Bayesian Learning about Ideal Points of U.S. Supreme Court Justices, 1953-1999[*]

Andrew D. Martin[†] Kevin M. Quinn[‡]

Center for Statistics and the Social Sciences
University of Washington
Working Paper no. 16
July 23, 2001

Abstract

At the heart of attitudinal and strategic explanations of judicial behavior is the assumption that justices have policy preferences. These preferences have been measured in a handful of ways, including using factor analysis and multidimensional scaling techniques (Schubert, 1965, 1974), looking at past votes in a single policy area (Epstein et al., 1989), content-analyzing newspaper editorials at the time of appointment to the Court (Segal and Cover, 1989), and recording the background characteristics of the justices (Tate and Handberg, 1991). In this manuscript we employ Markov chain Monte Carlo (MCMC) methods to fit Bayesian measurement models of judicial preferences for all justices serving on the U.S. Supreme Court from 1953 to 1999. We are particularly interested in considering to what extent ideal points of justices change throughout their tenure on the Court, and how the proposals over which they are voting also change across time. To do so, we fit four longitudinal item response models that include dynamic specifications for the ideal points and the case-specific parameters. Our results suggest that justices do not have constant ideal points, even after controlling for the types of cases that come before the Court.

---

[*] We would like to thank Michael Bailey, Scott Comparato, Lee Epstein, Mark Handcock, Peter Hoff, Rich Pacelle, Keith Poole, Adrian Raftery, Kate Stovel, and Steve Van Winkle for useful discussions along the way. All analyses were performed on Linux workstations using code written by the authors in `C++` using the Scythe Statistical Library (Martin and Quinn 2001b). We gratefully acknowledge financial support from the Center in Political Economy at Washington University. Data for replication can be obtained from the Supreme Court Database website at Michigan State University (`http://www.ssc.msu.edu/~pls/pljp/sctdata.html`). Documented `C++` code for these models will be available Fall 2001 at the authors' websites.

[†] Department of Political Science, Washington University, Campus Box 1063, One Brookings Drive, St. Louis, MO 63130-4899, `admartin@artsci.wustl.edu`

[‡] Department of Political Science and Center for Statistics in the Social Sciences, Box 354322, University of Washington, Seattle, WA 98195-4322, `quinn@stat.washington.edu`

1 Introduction

Whether one adopts an attitudinal (Segal and Spaeth, 1993) or strategic (Eskridge, 1991; Epstein and Knight, 1998) theoretical model of decision making on the Supreme Court, the key explanatory variables are the policy preferences of the justices. In this paper, we tackle the following question: Do ideal points of U.S. Supreme Court justices change over time? There are reasons to believe that they do not, as justices are sophisticated politicians, typically with vast experience on the bench. In reviewing the literature, Epstein et al. (1998) summarize the conventional wisdom in the judicial politics field: "[t]he occasional anomaly notwithstanding, most jurists evince consistent voting behavior over the course of their careers" (p. 801). On the other hand, it is plausible that as justices learn by deciding new and different cases, their world view – and thus their ideal points – change over time. Recent work by Epstein et al. (1998) demonstrate the the preferences of some justices may in fact change systematically over time. Of course, so too do the policy proposals that come to the Court, and the types of cases the justices hear. Does this mean that preferences change, or that the agenda changes?

The purpose of this paper is to present a handful of models that can be used to estimate the ideal points of Supreme Court justices. This is well-tread methodological ground; in every decade since the 1950s, scores of articles have been published that estimate ideal points for legislatures and courts using various techniques. The seminal work in the judicial politics subfield is Schubert's *The Judicial Mind* (1965). He presents a comprehensive look at the attitudes and ideologies of all Supreme Court justices serving from the end of the Roosevelt era to the beginning of the Johnson administration. His i-points (what we would now term ideal points) and j-points (the case stimuli), along with his C scale (political liberalism) and E scale (economics), became an important part of the judicial politics vernacular, and let to a "behavioral revolution" in the study of Supreme Court decision making. Perhaps the best example in the legislative politics subfield is the work by Keith Poole and Howard Rosenthal, who study the entirety of roll call voting in the United States House and Senate (Poole and Rosenthal, 1991a, 1997). Not only have their measures informed a generation of Congress scholars, but the computational tools brought to the problem were quite innovative.

As detailed below, we depart from the existing literature in two fundamental ways. First, we explicitly model the longitudinal nature of the data, allowing both ideal points and case-specific parameters to vary systematically over time. This is similar in spirit to the D-NOMINATE model proposed by Poole and Rosenthal, but our approach is quite different in terms of specification and estimation strategy. Second, we adopt a Bayesian approach. But for the development of Markov chain Monte Carlo (MCMC) methods in the late 1980s and early 1990s, the models we propose would have been intractable. Others have performed Bayesian inference for standard item response models (Albert, 1992; Patz and Junker, 1999) and item response models applied to congressional voting data (Clinton et al., 2000; Jackman, 2001), but to our knowledge no one has developed models for multiple cross-sections of votes that flexibly model the temporal dependence that one might expect to find.[1]

The main substantive findings of this paper are that even after controlling for the dynamics of issue creation in the lower courts, there is strong evidence to support the assertion that many justices' revealed preferences have changed (in some cases dramatically) over time. In many cases, these dynamics do not appear to be linear or quadratic functions of time. In addition, we find evidence that the rulings of lower courts that come under review by the Supreme Court have generally become increasingly liberal over time. Recently, the rulings of the Supreme Court have tended to become more conservative.

This paper proceeds as follows. In the following section we review other approaches scholars have employed to measure judicial preferences. This is followed in Section 3 by a presentation

---

[1] Bailey and Chang (2001) allow the ideal points of Supreme Court justices to vary across time in their measurement model of Supreme Court ideal points. Their model assumes linear or quadratic trajectories based on the conclusions of Epstein et al. (1998). As discussed below, the conclusions drawn by Epstein et al. (1998) are open to some criticism; and, as a result, so too are the Bailey and Chang (2001) estimates. Additionally, no standard errors are reported for the ideal point estimates, so it is unclear whether real trending is taking place. A better approach is to choose a functional form that allows the data to speak about which justices' preferences change over time, as well as the form of that change.

of the standard item response model which can be employed if justices are assumed to have constant preferences. In Section Four, we assume that ideal points are independent and identically distributed across time, and fit a model based on that assumption. This is followed by our first dynamic model, where we assume that a justice's ideal point follows a Gaussian random walk. In Section Six, we extend this model one step further, and jointly model the dynamics in the ideal points and case-specific parameters. The final section concludes with a discussion of directions for future research.

2 Measuring Judicial Preferences

Over the last fifty years many scholars have devoted a substantial amount of energy estimating ideal points for Supreme Court justices. The reason for this work is clear: regardless of theoretical approach, one requires a measure of the ideal policy position of each justice (or the Court, which is typically measured using the median justice) to explain judicial behavior. Here we remain agnostic as to whether justices behave attitudinally or strategically.[2] Like many previous scholars, we are interested in estimating the policy preferences of justices. Unlike many others, we are not interested in creating a set of preference measures that can be simply plugged into regression models of various sorts as either left-hand-side or right-hand-side variables. Indeed, we believe that such approaches are often inappropriate since the assumptions often used to construct the measures of preferences are inconsistent with aspects of the regression models used by future researchers. Instead, we believe that if one wants to make causal inferences about the effects of policy preferences, case characteristics, and other relevant independent variables on Supreme Court voting behavior, one must specify a structural model derived from theory and fit the resulting model. This is what we pursue in this paper.

The first generation of research relating to ideal point estimation grew from the work of Schubert (1959). Schubert (1965) applied various scaling models to votes on the Court, and finds two primary

---

[2] The attitudinal model is based on psychological stimulus-response models, while strategic approaches are based on economic rational actor models. Both can be viewed in terms of the spatial voting model (Downs, 1957; Enelow and Hinich, 1984), where each actor has an ideal point and votes for the policy alternative closest to that point. For expositional economy, we employ the parlance of the spatial voting model throughout this paper.

dimensions – political and economic liberalism – that explain votes on the merits for the Court from the late 1940s to the early 1960s. Nine years later, Schubert (1974) again finds that these two dimensions structure behavior through the Vinson and Warren Courts. The use of latent variable models to measure ideal points continued in the work of Rohde and Spaeth (1976), who find three dimensions that structure the Warren Court and first five terms of the Burger Court: freedom, equality, and what they term "New Dealism," and Spaeth (1990) who considers civil liberties subscales on the Burger Court. In this vein, Martin and Quinn (2001a) fit item response models to data from the seventh Burger Court. They develop models specifically suited to mediate the 'micro-committee problem' that plagues the study of the Supreme Court.

A similar approach relies on using past behavior in a policy area to explain votes in the same policy area in the future. Epstein et al. (1989) look at final case decisions in all criminal justice cases before the Supreme Court between 1946 and 1986, and gauge the preferences of the Court by using the percentage of pro-criminal rights decisions in the previous term. Similarly, Segal and Spaeth (1989) report the fraction of liberal decision on civil liberties and economic cases for all justices serving between 1953 and 1985. They find a great deal of variance within and between these two salient dimensions.

Other scholars have developed a handful of novel measures not based on the votes of the justices: Segal and Cover (1989) content analyze newspaper editorials at the time of appointment to the Court, and Tate and Handberg (1991) record background characteristics of the justices. The most widely used of these measures are the Segal and Cover (1989) scores, which were updated by Segal et al. (1995) to include the Bush appointees, and backdated for seven Roosevelt and four Truman nominees. Epstein and Mershon (1996) perform a "methodological audit" of this measure, and demonstrate that it should *only* be used for their designed purpose: explaining aggregated civil liberties votes. It is interesting to note that these scores explain less than 20% of the variance in economics cases, the other salient dimension documented in previous scholarship.

A key theoretical assumption on which both attitudinal and strategic models of Supreme Court behavior rest is the idea of preferences being fixed over time. The work of Schubert (1974), Baum (1988), Segal and Cover (1989), and many others suggest that the ideal points of justices rarely change. Epstein et al. (1998) challenge this notion, and seek to discover whether or not preferences do change. The preference measure they employ is the Baum-corrected percentage of each justice's votes in the liberal direction on civil liberties cases in each term (Baum, 1988). This correction subtracts the median difference between a justice in the previous term and the current term from all justices, and is geared to control for agenda changes. Of the sixteen justices under study, Epstein et al. find nine justices that exhibit changes over time. Not only did they include the justices which previous studies have demonstrated change over time, such as Black (Ulmer, 1973; Atkins and Sloope, 1986), Douglas (Ulmer, 1981), and Blackmun (Baum, 1992), but they also included some surprises like Frankfurter and Powell. Some justices – not surprisingly many of which are at the ends of the ideological spectrum – demonstrated no change, such as Brennan, Burger, and Marshall.

