairs through the ideal points show the 95% confidence intervals. The standard errors are small even for the second dimension. On the first dimension, 97 of 101 standard errors were 0.09 or less and on the second dimension, 90 of 101 standard errors were 0.10 or less. Given that W-NOMINATE constrains the legislator ideal points to lie within the unit circle, these standard errors are small relative to the estimated ideal points. For most senators the correlation between the estimated first– and second–dimension coordinates is very low. Normal theory confidence ellipses are shown for senators whose first–dimension coordinate is correlated with their second–dimension coordinate at $|\rho| > 0.15$. This correlation is a consequence of the constraint that legislator ideal points lie within the unit circle.

The bias of the NOMINATE estimates can be assessed by comparing ML estimates to the means of the bootstrap estimates. There is very little evidence of bias. The mean of the absolute differences between the estimated ideal points and the bootstrapped mean ideal points for the first dimension is 0.023 with a standard deviation of 0.017. The mean absolute difference for the second dimension is 0.029 with a standard deviation of .025. The diameter of the space is 2.0 units, so these mean differences are on the order of 1/100 the span of the space. One important caveat is that the potential bias that we are assessing is finite (or "small–sample") bias as opposed to specification bias.[8] The parametric bootstrap

---

[8]Note that in general ML models cannot be shown to be unbiased in finite samples. Typically, they can be shown to be consistent as the number of observations grows large. The question then is when is the sample large enough that the small sample bias is negligible? The parametric bootstrap provides a way of assessing the size of this small sample bias.
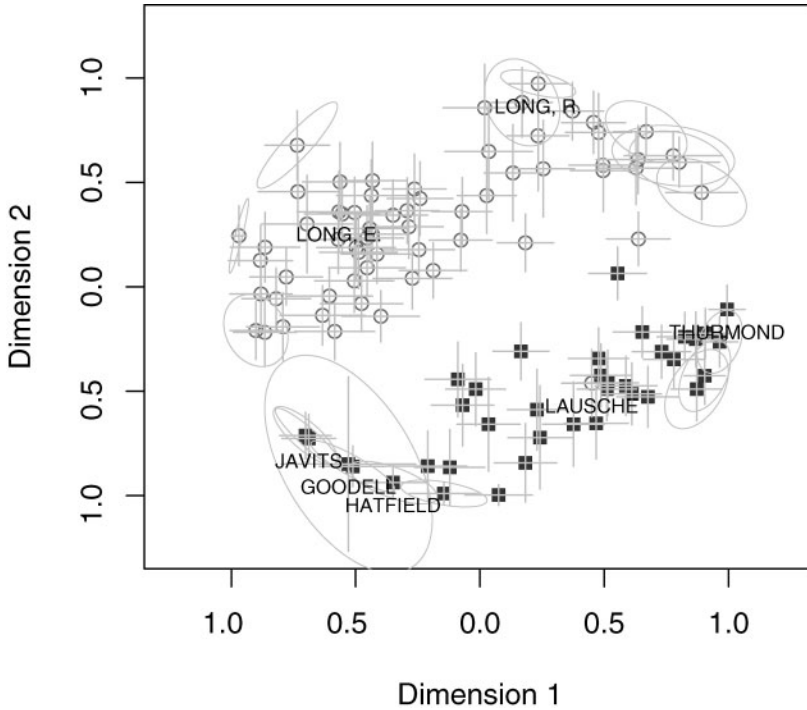
Fig. 1 Estimated legislator locations from the W-NOMINATE model, 90th Senate. The figure shows the estimated legislator locations in two dimensions based on roll calls taken in the 90th Senate. Squares represent Democrats; circles represent Republicans. The vertical and horizontal lines through each estimate show the 95% confidence intervals for each coordinate of a given legislator's position. For most senators the correlation between the estimated first– and second–dimension estimates is very low. Normal theory confidence ellipses are shown for senators whose first–dimension coordinate is correlated with their second–dimension coordinates at $|\rho| > 0.15$. This correlation results from the identifying constraint that senators be located in the unit circle.

conditions on the correctness of the model and therefore cannot detect bias that arises from incorrect specification. However, Lord (1983) and others have noted the existence of substantial small-sample biases in models similar to NOMINATE, so the fact that such bias is shown here to be small is an important finding.

The upper left-hand panel of Fig. 2 shows the bootstrapped standard errors for the two dimensions graphed against one another by senator. As expected, the standard errors on the first dimension are smaller than the second. Only seven senators had larger standard errors on the first dimension than they had on the second. Charles Goodell (R-NY) is a notable outlier on both dimensions. He was appointed to the Senate on September 10, 1968 to replace Robert F. Kennedy. Goodell voted on only 39 roll calls, of which only 31 were scalable (2.5% in minority or better).

