Improving Judicial Ideal Point Estimates with a More Realistic Model of Opinion Content*

Kevin M. Quinn
Department of Government
Harvard University
kevin_quinn@harvard.edu

Jong Hee Park
Department of Political Science
Washington University
jhpark@artsci.wustl.edu

Andrew D. Martin
Washington University School of Law
Washington University
admartin@wustl.edu

July 5, 2006

Abstract

We extend the dynamic item response theory model of Martin and Quinn (2002) by including information about the court of origin and the opinion writer using a dynamic item response theory (IRT) model with a hierarchical prior. The model is estimated using Markov chain Monte Carlo methods. This approach not only provides better estimates of quantities of interest, such as judicial ideal points, it also allows us to estimate both the status quo and policy alternative locations for each case, as well as the location of each circuit court. This application shows the value of incorporating institutional detail and other available information into statistical measurement models.

*The authors are listed in reverse alphabetical order as determined by lot. We gratefully acknowledge financial support from the National Science Foundation (Grants SES 03-50646, SES 03-50613, SES 01-35855, and SES 01-36679). We also thank Steve Van Winkle for help forming reasonable circuit court priors. Data for replication, posterior density samples, `R` code to replicate the figures, and documented `C++` code to fit the model will be made available at the authors' websites.

1 Introduction

It would not be an overstatement to claim that ideal point estimation has revolutionized the study of American political institutions. The most salient example are the (DW)-NOMINATE models developed by Poole and Rosenthal (1997). Not only do these models statistically operationalize the spatial model (Downs, 1957; Davis et al., 1970; Enelow and Hinich, 1984), they provide measures that other researchers can use to address a wide range of substantive questions. In a similar vein, Martin and Quinn (2002) estimate ideal points for U.S. Supreme Court justices, which are now frequently used in empirical applications.

The precise nature of these models varies from application to application. Poole and Rosenthal (1991) use Gaussian utility functions while Clinton et al. (2004) use quadratic utility functions. Estimation strategies vary with some researchers using classically-based optimization strategies (Poole and Rosenthal, 1997), while others use a fully Bayesian approach (Clinton et al., 2004). Dynamics are dealt with in a variety of fashions. Poole and Rosenthal (1991) assume linear trajectories through time, while Martin and Quinn (2002) use a random walk prior to model dynamics. Despite these differences, all of these models share a common assumption: that conditional on the ideal points the only source of information about the case (or bill) parameters for case (or bill) k comes from the observed votes on case (or bill) k. At a minimum, this will produce less efficient estimates than an estimation strategy that correctly makes use of background information about the nature of the case (or bill) parameters. Further, in some situations it is possible that this omission of background knowledge could bias estimates of both the case (or bill) parameters as well as the ideal points (Clinton and Meirowitz, 2001).

In this paper we focus our attention on the U.S. Supreme Court and extend the Martin and Quinn (2002) dynamic item response theory model by incorporating information about the circuit of origin and the opinion writer. Unlike Martin and Quinn (2002) we specify hierarchical prior distributions for the case parameters that reflect background knowledge of the judicial process. Not only does this model better reflect the underlying data generating process, it also provides more efficient estimates and allows for the direct estimation of other quantities of interest including the location of the status quo and policy alternative and the location of each Court of Appeals in the same ideological space.

In the following section we discuss alternative approaches to modeling the agenda within the ideal point estimation framework. Section 3 describes our data. In Section 4 we posit the dynamic IRT model with a hierarchical prior, highlighting the substantive nature of various technical assumptions. Section 5 describes the Markov chain Monte Carlo algorithm used to fit the model. We summarize our results in Section 6. The final section concludes.

2 Modeling the Judicial Agenda

In the legislative context, auxiliary information about the agenda process has been incorporated into ideal point models by Clinton and Meirowitz (2001, 2003, 2004), and Londregan (2000). Clinton and Meirowitz (2001, 2003) demonstrate that one can mis-estimate quantities of interest, including ideal points, by not including information about the legislative agenda. They note that "...the agenda should constrain nay location estimates" (2001, p. 243), thus putting more structure on the problem. Clinton and Meirowitz (2004) adopt a Bayesian approach and apply their agenda-constrained model to the "Compromise of 1790." Using a detailed case study of the legislative record, they are able to identify the location of the status quo for each roll call related to the compromise, thus isolating two salient issues: the location of the capitol and debt assumption. What makes this model work is the idea that an outcome of a previous roll call sets the status quo point for a subsequent roll call on the agenda tree. They do not constrain the alternative policy position. One substantive advantage of this approach is the ability to recover both the location of the status quo and policy alternative in an ideological space.

While the work of these authors highlights the importance of using background information when available, the direct application of their methods to Supreme Court decisions on the merits is problematic. Because many of the cases coming before the Court feature novel legal questions, it is not at all clear that past U.S. Supreme Court decisions generally determine the status quo policy position. Instead, it seems much more sensible to treat the status quo policy position as determined by the decision coming out of the lower court in question. If the U.S. Supreme Court decides to uphold the lower court ruling this ruling becomes national policy and should clearly be treated as one of the policy options before the justices. Further, unlike in the U.S. Congress where the primary author of legislation is usually unknown, in the context of the U.S. Supreme Court there is typically information about the identity of the opinion writer. Maltzman et al. (2000) discuss the strategic uses of opinion assignment and the strategies involved in opinion writing. This work suggests that opinion writers will attempt to write opinions consistent with their policy preferences subject to majority approval. In Section 4 we detail the precise manner in which we incorporate these two types of information to locate the status quo and alternative using a hierarchical prior.

What is the potential pay-off of adopting this modeling strategy? First, this model better represents the data generating process. By including auxiliary information through a hierarchical prior we obtain more efficient estimates of the ideal points and case parameters. Further, models that do not adjust in some way for agenda effects will potentially mis-estimate quantities of interest (Clinton and Meirowitz, 2001). Second, the model provides estimates of other things of interest besides just ideal points and cut-points. For each case we estimate the status quo and policy alternative. In addition, for each term we estimate the location of each circuit court in the same ideological space. Before laying out the structure of the model, we turn to the data.

3 The Data

We obtain our data from the *United States Supreme Court Judicial Database* (Spaeth, 2004). We select all cases decided from the 1953 term through the 1999 term.[1] A total of $J = 29$ justices served in these $T = 47$ terms. On a particular case, no more than nine justices cast a vote. There are a grand total of $K = 3450$ cases decided. The observed data matrix $\mathbf{Y}$ is a $(K \times J)$ matrix of votes and missing values. We code all votes as either being in favor of reversing ($y_{k,j} = 1$) or affirming ($y_{k,j} = 0$) the decision of a lower court. Overall, 63.4% of cases resulted in the Court reversing the decision of a lower court.

For each case we also cull covariates from the *United States Supreme Court Judicial Database*. To identify the court of origin for each case, we use the SOURCE variable from the database to identify the lower court. The Supreme Court not only hears cases from the federal Courts of Appeals, but also from state courts of last resort (in most states these are called Supreme Courts), as well as from other miscellaneous federal trial and appellate courts. We use the federal circuits to define a geographic region of origin, and classify every case that comes from that region—whether from the Court of Appeals, a lower federal court, or a state court—as coming from the the same region. This is a reasonable assumption: cases from lower courts in Texas or the Texas Supreme Court or Texas Court of Criminal Appeals are more likely to look like those from the Appeals Court of the Fifth Circuit than any other circuit. This is *not* to say that these cases are identical. Rather, as discussed in the following section, this implies that they come from a common distribution which might have a different mean across circuits. We create thirteen dichotomous variables: one for each of the eleven circuits (note that the Eleventh Circuit was established in October, 1981

[1]We use the case citation as the unity of analysis (ANALU=0). We include cases when the decision type (DEC_TYPE) is 1 (orally argued cases with signed opinions), 5 (cases with an equally divided vote), 6 (orally argued *per curiam* cases), or 7 (judgments of the Court). We exclude all unanimous cases from the analysis, which does not affect model estimates. This is the same selection as Martin and Quinn (2002).

when the Fifth Circuit was split). The twelfth variable represents the DC circuit. Other miscellaneous federal courts, cases that arose under original jurisdiction, etc. comprise the final category. These covariates are loaded into a matrix $\mathbf{C}_t$ indicating the originating lower court for each case in term t. Finally, for each case, we code the author of the majority opinion writer using the MOW variable when it is available. For some cases, such as *per curiam* decisions, the author is not identified.