There are some weaknesses of the Epstein et al. (1998) study which make its conclusions somewhat suspect. First, without explicitly modeling case-specific parameters, it is impossible to know whether the Baum correction adequately controls for the changing nature of the cases that are heard by the Court. Second, the dependent variable in the analysis (percent liberal on civil liberties cases) is treated as if it is a known quantity. However, with the small number of cases heard in each term, it is important to account for the inherent uncertainty in the measure. Third, this approach makes the theoretical assumption that ideal points are independent and identically distributed across terms. This is theoretically problematic, as we would expect some continuity in each justices' outlook. Finally, this approach is not based on a fully parameterized statistical model, which makes inference difficult.

As Epstein et al. (1998) argue, whether or not the ideal points of Supreme Court justices change over time is a very important issue. The answer to the question will dictate the manner in which future scholars should measure preferences. If it is the case that ideal points do change, this has profound implications for how all sorts of future research should be conducted, and it calls into question previous research findings. In this paper we pursue this question in the context of item response modeling. We begin by positing four theoretical models of judicial decision making across time, and then directly operationalize them into statistical models. Such a structural model has the property that if the theoretical model is correct, thus implying that the posited statistical assumptions hold, then such a model can explain votes in terms of preferences even though preferences are treated as latent variables.[3]

3 Modeling Supreme Court Decision Making

Our focus in this paper are the $T = 47$ terms of the Court from 1953 to 1999. In this preliminary model, we assume that during this time period each justice has a fixed ideal point.[4] We obtain our data from *The United States Supreme Court Judicial Database* (Spaeth, 1999). We select all non-unanimous cases that are either formally decided cases with written opinions after full oral argument or cases decided by an equally divided vote.[5] This results in votes for $J = 29$ justices who served in this time period. On a given case, at most nine justices cast a vote. A grand total of $K = 3210$ cases were decided, with an average of sixty-eight cases per term. The most active Courts sat during the 1972 and 1986 terms, deciding ninety-eight cases. The least active was the 1996 term that decided only forty-one cases.

The observed data matrix $\mathbf{Y}$ is thus a $(K \times J)$ matrix of votes and missing values. We can also partition this matrix by term t, and thus discuss the submatrices $\mathbf{Y}_t$. 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.[6] A total

---

[3] This focus on a unified structural model of voting and preference estimation is also found in Clinton et al. (2000).

[4] One key tenet of the strategic approach is that justices may behave in a non-sincere fashion. If it is the case justices are behaving strategically, their revealed preferences may not be the same as their sincere preferences. If one believes the attitudinal model, our estimates could be viewed as true ideal points. If one believes the strategic model, our estimates should be viewed as revealed preferences (which are likely to change over time due to strategic considerations).

[5] The unit of analysis is the case citation (ANALU=0). We select cases where DEC_TYPE equals 1 or 5, and drop all cases when VOTE equals 40, 50, 60, 70, 80, and 90. It would be possible to include unanimous decisions in our analysis if one is willing to fix *a priori* the discrimination and difficulty parameters for these cases.

[6] We use the disposition variable DIS in the Supreme Court Date Base to code affirmances and reversals. An affirmance (DIS=1) or a certification to a lower court (DIS=9) are coded as affirmances. Cases that are coded as reversed (DIS=2), reversed and remanded (DIS=3), vacated and remanded (DIS=4), five plus remanded (DIS=6), vacated (DIS=7), or petition denied or appeal dismissed (DIS=8) are coded as reversals.

of 2018 lower court cases were reversed, while 1162 were affirmed. Using the disposition variable in the *Supreme Court Database*, we are able to flip the vote variables to form the data matrix. The data matrix thus contains zeros, ones, and missing values (in fact, most of the data matrix is missing data, as only nine justices sit on the Court in a given term). As discussed below, we use data augmentation to deal with missing data.

3.1 The Attitudinal Model Formalized

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 D-dimensional issue space.[7] 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 status quo is greater than the utility the justice attaches to the alternative, regardless of the expected actions of the other actors.

To operationalize this model, we begin by writing down random utility functions. Let

$$u_{k,j}^{(a)} = -\|\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\|^2 + \delta_{k,j}^{(a)}$$

be the utility to justice $j \in J_k$ of voting to affirm on case $k \in K_t$ in term t, and

$$u_{k,j}^{(r)} = -\|\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\|^2 + \delta_{k,j}^{(r)}$$

be the utility to justice $j \in J_k$ of voting to reverse on case $k \in K_t$ in term t. $\boldsymbol{\theta}_j$ is justice j's ideal point in the D-dimensional issue space. Note that we assume that each justice's ideal point is constant across terms. We thus call this model the "constant ideal point model." $\mathbf{x}_k^{(a)}$ is the location of the policy under an affirmance vote, $\mathbf{x}_k^{(r)}$ is the location of the policy under a reversal, and $\delta_{k,j}^{(a)}$ and $\delta_{k,j}^{(r)}$ are Gaussian disturbances with zero mean and variances τ_a^2 and τ_r^2 respectively.

---

[7] We derive the model for a D-dimensional issue space, although we only present results for the one-dimensional case ($D = 1$). For examples of multi-dimensional models, see Martin and Quinn (2001a) and Jackman (2001).

Given this spatial model, justice j will vote to reverse on case k when $u_{k,j}^{(r)} > u_{k,j}^{(a)}$ or equivalently when $u_{k,j}^{(r)} - u_{k,j}^{(a)} > 0$. Let $z_{k,j}$ be the difference between $u_{k,j}^{(r)}$ and $u_{k,j}^{(a)}$. We can write and simplify this utility difference $z_{k,j}$ as follows:[8]

$$\begin{aligned}
z_{k,j} = u_{k,j}^{(r)} - u_{k,j}^{(a)} &= -\|\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\|^2 + \delta_{k,j}^{(r)} + \|\boldsymbol{\theta}_j - \mathbf{X}_k^{(a)}\|^2 - \delta_{k,j}^{(a)} \\
&= -\left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\right]' \left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\right] + \delta_{k,j}^{(r)} + \left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\right]' \left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\right] - \delta_{k,j}^{(a)} \\
&= -\boldsymbol{\theta}_j'\boldsymbol{\theta}_j + 2\boldsymbol{\theta}_j'\mathbf{x}_k^{(r)} - \mathbf{x}_k^{'(r)}\mathbf{x}_k^{(r)} + \delta_{k,j}^{(r)} + \boldsymbol{\theta}_j'\boldsymbol{\theta}_j - 2\boldsymbol{\theta}_j'\mathbf{x}_k^{(a)} + \mathbf{x}_k^{'(a)}\mathbf{x}_k^{(a)} - \delta_{k,j}^{(a)} \qquad (1) \\
&= \left[\mathbf{x}_k^{'(a)}\mathbf{x}_k^{(a)} - \mathbf{x}_k^{'(r)}\mathbf{x}_k^{(r)}\right] + 2\boldsymbol{\theta}_j'\left[\mathbf{x}_k^{(r)} - \mathbf{x}_k^{(a)}\right] + \left[\delta_{k,j}^{(r)} - \delta_{k,j}^{(a)}\right] \\
&= \alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j + \varepsilon_{k,j}
\end{aligned}$$

Here α_k is a scalar, $\boldsymbol{\beta}_k$ is a $D \times 1$ column vector, and $\boldsymbol{\theta}_j$ is a $D \times 1$ column vector. Given this model, justice j will vote to affirm on case k when $z_{k,j} > 0$.

3.2 A Multi-Dimensional Item Response Model

The process of translating the spatial voting model above (presented in Equation 1) into a statistical model begins by noting the relationship between the spatial voting model and our observed data.

For observed votes, we assume that:

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

Where:

$$z_{k,j} = \alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j + \varepsilon_{k,j} \qquad \varepsilon_{k,j} \overset{iid}{\sim} \mathcal{N}(0,1). \qquad (3)$$

Note that we have fixed the variance of $\varepsilon_{k,j}$ to one since this variance and the other model parameters are not separately identified in the likelihood. This is a typical assumption; e.g., one makes this assumption for the standard probit model for a dichotomous dependent variable.

To form a probability model, we assume that:

$$y_{k,j} | \alpha_k, \boldsymbol{\beta}_k, \boldsymbol{\theta}_j \overset{iid}{\sim} \mathcal{B}ernoulli(\pi_{k,j}).$$

---

[8] Clinton et al. (2000) derive the same expression in the context of a model of legislative voting and go on to show the link with standard item response models.

Taken together, Equations 2 and 3 imply that $\pi_{k,j} = \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Thus, the sampling density of the observed data is:

$$f(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) \propto \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j)^{y_{k,j}} \left[1 - \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j)\right]^{1-y_{k,j}}$$

We must define some additional notation to present our estimation strategy. Let$\boldsymbol{\alpha}$ denote the stacked $\alpha_k \ \forall \ k \in \{1, 2, \ldots, K\}$, $\boldsymbol{\beta}$ to denote the stacked $\boldsymbol{\beta}_k \ \forall \ k \in \{1, 2, \ldots, K\}$, and $\boldsymbol{\theta}$ denote the stacked $\boldsymbol{\theta}_j \ \forall \ j \in \{1, 2, \ldots, J\}$. Bayesian inference for the justices' ideal points $\boldsymbol{\theta}$ and the case parameters α and the $\boldsymbol{\beta}$ proceeds by summarizing the posterior density given by:

$$p(\alpha, \boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\alpha, \boldsymbol{\beta}, \boldsymbol{\theta}) p(\alpha, \boldsymbol{\beta}, \boldsymbol{\theta})$$

where $p(\alpha, \boldsymbol{\beta}, \boldsymbol{\theta})$ represent our prior beliefs about these parameters.

3.3 Identification and Prior Distributions

It is well known that item response models suffer from identification problems (see, for example, Albert, 1992; Johnson and Albert, 1999). The first problem is called scale invariance. The parameters of interest $\alpha, \boldsymbol{\beta}$, and $\boldsymbol{\theta}$ are identified only up to an arbitrary scale factor. Thus, one must anchor the scale on which ideal points are measured, just like one would fix a scale to measure any quantity. In the Bayesian context, one typically fixes the scale invariance problem by assuming a proper probability density function for the bliss points $\boldsymbol{\theta}$. The standard prior distribution for the ideal points is to assume $\boldsymbol{\theta}_j \overset{iid}{\sim} \mathcal{N}(\mathbf{t}_0, \mathbf{T}_0)$ for all justices $j = 1, \ldots, J$. We assume independent standard Normal priors on the ideal points ($\mathbf{t}_0 = \mathbf{0}$ and $\mathbf{T}_0 = \mathbf{I}_D$). For an alternative informative prior based on the ordering of the justices, see Martin and Quinn (2001a).