The remaining two panels in the first row of Fig. 2 show the conditional standard errors from W-NOMINATE versus the bootstrapped standard errors by dimension. As expected, the conditional standard errors are smaller than the bootstrapped standard errors, especially on the first dimension. On the first dimension the conditional standard errors are about half the magnitude of the bootstrapped standard errors, while on the second dimension the magnitude difference is not as large.
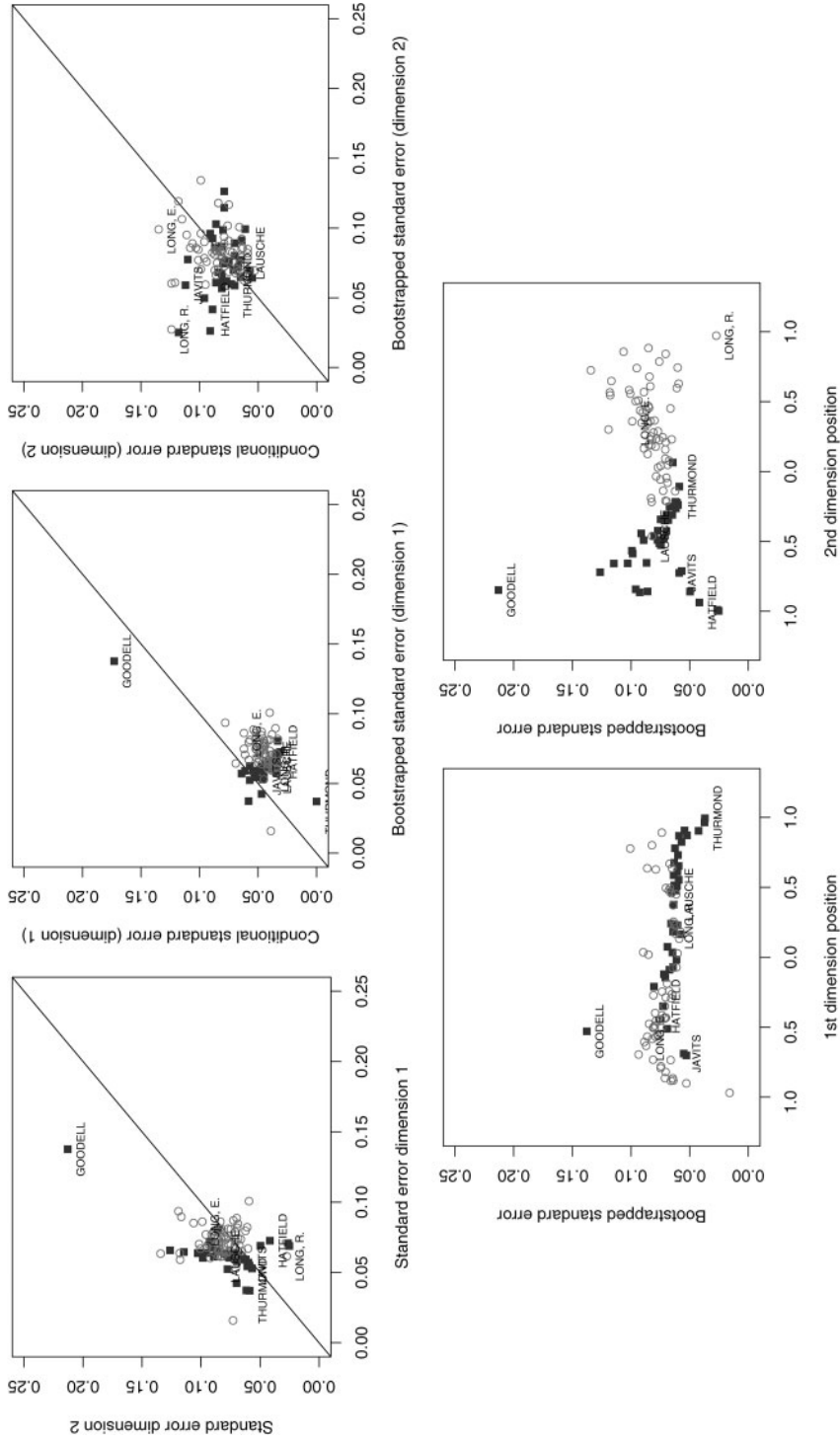
Fig. 2 Bootstrapped and conditional standard error estimates from the W-NOMINATE model, 90th Senate. The first row of panels shows bootstrapped standard error estimates plotted against each other and against the conditional standard error estimates that take the estimated roll call parameters as known. As expected, the conditional standard errors understate the degree of uncertainty in most cases, particularly for legislators with extreme positions. The second row of panels plots the bootstrap estimates of the standard errors of the scores against the the estimated legislator ideal points.
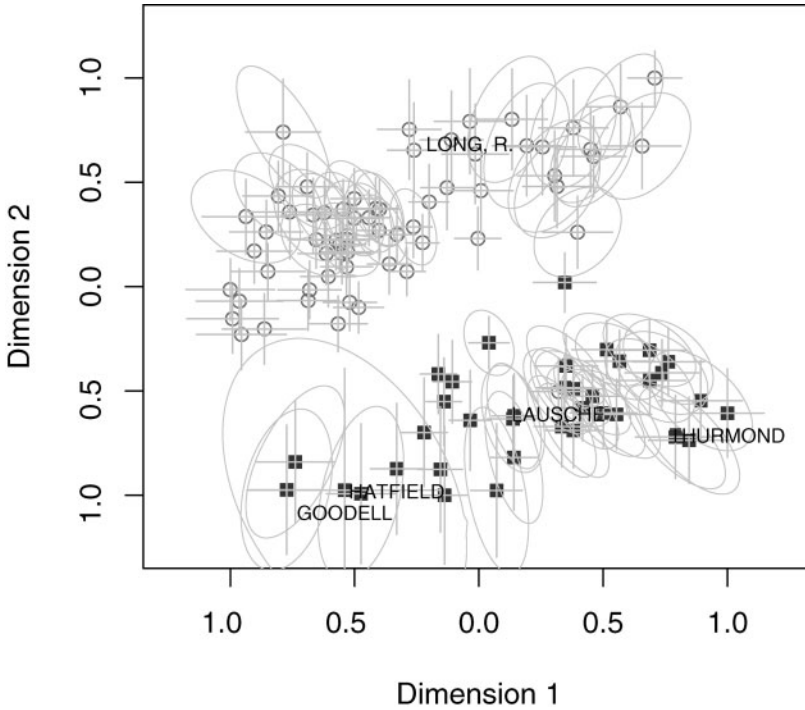
Fig. 3 Estimated legislator locations from the quadratic normal model, 90th Senate. The figure shows the estimated legislator locations in two dimensions based upon roll calls taken in the 90th Senate. Squares represent Democrats; circles represent Republicans. The vertical and horizontal lines through each estimate show the 95% confidence intervals for each coordinate of a given legislator's position. For most senators, the correlation between the estimated first– and second–dimension estimates is very low. Normal theory 95% confidence ellipses are shown for senators whose first–dimension coordinate is correlated with their second–dimension coordinates at $|\rho| > 0.3$.

The second row of panels in Fig. 2 display the senator coordinates on the two dimensions versus the respective bootstrapped standard errors. The unit circle constraint shows up clearly in the plots. Senators near the edges of the space have smaller standard errors.

Figures 3 and 4 show the bootstrap results for QN applied to the 90th Senate. Figure 3 shows the estimated ideal points for the 90th Senate from QN along with the bootstrapped standard errors. The crosshairs through the ideal points show the 95% confidence intervals. The standard errors for QN are smaller than those for W-NOMINATE. On the first dimension, 100 of 101 standard errors were 0.09 or less and on the second dimension, 92 of 101 standard errors were 0.10 or less. As was the case for W-NOMINATE, for most senators the correlation between the estimated first– and second–dimension coordinates is very low. Normal theory confidence ellipses are shown for senators whose first–dimension coordinate is correlated with their second–dimension coordinate at $|\rho| > .30$. This correlation is a consequence of the constraint that legislator ideal points lie within the unit circle.

