

Small Chamber Ideal Point Estimation

Michael Peress

*Department of Political Science, University of Rochester, 326 Harkness Hall,
University of Rochester, Rochester, NY 14627
e-mail: mperess@mail.rochester.edu (corresponding author)*

Ideal point estimation is a topic of central importance in political science. Published work relying on the ideal point estimates of Poole and Rosenthal for the U.S. Congress is too numerous to list. Recent work has applied ideal point estimation to the state legislatures, Latin American chambers, the Supreme Court, and many other chambers. Although most existing ideal point estimators perform well when the number of voters and the number of bills is large, some important applications involve small chambers. We develop an estimator that does not suffer from the incidental parameters problem and, hence, can be used to estimate ideal points in small chambers. Our Monte Carlo experiments show that our estimator offers an improvement over conventional estimators for small chambers. We apply our estimator to estimate the ideal points of Supreme Court justices in a multidimensional space.

1 Introduction

Ideal point estimation is a topic of central importance in political science. Published work relying on ideal point estimates by Poole and Rosenthal (1997) of the U.S. Congress are too numerous to list. Recent work has applied ideal point estimation to the state legislatures (Wright and Osborne 2002), Latin American chambers (Desposato 2004; Londregan 2000b), the Supreme Court (Martin and Quinn 2002), and many other chambers. The conventional methods—the maximum likelihood and Bayesian estimators—perform well when both the number of voters and the numbers of bills is large. Yet, many important applications of ideal point estimation require the analysis of small chambers.

The maximum likelihood estimator (Poole and Rosenthal 1997) requires estimating both legislator-specific and bill-specific parameters. This method is extremely effective in estimating ideal points in the U.S. House and Senate. It is much less effective in estimating ideal points in smaller chambers due to finite sample identification problems. For example, Clinton, Jackman, and Rivers (2003) report that, when run on data from the Supreme Court, Poole's W-NOMINATE program places all the justices at -1 or 1 .

Bayesian estimators have been proposed as solutions to this problem (Bafumi et al. 2005; Clinton, Jackman, and Rivers 2004; Martin and Quinn 2002). Bayesian estimators are able to deal with the finite sample identification problem because informative prior

Authors' note: The author would like to thank Michael Bailey, Tasos Kalandrakis, Jeff Lewis, Keith Poole, the anonymous referees, and the participants of seminars at the American Political Science Association meetings (Philadelphia 2006) and the Midwest Political Science Association meetings (Chicago 2007) for their helpful comments and suggestions.

© The Author 2009. Published by Oxford University Press on behalf of the Society for Political Methodology. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

distributions penalize extreme estimated parameter values. Bayesian estimators, however, do not solve the incidental parameters problem (Lancaster 2000).

The incidental parameters problem occurs in small chambers because both the maximum likelihood estimator and the Bayesian estimator require estimating a very large number of parameters (the bill-specific parameters) from a small amount of data. Inconsistency of the bill-specific parameter estimates is then translated into inconsistency of the estimated ideal points.¹

Ideal point models are a subclass of nonlinear panel data models, which are themselves subject to the incidental parameters problem. A number of solutions to the incidental parameters problem have been proposed. Neyman and Scott (1948) suggested employing a set of estimating equations that identify the parameters of interest but do not depend on the fixed effects. Andersen (1970, 1973) introduced conditional fixed-effects estimators, which provide a practical way of finding estimating equations that meet the requirements of Neyman and Scott. Unfortunately, conditional fixed-effects estimators are not universally applicable.² The only known conditional fixed-effects estimator for ideal point models requires unconventional and unattractive assumptions about the disturbance term (Heckman and Snyder 1997).

An alternative solution to the incidental parameters problem comes from random-effects estimators, which avoid the incidental parameters problem by integrating over the individual-specific parameters rather than estimating them (Kiefer and Wolfowitz 1956; Bock and Aiken 1981). The drawback of this approach is that one must make more restrictive assumptions about the individual-specific parameters.³ Londregan (2000a, 2000b) develops a random-effects estimator to estimate ideal points but requires the cut-points to be normally distributed. A second disadvantage of random-effects estimators is that they typically require numerical integration. Although this numerical integration can be performed using quadrature methods, the curse of dimensionality means that it is impractical to estimate ideal point models whose dimensionality is greater than one.

Bailey (2001) and Lewis (2001) have developed alternative random-effects estimators for the case where there are a large number of legislators but a small number of votes. Bailey's procedure focuses on recovering ideal point estimates, whereas Lewis' procedure focuses on recovering the distribution of ideal points across a number of districts. Although both Bailey's and Lewis' estimators have important applications, they will not be effective for the small chamber problem.

Bayesian and random-effects estimators provide the two most important approaches to estimating ideal points in small chambers. Each has a significant drawback—whereas the Bayesian estimator suffers from the incidental parameters problem, the random-effects estimator requires potentially restrictive assumptions about the distribution of bill-specific estimators, and current implementations require numerical integration. Londregan's random-effects estimator does not easily extend to higher dimensions.

In this paper, we will develop a more general random-effects estimator for ideal point estimation. This estimator will have a number of advantages. First, the estimator is more flexible (in the sense of allowing it to fit a larger class of models). Second, the estimator can

¹See Lancaster (2000) for a discussion of the incidental parameters problem as it relates to Bayesian estimation.

²Conditional fixed-effects estimators are known to exist for the fixed-effects nonlinear regression model, the fixed-effects logit model, and the fixed-effects poisson model. No conditional fixed-effects estimator is known for the fixed-effects probit model, and it is thought that one does not exist (Hsiao 1986).

³The analyst must typically restrict the functional form of the distribution of random effects, as well as the type of dependence between the random effects and other variables in the model.

be viewed as a first-order approximation based on matching the first two moments of the vector of latent variables defining the model. As a consequence, our estimator will be robust to misspecification of the distribution of random effects. Third, the estimator easily extends to the multidimensional case. Finally, the estimator can be implemented in a way that does not suffer from the curse of dimensionality since the relevant integrals in the likelihood function can be computed using simulation methods.