4 A Dynamic IRT Model with a Hierarchical Prior

Our statistical model is based on a theory of spatial voting along the lines of Clinton et al. (2004). We incorporate information of the court of origin and the opinion author into the model using informative, hierarchical priors. To model temporal dependence, we employ prior distributions that result in dynamic linear models (DLMs) for the ideal points and case parameters (Martin and Quinn, 2002).

Let $K_t \subset \{1, 2, \ldots, K\}$ denote the set of cases heard in term t. Similarly, let $J_k \subset \{1, 2, \ldots, J\}$ denote the set of justices who heard case k. We are interested in modeling the decisions made in terms $t = 1, \ldots, T$ on cases $k \in K_t$ by justices $j \in J_k$ in a uni-dimensional issue space.[2] Our assumption is that each justice's vote is an expressive action and depends only on the value the justice attaches to the policy positions of the status quo and the policy alternative. Put another way, a justice will vote to affirm the decision of the lower court if the utility the justice attaches to the alternative is greater than the utility the justice attaches to the status quo, regardless of the expected actions of the other actors.[3]

To operationalize this model, we begin by writing down random utility functions. Let $u^{(a)}_{t,k,j}$ be the utility to justice $j \in J_k$ of voting to affirm on case $k \in K_t$, and $u^{(r)}_{t,k,j}$ be the

[2]We derive the model for a uni-dimensional issue space. The extension to multiple issue dimensions is straightforward. For examples of multi-dimensional models, see Jackman (2001).

[3]One implication of this assumption is that sophisticated voting is not accounted for in the model.

utility to justice $j \in J_k$ of voting to reverse on case $k \in K_t$. $\theta_{t,j} \in \mathbb{R}$ is justice j's ideal point in term t. $x^{(a)}_{t,k} \in \mathbb{R}$ is the location of the policy under an affirmance vote, $x^{(r)}_{t,k} \in \mathbb{R}$ is the location of the policy under a reversal, and $\xi^{(a)}_{t,k,j}$ and $\xi^{(r)}_{t,k,j}$ are independent Gaussian disturbances with zero mean and variances τ_a^2 and τ_r^2 respectively.

Given this spatial model, justice j will vote to reverse on case k when $u^{(r)}_{t,k,j} > u^{(a)}_{t,k,j}$ or equivalently when $u^{(r)}_{t,k,j} - u^{(a)}_{t,k,j} > 0$. Let $z_{t,k,j}$ be the difference between $u^{(r)}_{t,k,j}$ and $u^{(a)}_{t,k,j}$. We can write and simplify this utility difference $z_{t,k,j}$ as follows:[4]

$$
\begin{aligned}
z_{t,k,j} &= u^{(r)}_{t,k,j} - u^{(a)}_{t,k,j} \\
&= -\|\theta_{t,j} - x^{(r)}_{t,k}\|^2 + \xi^{(r)}_{t,k,j} + \|\theta_{t,j} - x^{(a)}_{t,k}\|^2 - \xi^{(a)}_{t,k,j} \\
&= \left[x^{(a)}_{t,k}x^{(a)}_{t,k} - x^{(r)}_{t,k}x^{(r)}_{t,k}\right] + 2\theta'_{t,j}\left[x^{(r)}_{t,k} - x^{(a)}_{t,k}\right] + \epsilon_{t,k,j}
\end{aligned} \tag{1}
$$

where $\epsilon_{t,k,j} = \xi^{(r)}_{t,k,j} - \xi^{(a)}_{t,k,j}$ and $\epsilon_{t,k,j} \overset{iid}{\sim} \mathcal{N}(0, \tau_a^2 + \tau_r^2)$. For reasons of identification we assume $\tau_a^2 + \tau_r^2 = 1$. Thus, for observed votes $y_{t,k,j}$ and the above-defined utility difference $z_{t,k,j}$,

$$
y_{t,k,j} = \begin{cases} 1 & \text{if} \quad z_{t,k,j} > 0 \\ 0 & \text{if} \quad z_{t,k,j} \leq 0 \end{cases} \tag{2}
$$

We can write the sampling density for this model as:

$$
f(\mathbf{Y}|\mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \theta) = \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \Phi(\mu_{t,k,j})^{y_{t,k,j}} \left[1 - \Phi(\mu_{t,k,j})\right]^{1-y_{t,k,j}} \tag{3}
$$

where $\mu_{t,k,j} \equiv \left[x^{(a)}_{t,k}x^{(a)}_{t,k} - x^{(r)}_{t,k}x^{(r)}_{t,k}\right] + 2\theta'_{t,j}\left[x^{(r)}_{t,k} - x^{(a)}_{t,k}\right]$. An equivalent expression for the sampling density that will be of some use later in characterizing the posterior distribution and estimating the model is:

$$
\begin{aligned}
f(\mathbf{Y}|\mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \theta) = \int \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \big[&\mathbb{I}(z_{t,k,j} > 0)\mathbb{I}(y_{t,k,j} = 1) + \\
&\mathbb{I}(z_{t,k,j} \leq 0)\mathbb{I}(y_{t,k,j} = 0)\big] f_{\mathcal{N}}(z_{t,k,j}|\mu_{t,k,j}, 1) d\mathbf{Z}
\end{aligned} \tag{4}
$$

[4]Clinton et al. (2004) derive the same expression in the context of a model of legislative voting and go on to show the link with standard item response theory models.

where $\mathbb{I}$ is the indicator function, $\mu_{t,k,j}$ is as before, and $f_{\mathcal{N}}(a|b,c)$ is a normal density with mean b and variance c evaluated at a.

In what follows we use "dot" notation to express subsets of the data and parameter arrays. For instance, we use the notation $\mathbf{y}_{t,k,\cdot}$ to denote the vector of responses from all of the justices who, in term t, heard case k. Lack of subscripts entirely indicates we are dealing with the full data or parameter array; e.g., $\mathbf{Y}$ denotes all of the observed responses over all time periods, cases, and justices.

4.1 The Informative Prior on Case Parameters

At this point it is common to note the isomorphism to a two-parameter item response theory model and to parameterize in terms of a difficulty parameters $\alpha_k = [x_{t,k}^{(r)\prime} x_{t,k}^{(r)} - x_{t,k}^{(a)\prime} x_{t,k}^{(a)}]$ and discrimination parameters $\boldsymbol{\beta}_k = 2[x_{t,k}^{(r)} - x_{t,k}^{(a)}]$ (Clinton et al., 2004). We do not do that here. Instead, we adopt an informative prior for the affirmance and reversal positions and make inferences about these quantities. For the policy positions under affirmance, we assume:

$$\mathbf{x}_{t,\cdot}^{(a)} | \boldsymbol{\gamma}_t \sim \mathcal{N}(\mathbf{C}_t \boldsymbol{\gamma}_t, \sigma_{x^{(a)}}^2 \mathbf{I}) \tag{5}$$

and:

$$\boldsymbol{\gamma}_t \sim \mathcal{N}(\boldsymbol{\gamma}_{t-1}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}_t}) \tag{6}$$

where $\mathbf{C}_t$ is a matrix of dummy variables indicating the originating lower court for each case in term t. $\boldsymbol{\gamma}_t$ is a coefficient vector that estimates the mean position of the cases from each court of origin, $\sigma_{x^{(a)}}^2$ is the prior variance of each $x_{t,k}^{(a)}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_t}$ is the prior variance covariance matrix of $\boldsymbol{\gamma}_t$. As discussed above, this hierarchical prior does *not* require us to believe that all cases coming from a particular geographic region have the same status quo point. Rather, this assumes that they come from a common distribution and these distributions may vary across geographic areas and time. If it were the case that the estimated of $\boldsymbol{\gamma}_t$ where identical, then we would conclude that there were no differences among the circuits. We assume that $\sigma_{x^{(a)}}^2$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_t}$ are known *a priori*. We also assume that $\boldsymbol{\gamma}_0 \sim \mathcal{N}(\mathbf{g}_0, \mathbf{G}_0)$. This prior allows us to estimate the mean location of the policy proposals made by each circuit court.