An additional identification problem is called rotational invariance. For the one-dimensional case under consideration ($D = 1$), multiplying all of the model parameters by negative one would not change the value of the likelihood function. Substantively, the model cannot determine what direction is liberal or conservative; i.e., should William Rehnquist be given a large positive or large negative score? We also remedy this problem by using prior distributions. We assign zero prior probability that Marshall falls above zero, which forces us to estimate a conservatism measure, and thus fixes rotational invariance.[9]

So far we have discussed prior distributions for the bliss points $\boldsymbol{\theta}$, but it is also necessary to assign prior probabilities to the α and $\boldsymbol{\beta}$ parameters. The standard approach is to assume that these are drawn from normal distributions. In other words:

$$\begin{bmatrix} \alpha_k \\ \boldsymbol{\beta}_k \end{bmatrix} \sim \mathcal{N}_{D+1}(\mathbf{b}_0, \mathbf{B}_0) \;\; \forall \; k \in \{1, \ldots, K\}$$

In most applications, estimating these parameters is straightforward when there are many actors under study (whether students in the educational testing literature, or legislators in political science). However, when studying the Supreme Court, it becomes clear that estimating these parameters will be quite difficult because only nine justices vote on each case. For each case, we must estimate $(D + 1)$ case-specific parameters using essentially a probit model with nine observations. Our experience demonstrates that reliably estimating these parameters, using either frequentist or Bayesian methods, is impossible without additional information. For remedies to this 'micro-committee problem' for cross-sectional data, see Martin and Quinn (2001a). This problem is not unique to studying the Supreme Court. Londregan (2000b) studies legislative committees in Chile, and uses information about the agenda process to solve the problem.

We are interested in the posterior density of the standard item response model, which simplifies to the following given the priors above:

$$\begin{aligned} p(\boldsymbol{\theta}, \alpha, \boldsymbol{\beta} | \mathbf{Y}) \;\; & \propto f(\mathbf{Y} | \alpha, \boldsymbol{\beta}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) p(\alpha, \boldsymbol{\beta}) \\ & \propto \textstyle\prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \left\{ \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j)^{y_{k,j}} \left[1 - \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j) \right]^{1 - y_{k,j}} \right\} \times \\ & \qquad \textstyle\prod_{k=1}^{K} p(\boldsymbol{\eta}_k) \prod_{j=1}^{J} p(\boldsymbol{\theta}_j) \end{aligned} \tag{4}$$

For $\boldsymbol{\eta}_k = (\alpha_k \;\; \boldsymbol{\beta}_k')'$. As will become more apparent below, it will be useful to reexpress Equation 4 as:

$$\begin{aligned} p(\boldsymbol{\theta}, \alpha, \boldsymbol{\beta} | \mathbf{Y}) \;\; & = \textstyle\int p(\mathbf{Z}, \boldsymbol{\theta}, \alpha, \boldsymbol{\beta} | \mathbf{Y}) d\mathbf{Z} \\ & = \textstyle\int \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \{ f_{\mathcal{N}}(z_{k,j} | \mu_{k,j}, 1) \mathbb{I}(z_{k,j}, y_{k,j}) \} \prod_{k=1}^{K} p(\boldsymbol{\eta}_k) \prod_{j=1}^{J} p(\boldsymbol{\theta}_j) d\mathbf{Z} \end{aligned}$$

---

[9] Londregan (2000a) discusses an additional identification problem relating to likelihood based inference. Due to granularity of the voting data, estimates of the ideal points and case parameters are not jointly likelihood-identified. This problem does not plague the Bayesian approach due to the imposition of a prior distribution.

Where $\mu_{k,j} = \alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j$, and $\mathbb{I}(a, b)$ is an indicator function equal to 1 when $a > 0$ and $b = 1$, or when $a < 0$ and $b = 0$, or when b is unobserved, and is equal to 0 otherwise. As we will see below, conditioning on the latent utility differences allows us to recast the item response model for a dichotomous response variable as a factor analysis model for a (latent) continuous response. The integration over $\mathbf{Z}$ is easily accomplished in the simulation methods we employ by sampling from the joint posterior distribution $p(\mathbf{Z}, \boldsymbol{\theta}, \alpha, \boldsymbol{\beta})$, and then ignoring the draws of $\mathbf{Z}$ (Albert and Chib, 1993; Johnson and Albert, 1999).

3.4 Markov chain Monte Carlo for the Constant Ideal Point Model

The Gibbs sampling algorithm for this item response model contains three conditional distributions from which we iteratively sample. All of them take standard forms. To aid in exposition, we need to define two additional quantities. First, let $\boldsymbol{\theta}_j^* = (1\ \boldsymbol{\theta}_j')'$ denote a $(D + 1) \times 1$ column vector that contains a constant followed by the ideal point estimate. Second, we define $\boldsymbol{\theta}^*$ to be the $J \times (D+1)$ matrix formed by stacking the transpose of these elements for all j. The full conditional distributions are:

1. $f(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\eta})$. For terms $t = 1, 2, \ldots, T$, cases $k \in K_t$, and justices $j \in J_k$, we simulate the latent utilities from the following distribution:

$$f(z_{k,j}|y_{k,j}, \boldsymbol{\theta}_j, \alpha_k, \boldsymbol{\beta}_k) = \begin{cases} \mathcal{N}_{[0,\infty)}(\alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j, 1) & \text{if} \quad y_{k,j} = 1 \\ \mathcal{N}_{(-\infty,0]}(\alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j, 1) & \text{if} \quad y_{k,j} = 0 \\ \mathcal{N}(\alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j, 1) & \text{if} \quad y_{k,j} \text{ is unobserved} \end{cases}$$

   $\mathcal{N}_{[a,b]}$ denotes the Gaussian distribution truncated on the interval $[a, b]$. This is the standard set-up for estimating a probit model with missing data from Albert and Chib (1993) [see also Johnson and Albert, 1999].

2. $f(\boldsymbol{\eta}|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})$. For all cases $k \in K_t$ we simulate the case-specific parameters $\boldsymbol{\eta}_k$ from a $(D+1)$-variate Normal distribution:

$$f(\boldsymbol{\eta}_k|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}) = \mathcal{N}_{D+1}(\mathbf{e}, \mathbf{E})$$

Where $\mathbf{e} = \mathbf{E}\left[\boldsymbol{\theta}^{*\prime}\mathbf{z}_{k,\cdot} + \mathbf{B}_0^{-1}\mathbf{b}_0\right]$, $\mathbf{E} = \left[\boldsymbol{\theta}^{*\prime}\boldsymbol{\theta}^{*} + \mathbf{B}_0^{-1}\right]^{-1}$, and $\mathbf{z}_{k,\cdot}$ is the column vector of latent utilities for all justices on case k.

3. $f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\eta})$. For all justices $j \in \{1, 2, \ldots, J\}$ we simulate their ideal point from a D-variate Normal distribution:

$$f(\boldsymbol{\theta}_j|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\eta}) = \mathcal{N}_D(\mathbf{t}, \mathbf{T})$$

Where $\mathbf{t} = \mathbf{T}\left[\boldsymbol{\beta}'(\boldsymbol{\alpha} - \mathbf{z}'_{\cdot,j}) + \mathbf{T}_0^{-1}\mathbf{t}_0\right]$, $\mathbf{T} = \left[\boldsymbol{\beta}'\boldsymbol{\beta} + \mathbf{T}_0\right]^{-1}$, and $\mathbf{z}_{\cdot,j}$ is a $(1 \times K)$ row vector of latent utilities of member j across all cases and terms.

One iterates a large number of times from these conditional distributions, using draws from the previous conditionals as updating values. This is a generalization of the Albert (1992) and Johnson and Albert (1999) algorithms to a D-dimensional issue space (see also Clinton et al., 2000; Jackman, 2001).[10] This algorithm requires a substantial amount of looping. In fact, for each iteration of the sampler Step 1 requires 93090 draws, Step 2 requires 3210 draws, and Step 3 requires 29 draws. We thus cannot rely on an interpreted language to estimate the model. This model, and all other models presented in this paper, are estimated in `C++` using the Scythe Statistical Library (Martin and Quinn, 2001b).

3.5 Results for the Constant Ideal Point Model

We summarize the posterior densities of the ideal points $\boldsymbol{\theta}_j$ in Table 1.[11] The first two columns of the table contain the posterior means and posterior medians for the justices. In nearly all cases these are the same. It is worth noting that the scale ranges from Douglas – a notorious liberal – on the far left (-6.513) to Clarence Thomas on the far right (3.909). *Prima facie* these results are sensible; Marshall, Warren, Brennan, and Fortas are on the left, Harlan, Burger, Rehnquist, and Scalia are on the far right. The posterior standard deviations in the third column can be thought of like standard errors in the classical setting. For the most part, these standard deviations are

---

[10] Patz and Junker (1999) develop a similar algorithm for a one-dimensional model with a logit link function.

[11] These models were run for 200,000 Gibbs scans after 5000 burn-in scans. Standard diagnostics performed on the posterior samples suggests that the chain has reached steady state.

quite small compared to the scale on which the justices are placed. Moderates are estimated with the greatest precision (such as Stewart, Clark, O'Connor, and Souter), while for extremists the estimates are somewhat less precise. This is intuitive, as it is much easier to discriminate those at the middle of the ideological spectrum. The final two columns of the table contain the 95% Bayesian credible intervals for the ideal points, again illustrating the precision which with they are estimated.

How does this measure compare to existing measures of judicial preferences? We first correlate our measure with the percent conservative voting on civil rights, civil liberties, economics, and federalism cases across a justices' career (these measures, as well as the percent liberalism scores below, are taken from Epstein et al., 2001). This is in the spirit of the methodological audit of Epstein and Mershon (1996). The results are quite interesting: the posterior means correlate with these issue areas at -0.86, -0.84, -0.73, and -0.52 respectively ($J = 29$). These clearly out-perform the scores of Segal and Cover (1989), which correlate at 0.63, 0.60, 0.43, 0.33 ($J = 29$). If one is willing to assume that ideal points are constant across time, employing these posterior means or medians as measures would be appropriate, although it would be vital to at least average over the *a posteriori* uncertainty of the measures, and ideally write down a full structural model and treat ideal points as latent data (Clinton et al., 2000). The measure also correlates highly with Schubert's (1974) C-scale (0.74) and E-scale (0.58) [$J = 17$], and Rohde and Spaeth's (1976) freedom (-0.81), equality (-0.75), and New Deal (-0.85) scales [$J = 18$].

One advantage of performing Bayesian inference is that it is straightforward to summarize the posterior density of quantities of interest other than the parameters. One quantity of great substantive importance is the location of the median justice. We began by asking: Which justice is the over-all median justice from 1953-1999? Conditional on the constant ideal point model being true, there is a 50.6% chance that White is the median justice, and a 47.1% chance that Stewart is the median justice. Five other justices have non-zero posterior probabilities of being the over-all median: Clark, Reed, Minton, Jackson, and Souter.

The real quantity of interest, though, is the term-by-term median of the Court. These are valuable not only to study the policy outputs of the Court and bargaining on the Court, but also how the Court might relate to other institutions in the separation of powers system. In Figures 1 and 2 we display the estimated posterior probability that each justice is the median justice in a given term. Since the ideal points are constant, the median will only change when there is personnel change on the Court. At the beginning of our study Minton has the highest probability of being the median justice, who is followed by Clark, who is the median justice up until 1962, when Black becomes the median. Interestingly, by the mid-1960s, Chief Justice Earl Warren is the median justice. This changes in 1970, when White becomes the median justice with probability around 55% from 1970 to 1980, and then over 90% from 1980 to 1990. Throughout the 1990's, O'Connor or Kennedy are most likely to be the median justice. If one is willing to make the assumption that preferences are indeed constant, these results should be very compelling. We question this assumption, however, and now turn to results for models that allow ideal points to change over time.