Choosing between the Bayesian estimator and our small chamber estimator requires evaluating the relative costs of the incidental parameters problem and misspecification of the bill-specific parameters. We will show, through Monte Carlo experiments, that our estimator is more effective in recovering ideal points in small chambers than the Bayesian estimator.⁴ Our small chamber estimator is quite effective even when the distribution of bill-specific parameters deviates significantly from normality. We find that our estimator is much more robust to misspecification than the Bayesian estimator is to the incidental parameters problem.

An application to voting in the U.S. Supreme Court will demonstrate the robustness, stability, and broad applicability of our estimator. The application shows that our estimator performs well, and we can corroborate many of the findings from our Monte Carlo simulations. Our results indicate that the U.S. Supreme Court exhibits at least two important dimensions of conflict. In the two-dimensional case, our results suggest a liberal-conservative dimension and a “judicial activism” dimension.

2 Small Chamber Estimation

The quadratic-normal ideal point model can be described as follows. Let $y_{n,t} = 1$ denote a yea by individual n on bill t and let $y_{n,t} = 0$ denote a nay. There exist parameters α_n and δ_t such that $y_{n,t}^* = \delta_{t,0} + \delta_{t,1}\alpha_{n,1} + \delta_{t,2}\alpha_{n,2} + \dots + \delta_{t,D}\alpha_{n,D} + \epsilon_{n,t}$, where $\epsilon_{n,t} \sim N(0, 1)$. Furthermore, $y_{n,t} = 1$ if $y_{n,t}^* \geq 0$ and $y_{n,t} = 0$ otherwise. The equation for $y_{n,t}^*$ can be derived as a reduced-form representation of a random utility model with a quadratic utility function, where voting decisions depend on the positions of the bill and the status quo (Clinton, Jackman, and Rivers 2004; Martin and Quinn 2002; Poole and Rosenthal 1997).

Now, let us consider a slightly different representation of the same model. Once again suppose that $y_{n,t} = 1$ if $y_{n,t}^* \geq 0$ and $y_{n,t} = 0$ otherwise but suppose that $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*) \sim f(x; \theta_0)$. The probability of observing the vector $(y_{n,1}, y_{n,2}, \dots, y_{n,T})$ is then given by

$$\Pr(y_t; \theta) = \int_{x \in A_t} f(x; \theta) dx. \quad (1)$$

where $A_t = \{x \in \mathbb{R}^N : x_n \leq 0 \text{ if } y_{n,t} = 1, x_n > 0 \text{ if } y_{n,t} = 0\}$.

We can thus form the maximum likelihood estimator for this model using

$$\hat{\theta} = \operatorname{argmax}_{\theta} \left\{ \frac{1}{T} \sum_{t=1}^T \log \Pr(y_t; \theta) \right\}. \quad (2)$$

⁴Some may argue that it is inappropriate to evaluate a Bayesian estimator based on the property of consistency or based on its ability to recover the “true” parameter value in a Monte Carlo experiment. We will provide a justification for this approach in Section 3.

Notice that θ can be estimated consistently even when N is fixed and T goes to infinity, provided that θ is a finite dimensional parameter. Clearly, $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*) \sim f(x; \theta_0)$ holds for an ideal point model, so estimating α_0 simply involves finding the correct transformation from $\hat{\theta}$ to $\hat{\alpha}$.

In general, the transformation from $\hat{\theta}$ to $\hat{\alpha}$ depends on the distribution of δ_t . Suppose that $\delta_t \sim N(\mu_\delta, \Omega_\delta)$. In this case, we have that $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*) \sim N(\mu; \Omega)$ since a linear transformation of normal random variables has a normal distribution.

We can determine the correct transformation by matching the first two moments of $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*)$ to $(\mu; \Omega)$. For convenience, we will parameterize Ω_δ in terms of its' Cholesky factorization, $\Omega_\delta = L_\delta L_\delta'$, where L_δ is a lower triangular matrix. We can determine that

$$E[y_n^*] = \mu_n(\alpha, \mu_\delta) = (1, \alpha_n)' \mu_\delta, \tag{3a}$$

$$\text{Cov}(y_n^*, y_m^*) = \Omega_{n,m}(\alpha, \mu_\delta, L_\delta) = 1\{n = m\} + \alpha_n' L_\delta L_\delta' \alpha_m. \tag{3b}$$

This step is crucial to our analysis since it allows us to write the likelihood function in a way that does not depend on a large number of nuisance parameters. We can estimate the parameters of this model using maximum likelihood. We choose $(\hat{\alpha}, \hat{\mu}_\delta, \hat{L}_\delta)$ to maximize the objective function,

$$\begin{aligned} Q(\alpha, \mu_\delta, L_\delta) &= \frac{1}{T} \sum_{t=1}^T \log \left\{ \frac{1}{(2\pi)^{K/2} (\det \Omega(\alpha, \mu_\delta, L_\delta))^{1/2}} \right. \\ &\quad \left. \times \int_{A_t} \exp \left\{ -\frac{1}{2} \left(x - \mu(\alpha, \mu_\delta) \right)' \Omega(\alpha, \mu_\delta, L_\delta)^{-1} \left(x - \mu(\alpha, \mu_\delta) \right) \right\} dx \right\}. \end{aligned} \tag{4}$$

We impose restrictions on α for identification purposes (e.g., $\alpha_1 = -1$ and $\alpha_N = 1$, if $D = 1$). The resulting estimator will be consistent even when N is fixed and T goes to infinity.

The only difficulty here is that evaluating the likelihood involves computing rectangle probabilities of the multivariate normal distribution. We evaluate these integrals using the GHK simulator.⁵ The GHK simulator is an importance sampler that is an alternative to a raw frequency sampler for computing rectangles of the normal distribution. It has two advantages over the raw frequency sampler. First, it will not generate probabilities of zero and one in finite samples. Second, it leads to an objective function that varies continuously with the model parameters. We describe the implementation details of this estimator in the Appendix.