As mentioned above, we adopt an informative prior over the mean locations of the circuits $\boldsymbol{\gamma}_0$ at time $t = 0$; i.e., 1953 for all of the circuits except the 11th. We assume that the most liberal circuits in 1953 were the 2nd, 1st, and 9th, with prior means -0.5, 0.0, and 0.5 respectively. The DC circuit is assigned a prior mean of 0.6. The moderate circuits—the 6th, 7th, and 8th—all are assigned prior means of 0.7. For the conservative circuits, the 10th is assigned mean 0.8, the 4th 1.0, the 3rd 1.1, and the 5th 1.4. The prior variance for these quantities is $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_0} = 0.1\ \mathbf{I}$. In our dynamic specification, we must only assign priors for the initial state—the rest of the dynamics are determined by the data and the prior distribution at $t = 0$.

We further expect there to be a structural break in 1983. We model this by assuming $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_{1983}} = 0.1\ \mathbf{I}$, while $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_t} = 0.01\ \mathbf{I}$ for all other time periods. In essence this allows for a discontinuity in the evolution of the circuit court means (the $\boldsymbol{\gamma}_t$s) to occur in 1983. This allows for two years of percolation time from both the newly created 11th circuit (when the 5th was split in October, 1981), and for the effect of Reagan's initial appointees. It is well known that President Reagan was the first president to aggressively screen judges for the lower bench, in effect neglecting the past norm of senatorial courtesy. Our qualitative results do not depend on this assumption.

For the reversal point $x_{t,k}^{(r)}$, we assume a hierarchical prior that depends on the ideal point of the reversal opinion writer (if known):

$$x_{t,k}^{(r)} \sim \begin{cases} \mathcal{N}(\theta_{t,w_k}, \sigma^2_{x^{(r)}}) & \text{if opinion writer known} \\ \mathcal{N}(0, \sigma^2_{x^{(r)}}) & \text{if opinion writer not known} \end{cases} \tag{7}$$

where w_k is the index of the justice who wrote $x_{t,k}^{(r)}$ and $\sigma^2_{x^{(r)}}$ is the prior variance of $x_{t,k}^{(r)}$ which is assumed known. What this prior substantively means is that if the opinion writer is known, the reversal opinion likely resides near her ideal point. In the cases when the Court affirms, the opinion writer supporting the reversal position is not known, so we assume *a* *priori* that the reversal point has mean zero and variance $\sigma^2_{x^{(r)}}$.[5] This completes the prior on the case parameters.

4.2 Dynamic Prior on Ideal Points

To form the prior on the ideal points, we model the dynamics of the ideal points with a separate random walk prior for each justice (Martin and Quinn, 2002):

$$\theta_{t,j} \sim \mathcal{N}(\theta_{t-1,j}, \Delta_{\theta_{t,j}}) \quad \text{for } t = \underline{T}_j, \ldots, \overline{T}_j \text{ and} j = 1, \ldots, J \tag{8}$$

$\underline{T}_j$ is the first term justice j served, and $\overline{T}_j$ is the last term j served. We do not estimate ideal points for terms in which a justice did not serve. $\Delta_{\theta_{t,j}}$ is an evolution variance parameter the magnitude of which governs how much smoothing takes place over time. As $\Delta_{\theta_{t,j}}$ approaches 0, the model approaches a model in which a justice's ideal points are constant across time. By contrast, as $\Delta_{\theta_{t,j}}$ approaches ∞, the model approaches a model in which the votes in each term are modeled independently of votes in other terms.[6]

In an abuse of notation, we let term $t = 0$ denote term $\underline{T}_j - 1$ for each justice j. For each justice we assume

$$\theta_{0,j} \sim \mathcal{N}(m_{0,j}, C_{0,j}) \tag{9}$$

Here $m_{0,j}$ is set equal to 0 for all justices except: Harlan, Douglas, Marshall, Brennan, Frankfurter, Fortas, Rehnquist, Scalia, and Thomas. Their prior means at time zero were set to 1, -3, -2, -2, 1, -1, 2, 2.5, and 2.5 respectively. $C_{0,j}$ was set to 1 for all justices except the aforementioned nine, whose prior variances at time zero were set to 0.1.

[5]To assess the sensitivity of our results to this prior specification we have fit models in which the prior distribution of $\mathbf{x}_k^{(r)}$ is uniform between $\mathbf{x}_k^{(a)}$ and the ideal point of the justice who wrote $\mathbf{x}_k^{(r)}$. We have found the dynamics to be strikingly similar. This suggests that the data is in fact quite informative, and thus the prior is not driving the inference.

[6]Following Martin and Quinn (2002), $\Delta_{\theta_{t,j}}$ is set equal to 0.1 for all justices except Douglas. Because of the small number of cases that Douglas heard towards the end of his career, it was necessary to use a more informative value of $\sigma^2_{\theta_{t,j}} = 0.001$ in later terms of Douglas' career.

4.3 The Posterior Distribution

The posterior density of the model parameters is proportional to the sampling density times the prior density. Putting these pieces together we are able to write the posterior (up to a constant of proportionality) as:

$$
\begin{aligned}
p(\theta, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma}|\mathbf{Y}) \propto\ & p(\mathbf{x}^{(a)}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\mathbf{x}^{(r)}|\theta)p(\theta)\times \\
& \int \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \big[\mathbb{I}(z_{t,k,j} > 0)\mathbb{I}(y_{t,k,j} = 1)+ \\
& \mathbb{I}(z_{t,k,j} \leq 0)\mathbb{I}(y_{t,k,j} = 0)\big] f_{\mathcal{N}}(z_{t,k,j}|\mu_{t,k,j}, 1) d\mathbf{Z}
\end{aligned} \tag{10}
$$

Treating $\mathbf{Z}$ as latent data we can also write:

$$
\begin{aligned}
p(\theta, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma}, \mathbf{Z}|\mathbf{Y}) \propto\ & p(\mathbf{x}^{(a)}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\mathbf{x}^{(r)}|\theta)p(\theta)\times \\
& \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \big[\mathbb{I}(z_{t,k,j} > 0)\mathbb{I}(y_{t,k,j} = 1)+ \\
& \mathbb{I}(z_{t,k,j} \leq 0)\mathbb{I}(y_{t,k,j} = 0)\big] f_{\mathcal{N}}(z_{t,k,j}|\mu_{t,k,j}, 1)
\end{aligned} \tag{11}
$$

As we show in the following section, this latent data formulation allows us to fit the model using data augmentation (Tanner and Wong, 1987).