4 Independent Ideal Point Model

The model fit in the previous section presumes that each justices' ideal point $\boldsymbol{\theta}_j$ is constant throughout their career on the high bench. This is a very strict assumption, and as suggested by Epstein et al. (1998), Baum (1988), and Ulmer (1981), may be entirely inappropriate. One alternative is to assume that a justices' ideal point $\boldsymbol{\theta}_j$ are independent and identically distributed across the course of their careers. Substantively this is a bit troubling, as it in essence implies that judges have no memory about their past behavior, and that their current behavior does not in any way depend on the past. However, this independent ideal point model serves as a baseline for our final two models.

4.1 A Modified Gibbs Sampling Algorithm

This model is essentially the same as the constant ideal point model except for the fact that the $\boldsymbol{\theta}_j$ are indexed by time $\boldsymbol{\theta}_{t,j}$. Our goal is now to estimate an ideal point for each term in which a particular justice served. We again assume Normal priors for the ideal points: $\boldsymbol{\theta}_{t,j} \overset{iid}{\sim} \mathcal{N}(\mathbf{t}_{0,t}, \mathbf{T}_{0,t})$ ($\mathbf{t}_{0,t} = \mathbf{0} \ \forall \ t$ and $\mathbf{T}_{0,t} = \mathbf{I}_D \ \forall \ t$). One could just as easily employ informative Gaussian priors on the ideal points, such as having different prior means for each justice. To estimate these additional parameters we only have to change the MCMC algorithm above in some minor ways. In Step 1, one has to merely substitute $\boldsymbol{\theta}_{t,j}$ for $\boldsymbol{\theta}_j$ in the update. In Step 2, one stacks $\boldsymbol{\theta}_{t,j}$ into a term-specific $\boldsymbol{\theta}^*_t$, which are employed in the updates. To complete the sampler, Step 3 requires simulating the $\boldsymbol{\theta}_{t,j}$ term-by-term. The conditional distribution is:

$$f(\boldsymbol{\theta}_{t,j}|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\eta}) = \mathcal{N}_D(\mathbf{t}, \mathbf{T})$$

Where $\mathbf{t} = \mathbf{T}\left[\boldsymbol{\beta}'_t(\alpha_t - \mathbf{z}'_{t,\cdot,j}) + \mathbf{T}^{-1}_{0,t}\mathbf{t}_{0,t}\right]$ and $\mathbf{T} = \left[\boldsymbol{\beta}'_t\boldsymbol{\beta}_t + \mathbf{T}_{0,t}\right]^{-1}$. Note that $\mathbf{z}_{t,\cdot,j}$ contains the stacked utilities for all cases in term t for justice j, and α_t and $\boldsymbol{\beta}_t$ contain the stacked case parameters in term t.

4.2 Results from the Independent Ideal Point Model

We summarize the posterior densities of the $\boldsymbol{\theta}_{t,j}$ in Figures 3 and 4. For each justice, the thick, dark line denotes the posterior mean, and the light lines ± 2 posterior standard deviations.[12] What is immediately apparent in these figures is that the term-to-term changes for a particular justice are quite small. If it were the case that ideal points were in fact independent and identically distributed, we would expect there to be some large term-to-term fluctuations. The implication is that there is a persistent component to these ideal points that should be modeled to get better estimates, not only of the ideal points, but also case parameters and other quantities of interest.

We do see some interesting trending in these plots. Harlan seems to get more conservative toward the middle of his career, and then becomes moderate. Marshall trends more and more toward liberal throughout his career, as does Blackmun (a Nixon appointee who by the time he retired was one of the most liberal justices on the high bench). Stevens and Souter also look to become more liberal after 1990, after President Bush bolstered the conservative wing with the

---

[12] This model was run for 200,000 scans after 5000 burn-in scans. Standard convergence tests suggest that the chain has reached steady state.

appointment of Clarence Thomas. Consistent with the findings of Ulmer (1973) and Atkins and Sloope (1986), Black seems to trend toward conservatism. Douglas, however, does not appear to become more liberal over time (c.f., Ulmer, 1981). These figures suggest that for at least some justices, ideal points change substantially over time. However, the statistical assumptions on which this model is based are faulty, and we must turn to a modeling strategy that allows for persistence.

5 Dynamic Ideal Point Model

Assuming that each a justice's ideal point at time t is independent of her ideal point at $t-1$ is clearly problematic. And, as demonstrated in the previous section, leads to rather unsatisfactory estimates of the ideal points of the justices. A much better substantive assumption is that the ideal points follow a random walk process. This strikes a balance between the overly strong assumptions of both the constant ideal point model and the independent ideal point model, and allows for a wide range of temporal patterns. In this section, we propose such a model that uses a state-space representation to model the temporal evolution of the ideal points. We employ the forward filtering, backward sampling algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994) [also discussed in West and Harrison (1997), Section 15.2.3] to estimate the ideal points.[13]

This model begins with the same latent utility specification:[14]

$$z_{t,k,j} = \alpha_k + \beta_k' \theta_{t,j} + \varepsilon_{t,k,j} \qquad \varepsilon_{t,k,j} \overset{iid}{\sim} \mathcal{N}(0,1)$$

We adopt the standard multivariate normal prior over the case-specific parameters:

$$\begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I}) \;\; \forall \; k \in \{1, 2, \ldots, K\}$$

The model differs in the structure of the prior on $\theta_{t,j}$. For time period zero (which is unobserved), we assume that:

$$\theta_{0,j} \sim \mathcal{N}(s_{0,j}, S_{0,j})$$

---

[13] For an excellent introduction to time series and dynamic modeling in the Bayesian context we recommend West and Harrison (1997). Chapters 2-4 contain a discussion of dynamic linear models (DLMs). Chapter 15 presents MCMC algorithms for DLMs. Our exposition here follows their account closely.

[14] Here we restrict ourselves to the $D = 1$ case for concreteness. Since the usual dynamic linear model DLM results apply equally to scalar and vector state variables (West and Harrison, 1997), it is easy to re-specify and fit the following model for the $D > 1$ case.

We model the dynamics of the ideal points with a separate random walk prior for each justice:

$$\theta_{t,j} \sim \mathcal{N}(\theta_{t-1,j}, \sigma^2_{\theta_{t,j}}) \quad \text{for } t = \underline{T}_j, \ldots, \overline{T}_j \text{ and justice } j \text{ on the Court}$$

$\underline{T}_j$ 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. $s_{j,0}$ 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. $S_{0,j}$ was set to 1 for all justices except the aforementioned nine. Their prior variances at time zero were set to 0.1. These informative priors were used to identify the model, thus ameliorating both the scale and rotational invariance problems. $\sigma^2_{\theta_{t,j}}$ was 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 $\sigma^2_{\theta_{t,j}} = 0.001$ to help identify his sequence of ideal points.

5.1 Estimation of Dynamic Ideal Points via MCMC

Estimating this model is also a straightforward application of the Gibbs sampling algorithm. The first two conditional distributions are precisely the same as those for the independent ideal point model. The only update that changes is that for the $\theta_{t,j}$. Here we employ a state-space representation and use the forward filtering, backward sampling algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994) to sample from $f(\boldsymbol{\theta}_{\cdot,j}|\alpha, \boldsymbol{\beta}, \mathbf{Z})$, where $\boldsymbol{\theta}_{\cdot,j}$ now represents the vector of j's ideal points over his/her career (recall that we are estimating a one-dimensional dynamic model). Let $\boldsymbol{\Theta}_j = \{\boldsymbol{\theta}_{0,j}, \boldsymbol{\theta}_{\underline{T}_j,j}, \boldsymbol{\theta}_{\underline{T}_j+1,j}, \ldots, \boldsymbol{\theta}_{\overline{T}_j,j}\}$. Further, let D_t denote all information available at time t. Given the assumed Markov dependence in the prior for $\boldsymbol{\theta}_{\cdot,j}$ is easy to show that the full conditional posterior density for $\boldsymbol{\Theta}_j$ given D_t can be decomposed as follows:

$$p(\boldsymbol{\Theta}_j|D_t) = p(\boldsymbol{\theta}_{\overline{T}_j,j}|D_{\overline{T}_j})p(\boldsymbol{\theta}_{\overline{T}_j-1,j}|\boldsymbol{\theta}_{\overline{T}_j,j}, D_{\overline{T}_j-1}) \cdots p(\boldsymbol{\theta}_{\underline{T}_j,j}|\boldsymbol{\theta}_{\underline{T}_j+1,j}, D_{\underline{T}_j})p(\boldsymbol{\theta}_{0,j}|\boldsymbol{\theta}_{\underline{T}_j,j}, D_0)$$

This forms the basis of the forward filtering, backward sampling sampling algorithm. In this context, this amounts to a three step process for each justice.

1. For $t = \underline{T}_j, \ldots, \overline{T}_j$ calculate the quantities a_t, R_t, s_t, S_t.

2. Sample $\theta_{\overline{T}_j,j}$ from $\mathcal{N}(s_{\overline{T}_j,j}, S_{\overline{T}_j,j})$.

3. For $t = (\overline{T}_j - 1), (\overline{T}_j - 2), \ldots, \underline{T}_j$ sample $\theta_{t,j}$ from $\mathcal{N}(h_t, H_t)$.

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

$$\begin{aligned} a_t &= s_{t-1} \\ R_t &= S_{t-1} + \sigma^2_{\theta_{t,j}} \end{aligned}$$

Next is s_t, which is the posterior mean of $\theta_{t,j}$ given the information available at time t (not the marginal mean given all the data), and S_t, which is the posterior variance of $\theta_{t,j}$ given the information available at time t (not the marginal variance given all the data). These quantities are defined as:

$$\begin{aligned} s_t &= a_t + \mathbf{A}_t \mathbf{e}_t \\ S_t &= R_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t' \end{aligned}$$

Where $\mathbf{A}_t = R_t \boldsymbol{\beta}_k' \mathbf{Q}_t^{-1}$, $\mathbf{e}_t = \mathbf{z}_{t,\cdot,j} - \boldsymbol{\alpha}_k - \boldsymbol{\beta}_k a_t$, and $\mathbf{Q}_t = \boldsymbol{\beta}_k R_t \boldsymbol{\beta}_k' + \mathbf{I}$. Step 2 of the algorithm is just a draw using the quantities above.

Step 3 of the algorithm is the backward sampling part of the algorithm. Using the quantities computed above, we sample down from $t = \overline{T}_j - 1$ to $t = \underline{T}_j$, backwards through the justices' career. The mean and variance for the Normal update is:

$$\begin{aligned} h_t &= s_t + B_t(\theta_{t+1,j} - a_{t+1}) \\ H_t &= S_t - B_t^2 R_{t+1} \end{aligned}$$

Where $B_t = S_t R_{t+1}^{-1}$, the the posterior variance of $\theta_{t,j}$ given the information available at time t divided by its prior variance given the information available at time $t-1$. The benefit of this algorithm over more direct Gibbs sampling approaches is that it allows us to sample $\mathbf{\Theta}_j$ directly in one piece rather than sampling component by component. Given the generally high correlations between the components of $\mathbf{\Theta}_j$ (which is why we are fitting this dynamic model to begin with), component by component approaches mix very slowly. This algorithm is even more computationally intensive than that in the previous section, as for each scan of the sampler we use this forward filtering, backward sampling approach for each justice.