The approach we mention here will generate consistent estimates as long as $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*)$ has the multivariate normal distribution. This assumption is met when both δ_t and $\epsilon_{n,t}$ are normally distributed but will hold more generally. Even when $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*)$ are not normally distributed, our approach can be viewed as a first-order approximation based on matching the first two moments of

⁵An alternative approach is to evaluate these integrals using quadrature methods (as Londregan 2001a does). This method suffers from the curse of dimensionality, as we need to compute a $D+1$ -dimensional integral to evaluate the likelihood function. We experimented with quadrature methods in our Monte Carlo experiments and found that even when $D = 1$ the quadrature approach was slower and less accurate. The dominance of the GHK approach would be even more dramatic when estimating larger dimensional models.

$(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*)$ to the first two moments of the normal distribution. The normal distribution can in turn be viewed as a quadratic approximation to the log of the characteristic function.⁶ In principle, we can take a third-order expansion of the log of the characteristic function and match the first three moments of $(y_{n,1}^*, y_{n,2}^*, \dots, y_{n,T}^*)$ to the first three moments of the distribution that has a cubic characteristic function.⁷ Higher order expansions can yield more precise approximations, but common sample sizes will limit the applicability of even a third-order approximation. Given the relative success of our estimator, we chose not to explore these higher order expansions further.

Our approach can easily allow for the distribution of δ_t to depend on some independent variables.⁸ For example, we could have $\delta_t \sim N(\beta_\delta X_t, \Omega_\delta)$ in which case

$$E[y_n^*] = (1, \alpha_n)' \beta_\delta X_t, \quad (5a)$$

$$\text{Cov}(y_n^*, y_m^*) = 1\{n = m\} + \alpha_n' \Omega_\delta \alpha_m. \quad (5b)$$

The likelihood can be formed in a similar way as suggested above. The inclusion of individual-specific covariates is essential in Bailey (2001) and Lewis (2001) but is not essential here. If good covariates are available, then including them may improve efficiency.

As we show later, the Bayesian estimator is particularly vulnerable to the problem of determining the dimensionality of the data. Since the small chamber estimator is a (standard) maximum likelihood estimator, we can rely on the theory of likelihood ratio tests. It is useful, however, to have different measures of model fit, comparable to the percent of votes correctly predicted and the geometric mean probability, that are available for the large chamber estimators (Poole and Rosenthal 1997). Such measures are not directly available because the small chamber estimator does not produce estimates of the bill-specific parameters. To remedy this problem, after the ideal points are estimated, we estimate an individual probit for each vote, producing pseudoestimates of the bill-specific parameters.⁹ Based on these pseudoestimates, we can determine the percentage of votes that are correctly predicted and determine the geometric mean probability. These measures will allow us to investigate the dimensionality of the data.

3 Monte Carlo Evidence

Theory tells us that the small chamber estimator will be consistent when N is held fixed and T goes to infinity, provided that the vector y_t^* is normally distributed. The Bayesian estimator requires that both N and T go to infinity.¹⁰ In this section, we will evaluate the small chamber estimator in comparison to the Bayesian estimator. In particular, we will argue

⁶This partially explains our later finding of robustness to misspecification.

⁷This approach follows the same logic employed when performing an Edgeworth expansion (Hall 1997).

⁸Londregan (2001a, 2001b) includes covariates in his random-effects model. Quinn, Park, and Martin (2007) include covariates in a Bayesian ideal point estimator in order to obtain more precise estimates.

⁹We deal with finite sample identification problems by using penalized maximum likelihood (Firth 1993; Zorn 2005). An alternative approach would be to select the bill-specific parameters to maximize the classification success on each vote, given the estimated ideal points. It is, in fact, possible to solve this problem by formulating it as a mixed-integer programming problem (Liittschwager and Wang 1978), but using individual probits yields classification rates that are more comparable with those produced by conventional ideal point estimators (which do not maximize classification success).

¹⁰More specifically, the Bayesian estimator will be consistent only when both N and T go to infinity (Lewis 2001).

Some may argue that consistency is not a relevant property for evaluating Bayesian estimators, but we ultimately base our conclusions on the finite sample properties of the estimators.

that when the chamber size is small, the Bayesian estimator will perform poorly, particularly with respect to the quality of inferences. Hence, the incidental parameters problem for the Bayesian estimator is more severe than the misspecification problem is for the small chamber estimator.

Although Monte Carlo experiments have not been widely applied to the evaluation of Bayesian estimators in the political science literature, this approach is widely accepted, even among some “hard-core” Bayesians. For example, Geweke, Keane, and Runkle (1994, 1997) compare Bayesian, Simulated Maximum Likelihood, and Simulated Method of Moments estimators for the multinomial probit and panel probit models. Applications in political science include Bailey (2001) and Carroll et al. (forthcoming).

Selecting a distribution for δ_t that deviates from normality allows us to assess the robustness of the small chamber estimator to deviations from idealized assumptions. Our experience with estimated bill-specific parameters in other applications leads us to expect that, in most cases, the distribution will be relatively symmetric but may be unimodal, flat, or bimodal. We consider two choices for the bill-specific parameters. In the “easier” example, we select the bill specific parameters to be uniformly distributed. In the “harder” example, we select the bill-specific parameters to have a skewed distribution and exhibit nonlinear dependence.

We evaluate the estimators using four criteria. We computed the median bias, the root median squared error, the standard error accuracy (overconfidence), and the coverage of a 95% confidence interval. The median bias is the median difference between the estimate and the true parameter value. The root median squared error is the square root of the median squared distance between the estimate and the true parameter value. Overconfidence is the median estimated standard error, divided by the root mean squared error in the sample. The coverage of a 95% confidence interval is the percentage of times the true parameter value falls within the estimated confidence interval.