5 The Markov Chain Monte Carlo Algorithm

To fit this model we use Markov chain Monte Carlo methods. We iteratively sample from the following full conditional distributions to characterize the posterior distribution:

- $[\mathbf{Z}|\theta, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma}, \mathbf{Y}]$
- $[\boldsymbol{\gamma}|\mathbf{Z}, \theta, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{Y}]$
- $[\theta|\mathbf{Z}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{Y}]$
- $[\mathbf{x}^{(a)}, \mathbf{x}^{(r)}|\mathbf{Z}, \theta, \boldsymbol{\gamma}, \mathbf{Y}]$

Iteratively sampling from these full conditional distributions will yield a series of draws of $(\theta, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma})$ that are approximately from $p(\theta, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma}|\mathbf{Y})$. In the remainder of this Section, we detail the methods we use to sample from each of the full conditional distributions.

5.1 Sampling From $[\mathbf{Z}|\boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma}, \mathbf{Y}]$

The distribution of Z given the other model parameters is made up of conditionally independent truncated normal distributions (Albert and Chib, 1993; Johnson and Albert, 1999). More specifically,

$$\mathbf{z}_{t,k,j}|\boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma} \quad \overset{ind.}{\sim} \begin{cases} \mathcal{N}_{[0,\infty)}(\mu_{t,k,j}, 1) & \text{if} \quad y_{t,k,j} = 1 \\ \mathcal{N}_{(-\infty,0]}(\mu_{t,k,j}, 1) & \text{if} \quad y_{t,k,j} = 0 \\ \mathcal{N}(\mu_{t,k,j}, 1) & \text{if} \quad y_{t,k,j} \text{ is unobserved} \end{cases} \tag{12}$$

where $\mu_{t,k,j}$ is as above and $\mathcal{N}_{[a,b]}(c, d)$ represents the normal distribution with mean c and variance d truncated to the interval $[a, b]$.

5.2 Sampling From $[\boldsymbol{\gamma}|\mathbf{Z}, \boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{Y}]$

We begin by noting that because of the conditional independence assumptions in place we can write

$$p(\boldsymbol{\gamma}|\mathbf{Z}, \boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{Y}) \propto p(\mathbf{x}^{(a)}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})$$

and that this corresponds to a Gaussian dynamic linear model (West and Harrison, 1997). As such, sampling from this full conditional distribution can be accomplished by standard methods for DLMs. For computational efficiency we use the forward filtering, backward sampling algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994).

Let D_t denote all of the information (data and parameters) available at time t. Given the assumed Markov dependence in the prior for $\boldsymbol{\gamma}$ one can show that the full conditional for $\boldsymbol{\gamma}$ given D_T is:

$$p(\boldsymbol{\gamma}|D_T) = p(\boldsymbol{\gamma}_T|D_T)p(\boldsymbol{\gamma}_{T-1}|\boldsymbol{\gamma}_T, D_{T-1}) \cdots p(\boldsymbol{\gamma}_1|\boldsymbol{\gamma}_2, D_1)p(\boldsymbol{\gamma}_0|\boldsymbol{\gamma}_1, D_0)$$

This expression is the key idea that motivates the forward filtering backward sampling algorithm. The sampling process consists of three steps.

1. For $t = 1, \ldots, T$ calculate the quantities $\mathbf{a}_t, \mathbf{R}_t, \mathbf{g}_t, \mathbf{G}_t, \mathbf{h}_t, \mathbf{H}_t$ (defined below).

2. Sample $\boldsymbol{\gamma}_T$ from $\mathcal{N}(\mathbf{g}_T, \mathbf{G}_T)$; i.e., from $p(\boldsymbol{\gamma}_T | D_T)$.

3. For $t = (T-1), (T-2), \ldots, 1$ sample $\boldsymbol{\gamma}_t$ from $\mathcal{N}(\mathbf{h}_t, \mathbf{H}_t)$; i.e., from $p(\boldsymbol{\gamma}_t | \boldsymbol{\gamma}_{t+1}, D_t)$.

In the forward filtering stage of the algorithm (Step 1), it is necessary to compute the following quantities. First is $\mathbf{a}_t$, which is the prior mean of $\boldsymbol{\gamma}_t$ given the information available at time $t-1$, and $\mathbf{R}_t$ which is the prior variance of $\boldsymbol{\gamma}_t$ given the information available at time $t-1$. These quantities are defined as:

$$\mathbf{a}_t = \mathbf{g}_{t-1}$$

and

$$\mathbf{R}_t = \mathbf{G}_{t-1} + \boldsymbol{\Sigma}_{\gamma_t}.$$

Next is $\mathbf{g}_t$, which is the posterior mean of $\boldsymbol{\gamma}_t$ given the information available at time t, and $\mathbf{G}_t$, which is the posterior variance of $\boldsymbol{\gamma}_t$ given the information available at time t. These quantities are defined as:

$$\mathbf{g}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t$$

and

$$\mathbf{G}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t'$$

where $\mathbf{A}_t = \mathbf{R}_t \mathbf{C}_t \mathbf{Q}_t^{-1}$, $\mathbf{e}_t = \mathbf{x}_{t,\cdot}^{(a)} - \mathbf{C}_t \mathbf{a}_t$, and $\mathbf{Q}_t = \mathbf{C}_t \mathbf{R}_t \mathbf{C}_t' + \sigma^2_{x^{(a)}} \mathbf{I}$. Step 2 involves sampling $\boldsymbol{\gamma}_T$ from a normal distribution with mean $\mathbf{g}_t$ and variance-covariance matrix $\mathbf{G}_t$

Step 3 of the algorithm is the backward sampling part of the algorithm. Using the quantities computed above, we sample $\boldsymbol{\gamma}_t$ backward in time from $t = T-1$ to $t = 1$. The mean and variance for the tth normal draw are:

$$\mathbf{h}_t = \mathbf{g}_t + \mathbf{B}_t(\boldsymbol{\gamma}_{t+1} - \mathbf{a}_{t+1})$$

and

$$\mathbf{H}_t = \mathbf{G}_t - \mathbf{B}_t \mathbf{R}_{t+1} \mathbf{B}_t'$$

respectively. The benefit of this algorithm over more direct Gibbs sampling approaches is that it allows us to sample $\boldsymbol{\gamma}$ directly in one piece rather than sampling component by component. This greatly improves the mixing of the Markov chain.

5.3 Sampling From $[\boldsymbol{\theta}|\mathbf{Z}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{Y}]$

To begin, note that the full conditional for justice j's ideal points is independent of justice j''s ideal points for all j and j'. This allows us to sample each justice's ideal points independently of one another. The target density for justice j is:

$$p(\theta_{\cdot,j}|\mathbf{z}_{\cdot,\cdot,j}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{y}_{\cdot,\cdot,j}) \propto p(\mathbf{z}_{\cdot,\cdot,j}|\theta_{\cdot,j}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{y}_{\cdot,\cdot,j}) p(\mathbf{x}_j^{(r)}|\theta_{\cdot,j}) p(\theta_{\cdot,j}), \tag{13}$$

Where $\mathbf{x}_j^{(r)}$ denotes the reversal opinions that j authored.

An independent Metropolis-Hasting algorithm is implemented by sampling a candidate value of $\theta_{\cdot,j}$ (denoted $\theta_j^{(can)}$) from a candidate generating density, $q(\theta_j^{(can)})$. We let $\theta_j^{(cur)}$ denote the current value of $\theta_{\cdot,j}$. The candidate value is accepted with probability:

$$\min\left\{\frac{p(\theta_j^{(can)}|\mathbf{z}_{\cdot,\cdot,j}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}), \mathbf{y}_{\cdot,\cdot,j}}{p(\theta_j^{(cur)}|\mathbf{z}_{\cdot,\cdot,j}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}), \mathbf{y}_{\cdot,\cdot,j}} \frac{q(\theta_j^{(cur)})}{q(\theta_j^{(can)})},\ 1\right\} \tag{14}$$

If $\theta_j^{(can)}$ is not accepted the value of $\theta_j^{(cur)}$ is used as the draw.