5.2 Results from the Dynamic Ideal Point Model

We now turn to the results for the dynamic ideal point model.[15] In Figures 5 and 6 we summarize the posterior densities of the justices' ideal points throughout their careers. Again the thick, dark line represents the posterior mean, and the light lines are ± 2 posterior standard deviations away from the posterior mean. The most striking difference between these results and those for the independent ideal point model is that the trajectories are smoother. This, of course, is a result of the random walk process used to model the ideal points. Again we see some trending. Harlan again follows a parabolic trajectory. Black, Scalia, and Thomas trend towards conservatism, while Marshall, Brennan, Blackmun, Stevens, and Souter trend towards liberalism. Again Blackmun is the most striking example, beginning well right of center and retiring well left of center. It seems that Epstein et al. (1998) are correct in their conclusion that some justices are moving over time, although these results suggest that the number of justices with non-constant ideal points is not nearly as great as they conclude.

How does this measure fare against the Epstein and Mershon (1996) validity test? In short, very well. We present the term-by-term correlations between the posterior mean ideal points from the dynamic model with percent conservative decisions in Figure 9. To make the findings

---

[15] This model was run for 100,000 iterations after 5000 burn-in iterations. Standard convergence tests suggest that the chain has reached steady state.

consistent with Epstein and Mershon (1996), the term-by-term measures are correlated with the percent conservative decisions for the entire justices' career. This is quite sensible, as term-by-term percentages are often based on a very small number of cases. Looking across the cells it is clear that the model does exceedingly well, particularly for civil liberties and civil rights case. It also does quite well in economics cases, which is the major weakness of the Segal and Cover scores (1989). It also does well in federalism cases except for a few terms around 1970, and throughout the 1990s. This is suggestive of perhaps a second issue dimension that structures decision making in federalism cases during this time period. Not coincidentally, these are both time periods when "new federalism" was being considered in the executive and legislative branches respectively. When comparing these correlations with those from the constant model, it is clear that the dynamic model does a superior job.

For the dynamic ideal point model we also explore the identity of the median justice in each term. These posterior probabilities are presented in Figures 7 and 8. Frankfurter is the median justice in 1953, followed by Clark, then Stewart, then Goldberg, then Black. Brennan is likely the median in a handful of terms in the 1960s. From the late 1960s to the early 1990s, White is the median justice in a good number of terms, although Stewart, Blackmun, and Powell are medians in certain terms. Finally, in the 1990s, O'Connor and Kennedy again appear to be the median justice, except for the 1989 term when it is likely Souter.

The identity of the median justice is interesting, but so too is the location of the median on the estimated issue dimension. In Figure 10 we summarize the posterior density of the location of the median justice in each term. During the early Warren Courts, until 1961 when White joined the Court, the median is quite conservative. This changed in the 1962 term, with the addition of Goldberg, and later on in the decade Fortas and Marshall. There is a break from liberalism in the 1969 term, when Nixon nominated Burger to replace Warren. With that nomination, and then the nomination of Blackmun in 1970, and Powell and Rehnquist in in 1971, the Court became quite conservative. It trended toward a more moderate position, but that changed with the O'Connor appointment by President Reagan. The Court became more conservative during the late Reagan years, grew somewhat more moderate during the Bush administration, and then became somewhat more conservative during the Clinton years. The most conservative Court sat in the 1988 term. The most liberal sat in 1967. On face, these results are quite compelling. Perhaps it is the case, though, the agenda changes over time account not only for the trends in justices' ideal points, but also in the location of the median justice. In the following section we model the case parameters over time to not only study the content of the cases reaching the Court, but also to see if controlling for them impacts the inferences we draw about the ideal points of the justices.

6 Dynamic Ideal Point and Case Parameter Model

The results from the previous section are quite interesting. Many justices seem to be trending over time. But perhaps this is an artifact of the types of cases that are being heard by the Court rather than true preference change. The results above are conditioned on the fact that the α_k and $\boldsymbol{\beta}_k$ case parameters are *a priori* independent and identically distributed from standard Normal distributions. Just as this assumption was problematic for the ideal points, it is also troublesome for the case parameters. Additionally, we have information available to us – the federal circuit from which the case originated – that we can employ to better model these parameters.

Recall that the utility difference between a reversal vote and an affirmative vote for justice j on case k in term t can be written as:

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

While the $x^{(r)}_{t,k}$ and the $x^{(a)}_{t,k}$ parameters are not identified in the likelihood, they can be identified in the posterior if informative priors are used for these parameters as well as the ideal points. This is the approach we employ here. As in the previous section, we assume independent random walk priors for the ideal points:

$$\theta_{t,j} \sim \mathcal{N}(\theta_{t-1,j}, \sigma^2_{\theta_{t,j}})$$

Where $\sigma^2_{\theta_{t,j}}$ is assumed known for all j, t. Just as before, we assume $\theta_{0,j} \sim \mathcal{N}(s_{0,j}, S_{0,j})$

Instead of assuming priors over the case-specific parameters (α_k and $\boldsymbol{\beta}_k$), we now parameterize the prior in terms of the affirm and reversal points. For the policy positions under affirmance, we assume:

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

And:

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

Where $\mathbf{x}_{t,\cdot}^{(a)}$ is the vector of stacked $x_{t,k}^{(a)}$ $\forall k \in K_t$, and $\mathbf{C}_t$ is a matrix of dummy variables indicating the originating lower court for each case in term t.[16] $\boldsymbol{\gamma}_t$ is a coefficient vector that estimates the mean position of the cases originating from each court of origin, $\sigma^2_{x^{(a)}}$ is the prior variance of each $x_{t,k}^{(a)}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_t}$ is the prior variance covariance matrix of $\boldsymbol{\gamma}_t$. We assume that $\sigma^2_{x^{(a)}}$ 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.

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):

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

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 is saying substantively is that if the opinion writer is known, the reversal opinion likely resides near her ideal point. When the majority of the Court reverses an opinion of a lower Court, Spaeth (1999) codes the majority opinion writer. 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 $\sigma^2_{x^{(r)}}$.

---

[16] We cull these covariates from the *Supreme Court Database*. We assume that political geography matters, and thus cases the originate in lower federal courts or state courts in the twelve federal circuits are similar. This is a reasonable assumption: cases from lower courts in Texas or the Texas Supreme Court are likely to look like those from the Appeals Court of the Fifth Circuit. We thus employ the SOURCE variable from the Supreme Court Data Base, and use it to create thirteen dummy variables: one for each of the eleven circuits (note that the Eleventh Circuit was established in October, 1981 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 (there is no constant term).

6.1 Conditional Distributions

Conditional on the latent utility differences $\mathbf{Z}$, we can write the posterior density for this model as:

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

Where $\mathbf{x}^{(a)}$ and $\mathbf{x}^{(r)}$ denote the stacked affirmance and reversal positions for all k respectively. The full conditional densities are:

$$\begin{aligned}
p(\mathbf{z}_{t,k,j}|\boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}, \boldsymbol{\gamma}) &= \begin{cases} \mathcal{N}_{[0,\infty)}(\mu_{t,k,j}, 1) & \text{if} \quad y_{k,j,t} = 1 \\ \mathcal{N}_{(-\infty,0]}(\mu_{t,k,j}, 1) & \text{if} \quad y_{k,j,t} = 0 \\ \mathcal{N}(\mu_{t,k,j}, 1) & \text{if} \quad y_{k,j,t} \text{ is unobserved} \end{cases} & (5) \\
p(\boldsymbol{\gamma}|\mathbf{Z}, \boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}) &\propto p(\mathbf{x}^{(a)}|\boldsymbol{\gamma}) p(\boldsymbol{\gamma}) & (6) \\
p(\boldsymbol{\theta}|\mathbf{Z}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}) &\propto f(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}) p(\mathbf{x}^{(r)}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) & (7) \\
p(\mathbf{x}^{(a)}, \mathbf{x}^{(r)}|\mathbf{Z}, \boldsymbol{\gamma}, \boldsymbol{\theta}) &\propto f(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)}) p(\mathbf{x}^{(a)}|\boldsymbol{\gamma}) p(\mathbf{x}^{(r)}|\boldsymbol{\theta}) & (8)
\end{aligned}$$

Where the mean

$$\mu_{t,k,j} = \left[x^{(a)}_{t,k} x^{(a)}_{t,k} - x^{(r)}_{t,k} x^{(r)}_{t,k}\right] + 2\boldsymbol{\theta}'_{t,j} \left[x^{(r)}_{t,k} - x^{(a)}_{t,k}\right]$$

is now defined as a function of the reversal and affirmance points instead of the case-specific parameters. The densities in Equation 5 are all easily simulated from (see the baseline constant ideal point model). The density in Equation 6 is the product of a Gaussian density and a random walk prior, and so can be simulated from using standard methods for dynamic linear models (now with covariates), as discussed in the previous section (West and Harrison, 1997). The densities in Equations 7 and 8 are nonstandard, and are best simulated from using Metropolis-Hastings steps.

6.2 Sampling Form $p(\boldsymbol{\theta}|\mathbf{Z}, \boldsymbol{\gamma}, \mathbf{x}^{(a)}, \mathbf{x}^{(r)})$

The full conditional for $\boldsymbol{\theta}_{\cdot,j}$ is similar to the corresponding full conditional for the dynamic ideal point model discussed in the previous section, except that now the term $p(\mathbf{x}^{(r)}|\boldsymbol{\theta})$ appears. This additional term makes the resulting full conditional non-standard. As a result, the forward filtering, backward sampling algorithm for sampling $\boldsymbol{\theta}$ no longer applies directly. Instead, we use a independent Metropolis-Hastings step [see Chib and Greenberg (1995), Robert and Casella (1999)].