We choose to use robust measures (the median bias and root median squared error) rather than the more common mean bias and root mean squared error because both estimators occasionally produce estimates that are widely off.¹¹ One or two extreme observations tend to dominate the mean bias and root mean squared error, so these measures may not accurately reflect the typical behavior of these estimators.

In the first experiment, we consider $N = 4$ and vary $T = 201, 501, 1001$. We use the true values $\alpha_0 = (-1, -0.25, 0.5, 1)$. We assume that $\delta_{t,0} = \frac{t-1}{T-1}$ and $\delta_{t,1} \sim \text{Uniform}(-1, 1)$ for $t = 1, \dots, T$. Results are given in Table 1. For low values of T , the Bayesian estimator has lower bias than the small chamber estimator. When T becomes larger, the bias of the small chamber estimator improves, whereas the bias of the Bayesian estimator does not. The Bayesian estimator has lower root median squared error for low values of T , but the small chamber estimator begins to outperform the Bayesian estimator when T becomes larger. Overall, the point estimates produced by these estimators are of comparable quality for typical values of T .

The inferential properties of the Bayesian estimator are quite poor and do not improve when T alone is increased. The Bayesian estimator is very overconfident, although this improves somewhat as T increases. Consistent with this, 95% confidence intervals had coverage around 60% for all the sample sizes we considered. The inferential properties of the small chamber estimator are substantially better. Although the estimator is overconfident when $T = 201$, the standard errors become more accurate as T increases. The

¹¹This is due to the fragile identification of these models when N is small.

Table 1 Monte Carlo results ($D = 1, N = 4$)

<i>T</i>	<i>Median bias</i>		<i>Root median squared error</i>		<i>Overconfidence</i>		<i>Coverage (95%)</i>		<i>Correctly ordered (%)</i>
	<i>Alpha2</i>	<i>Alpha3</i>	<i>Alpha2</i>	<i>Alpha3</i>	<i>Alpha2 (%)</i>	<i>Alpha3 (%)</i>	<i>Alpha2 (%)</i>	<i>Alpha3 (%)</i>	
	Bayesian estimator								
201	-0.009	0.077	0.383	0.337	490	347	63.9	67.0	70.0
501	0.051	0.161	0.274	0.245	244	276	60.1	61.1	89.0
1001	0.105	0.140	0.246	0.197	261	250	53.2	53.7	97.7
Small chamber estimator									
201	-0.006	0.137	0.543	0.464	490	347	73.7	77.8	69.4
501	-0.077	0.008	0.365	0.346	244	152	83.4	82.2	90.0
1001	-0.029	-0.023	0.232	0.240	112	119	90.0	90.1	97.1

Note. Reported results are averages over $R = 1000$ replications.

coverage of a 95% confidence interval is around 75% when $T = 201$ but becomes substantially better as T increases. Both estimators are quite successful in recovering the ordering of the legislators.

Next, consider $\delta_{t,0} = -1 + e_t$ and $\delta_{t,1} = \delta_{t,0}^2(u_t - c_t)$, where e_t , u_t , and c_t are independent draws from the exponential distribution, the standard uniform distribution, and the chi-squared distribution with three degrees of freedom. This presents a higher degree of misspecification for the small chamber estimator. The results are presented in Table 2. We find here that the performance of both estimators deteriorates, but the small chamber estimator clearly dominates here. The small chamber estimator has lower median bias and root median squared error. Although the inferences produced by the small chamber estimator are not spectacular, they are clearly an improvement over the Bayesian estimator. The median bias, standard error accuracy, and coverage for the small chamber estimator do not improve as the sample size increases because the estimator is so heavily misspecified here. Fortunately, both estimators are relatively successful in recovering the ordering of the candidates, though the small chamber estimator is more successful.

Table 2 Monte Carlo results for difficult specification ($D = 1, N = 4$)

<i>T</i>	<i>Median bias</i>		<i>Root median squared error</i>		<i>Overconfidence</i>		<i>Coverage (95%)</i>		<i>Correctly ordered (%)</i>
	<i>Alpha2</i>	<i>Alpha3</i>	<i>Alpha2</i>	<i>Alpha3</i>	<i>Alpha2 (%)</i>	<i>Alpha3 (%)</i>	<i>Alpha2 (%)</i>	<i>Alpha3 (%)</i>	
	Bayesian estimator								
201	-0.099	0.351	0.145	0.351	163	286	79.3	26.5	83.5
501	-0.112	0.351	0.121	0.351	191	419	67.8	3.2	94.9
1001	-0.114	0.355	0.114	0.355	221	611	55.6	0.0	98.5
Small chamber estimator									
201	-0.142	0.127	0.175	0.189	124	115	85.3	91.8	93.4
501	-0.122	0.133	0.128	0.144	138	131	82.3	86.2	99.1
1001	-0.126	0.131	0.127	0.134	175	157	69.6	76.4	100.0

Note. Reported results are averages over $R = 1000$ replications.

Table 3 Monte Carlo results ($D = 1, N = 9$)

T	<i>Median bias (average absolute value)</i>	<i>Root median squared error (average)</i>	<i>Overconfidence (average) (%)</i>	<i>Coverage (95%) (average) (%)</i>
Bayesian estimator				
201	0.016	0.310	230.3	81.0
501	0.018	0.196	160.0	79.7
Small chamber estimator				
201	0.021	0.327	175.1	93.1
501	0.016	0.186	100.5	96.1

Note. Reported results are averages over $R=1000$ replications.

The results for the highly misspecified model may initially be surprising because the validity of the Bayesian estimator does not rely on normality of the bill-specific parameters. However, a skewed distribution for the bill-specific parameters serves to exacerbate the incidental parameters problem. The prior distributions compensate for the fact that a minimal amount of data is available to estimate each bill-specific parameters, but when the distribution of the bill-specific parameters deviates from normality, normal prior distributions provide misleading information.