Figures 1 and 3 are very similar. QN and W-NOMINATE recover essentially the same configuration of ideal points. Regressing the first–dimension W-NOMINATE coordinates on the first–dimension QN coordinates produces an r-square of .982 (W-NOM_1^{st} = $-.025 + 1.316$*QN_1^{st}) and an r-square of .958 for the corresponding second–dimension coordinates (W-NOM_2^{nd} = $.010 + 1.324$*QN_2^{nd}). The fact that the W-NOMINATE
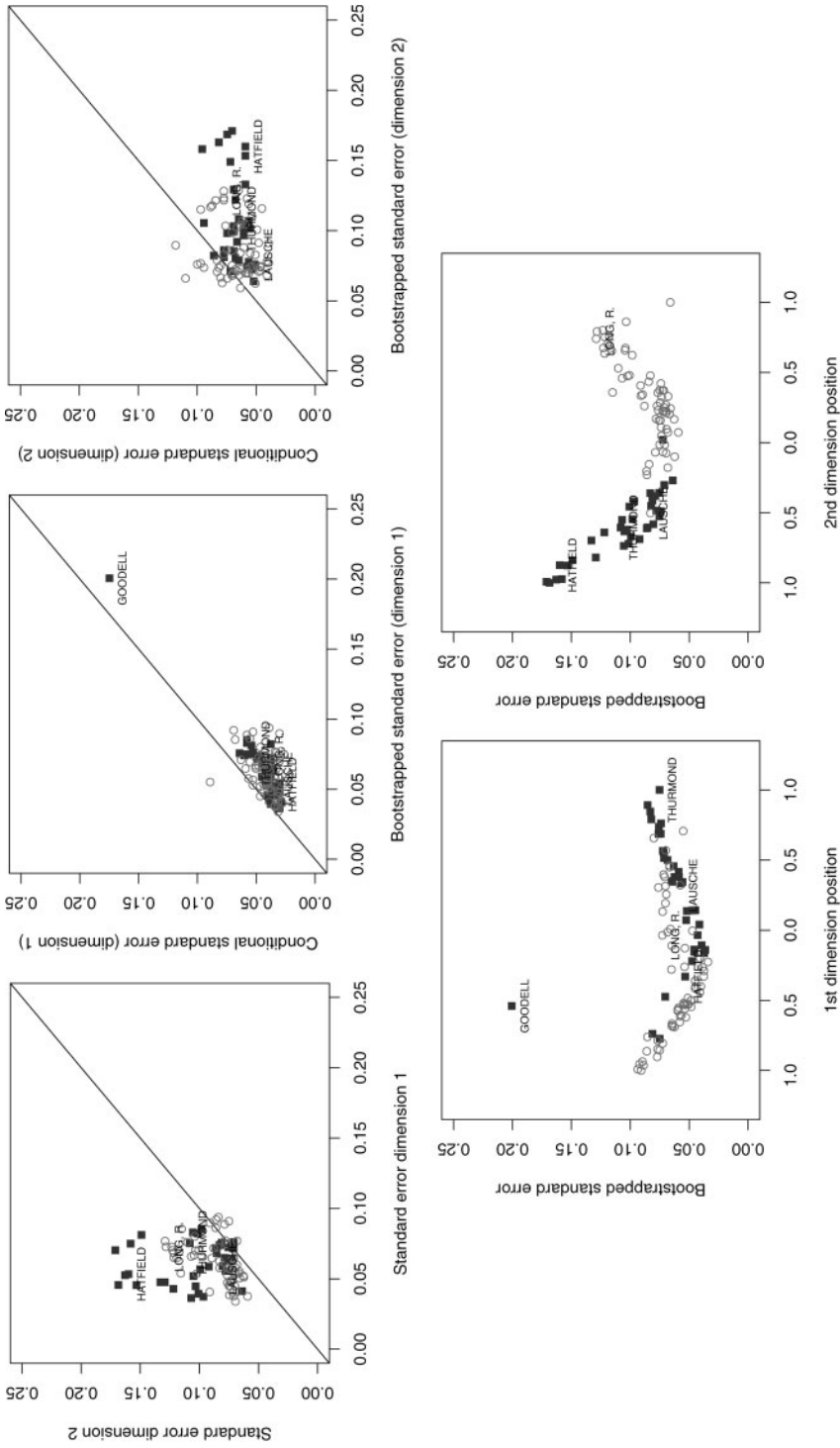
Fig. 4 Bootstrapped and conditional standard error estimates from the quadratic normal model, 90th Senate. The first row of panels shows bootstrapped standard error estimates plotted against each other and against the conditional standard error estimates that take the estimated roll call parameters as known. As expected, the conditional standard errors understate the degree of uncertainty in most cases, particularly for legislators with extreme positions. Not shown on the right panel is Senator Goodell, whose coordinates would be (0.30, 0.71). The second row of panels plots the bootstrap estimates of the standard errors of the QN scores against the estimated senator locations. Not shown on the right panel is Senator Goodell, whose coordinates in that plot would be (−0.97, 0.30).

114

configuration is slightly "inflated" vis–à–vis the QN configuration is due to W-NOMINATE setting the most extreme legislators at opposite ends of the first dimension to −1 and +1. In a one-dimensional scaling there will *always* be at least one legislator at −1 and at least one legislator at +1. When a second dimension is estimated, some of those legislators at or near −1 or +1 may end up on the rim of the circle. QN estimates both dimensions simultaneously and only constrains legislators to lie within the unit hypersphere. Hence a typical QN configuration will not be as "inflated" as the corresponding configuration from W-NOMINATE. To make the results more comparable across methods, the QN results are rescaled such that the most extreme members on each end of each dimension are placed at −1 or 1.[9]

Again, there is very little evidence of finite-sample bias. The mean of the absolute differences between the estimated ideal points and the bootstrapped mean ideal points for the QN first dimensions is 0.022 with a standard deviation of 0.013. The mean absolute difference for the second dimension is 0.029 with a standard deviation of .025. These mean differences are on the order of 1/100 the span of the space.

The upper left-hand panel of Fig. 4 shows the bootstrapped standard errors graphed against one another by senator. Just as with W-NOMINATE, the QN standard errors on the first dimension are smaller than the second. Once again, Charles Goodell (R-NY) is a notable outlier on both dimensions.

The next two panels of Fig. 4 show the conditional standard errors from QN versus the bootstrapped standard errors by dimension. As expected, the conditional standard errors are smaller than the bootstrapped standard errors. On both dimensions the conditional standard errors are about half the magnitude of the bootstrapped standard errors.

The bottom row of panels in Fig. 4 display the senator coordinates on the two dimensions versus the respective bootstrapped standard errors. The patterns in Fig. 4 for QN are just the opposite of those for W-NOMINATE in Fig. 2. In W-NOMINATE, legislators are constrained to lie on the −1 to +1 interval when the first dimension is estimated. Because W-NOMINATE estimates one dimension at a time, this has the effect that legislators who are extreme on the first dimension tend to lie near the unit circle when the second dimension is estimated. Since they cannot wander out of the unit circle when the second dimension is estimated, extremists have little "wiggle room" in the W-NOMINATE framework. In contrast, in QN the dimensions are estimated simultaneously. The different constraint structure in QN means that extremists have more wiggle room. Hence legislators furthest from the center (recall that we have scaled the QN coordinates to −1 to +1 for graphical purposes only) tend to have slightly larger standard errors. However, this difference between the two procedures is not really that great. Note that the standard errors are small for both procedures—especially for the bulk of legislators who are not extremists.