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:

$$p(\boldsymbol{\theta}_{\cdot,j}|\mathbf{z}_{\cdot,j},\boldsymbol{\gamma},\mathbf{x}^{(a)},\mathbf{x}^{(r)}) \propto f(\mathbf{z}_{\cdot,j}|\boldsymbol{\theta}_{\cdot,j},\mathbf{x}^{(a)},\mathbf{x}^{(r)})p(\mathbf{x}^{(r)}_j|\boldsymbol{\theta}_{\cdot,j})p(\boldsymbol{\theta}_{\cdot,j}), \tag{9}$$

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

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

$$\min\left\{\frac{p(\boldsymbol{\theta}^{(can)}_j|\mathbf{z}_{\cdot,j},\boldsymbol{\gamma},\mathbf{x}^{(a)},\mathbf{x}^{(r)})}{p(\boldsymbol{\theta}^{(cur)}_j|\mathbf{z}_{\cdot,j},\boldsymbol{\gamma},\mathbf{x}^{(a)},\mathbf{x}^{(r)})}\frac{q(\boldsymbol{\theta}^{(cur)}_j)}{q(\boldsymbol{\theta}^{(can)}_j)},\ 1\right\} \tag{10}$$

Our choice of candidate generating density is the following:

$$q(\boldsymbol{\theta}^{(can)}_j) \propto f(\mathbf{z}_j|\boldsymbol{\theta}^{(can)}_j,\mathbf{x}^{(a)},\mathbf{x}^{(r)})p(\boldsymbol{\theta}^{(can)}_j) \tag{11}$$

Note that a candidate value can be sampled from this density using exactly the same forward filtering, backward sampling algorithm used in dynamic ideal point model in the previous section. Substituting this choice of $q(\cdot)$ into Equation 10 reveals that the acceptance probability becomes:

$$\min\left\{\frac{p(\mathbf{x}^{(r)}_j|\boldsymbol{\theta}^{(can)}_j)}{p(\mathbf{x}^{(r)}_j|\boldsymbol{\theta}^{(cur)}_j)},\ 1\right\} \tag{12}$$

6.3 Sampling From $p(\mathbf{x}^{(a)},\mathbf{x}^{(r)}|\mathbf{Z},\boldsymbol{\gamma},\boldsymbol{\theta})$

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

$$p(x^{(a)}_{t,k}, x^{(r)}_{t,k}|\mathbf{z}_{t,k},\boldsymbol{\gamma},\boldsymbol{\theta}_{t,\cdot}) \propto \prod_{j=1}^{J_t} f_{\mathcal{N}}(z_{t,k,j}|\mu_{t,k,j},1)p(\mathbf{x}^{(a)}|\boldsymbol{\gamma})p(\mathbf{x}^{(r)}|\boldsymbol{\theta}_{t,\cdot})$$

Where $\mu_{t,k,j}$ is define above, and $\boldsymbol{\theta}_{t,\cdot}$ denotes the stacked $\boldsymbol{\theta}_{t,j}$ for all justices j. With the symmetric candidate generating densities the Metropolis acceptance probability is:

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

To estimate the full dynamic ideal point and case parameter model, we iteratively sample from these three blocks along with the block for the latent utilities.

6.4 Results from the Dynamic Ideal Point and Case Parameter Model

We report the estimated ideal points for the justices in Figures 11 and 12.[17] The prior probabilities for the $\boldsymbol{\theta}_{t,j}$ are the same as those for the dynamic ideal point model above. We adopt an informative prior over the mean locations of the circuits $\boldsymbol{\gamma}_0$ at time $t = 0$. We base our priors on a conversation with an expert on the federal appeals courts: Steve Van Winkle (SUNY Stony Brook). The most liberal circuits are 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 $t = 0$. Based on our conversation with the expert, 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 is allowing 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.[18]

---

[17] This model was run for 200,000 iterations after 20,000 burn-in iterations. Standard convergence tests suggest that the chain has reached steady state. This took approximately six days on a dedicated Pentium 933 Linux workstation using compiled and optimized `C++` code.

[18] Due to time constraints we have not been able to adequately assess prior sensitivity for this model. We have, however, fit models in which the prior distribution of $x_k^{(r)}$ is uniform between $x_k^{(a)}$ and the ideal point of the justice

The results are quite striking. Again, many justices seem to trend over time. Just as with the dynamic model discussed above, Harlan demonstrates a parabolic trajectory, Black trends to conservatism, as does Frankfurter and Scalia. Many justices trend toward liberalism in both models: Marshall, Brennan, Blackmun, Stevens, Souter, and Ginsburg. For the most part, these dynamics mirror those of the dynamic ideal point model. This suggests that after controlling for case stimuli, many the ideal points of justices do in fact trend over time. How do these results compare to Epstein et al. (1998)? Of the seven justices they assert have constant preferences over time, our results differ for Brennan, Harlan, Marshall, and Stewart. Of the remaining nine that they conclude trend over time, our results differ for Clark, White, and Warren. It seems clear that without modeling case-specific parameters and controlling for case stimuli, it is quite difficult to ascertain whether or not preferences change over time. Nonetheless, our major substantive conclusion agrees with that of Epstein et al. (1998): preferences of many Supreme Court justices do in fact change over time.

It is interesting to look at the position of the median justice along the estimated dimension from 1953 to 1999 (the posterior probabilities that a particular justice is the median is substantively quite similar to that in the previous section). We plot the posterior densities of the median justice in Figure 13. 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 control for temporal changes in the inputs to the system.

It is also of interest to know where the lower federal courts are setting the status quo point for

---

who wrote $x_k^{(r)}$. 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.

the Supreme Court; i.e., the policy reversion if the Court votes to affirm. We plot these posterior densities in Figure 14. Again, the thick, dark line denotes the posterior mean, and the light lines are ± 2 posterior standard deviations away from the posterior mean. 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

The results from these four models are quite informative, and have taught us a great deal about the Supreme Court. Some of the findings support those of previous research. Others are substantively more unique. The most important substantive conclusion to draw is that the ideal points of Supreme Court justices do change over time. This is not a universal phenomenon, but it is certainly the case that the preferences of some justices change over time *even when controlling for the types of cases that come before the Court.* This implies that a constant measure of judicial preferences – such as the measure of Segal and Cover (1989) – is not appropriate for explaining longitudinal judicial decision making. Our findings also demonstrate that the final conclusion of Epstein et al. (1998) that judicial preferences change is correct, but our results demonstrate that the identity of the changing justices and the temporal patterns that they exhibit are very different. In addition to presenting posterior standard deviations (standard errors) for all measures, we also learned the identity and the location of the median justice on the Court during each term. Finally, we learned about the types of decisions that were selected out of each of the federal circuits by the Justices on the Supreme Court.

We have also learned a great deal about the modeling strategies, in particular for dynamical systems. First, it is clearly important to account for persistence, as the results of both dynamic models illustrate. At a minimum, to understand how any social system works, it is vital to control for different inputs to the system. Since the Supreme Court is an appellate court, it is necessary to control for the cases coming before it. Second, we have shown two advantages of adopting a Bayesian approach: (a) we were able to incorporate prior information to not only identify a statistical model, but also to bring additional information, such as the ideological predispositions of the federal appeals courts, to bear on the problem of estimating judicial ideal points; (b) we were able to estimate complex and flexible models using Markov chain Monte Carlo methods. Finally, we have demonstrated the benefits of using a compiled language, such as `C++`, to estimate computationally intensive, yet substantively meaningful models.

This research is only the first step of a research agenda. Others might be tempted to use our estimates as independent or dependent variables in other analyses. The better approach, which we leave for future research, is to write down a full structural decision making model, treating the ideal points as missing data. Such an approach is straightforward in a Bayesian context, and conditional on the modeling assumptions being true, allows for one to test various theories of decision making *while also estimating the ideal points of the actor under study.* Nonetheless, we have provided estimates that we hope others will find useful in their research. Another important step is explaining the causes of the dynamics we observe. Why do justices' ideal points change over time? Is this some psychological process, or do revealed preferences change with different compositions of the Court, or perhaps because of separation of powers concerns? 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 can incorporate suitable covariates in the statistical models to test for these explanations, or perhaps employ existing data about decision making in the Federal 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 publicly available.

References

Albert, James H. 1992. "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling." *Journal of Educational Statistics* 17(Fall):251–269.

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.

Atkins, Burton M., and Terry Sloope. 1986. "The 'New' Hugo Black and the Warren Court." *Polity* 18:621–637.

Bailey, Michael, and Kelly Chang. 2001. "Comparing Presidents, Senators, and Justices: Interinstitutional Preference Estimation." *Journal of Law, Economics, and Organization,* forthcoming .

Baum, Lawrence. 1988. "Measure Policy Change in the United States Supreme Court." *American Political Science Review* 82(September):905–912.

Baum, Lawrence. 1992. "Membership Change and Collective Voting Change in the United States Supreme Court." *Journal of Politics* 54(February):3–24.

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

Chib, Siddhartha, and Edward Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician* 49(November):327–336.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2000. "The Statistical Analysis of Legislative Behavior: A Unified Approach." Paper presented at the Annual Meeting of the Political Methodology Society.

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.

Epstein, Lee, Valerie Hoekstra, Jeffrey A. Segal, and Harold J. Spaeth. 1998. "Do Political Preferences Change: A Longitudinal Study of U.S. Supreme Court Justices." *Journal of Politics* 60(Augut):801–818.

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.

Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40(February):261–294.

Epstein, Lee, Jeffrey A. Segal, Harold J. Spaeth, and Thomas G. Walker. 2001. *The Supreme Court Compendium: Data, Decisions, and Developments, Second Edition*. Washington, DC: CQ Press.

Epstein, Lee, Thomas G. Walker, and William J. Dixon. 1989. "The Supreme Court and Criminal Justice Disputes: A Neo-Institutional Perspective." *American Journal of Political Science* 33(November):825–841.

Eskridge, William N. 1991. "Overriding Supreme Court Statutory Interpretation Decisions." *Yale Law Journal* 101(November):331–417.

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,* forthcoming .

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

Londregan, John. 2000a. "Estimating Legislators' Preferred Points." *Political Analysis* 8(Winter):35–56.

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

Martin, Andrew D., and Kevin M. Quinn. 2001a. "Estimating Latent Structures of Voting for Micro-Committees, with Application to the U.S. Supreme Court." Paper presented at the Annual Meeting of the Midwest Political Science Association.

Martin, Andrew D., and Kevin M. Quinn. 2001b. "Scythe Statistical Library, Release 0.1." `http://scythe.wustl.edu/`.

Patz, Richard J., and Brian W. Junker. 1999. "A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models." *Journal of Educational and Behavioral Statistics* 24(Summer):146–178.

Poole, Keith T., and Howard Rosenthal. 1991a. "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.

Robert, Christian P., and George Casella. 1999. *Monte Carlo Statistical Methods*. New York: Springer.

Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco: W. H. Freeman.

Schubert, Glendon. 1959. *Quantitative Analysis of Judicial Behavior*. Glencoe, IL: Free Press.

Schubert, Glendon. 1965. *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963*. Evanston: Northwestern University Press.

Schubert, Glendon. 1974. *The Judicial Mind Revisited: Psychometric Analysis of Supreme Court Ideology*. London: Oxford University Press.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83(June):557–565.

Segal, Jeffrey A., Lee Epstein, Charles M. Cameron, and Harold J. Spaeth. 1995. "Ideological Values and the Votes of U.S. Supreme Court Justices Revisited." *Journal of Politics* 57(August):812–823.

Segal, Jeffrey A., and Harold J. Spaeth. 1989. "Decisional Trends on the Warren and Burger Courts: Results from the Supreme Court Data Base Project." *Judicature* 73(June-July):103–107.

Segal, Jeffrey A., and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge: Cambridge University Press.

Spaeth, Harold J. 1990. "The Dimensions of Civil Liberties Decision Making on the Burger Court." In *Studies in U. S. Supreme Court Behavior* (Harold J. Spaeth, and Saul Brenner, editors), New York: Garland, pp. 184–194.

Spaeth, Harold J. 1999. *United States Supreme Court Judicial Database, 1953-1998 Terms [Computer File]*. Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 15th edition.