Our choice of candidate generating density is the following:

$$q(\theta_j^{(can)}) \propto p(\mathbf{z}_{\cdot,\cdot,j}|\theta_j^{(can)}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \mathbf{y}_{\cdot,\cdot,j}) p(\theta_j^{(can)}) \tag{15}$$

Note that a candidate value can be sampled from this density using a version of the forward filtering, backward sampling algorithm used in the previous section applied to $\theta_\mathbf{j}$. Substituting this choice of $q(\cdot)$ into Equation 14 reveals that the acceptance probability becomes:

$$\min\left\{\frac{p(\mathbf{x}_j^{(r)}|\theta_j^{(can)})}{p(\mathbf{x}_j^{(r)}|\theta_j^{(cur)})},\ 1\right\} \tag{16}$$

5.4 *Sampling From* $[\mathbf{x}^{(a)}, \mathbf{x}^{(r)}|\mathbf{Z}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{Y}]$

We use a random walk Metropolis-Hastings step to sample from the full conditional distribution of $\{x_{t,k}^{(a)}, x_{t,k}^{(r)}\}$ given the other model parameters. The candidate generating density is a bivariate student-t density centered at the current value of $\{x_{t,k}^{(a)}, x_{t,k}^{(r)}\}$. Given the latent $\mathbf{z}_{t,k,\cdot}$, the target density is (up to a constant of proportionality):

$$p(x_{t,k}^{(a)}, x_{t,k}^{(r)}|\mathbf{z}_{t,k,\cdot}, \boldsymbol{\gamma}, \theta_{t,\cdot}, \mathbf{y}_{t,k,\cdot}) \propto \prod_{j \in J_k} f_{\mathcal{N}}(z_{t,k,j}|\mu_{t,k,j}, 1)p(\mathbf{x}^{(a)}|\boldsymbol{\gamma})p(\mathbf{x}^{(r)}|\theta_{t,\cdot})$$

where $\mu_{t,k,j}$ is define as above. With the symmetric candidate generating densities the Metropolis acceptance probability is:

$$\min\left\{\frac{p(x_{t,k}^{(a)(can)}, x_{t,k}^{(r)(can)}|\mathbf{z}_{t,k,\cdot}, \boldsymbol{\gamma}, \theta_{t,\cdot}, \mathbf{y}_{t,k,\cdot})}{p(x_{t,k}^{(a)(cur)}, x_{t,k}^{(r)(cur)}|\mathbf{z}_{t,k,\cdot}, \boldsymbol{\gamma}, \theta_{t,\cdot}, \mathbf{y}_{t,k,\cdot})},\ 1\right\}$$

6 Results

We report the estimated ideal points for the justices in Figures 1 and 2.[7] How do the ideal point measures compare with the Martin-Quinn (2002) scores? The results look very similar to those of Martin and Quinn (2002); the ideal points of many justices seem to vary over time. Just as Martin and Quinn (2002)'s dynamic model showed, Harlan demonstrates a parabolic trajectory, Black trends to conservatism, as does Frankfurter and Scalia. Justices such as Marshall, Brennan, Blackmun, Stevens, Souter, and Ginsburg trend toward liberalism also. For the most part, these dynamics mirror those of the dynamic ideal point model. This suggests that after controlling for case stimuli in the manner performed here, the ideal points of many justices do in fact trend over time.

[7]This model was run for 200,000 iterations after 20,000 burn-in iterations. Standard convergence tests suggest that the chain has converged to the the target distribution.

We perform a more systematic comparison in two ways. First, we correlate the estimates between the two models. Overall, the correlation between is 0.917, demonstrating a strong, linear relationship. In Figure 3 we break down these correlations justice-by-justice, and present correlations between the Martin-Quinn scores and those estimated from the dynamic IRT model with a hierarchical prior. All justices except three have correlations that exceed 0.9. Extremists seem to show the highest correlations, while centrists are somewhat lower. This figure suggests that even after controlling for the agenda in a more principled fashion, the resulting ideal point estimates are essentially the same. This suggests that, when put on the right-hand-side of a regression model, the ideal point estimates in this paper and those of Martin and Quinn (2002) will produce nearly identical results. Thus this paper does not invalidate the use of Martin-Quinn scores as measures of judicial ideology. It is worth noting that this similarity is application-specific and will not be the case *in general*. Indeed, Clinton and Meirowitz (2001) make the case that results based on a model of the agenda will generally be different than results based on agnostic assumptions about the agenda process.

Second, we compare the efficiency of the estimates between the two models. We plot the posterior standard deviations for the ideal point estimates in Figure 4. The results are striking. By including additional information about the court of origin and the opinion writer, we get far more precise estimates of ideal points. Indeed, all of the points in the scatterplot fall below the 45 degree line. This shows one concrete advantage of the more elaborate model—we obtain far more precise estimates of ideal points.

Of course the estimated ideal points are not the totality of interesting quantities. Following Martin and Quinn (2002), we look at the position of the median justice along the estimated dimension from 1953 to 199. We plot the posterior densities of the median justice in Figure 5. The results here are quite interesting. The early Warren courts were, on an absolute scale, the most conservative. The late Warren courts were the most liberal. The Burger and Rehnquist courts were certainly more conservative than the Warren courts, but, particularly in the 1990s, not as conservative as we might have expected (and which was suggested by the other models). This is not entirely surprising, as either O'Connor (or perhaps Kennedy) is likely the median justice. Further, many of the cases coming before the Court were decided by lower courts in fairly moderate directions. This is another reason to dynamically model the case parameters across time; to compare justices across time on an absolute scale, one has to adjust for temporal changes in the inputs to the system.

Another advantage of our modeling approach is the ability to place the lower federal courts in the same ideological space as the justices. The model assumes that the lower federal courts are setting the status quo point for the Supreme Court; i.e., the policy reversion if the Court votes to affirm. We plot summaries of the posterior distribution of $\boldsymbol{\gamma}_t$ in Figure 6. Many of the circuits exhibit a similar pattern: tending toward conservatism in the 1950s, becoming more liberal throughout the 1960s and 1970s, and slowly returning to conservatism in the 1980s, and turning slightly more liberal in the late 1990s. This pattern, which follows presidential politics quite well (with a time lag), is apparent for the 1st (New England), 2nd (New York, Vermont, and Connecticut), 4th (the mid-Atlantic states), 9th (the west coast), and DC circuits.

Some circuits trend toward liberalism until the early-to-mid 1980s, and then either level out or start becoming increasingly conservative. The 3rd (Pennsylvania, New Jersey, and Delaware), the 5th (the deep South before 1981, and Texas, Louisiana, and Mississippi thereafter), the 6th (the Midwest), the 8th (the plains states), and the 10th (the Rocky Mountain states) follow this pattern. Overall, the 4th, 5th, 10th and 11th circuits are most conservative. It is also worth noting that while imposing a structural break in 1983 *a priori*, there is little evidence of such a break in the data. One could argue that Reagan's disregard for Senatorial courtesy was either counter-balanced by the increasing liberalization of society, or only manifested itself after a considerable time lag.

7 Discussion

This paper shows that incorporating background information in statistical measurement models affords a number of advantages. The dynamic IRT model with a hierarchical prior offered here harnesses information about the docket and opinion writer to more reliably model the data generating process. In so doing, the model not only provides more efficient estimates of judicial ideal points, it also places the circuits in the same ideological space as the justices.

This application shows some of the advantages of adopting a Bayesian approach to estimation. We were able to incorporate prior information to bring additional information, such as the ideological predispositions of the federal appeals courts, to bear on the problem of estimating judicial ideal points. We were also able to estimate complex and flexible models using Markov chain Monte Carlo methods that would have been otherwise intractable.