We next consider a similar experiment with $N = 9$. In this case, we set $\alpha_0 = (-1, -0.75, -0.5, -0.25, 0.0, 0.25, 0.5, 0.75, 1)$, using $\delta_{t,0} = \frac{t-1}{T-1}$ and $\delta_{t,1} \sim \text{Uniform}(-1, 1)$ for $t = 1, \dots, T$. The simulation results are presented in Table 3. Both estimators perform quite well in terms of bias and result in comparable root median squared error. The point estimates produced by either estimator are therefore of comparable quality. The small chamber estimator once again leads to improved inferences. The Bayesian estimator is overconfident and confidence intervals undercover. The small chamber estimator performs better here, with accurate standard errors and near-perfect coverage when $T = 501$.

Finally, we consider an experiment where $N = 9$ and $D = 2$. We set

$$\alpha_0' = \begin{bmatrix} -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}. \tag{6}$$

and we consider $T = 201, 501$. We use $\delta_{t,0} = \frac{t-1}{T-1}$, $\delta_{t,1} \sim \text{Uniform}(-1, 1)$, and $\delta_{t,2} \sim \text{Uniform}(-1, 1)$ for $t = 1, \dots, T$. These results are presented in Table 4. Once again, the estimators are comparable in the quality of their point estimates. The small chamber estimator has superior inferential properties. Although the Bayesian estimator is overconfident and undercovers, the small chamber estimator provides quality standard errors and near-perfect coverage.

Table 4 Monte Carlo results ($D = 2, N = 9$)

T	<i>Median bias (average absolute value)</i>	<i>Root median squared error (average)</i>	<i>Overconfidence (average) (%)</i>	<i>Coverage (95%) (average) (%)</i>
Bayesian estimator				
201	0.046	0.327	245.4	85.8
501	0.029	0.209	142.1	84
Small chamber estimator				
201	0.038	0.380	117.2	96
501	0.026	0.232	94.4	95.7

Note. Reported results are averages over $R = 1000$ replications.

Our results indicate that we should, indeed, worry about the incidental parameters problem. The point estimates of the Bayesian estimator are surprisingly good even in relatively small chambers, provided that the information provided by the prior distributions of the bill-specific parameters is not too misleading. The incidental parameters problem affects the Bayesian estimator most dramatically in the quality of inferences. We found that the problems were severe when $N = 4$, while the performance of Bayesian inferences was less than stellar when $N = 9$. These problems do not go away when T alone is increased.

The computational burden of the small chamber estimator is lower in small sample sizes. The small chamber estimator does not scale as well as the Bayesian estimator (or the maximum likelihood estimator), however. Moreover, although the small chamber estimator is superior at small chamber sizes, eventually the Bayesian estimator will achieve better performance. This suggests a cutoff from switching between the small chamber and Bayesian estimators. We think that somewhere between $N = 15$ and $N = 30$ serves as a reasonable cutoff for abandoning the small chamber estimator in favor of alternative methods. The cutoff depends on how “wild” the distribution of cutpoints is. For very wild distributions, our results suggest a higher cutoff before switching to large chamber estimators.

4 Voting in the U.S. Supreme Court

In this section, we will apply our estimator to Spaeth’s (1999) Supreme Court Database and compare our estimator to the Bayesian estimator. We consider a data set including nine justices and 344 votes from the last 5 years of the final Rehnquist Court (2001–05). We estimate models both with and without covariates. Following the suggestion of Quinn, Park, and Martin (2007), we create dummy variables by dividing the country into 13 regions. A court is assigned to region X if the case originated from the X th circuit court or from a state court within a state served by the X th circuit. We also create dummy variables for the DC circuit court and federal courts.

The estimates of the one-dimensional model are reported in Table 5. We present results for the small chamber estimator, the small chamber estimator with covariates, and the Bayesian estimator.¹² The Bayesian estimator and both small chamber estimators agree on the ordering of the justices, after accounting for uncertainty. The estimates generally conform to conventional wisdom about the ordering of the justices as well as estimates of Martin and Quinn (2002), which were obtained for a different time period. In particular, O’Connor is viewed as the median member of the last Rehnquist Court, with Kennedy having a somewhat more conservative voting record.

Our Monte Carlo results indicated that there was no strong reason to favor the point estimates of either estimator at these sample sizes. Our Monte Carlo results, however, suggested that the standard errors produced by the small chamber estimator are generally accurate and the standard errors produced by the Bayesian estimator are generally overconfident. Indeed, we find that the standard errors of the small chamber estimator are substantially larger, suggesting that the Bayesian estimator is not adequately accounting for uncertainty.

For the small chamber estimator with covariates, we found that about half of the covariates were statistically significant at the 5% level. This indicates that there are differences in the mean bill-specific parameters across jurisdictions. Consistent with, Quinn, Park, and Martin (2007), the inclusion of covariates allows for more precise estimates.

¹²Since the Bayesian estimator is a fixed-effects estimator, the inclusion of covariates would have only a limited effect on the ideal point estimates. Hence, we did not consider the Bayesian estimator with covariates here.

Table 5 One-dimensional ideal point estimates for the Supreme Court

<i>Justice</i>	<i>Small chamber estimator</i>		<i>Small chamber estimator (with covariates)</i>		<i>Bayesian estimator</i>	
	<i>Alpha (SE)</i>	<i>Rank</i>	<i>Alpha (SE)</i>	<i>Rank</i>	<i>Alpha (SE)</i>	<i>Rank</i>
Rehnquist	0.341 (0.293)	7	0.434 (0.207)	7	0.677 (0.072)	7
Stevens	-1.000 (0.000)	1	-1.000 (0.000)	1	-1.000 (0.000)	1
O'Connor	-0.120 (0.321)	5	0.007 (0.226)	5	-0.270 (0.060)	5
Scalia	0.876 (0.341)	8	0.870 (0.224)	8	0.977 (0.067)	8
Kennedy	0.256 (0.292)	6	0.344 (0.206)	6	0.524 (0.067)	6
Souter	-0.487 (0.359)	4	-0.311 (0.243)	4	-0.785 (0.072)	4
Thomas	1.000 (0.000)	9	1.000 (0.000)	9	1.000 (0.000)	9
Ginsburg	-0.560 (0.386)	2	-0.381 (0.252)	2	-0.822 (0.070)	3
Breyer	-0.539 (0.409)	3	-0.365 (0.269)	3	-0.826 (0.071)	2

Note. Justices in bold were constrained for identification purposes.