4   Comparing W-NOMINATE, QN, and IDEAL

In this section we compare the bootstrapped NOMINATE and QN estimator to the IDEAL model of Clinton et al. (2002). The IDEAL model begins with the same random utility model as QN. However, IDEAL is a Bayesian estimator that is estimated using Markov chain Monte Carlo (MCMC) models. MCMC models yield estimates of the complete joint posterior distribution of the estimates from which point estimates and confidence intervals

---

[9]This is accomplished by simple linear transformations that are applied to the ML estimates and to the bootstrapped standard errors.

and other quantities of interest can be computed (see Jackman 2000a, 2000b, 2001). MCMC models are, however, computationally intensive and time consuming to estimate even for modern computers. Until now, however, MCMC was only the way to obtain unconditional estimates of parameter uncertainty.

Unfortunately, comparisons among ideal point estimators cannot be made directly. Because the scale and location of the issue dimensions are arbitrary up to rotations of axes, shifts of axis location, and stretches of scale, disagreement between the first–dimension positions as measured by any two methods cannot be taken as direct evidence of fundamental disagreement between the two methods. This problem is further exacerbated when, as is the object here, we wish to make comparisons between the uncertainties associated with the ideal point estimates made by different methods. Beyond any differences in the orientation, location, and metric of the issue space, the definition of the issue space affects how the uncertainty is distributed across the ideal point estimates. For example, if the space is identified by fixing the location of three of the ideal points (in two dimensions), those three members will, by definition, have no uncertainty associated with their locations. Any sampling variability that is associated with the voting records of the members whose locations are fixed will be accounted for through the uncertainty in the estimates of the remaining members whose positions are measured *relative* to those members whose positions are fixed a priori. What is identified in spatial models are the distances between member locations, thus the relevant uncertainties in ideal point estimates are ultimately uncertainties in the distances between member locations.

In order to make comparisons between IDEAL and either QN or NOMINATE, we must first account for any arbitrary differences between the issue space recovered by IDEAL and the space recovered by the other two methods. In what follows, we remove any arbitrary scale difference between IDEAL and NOMINATE by postprocessing the posterior draws from a specification of IDEAL that does not uniquely determine the issue space. The postprocessing involves linearly transforming each posterior draw from IDEAL to minimize the squared distance between each posterior draw and the NOMINATE estimates.[10] This approach is similar to fixing the means and variance along each dimension and fixing the orientation of the axes.

The IDEAL estimates and associated uncertainties are shown in Fig. 5. Note that the overall pattern is very similar to that found for NOMINATE and for QN. Indeed, as shown in Table 1, the correlations among the estimates across the three methods exceed 0.99 for the first–dimension positions and 0.97 for the second–dimension scores.

Figure 6 shows plots of the estimated standard errors (a posteriori standard deviations) from IDEAL plotted against each other and against the first– and second–dimension point (mean a posteriori) estimates. These plots are similar to Fig. 4 even though QN uses constraints.

Table 1 reports the correlations among the point estimates and standard errors for the three methods and Fig. 7 plots the standard error estimates of the three methods against one another. Looking first at the estimates themselves, we find a very high level of agreement among all three methods. The point estimates correlate over 0.97. This result is consistent with previous comparisons of roll scaling techniques (see Heckman and Snyder 1997; Poole and Rosenthal 1997; Poole 2000, 2001).

---

[10]This amounts to running regression of each of the NOMINATE coordinates on both of the IDEAL coordinates for each of the posterior draws from IDEAL. The original draws from IDEAL are then replaced with the predicted values from the regressions. Identification of the IDEAL model by postprocessing the MCMC output is Jackman's preferred method of identification when multiple dimensions are to be estimated (personal correspondence).
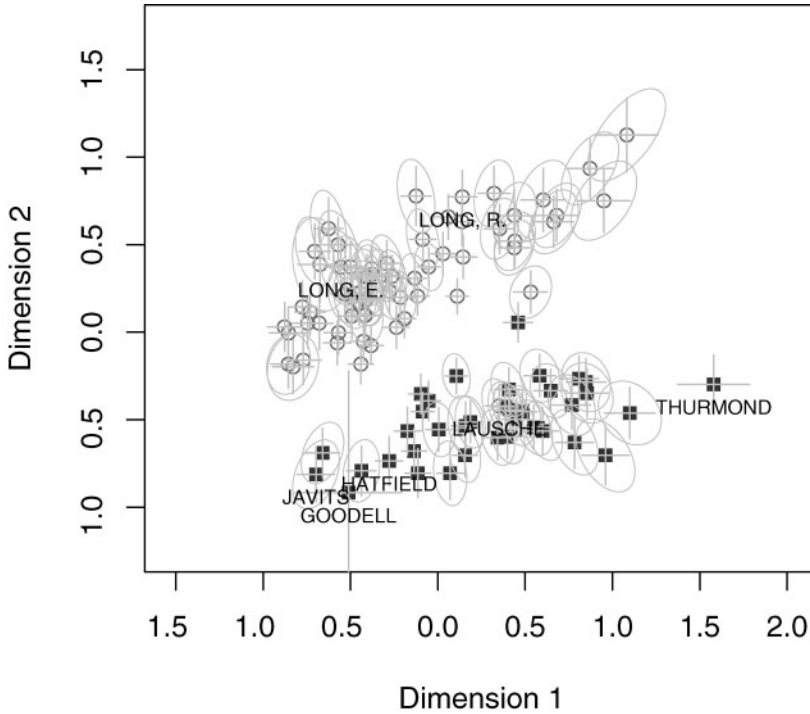
Fig. 5 Estimated legislator locations from IDEAL, 90th Senate. The figure shows the estimated legislator locations in two dimensions based upon roll calls taken in the 90th Senate. Squares represent Democrats; circles represent Republicans. The vertical and horizontal lines through each estimate show the 95% credible intervals for each coordinate of a given legislator's position. For most senators, the posterior correlation between the estimated first– and second–dimension estimates is very low. Normal theory 95% credible ellipses are shown for senators whose first–dimension coordinate is correlated with their second–dimension coordinates at $|\rho| > 0.3$.