Substantively we have confirmed Martin and Quinn (2002)'s conclusion that ideal points of Supreme Court justices do change over time. More specifically, we find that after explicitly modeling the location of the policy options open to justices we recover ideal point estimates very similar to those of Martin and Quinn (2002). Perhaps the major benefit of the approach taken in this paper is that it produces an estimate of the mean policy location of opinions coming out of each of the circuits. Since these estimates reflect only the cases from each circuit that were actually heard by the U.S. Supreme Court this estimates should not be interpreted as pure measures of lower court ideology. Nonetheless, they do provide a window to assess how the high Court has shaped its docket over time.

The results beg the question: why do the circuits change ideologically? Is this simply due to replacement on the bench, or due to strategic concerns about *en banc* or Supreme Court review? We leave this question for future research. But our modeling approach could be adapted to address these questions. One could incorporate suitable covariates in the statistical models to test for these explanations, or perhaps employ existing data about decision making in the Courts of Appeal. Finally, there is additional data that can be brought to the problem. In particular, having data about the *cert* process would be quite informative about not only the status quo and alternative points, but also how a minority of the Court goes about setting the agenda. Such data is available in a limited form, but to date no comprehensive data base of *cert* votes is publicly available.

References

Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(June):669–679.

Carter, C. K., and R. Kohn. 1994. "On Gibbs Sampling for State-Space Models." *Biometrika* 81(September):541–533.

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

Clinton, Joshua D., and Adam Meirowitz. 2001. "Agenda Constrained Legislator Ideal Points and the Spatial Voting Model." *Political Analysis* 9(Summer):242–259.

Clinton, Joshua D., and Adam Meirowitz. 2003. "Integrating Voting Theory and Roll Call Analysis: A Framework." *Political Analysis* 11(Fall):381–396.

Clinton, Joshua D., and Adam Meirowitz. 2004. "Testing Explanations of Strategic Voting in Legislatures: A Reexamination of the Compromise of 1790." *American Journal of Political Science* 48(October):675–689.

Davis, Otto A., Melvin J. Hinich, and Peter C. Ordeshook. 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *The American Political Science Review* 64(2):426–448.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Enelow, James, and Melvin Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge: Cambridge University Press.

Frühwirth-Schnatter, Sylvia. 1994. "Data Augmentation and Dynamic Linear Models." *Journal of Time Series Analysis* 15:183–202.

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

Johnson, Valen E., and James H. Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Londregan, John. 2000. *Legislative Institutions and Ideology in Chile*. Cambridge: Cambridge University Press.

Maltzman, Forrest, James F. Spriggs, II, and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. New York: 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(Spring):134–153.

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

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll-Call Voting*. Oxford: Oxford University Press.

Spaeth, Harold J. 2004. "United States Supreme Court Judicial Database, 1953-2003 Terms [Computer File]." `http://www.as.uky.edu/polisci/ulmerproject/`.

Tanner, M. A., and W. Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82:528–550.

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer.

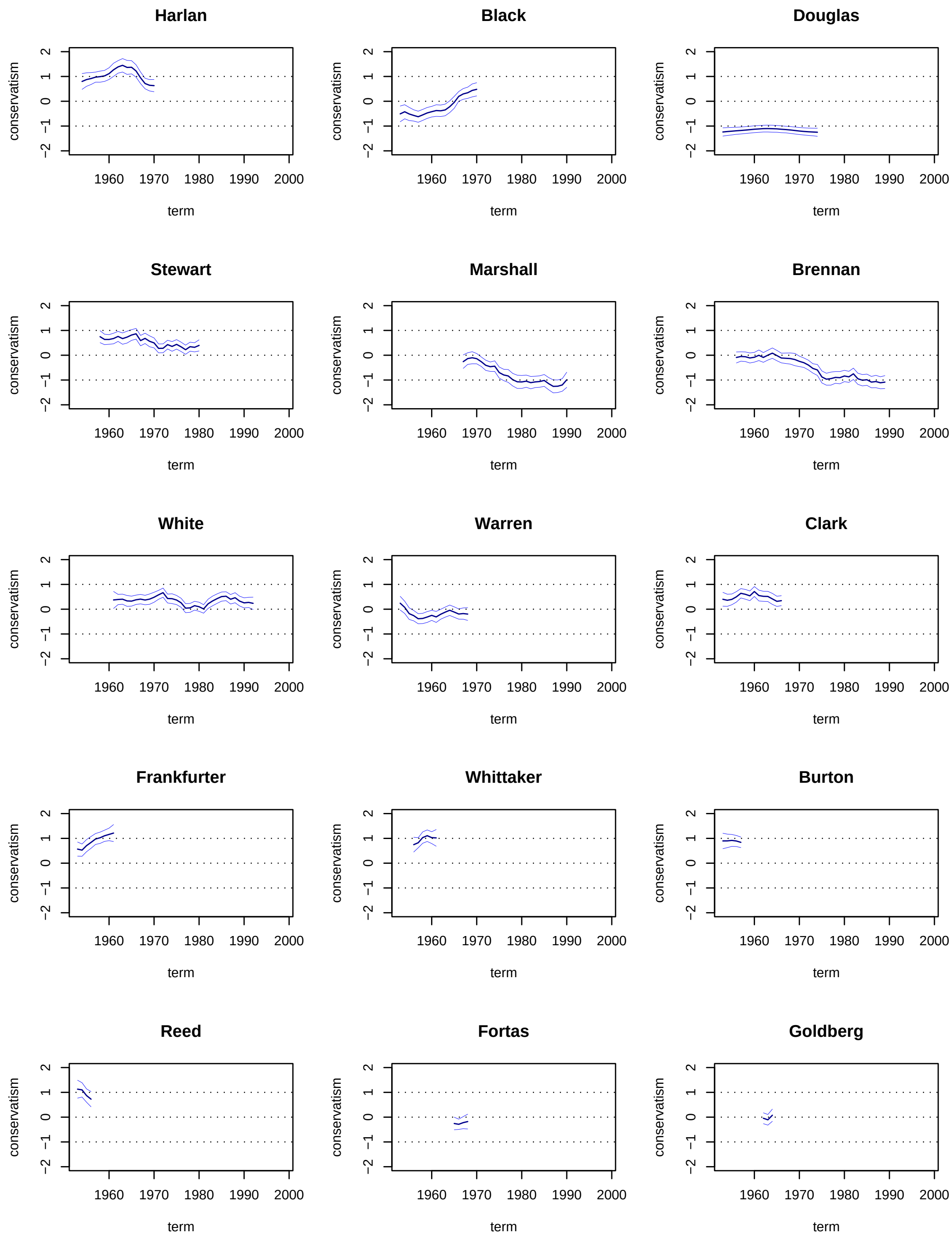

Figure 1: Posterior density summary of the ideal points of each justice for the terms in which they served for the dynamic IRT model with a hierarchical prior, Justices Harlan-Goldberg. The thick, dark line denotes the posterior mean, and the light lines are ± 2 posterior standard deviations away from the posterior mean.