Tate, C. Neal, and Roger Handberg. 1991. "Time Binding and Theory Building in Personal Attribute Models of Supreme Court Voting Behavior, 1916-88." *American Journal of Political Science* 35(May):460–480.

Ulmer, S. Sidney. 1973. "The Longitudinal Behavior of Hugo Lafayette Black: Parabolic Support for Civil Liberties, 1937-1971." *Florida State University Law Review* 1.

Ulmer, S. Sidney. 1981. "Parabolic Support of Civil Liberties Claims: The Case of William O. Douglas." In *Courts, Law, and Judicial Processes* (S. Sidney Ulmer, editor), New York: Free Press.

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

Justice	Post. Mean	Post. Median	Post. Std.	BCI 2.5%	BCI 97.5%
Harlan	1.545	1.540	0.124	1.315	1.809
Black	-1.304	-1.303	0.101	-1.510	-1.109
Douglas	-6.513	-6.504	0.593	-7.691	-5.363
Stewart	0.403	0.402	0.053	0.299	0.511
Marshall	-2.002	-1.998	0.113	-2.231	-1.784
Brennan	-1.643	-1.640	0.077	-1.797	-1.496
White	0.407	0.407	0.044	0.322	0.496
Warren	-1.478	-1.476	0.109	-1.696	-1.277
Clark	0.160	0.159	0.068	0.026	0.295
Frankfurter	1.112	1.108	0.120	0.879	1.354
Whittaker	1.078	1.073	0.143	0.806	1.374
Burton	1.183	1.178	0.165	0.879	1.517
Reed	1.140	1.129	0.213	0.741	1.586
Fortas	-1.609	-1.597	0.248	-2.134	-1.151
Goldberg	-1.011	-1.005	0.194	-1.414	-0.656
Minton	0.987	0.979	0.225	0.556	1.454
Jackson	1.132	1.090	0.494	0.276	2.308
Burger	1.468	1.466	0.097	1.287	1.657
Blackmun	-0.073	-0.073	0.043	-0.157	0.009
Powell	0.809	0.807	0.067	0.679	0.943
Rehnquist	2.914	2.906	0.176	2.582	3.277
Stevens	-0.553	-0.552	0.052	-0.658	-0.453
O'Connor	1.309	1.309	0.085	1.145	1.475
Scalia	2.433	2.422	0.200	2.071	2.848
Kennedy	1.293	1.289	0.104	1.103	1.507
Souter	0.209	0.209	0.083	0.048	0.367
Thomas	3.909	3.888	0.432	3.105	4.781
Ginsburg	-0.227	-0.225	0.109	-0.444	-0.016
Breyer	-0.180	-0.179	0.117	-0.419	0.044

Table 1: Posterior density summary of ideal points of U.S. Supreme Court Justices, 1953-1999, for the constant ideal point model.

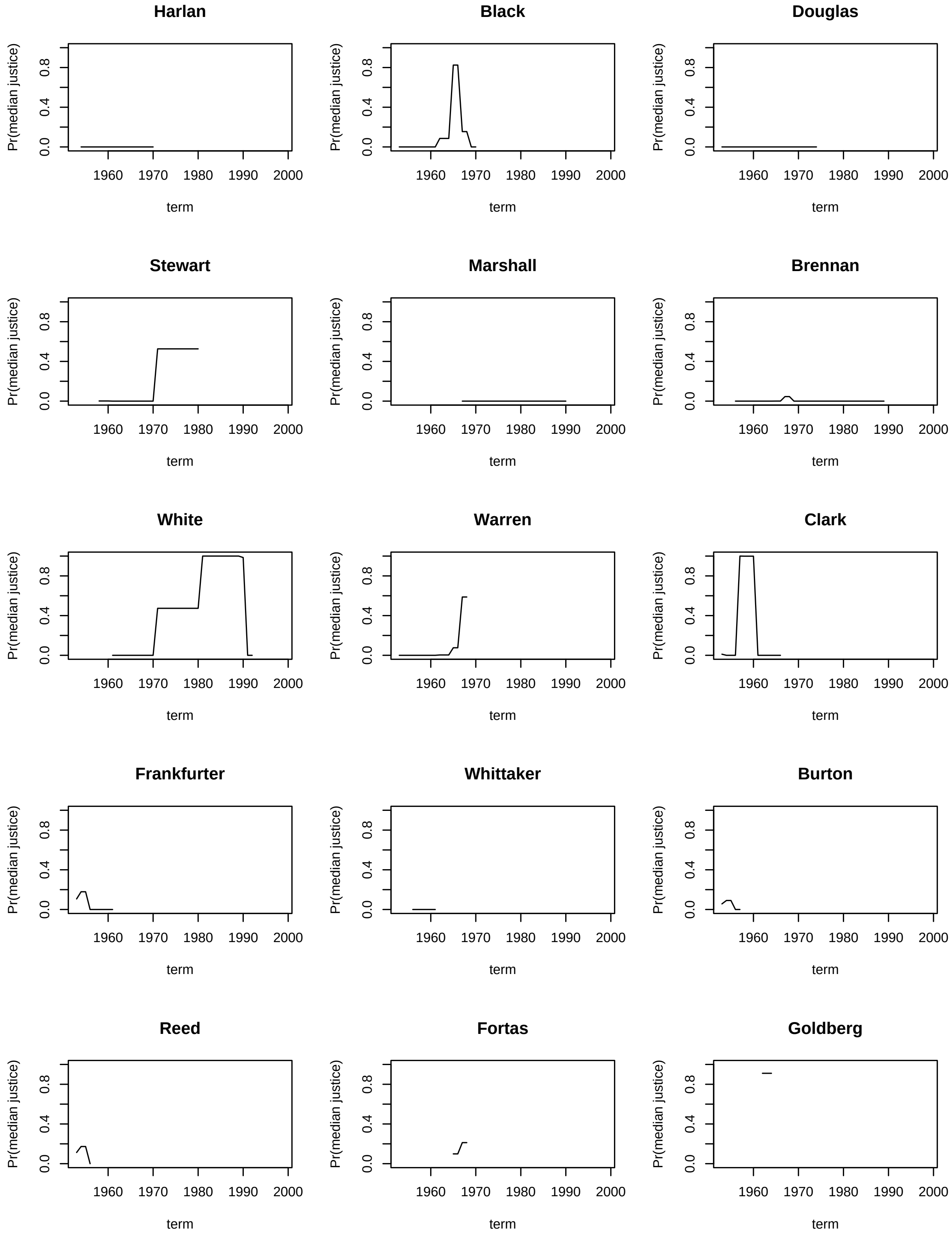


Figure 1: Estimated posterior probabilities that each justice in a given term is the median justice on the Court for the constant ideal point model, Justices Harlan-Goldberg.

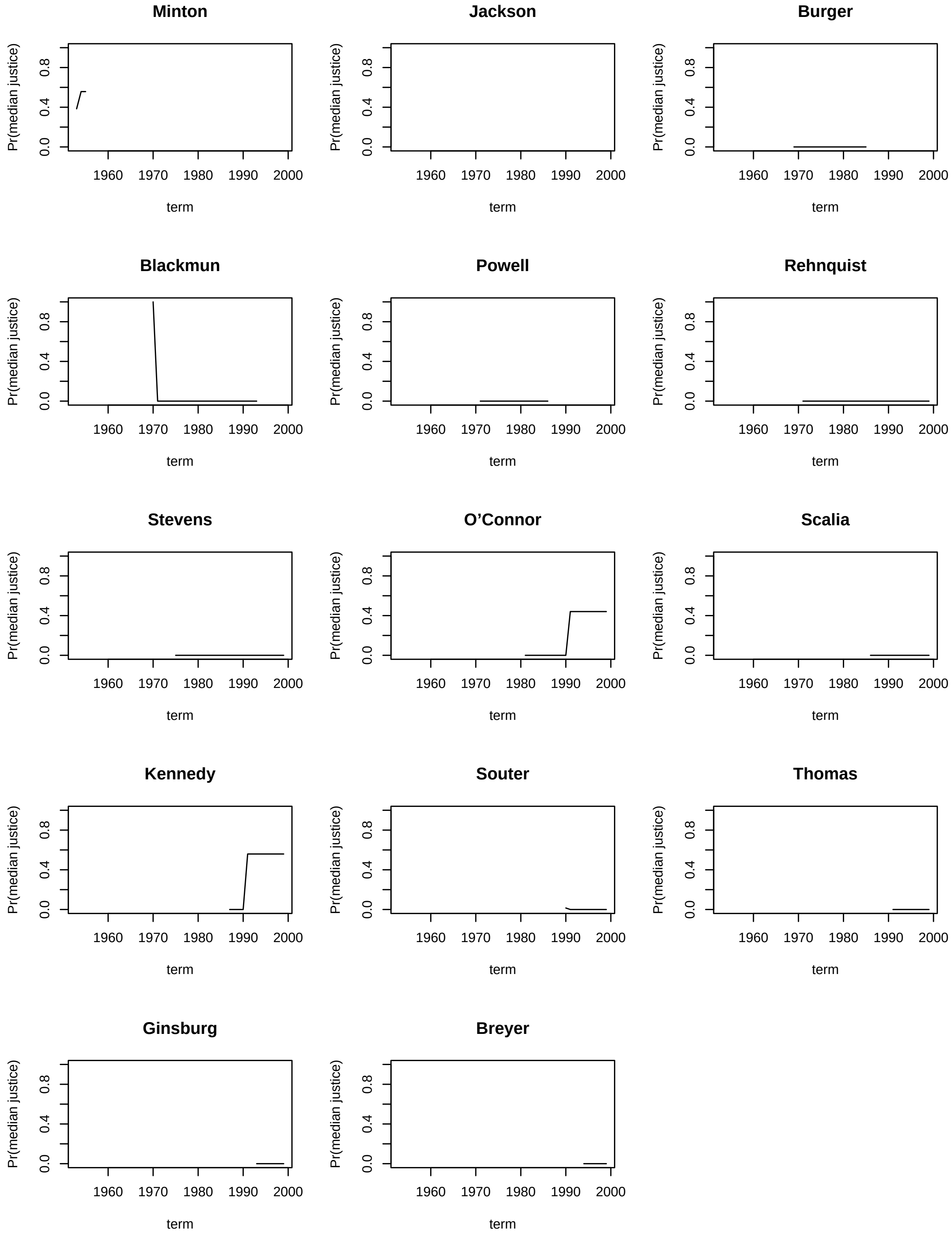


Figure 2: Estimated posterior probabilities that each justice in a given term is the median justice on the Court for the constant ideal point model, Justices Minton-Breyer.