As shown in Table 2, the average standard errors are small for all three models. Even though the range of the ideal point estimates has been normalized, there is still some variation in the standard deviations in the estimates across methods as seen in the second column of the table. The third column shows a rough "signal–to–noise" ratio for the

Table 1 Correlation among ideal-point estimates and among their standard errors from three different estimators, 90th Senate

	Quadratic normal and IDEAL	NOMINATE and IDEAL	NOMINATE and quadratic normal
First dimension:			
Estimate	0.99	0.99	0.99
Standard error	0.25	0.70	0.16
Second dimension:			
Estimate	0.97	0.99	0.98
Standard error	0.53	0.46	0.16

Note. Correlations are shown among the ideal point estimates and standard errors of senators' ideal points as estimated by W-NOMINATE, QN, and IDEAL. For IDEAL, the "standard errors" are posterior standard deviations.
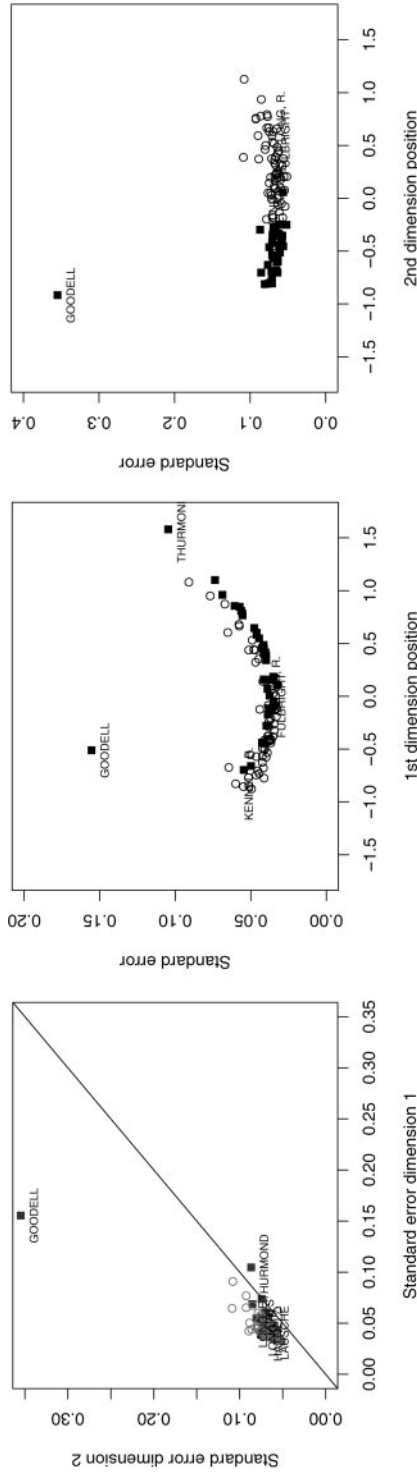
Fig. 6 Estimated positions and standard errors from IDEAL, 90th Senate. The first panel plots posterior uncertainties in the first–dimension ideal points against the uncertainties in the second–dimension ideal points; as expected, the first–dimension ideal points are generally more precisely estimated (most of the points lie above the 45° line). The second and third panels plot the first– and second–dimensional ideal points against their posterior standard deviations (standard errors). Extreme members are found to have larger standard errors in IDEAL.
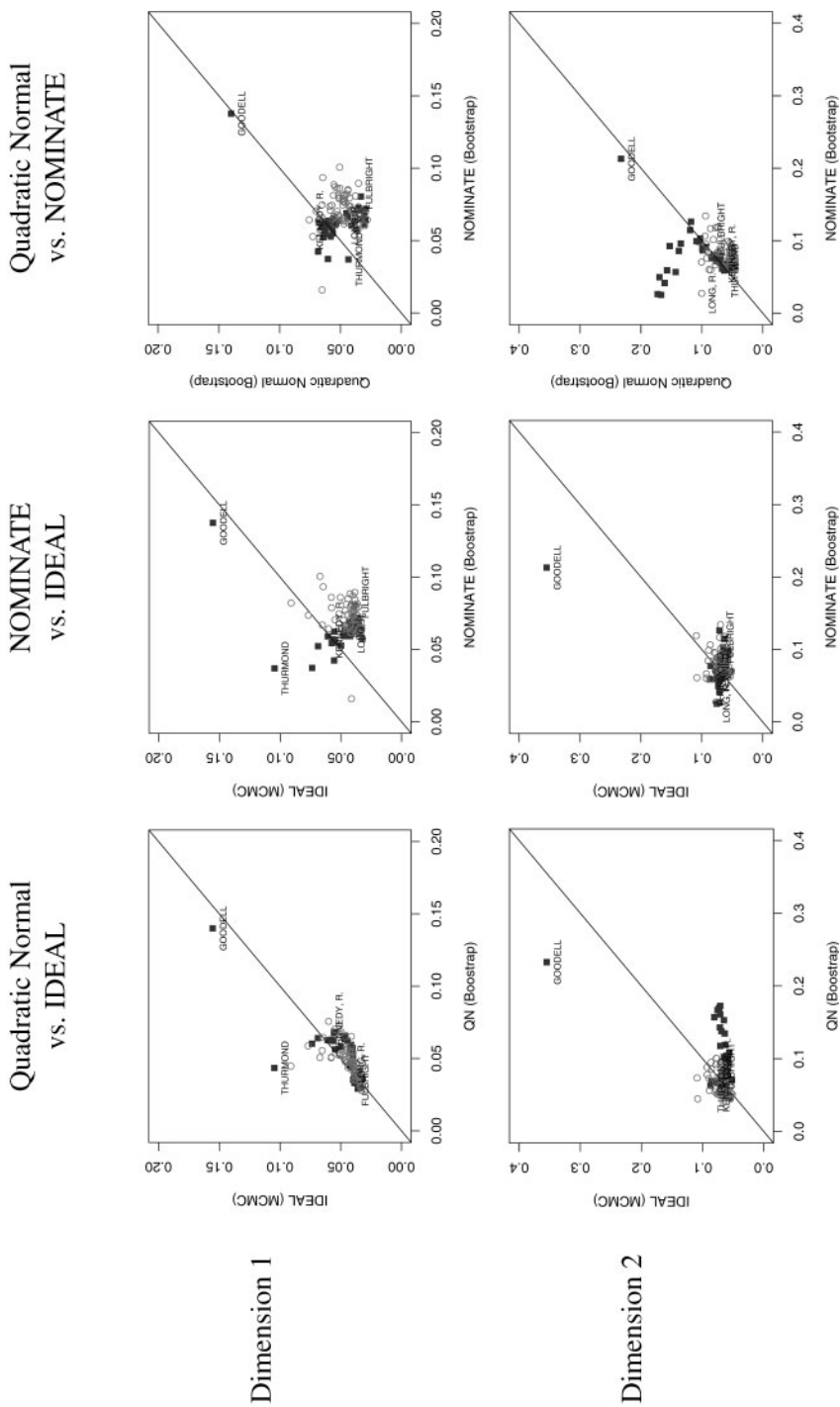
Fig. 7 Comparisons of two-dimensional ideal point estimates and their standard errors, 90th Senate. Standard error estimates made by a given pair of methods and for a given issue dimension are plotted against each other. Notice that aside from identifying Goodell's location as highly uncertain (due to his abbreviated voting record), there is relatively little agreement among the three methods as to which members are more or less precisely estimated.