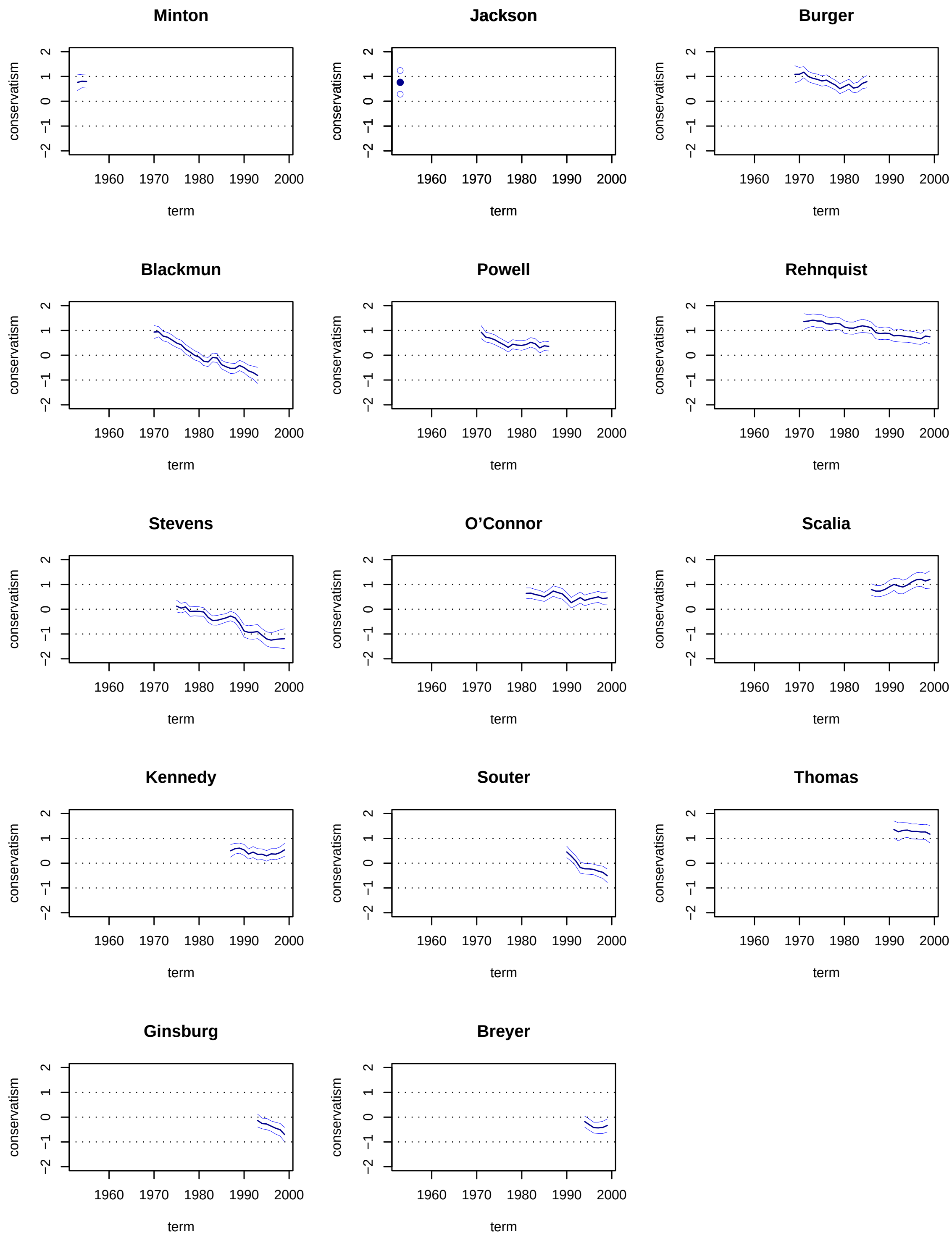

Figure 2: Posterior density summary of the ideal points of each justice for the terms in which they served for the dynamic IRT model with a hierarchical prior, Justices Minton-Breyer. The thick, dark line denotes the posterior mean, and the light lines are ± 2 two posterior standard deviations away from the posterior mean.

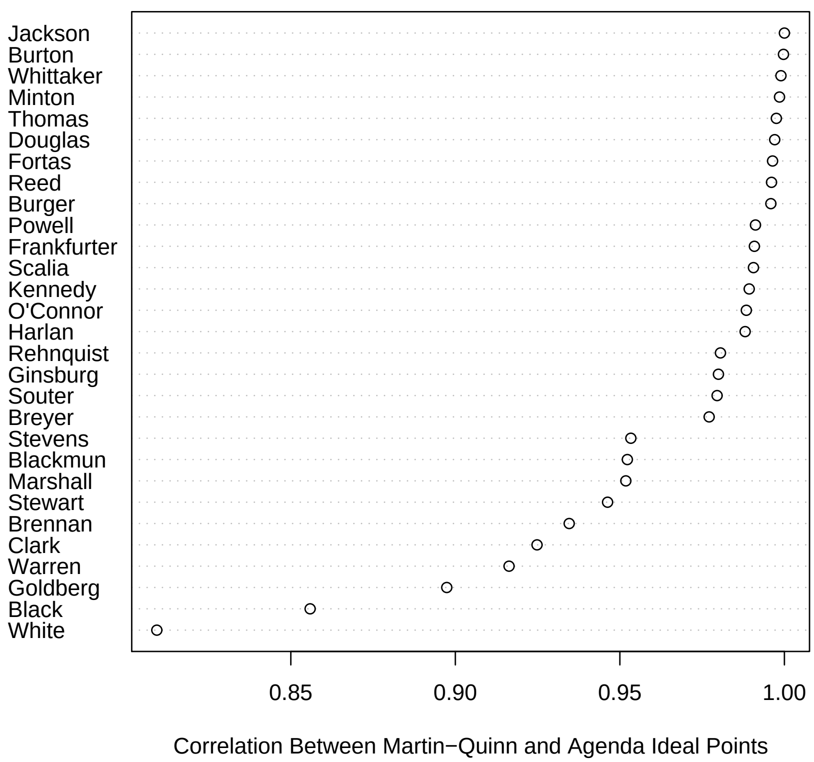

Figure 3: Justice-by-justice ideal point correlations between the Martin-Quinn (2002) posterior means and the the dynamic IRT model with a hierarchical prior (agenda) posterior means.

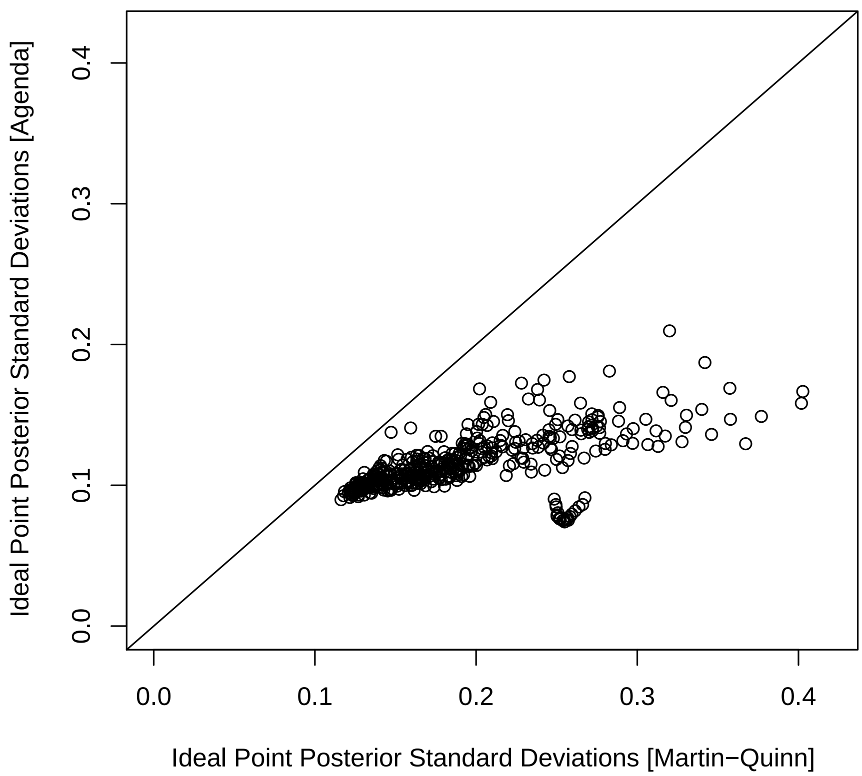

Figure 4: Ideal point posterior standard deviations scatterplot for the Martin-Quinn (2002) posterior means and the the dynamic IRT model with a hierarchical prior (agenda).