Poole (2005) suggests that Justice Stevens' ideal point is hard to pin down because he makes so few voting errors. To further assess the robustness of our estimator, we reestimated a one-dimensional model on all the justices excluding Stevens. The ideal point estimates were nearly identical and result in a correlation of 99.4%. As a general rule, we found that both ideal points and estimated standard errors were robust to the deletion of individual voters, provided that the normalization remained constant.

We next estimated a two-dimensional model. The results for the two-dimensional model are given in Figs 1 and 2. We normalized Stevens at $(-1, 1)$, Breyer at $(-1, -1)$, and Scalia at $(1, 0)$. The normalization was chosen such that, for the small chamber estimator, the modal cutting line was perpendicular to the first dimension.¹³ Thus, α_1 denotes the primary dimension of conflict and α_2 denotes the secondary dimension of conflict. The two estimators agree over the relative placement of all justices except Stevens and Breyer. Stevens and Breyer are placed at more extreme positions relative to the other justices by the small chamber estimator. Following Poole's (2005) argument, however, the positions of these justices are the most difficult to pin down. In the two-dimensional case, justice Stevens makes only two voting errors and Justice Breyer makes no voting errors. The relative positions of the remaining justices are common across both estimators.

We can clearly identify the first dimension of conflict as a liberal-conservative dimension based on the ordering the justices along this dimension. Identifying the remaining dimension is a difficult task, but there is some evidence to suggest that the second dimension captures "judicial activism."¹⁴ We use a variable from the Supreme Court Database that indicates whether the majority opinion declares a federal or state law unconstitutional. In Fig. 3, we plot the distribution of cutting line angles for cases that featured no declaration of unconstitutionality. We can see that this distribution has a mode near zero degrees. In Fig. 4, we plot the distribution of cutting line angles for cases that featured a declaration of

¹³Cutting line angles are computed using the formula, $\theta_i = \frac{180}{\pi} \tan^{-1}(\delta_{i,1}/\delta_{i,2}) + 90$. An angle of 0 degrees indicates a vote that divides the justices along the first dimension. An angle of 90 degrees indicates a vote that divides the justices along the second dimension.

¹⁴Judicial activism is a loaded word with a strong negative connotation, but we use it here to signify an increased willingness to overturn laws that may conflict with constitutional provisions or a decreased deference to legislative bodies.

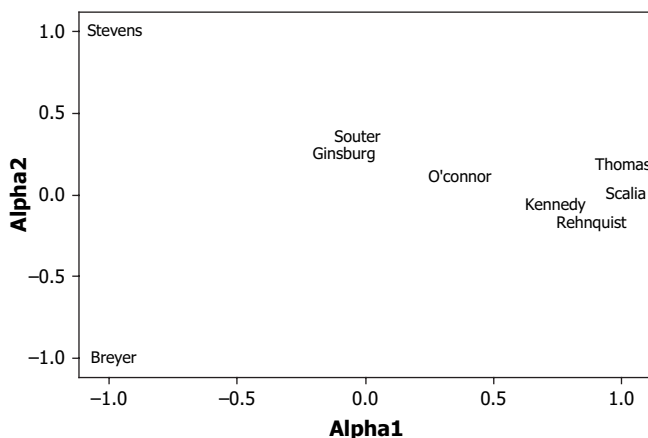


Fig. 1 Two-dimensional small chamber ideal point estimates for the Supreme Court.

unconstitutionality. This distribution has a mode near 90 degrees. Our results therefore suggest that the second dimension may relate to deference to legislative bodies. Some caution is warranted, however, because although the patterns in Figs 3 and 4 are strong, few cases featured a declaration of unconstitutionality and, hence, the patterns we observe may not be representative.

To assess the dimensionality of voting in the Supreme Court, we estimate a number of higher dimensional models. To assess the relative fit of each model, we consider measures based on the objective function (the log-likelihood), the percent of correct predictions, the geometric mean probability, and the Bayesian information criterion. For the small chamber estimator, we also consider likelihood ratio tests. We estimate the models with between 1 and 4 dimensions. These results are reported in Table 6. Consider first the small chamber estimator. The objective function levels off after the second dimension. The percent correctly predicted and the geometric mean probability level off after the third dimension. The likelihood ratio test suggests at least four dimensions, whereas the Bayesian information criterion selects two dimensions. Although the different methods do not agree on dimensionality, they agree that more than one dimension of conflict is necessary to explain

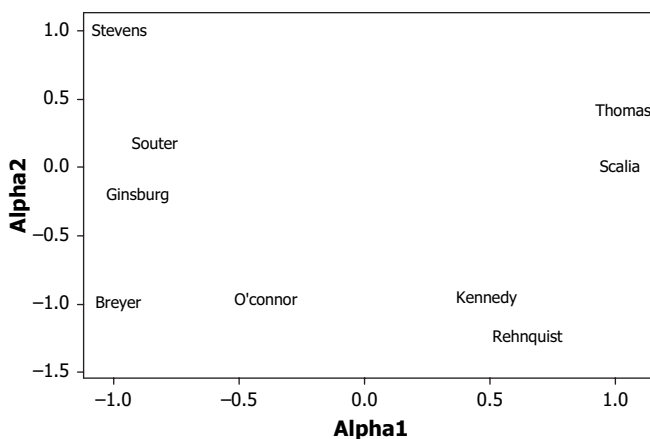


Fig. 2 Two-dimensional Bayesian ideal point estimates for the Supreme Court.