119

Table 2   Uncertainty in ideal point estimates, 90th Senate

	Average SE	Standard deviation of scores	Signal-to-noise ratio
First dimension:			
NOMINATE	0.07	0.56	8.1
Quadratic-normal	0.05	0.42	8.4
IDEAL	0.05	0.55	11.8
Second dimension:			
NOMINATE	0.08	0.51	6.3
Quadratic-normal	0.08	0.35	4.6
IDEAL	0.07	0.48	6.7

Note. Average standard errors are the average of the bootstrapped standard error estimates across senators for the NOMINATE and quadratic normal models and the average posterior standard deviations for the IDEAL models. The second column shows the standard deviation of the estimated scores across the 100 senators. The signal-to-noise ratio is the ratio of the first column to the second.

estimates, by dividing the standard deviation in the point estimates (the signal) by the average standard error of the estimates (the noise). Even with these adjustments, however, the table does not allow a direct comparison of the uncertainty across methods. IDEAL loads a considerable amount of uncertainty in the locations of a few extreme members, while QN and particularly NOMINATE spread the uncertainty more evenly across members.

All three methods yield very similar overall estimates of uncertainty. The posterior standard deviations of IDEAL and the bootstrapped standard errors generally fall between 0.03 and 0.12. However, there is less agreement as to which senators are more or less reliably measured. Table 1 reports the correlations among the point estimates and standard errors. The positive correlations found between the standard error estimates across methods are largely the result of the high level of uncertainty in the location of Goodell.

Differences in which members are more precisely estimated are due to differences in the utility functions used by each model and in the identifying restrictions. Under the NOMINATE model, which uses Gaussian utility functions and identifies the space by limiting ideal points to fall inside the unit circle, extreme members are estimated to have small standard errors. In every bootstrapped sample the same members are found to be the most extreme and the most extreme position is by definition within the unit circle. If in a given bootstrap sample the left-most member is estimated to be relatively more extreme than in the average bootstrap sample, it is the rest of the Senate, and not the extreme member, whose ideal points shift. On the other hand, IDEAL does not constrain the policy space to a particular interval or region (even after postprocessing). Thus, if the left-most member is relatively more extreme in a given posterior draw, that relative extremism can be reflected in her own ideal point. Additionally, under the quadratic utility function employed by IDEAL and QN, the likelihood of casting a vote for the left alternative is monotonically increasing as a senator's ideal point is moved to the left for any bill parameters. However, under the Gaussian utility model employed by NOMINATE, moving to the left does not always produce a higher probability of choosing the left alternative. Thus the NOMINATE model does not have the same centripetal pressure to push extreme senators farther from moderate senators at an increasing rate as their voting records approach perfect ideological voting. These two effects account for the variation in standard error estimates seen in Fig. 7.
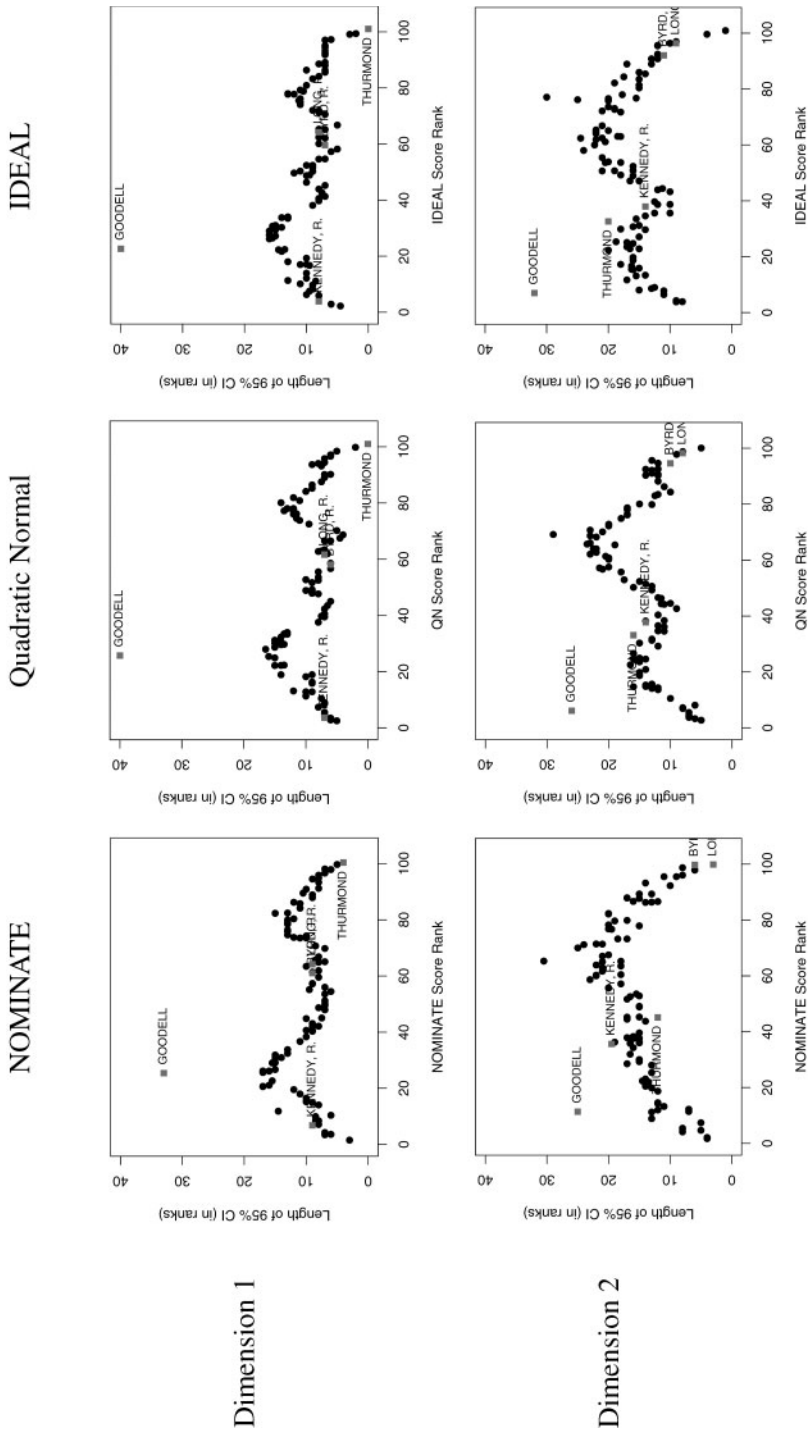
Fig. 8 Size of rank confidence interval versus estimated rank for three ideal point models, 90th Senate. The figure plots the estimated rank position of each senator in the 90th Senate against the length of the bootstrapped confidence interval of each senator's rank position estimate. In all cases, the moderates and extremists have very small confidence intervals, while confidence intervals are much larger for those between the center and the wings.