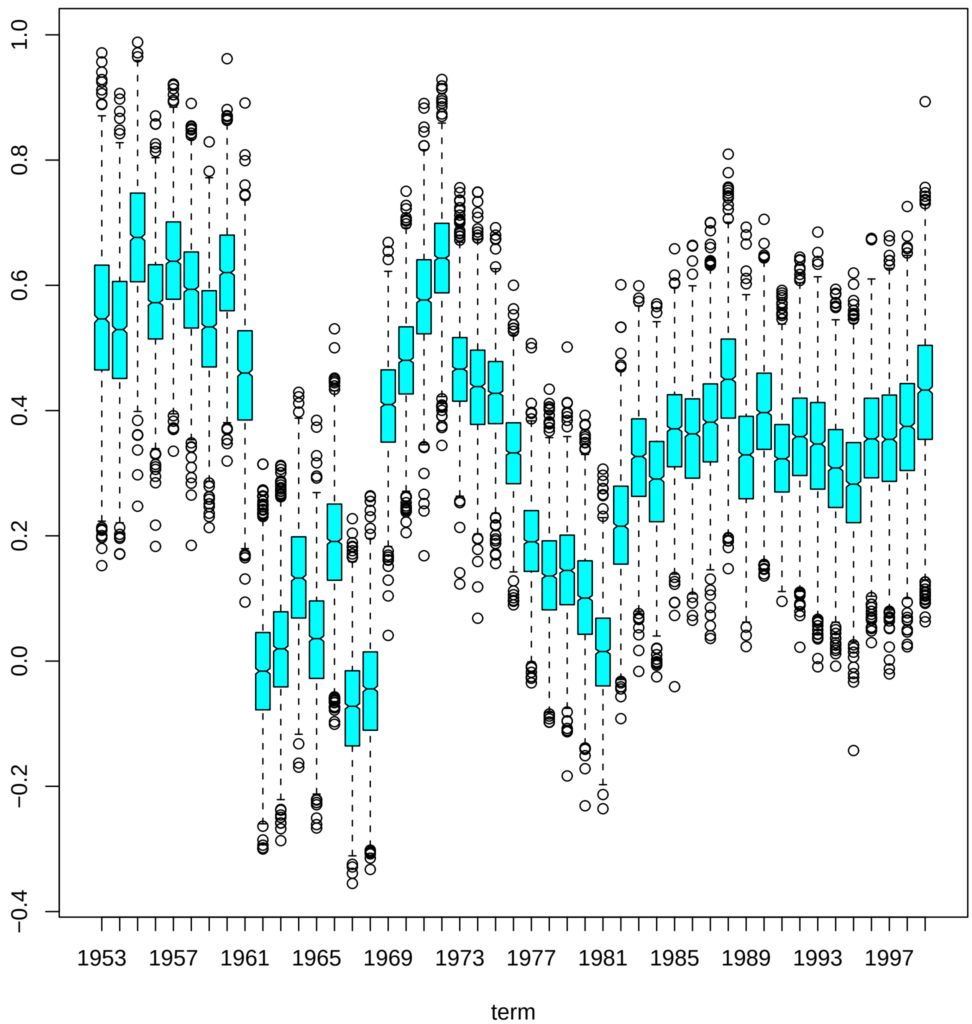

Figure 5: Estimated posterior distribution of the location of the median justice for the dynamic ideal point and case parameter model.

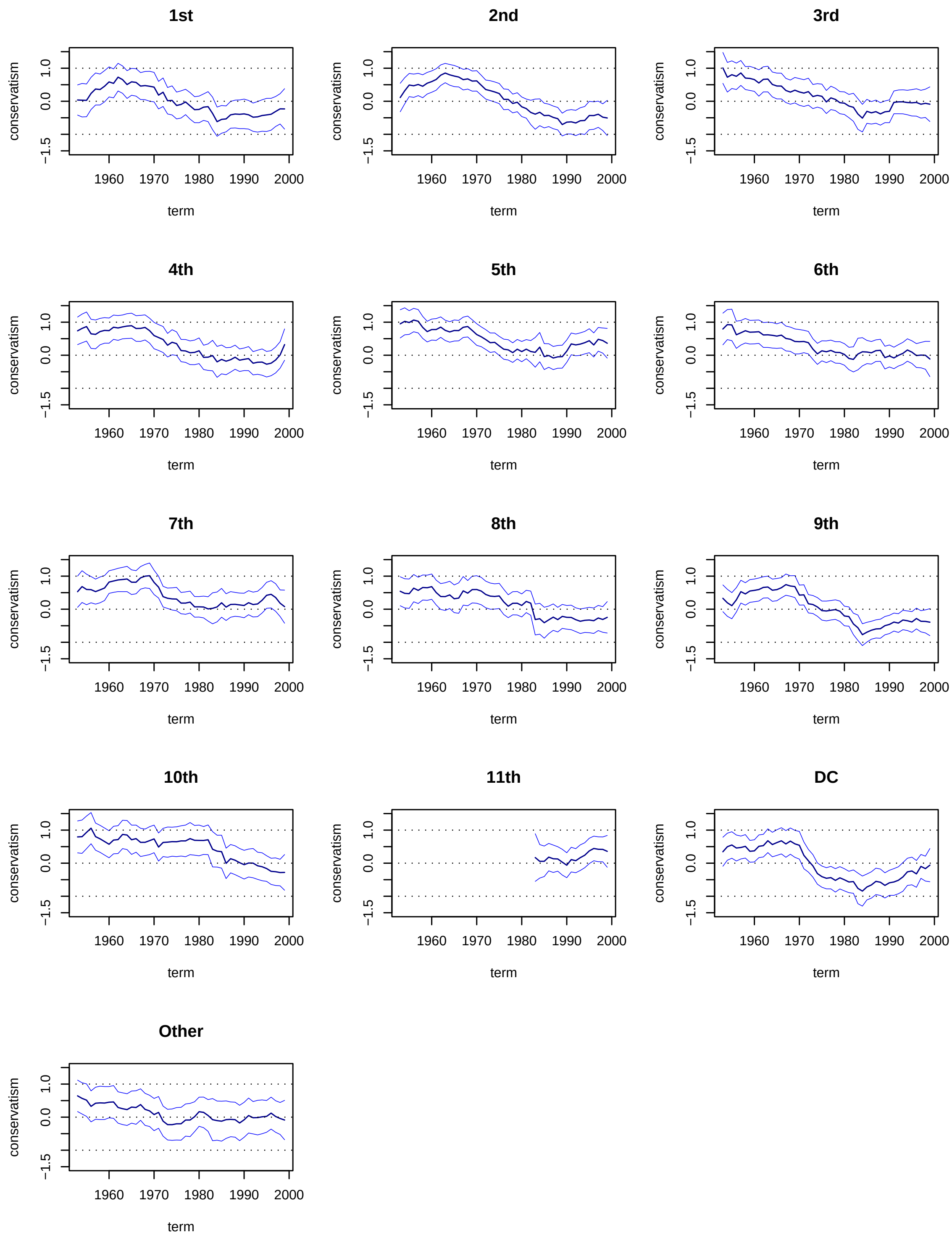

Figure 6: Estimated position of the prior mean ($\gamma_{c,t}$) of affirmance points [$\mathbf{x}_t^{(a)}$] from the lower court of origin c for the dynamic ideal point and case parameter model. The thick, dark line denotes the posterior mean, and the light lines are ± 2 posterior standard deviations away from the posterior mean.