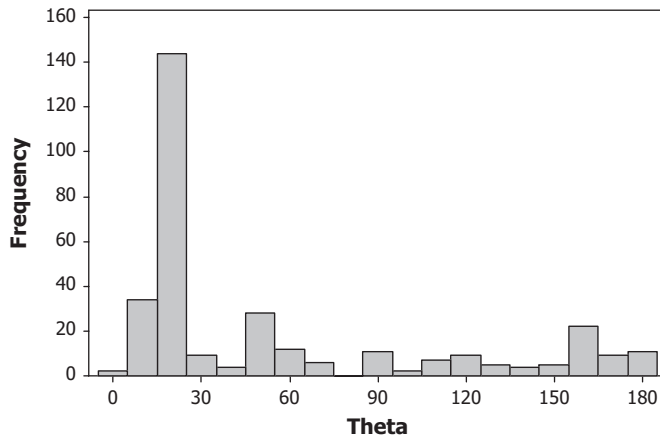


Fig. 3 Two-dimensional cutting line angles for Supreme Court cases with no declaration of unconstitutionality.

voting in the Supreme Court. This is consistent with our finding that the second dimension is important in explaining voting behavior in cases where a law conflicts with a state or federal constitution.

Assessing dimensionality is particularly problematic when employing the Bayesian estimator. None of the measures of model fit are monotonic in the dimensionality of the model and the Bayesian information criterion exhibits at least two local minima. The Bayesian estimator deals with finite sample identification problems via the prior. The prior essentially penalizes parameter values that deviate from 0 (the prior mean). This penalty is particularly strong when the chamber size is small because the data will not overwhelm the prior. Leaving the prior variance fixed and increasing the estimated dimension will lead the prior to be more constraining on the ideal point estimates, often leading to a poorer fit.¹⁵ Here, we again see

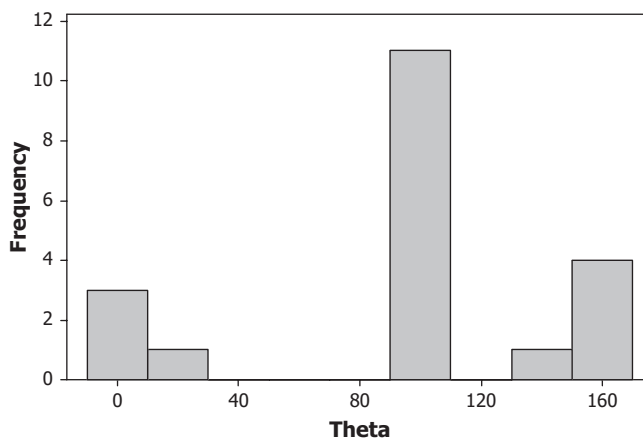


Fig. 4 Two-dimensional cutting line angles for Supreme Court cases with a declaration of unconstitutionality.

¹⁵We reiterate that this result is specific to small chamber ideal point estimation. The Bayesian estimator will *not* suffer from this problem in large chambers because the data will dominate the prior.

Table 6 Model fit statistics for the small chamber and Bayesian estimators

<i>D</i>	<i>K</i>	<i>Correctly predicted (%)</i>	<i>Geometric mean probability (%)</i>	<i>Log-likelihood</i>	<i>Bayesian information criterion</i>	<i>Likelihood ratio statistics</i>	<i>Degrees of freedom</i>	<i>p Value</i>
Small chamber estimator (no covariates)								
1	12	90.9	81.8	-1185.9	2441.9			
2	21	93	83.7	-1131.5	2385.6	108.8	9	0
3	29	97.2	92.5	-1109.7	2388.8	43.5	8	0
4	36	99.2	95.5	-1101.2	2412.7	17	7	0.017
Bayesian estimator								
1	12	89.9	85.4	-71.5	213.1			
2	21	91.9	88.1	-56.6	235.8	29.8	9	0
3	29	82.3	79.8	-613	1395.4	-1112.9	8	1
4	36	95.5	91.6	-38.1	286.4	1149.9	7	0

the consequences of the incidental parameters problem for the Bayesian estimator in small chambers. The small chamber estimator provides a solution to this problem, and we are able to obtain useful information about the underlying dimensionality of the data.¹⁶

5 Discussion

In this paper, we have developed an estimator for ideal point models that is consistent when N is fixed and T goes to infinity. This type of estimator is uniquely suited for estimating ideal points in small chambers. The Bayesian estimator suffers from the incidental parameters problem and is not consistent if N is held fixed. In practice, the Bayesian estimator produces surprisingly good point estimates under reasonably favorable conditions, but the quality of the inferences is quite poor if N is small. The drawbacks of the Bayesian estimator are more apparent in higher dimensional models.

Our small chamber estimator has reasonably good properties even when N is quite small and improves substantially as T increases. Our estimator outperformed the Bayesian estimator in our Monte Carlo experiments and is quite robust to misspecification. In fact, in harder applications, the benefit of the small chamber estimator was even larger. Finally, we applied our estimator to decision making in the U.S. Supreme Court. The application showed that the estimator performs well.

Our results should not be read as a general statement against W-NOMINATE or the Bayesian estimator. We simply argue that these estimators are not appropriate for small chambers. Nor should our results be read as a general criticism of Bayesian estimation. Rather, our results compare one random-effects estimator (our small chamber estimator) to one fixed-effects estimator (the Bayesian estimator). Just as a Bayesian analog of the W-NOMINATE estimator by Poole and Rosenthal (1997) has been developed, one could develop a Bayesian analog of our small chamber estimator. Moreover, this paper experimented with one possible

¹⁶We note that exactly how to evaluate model fit remains controversial and one may still argue for the one-dimensional model on the basis of parsimony. Our point is simply that the conventional measures are not useful for assessing dimensionality when the Bayesian estimator is applied to small chambers.

implementation of a random-effects estimator for small chamber ideal point estimation. We encourage future work to experiment with alternative implementations (including Bayesian implementations) of random-effects estimators applied to small chambers.

Appendix: Computational Details

In this section, we detail the implementation of the Bayesian and small chamber estimators used in this paper.