Table 3 Uncertainty in ideal point estimates, 105th Senate

	Average SE	*Standard deviation of scores*	*Signal-to-noise ratio*
NOMINATE	0.059	0.671	11.37
Quadratic-normal	0.033	0.740	22.30
IDEAL	0.043	0.555	12.90

Note. Average standard errors are the average of the bootstrapped standard error estimates across senators for the NOMINATE and quadratic normal models and the average posterior standard deviations for the IDEAL models. The second column shows the standard deviation of the estimated scores across the 100 senators. The signal-to-noise ratio is the ratio of the first column to the second.

In order to more directly compare the uncertainty of the estimates generated by each method, we remove differences in scale by considering the estimated rank order of senators along each dimension and confidence intervals for these rank orderings. By looking at ranks (an inherently relative measure of location), much of the apparent difference in the uncertainty associated with each member's location across estimators vanishes. The bootstrap confidence intervals for the ranks are easily computed. For each bootstrap sample, the ideal point estimates are ranked. The 0.025 and 0.975 quantiles of each senator's rank position across the 1000 bootstrapped samples are taken as lower and upper bounds of a 95% confidence interval for each senator's rank position. The results of this analysis are shown in Fig. 8. Because ranks are inherently scale-free and relative, we see much less variation across the three methods. In terms of ranks, all three methods provide strikingly similar estimates of uncertainty. This finding demonstrates how variation in the estimation uncertainty across senators' locations is largely a function of the constraints that must be imposed in order to identify the scale of the issue dimension. Once the choice of scale is removed (as when ranks are considered), the more fundamental variation is revealed. Those at the end of the scale can be said with great confidence to be at the ends; those in the middle are similarly pinned down. More difficult to disentangle are the those liberal and conservative members located near the median of each party's caucus.

Overall, we find that all three methods provide very similar estimates of senators' ideal points and produce very similar uncertainty estimates once arbitrary differences in the definition of the policy space and identifying restrictions are accounted for. Remaining differences are attributable to differences between the assumed utility functions and an inability to completely account for differences in identifying restrictions.

5   Beyond Standard Errors: Bootstrapping Auxiliary Quantities of Interest

One particularly useful aspect of the bootstrap is that it allows us to quickly and easily compute confidence intervals and other measures of uncertainty for many of the auxiliary quantities of interest that can be inferred from these models. In the previous section, we used the bootstrap to estimate confidence intervals for the rank-order positions of members of the 90th Senate along each of two issue dimensions. In this section, we present a few examples of how the uncertainties of other quantities of interest derived from ideal point estimates can be ascertained through the use of the parametric bootstrap. For these examples, we turn to data from the 105th Senate in part because the names of the central actors will be more familiar to the reader and because the 105th Senate is nearly unidimensional in contrast to the two-dimensional 90th Senate. Unidimensionality is relevant here because the quantities of interest that we consider, such as the location of the median voter in the chamber, the
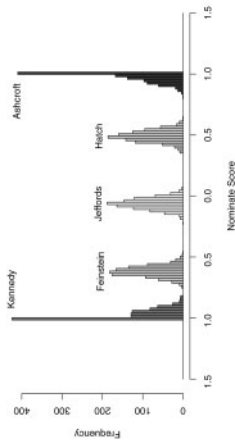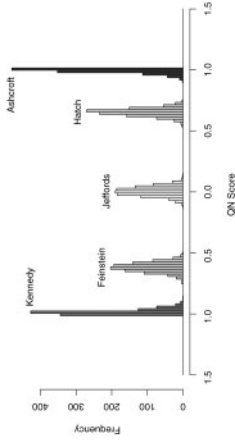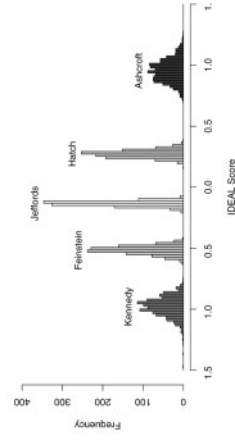
Fig. 9 Sampling or posterior distributions for three ideal point models, 105th Senate. The figure shows the estimated sampling distributions of the ideal point estimates of five members of the 105th Senate based on three models/estimators.

123

Table 4 Location of the median senator, 105th Senate

Method	Estimate	SE	95% CI
NOMINATE	0.14	0.04	(0.06,0.22)
QN	0.26	0.03	(0.20,0.32)
IDEAL	−0.00	0.02	(−0.03,0.03)

Note. The estimated median senator's location is shown for each of three models.

location of the filibuster pivot, and the identity of the median and filibuster pivot, are theoretically compelling in the context of a unidimensional setting.

As was the case for the 90th Senate, ideal point estimates for the 105th Senate are very highly correlated (greater than 0.98 correlation). Table 3 shows the average standard errors for estimates of one-dimensional ideal points for the 105th Senate.[11] As was the case in the presentation of the 90th Senate, all three models present similar estimates of estimation uncertainty. The signal-to-noise ratios are somewhat higher here due to the unidimensionality and the high degree of ideological voting seen in the 105th Senate.

Beyond simply estimating standard errors and confidence intervals, the bootstrap provides estimates of the complete sampling distribution of the model parameters. Figure 9 shows histograms of the (bootstrapped) sampling distribution of the ideal point estimates of five senators. The figure shows how the constraints in NOMINATE and QN limit the variability of the estimates of extremists such as Kennedy and Ashcroft, both of whom have asymmetric sampling distributions under QN and NOMINATE. However, the posterior distributions from IDEAL show the greatest uncertainty in the locations of the extremists.

Table 4 shows the position of the estimated chamber median for the 105th Senate.[12] The bootstrapped standard errors are calculated by finding the position of the median senator in each bootstrap estimate and then taking the standard deviation across those medians. Note that the standard error of this estimate is smaller than the average standard error for the individual members. This is true both because the median is among the members whose ideal point is estimated to be quite small and because the identity of the median is not fixed across bootstrap estimates (or posterior draws in the case of IDEAL). Thus, in samples in which a potential median voter is estimated to have a relatively more extreme position than other nearby members, some other senator will be estimated to be the median.

Table 5 Location of the filibuster pivot, 105th Senate

Method	Estimate	SE	95% CI
NOMINATE	−0.52	0.04	(−0.59,−0.45)
QN	−0.47	0.03	(−0.54,−0.44)
IDEAL	−0.41	0.02	(−0.46,−0.38)

Note. Table shows the estimated location and standard error of the filibuster pivot in the 105th Senate.

---

[11]The IDEAL estimates presented in this section are postprocessed as in the previous section. In this case, only one dimension is extracted and the postprocessing involves simply removing (arbitrary) variation in scale location across posterior draws.

[12]Londregan and Snyder (1994) use a similar bootstrap procedure to test for committee outliers. However, their estimates of senators' locations are based on interest groups' rating scores.

Table 6 Who was the median senator in the 105th senate?

	NOMINATE	QN	IDEAL
Chafee	0.005	0.003	0.001
Snowe	0.157	0.068	0.089
D'Amato	0.243	0.201	0.301
Collins	0.595	0.728	0.609

Note. Table shows for NOMINATE and QN the bootstrapped sampling distribution over the identity of the median senator in the 105th Senate. For the Bayesian IDEAL model, each value is the posterior probability that a given member is the median.

Similar results are found for the location of the filibuster pivot and are shown in Table 5. Interestingly, while there is greater uncertainty in the location of members close to the estimated filibuster pivot (the location of the 40th most liberal member), there are many more members who are in close proximity to it. Thus the standard error of the estimated pivot position is similar to that of the estimated median.

Tables 6 and 7 show the estimated sampling distributions over the identities of the median and filibuster pivot Senator. For the Bayesian IDEAL estimator, this can be directly interpreted as the posterior probability that a given senator was in fact the median or the pivot. For the frequentist QN and NOMINATE models, we cannot correctly make the same interpretation. Nor, however, can these probabilities be construed as p-values for tests of the hypothesis that a given senator is the median or filibuster pivot because these probabilities are conditional on the estimated model and not on the validity of the null hypothesis. However, they do give us a measure of confidence in the assertion that a particular member was indeed the filibuster pivot or median.

Table 7 Who was the filibuster pivot in the 105th Senate?

	NOMINATE	QN	IDEAL
Bryan	*	*	0.016
Kohl	*	0.013	0.048
Bob Kerrey	0.011	0.004	0.002
Biden	0.012	0.004	0.008
Moynihan	0.022	0.027	0.031
Reid	0.027	0.014	0.015
Robb	0.031	0.064	0.079
Ford	0.033	0.061	0.092
Lieberman	0.056	0.093	0.121
Landrieu	0.068	0.018	0.005
Bob Graham	0.103	0.173	0.193
Hollings	0.161	0.198	0.226
Baucus	0.188	0.234	0.145
Byrd	0.213	0.075	0.019

Note. Table shows for NOMINATE and QN the bootstrapped sampling distribution over the identity of the filibuster pivot Senator in the 105th Senate. For the Bayesian IDEAL model, each value is the posterior probability that a given member is the filibuster pivot.
*$P < .001$.

6   Conclusion

Our results show that the parametric bootstrap is an effective method to obtain standard errors for the parameters of ideal point estimation methods. We show that in general the standard errors for NOMINATE and QN scores are relatively small. We also show how the bootstrap can be used to provide estimates of the many quantities of interest associated with ideal point estimates.

In future work we will describe how bootstrapped estimates can be used to correct for measurement error in regression models in which NOMINATE scores are used as dependent or independent variables.[13] We also plan to make bootstrapped NOMINATE estimates available to researchers so that in future work scholars can incorporate the call of methodologists such as Herron and Shotts (2003) to explicitly account for the estimation uncertainty in the variables that are used in subsequent analysis.

References

Bollen, Kenneth A., and Robert A. Stein. 1992. "Bootstrapping Goodness-of-Fit Measures in Structural Equation Models." *Sociological Methods and Research,* 21:205–229.

Brownstone, David, and Robert G. Valletta. 1996. "Modeling Earnings Measurement Error: A Multiple Imputation Approach." *Review of Economics and Statistics* 78:705–717.

Clinton, Joshua, Simon D. Jackman, and Douglas Rivers. 2002. "The Statistical Analysis of Roll Call Data." Working paper. Stanford, CA: Stanford University.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7:1–26.

Efron, Bradley. 1982. "Maximum Likelihood and Decision Theory." *Annals of Statistics* 10:340–356.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Hall, Peter. 1985. "Resampling a Coverage Pattern." *Stochastic Processes and Their Applications* 20:231–246.

Hall, Peter. 1994. "Methodology and Theory for the Bootstrap." In *Handbook of Econometrics,* Vol. IV, eds. R. F. Engle and D. L. McFadden. New York: Elsevier, pp. 2342–2379.

Heckman, James J., and James M. Snyder. 1997. "Linear Probability Model s of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics* 28:142–189.

Herron, Michael C., and Kenneth W. Shotts. 2003. "Using Ecological Inference Point Estimates as Dependent Variables in Second-Stage Linear Regressions." *Political Analysis* 11:44–64.

Jackman, Simon D. 2000a. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44:375–404.

Jackman, Simon D. 2000b. "Estimation and Inference Are 'Missing Data' Problems: Unifying Social Science Statistics via Bayesian Simulation." *Political Analysis* 8:307–332.

Jackman, Simon D. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking." *Political Analysis* 9:227–241.

Londregan, John. 2000. "Estimating Legislator's Preferred Points." *Political Analysis* 8:35–56.

Londregan, John, and James M. Snyder. 1994. "Comparing Committee and Floor Preferences." *Legislative Studies Quarterly* 19:233–266.

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

McFadden, Daniel. 1976. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5:363–390.

Mooney, Chrisopher Z. 1996. "Bootstrap Statistical Inference: Examples and Evaluations of Political Science." *American Journal of Political Science* 40:570–602.

Poole, Keith T. 2000. "Non-Parametric Unfolding of Binary Choice Data." *Political Analysis* 8:211–237.

Poole, Keith T. 2001. "The Geometry of Multidimensional Quadratic Utility in Models of Parliamentary Roll Call Voting." *Political Analysis* 9:211–226.

Poole, Keith T., and Howard Rosenthal. 1985. "A Spatial Model For Legislative Roll Call Analysis." *American Journal of Political Science* 29:357–384.

---

[13]Brownstone and Valletta (1996) take a similar approach in a related area.

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. 1997. *Congress: A Political-Economic History of Roll Call Voting.* New York: Oxford University Press.

Schonemann, Peter H. 1966. "A Generalized Solution of the Orthogonal Procrustes Problem." *Psychometrika* 31:1–10.

Young, G. Alastair. 1994. "Bootstrap: More than a Stab in the Dark?" *Statistical Science* 9(3):382–415.