Bayesian Estimator

Our implementation of the Bayesian estimator is quite standard, closely following the approach of Martin and Quinn (2002) and Clinton, Jackman, and Rivers (2004). We use independent diffuse normal priors for all models parameters with mean zero and variance 1000. In the Monte Carlo experiments, we used 3000 burn-in iterations and 3000 Gibbs iterations, and we do not thin the output. In the results reported in section 4, we used 10,000 burn-in iterations and 20,000 Gibbs iterations.

We only differ from Clinton, Jackman, and Rivers in how we deal with identification. After our Markov chain has run, we transform each draw so that it conforms with the parameter constraints we are imposing. This corresponds to the postprocessing option Simon Jackman recommends when using the “ideal” software package. We then use the posterior mean to compute estimates of the ideal points¹⁷ and we use the 2.5% and 97.5% quantiles to form 95% confidence intervals. We tested our code on a number of data sets, finding that the results correlated with previous results at extremely high levels (thus indicating that our code is functioning correctly).

To provide one final test that our code is working properly, we preformed a Monte Carlo simulation with $N = 100$ legislators and $T = 500$ bills. The large sample properties of the estimator suggest that under these conditions, the estimator should have (1) negligible bias, (2) no under or overconfidence, and (3) near-perfect coverage. We found that this was, indeed, the case, indicating that we have correctly implemented the estimator and that 3000 burn-in and Gibbs iterations (which might otherwise be considered stingy) are sufficient to properly explore the posterior distribution.

Small Chamber Estimator

To implement the small chamber estimator, we need to be able to compute rectangles of the multivariate normal distribution. We can compute these integrals through simulation using the GHK method (Geweke, Keane, and Runkle 1994). This method has already been successfully applied to estimate the multinomial probit model and related models (Peress 2007). We use $S = 10$ simulations in our Monte Carlo simulations and $S = 50$ simulations in our application to the U.S. Supreme Court.¹⁸

To provide one final test that our code is working properly, we preformed a Monte Carlo simulation with $N = 4$ legislators, $T = 2000$ votes, and normally distributed bill-specific parameters. The large sample properties of the estimator suggest that under these conditions, the estimator should have (1) negligible bias, (2) no under- or overconfidence, and (3)

¹⁷We found that our results were not sensitive to the choice of the posterior mean over the posterior median as our point estimator.

¹⁸ $S = 10$ has been found to be sufficient in other work (Geweke, Keane, and Runkle 1994).

near-perfect coverage. We found that this was, indeed, the case indicating that we have correctly implemented the estimator and that $S = 10$ simulations are sufficient to accurately compute the likelihood function.

References

- Andersen, Erling Bernhard. 1970. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society* 32:283–301.
- . 1973. Conditional inference and models for measuring. Copenhagen, Denmark: Mentalhygiejnisk Forsknings Institut.
- Bafumi, James, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. Practical issues in implementing and understanding Bayesian ideal point estimation. *Political Analysis* 13:171–87.
- Bailey, Michael A. 2001. Ideal point estimation with a small number of votes: A random effects approach. *Political Analysis* 9:192–210.
- Bock, R. Darell, and Murray Aitken. 1981. Marginal maximum likelihood estimation of item parameters: Application of the EM algorithm. *Psychometrika* 46:443–59.
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. Forthcoming. Comparing NOMINATE and IDEAL: Points of difference and Monte Carlo tests. *Legislative Studies Quarterly*.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2003. The statistical analysis of roll call data. *Working Paper*.
- . 2004. The statistical analysis of roll call data. *American Political Science Review* 98:355–70.
- Desposato, Scott. 2004. The impact of party-switching on legislative behavior in Brazil. *Working Paper*.
- Firth, David. 1993. Bias reduction of maximum likelihood estimates. *Biometrika* 80:27–38.
- Geweke, John, Michael Keane, and David Runkle. 1994. Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics* 76:609–32.
- . 1997. Statistical inference in the multiperiod multinomial probit model. *Journal of Econometrics* 81:125–66.
- Hall, Peter. 1997. *The bootstrap and the Edgeworth expansion*. New York: Springer.
- Heckman, James J., and James M. Snyder. 1997. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics* 28:S142–89.
- Hsiao, Cheng. 1986. *Analysis of panel data*. Cambridge: Cambridge University Press.
- Kiefer, J., and J. Wolfowitz. 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics* 27:887–906.
- Lancaster, Tony. 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95:391–413.
- Lewis, Jeffrey B. 2001. Estimating voter preference distributions from individual level voting data. *Political Analysis* 9:275–97.
- Liittschwager, J. M., and C. Wang. 1978. Integer programming solution of a classification problem. *Management Science* 24:1515–25.
- Londregan, John. 2000a. Estimating legislator's preferred points. *Political Analysis* 8:35–56.
- . 2000b. *Legislative institutions and ideology in Chile's democratic transition*. Cambridge: Cambridge University Press.
- Martin, Andrew D., and Kevin M. Quinn. 2002. Dynamic ideal point estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* 10:134–53.
- Neyman, J., and Elizabeth L. Scott. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16:1–32.
- Peress, Michael. 2007. Securing the base: Electoral competition under variable turnout. *Working Paper*.
- Poole, Keith T. 2005. *Spatial models of parliamentary voting*. New York: Cambridge University Press.
- Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A political economic history of roll call voting*. New York: Oxford University Press.
- Quinn, Kevin M., Jong Hee Park, and Andrew D. Martin. 2007. Improving judicial ideal point estimates with a more realistic model of opinion content. *Working Paper*.
- Spaeth, Harold J. 1999. United State Supreme Court judicial database, 1953–1998 terms [Computer File]. 15th ed. Ann Arbor, MI: Inter University Consortium of Political and Social Research.
- Wright, Gerald C., and Tracy Osborne. 2002. Party and roll call voting in the American legislature. *Working Paper*.
- Zorn, Christopher. 2005. A solution to separation in binary response models. *Political Analysis* 13:157–70.