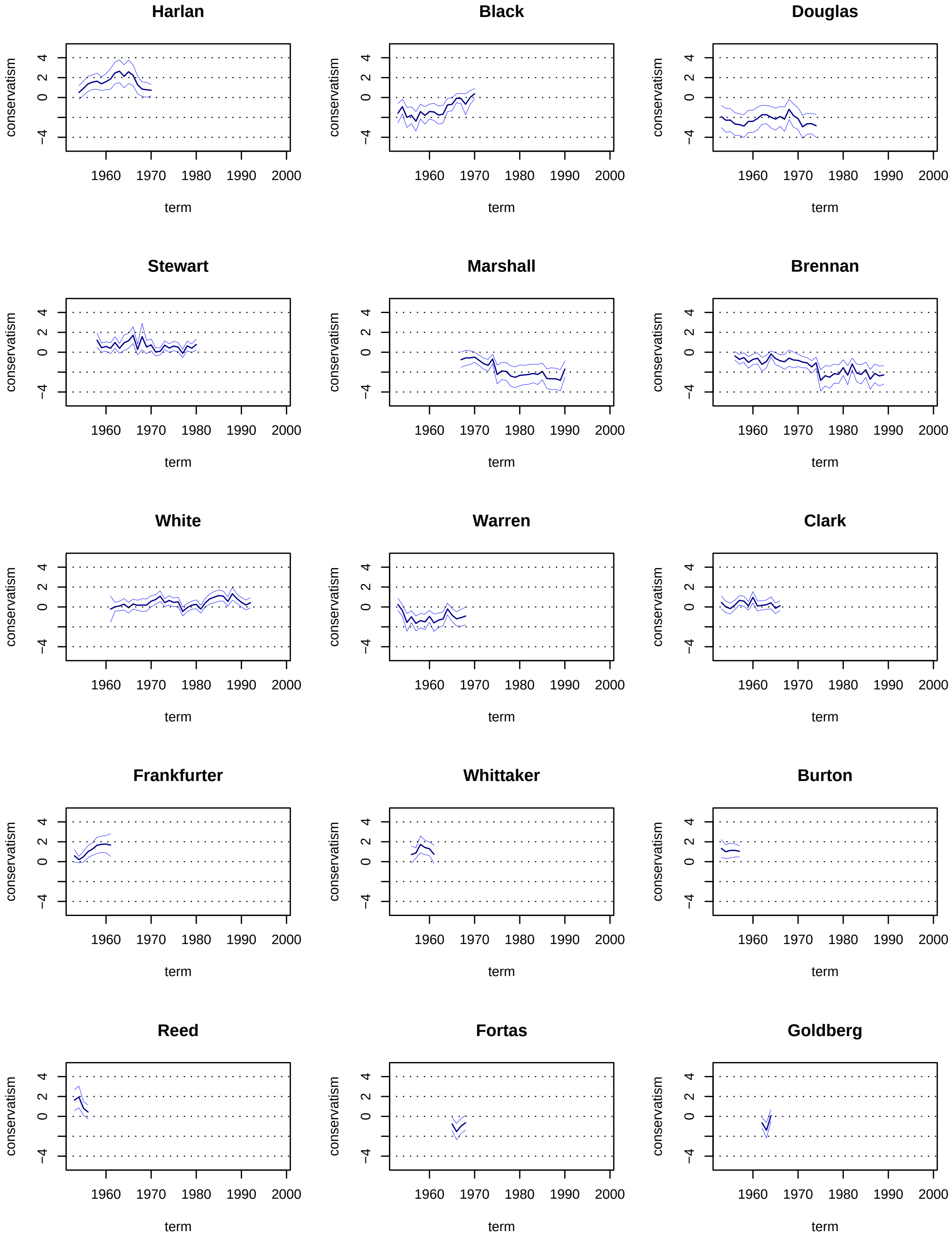


Figure 3: Posterior density summary of the ideal points of each justice for the terms in which they served for the independent ideal point model, Justices Harlan-Goldberg.

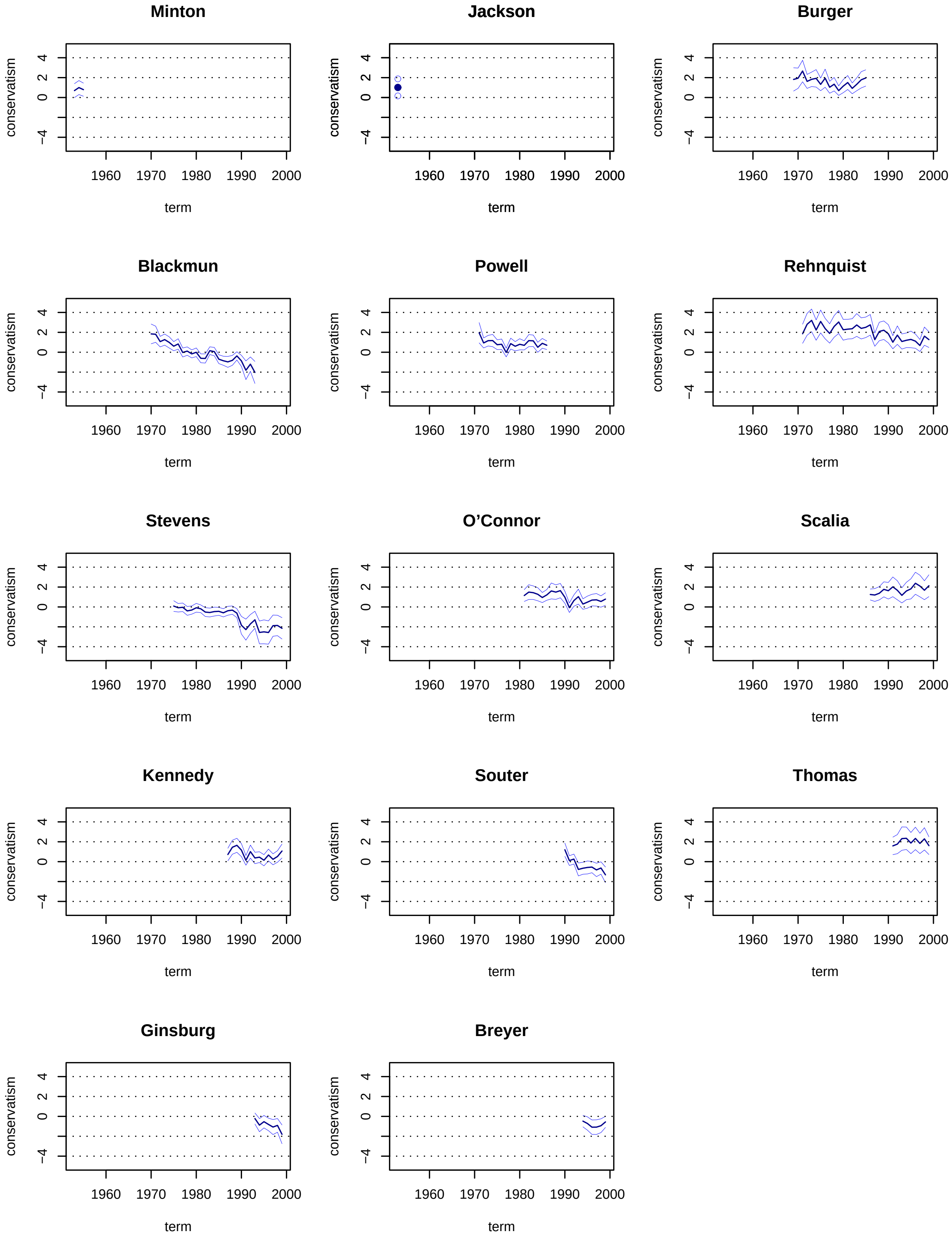


Figure 4: Posterior density summary of the ideal points of each justice for the terms in which they served for the independent ideal point model, Justices Minton-Breyer.

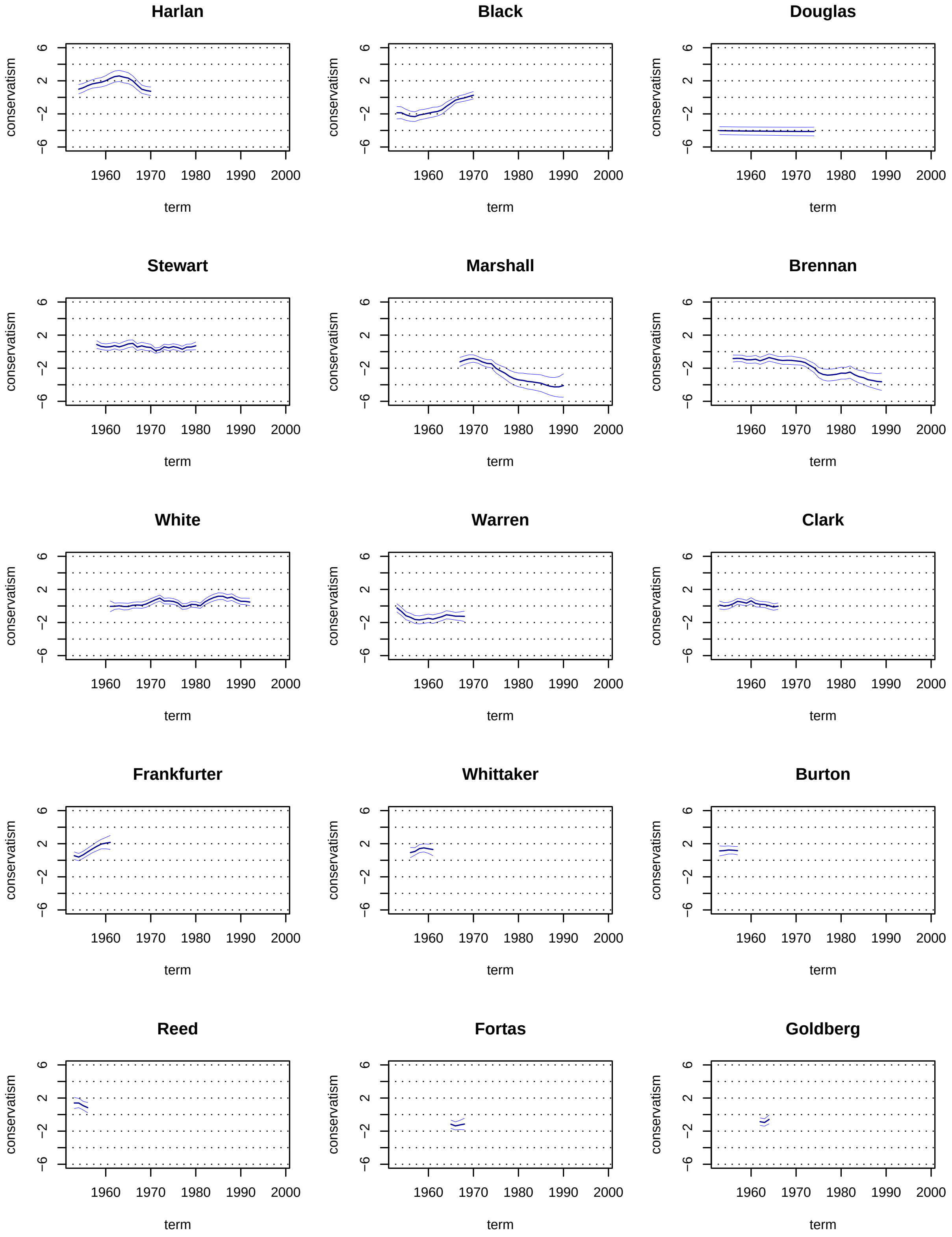
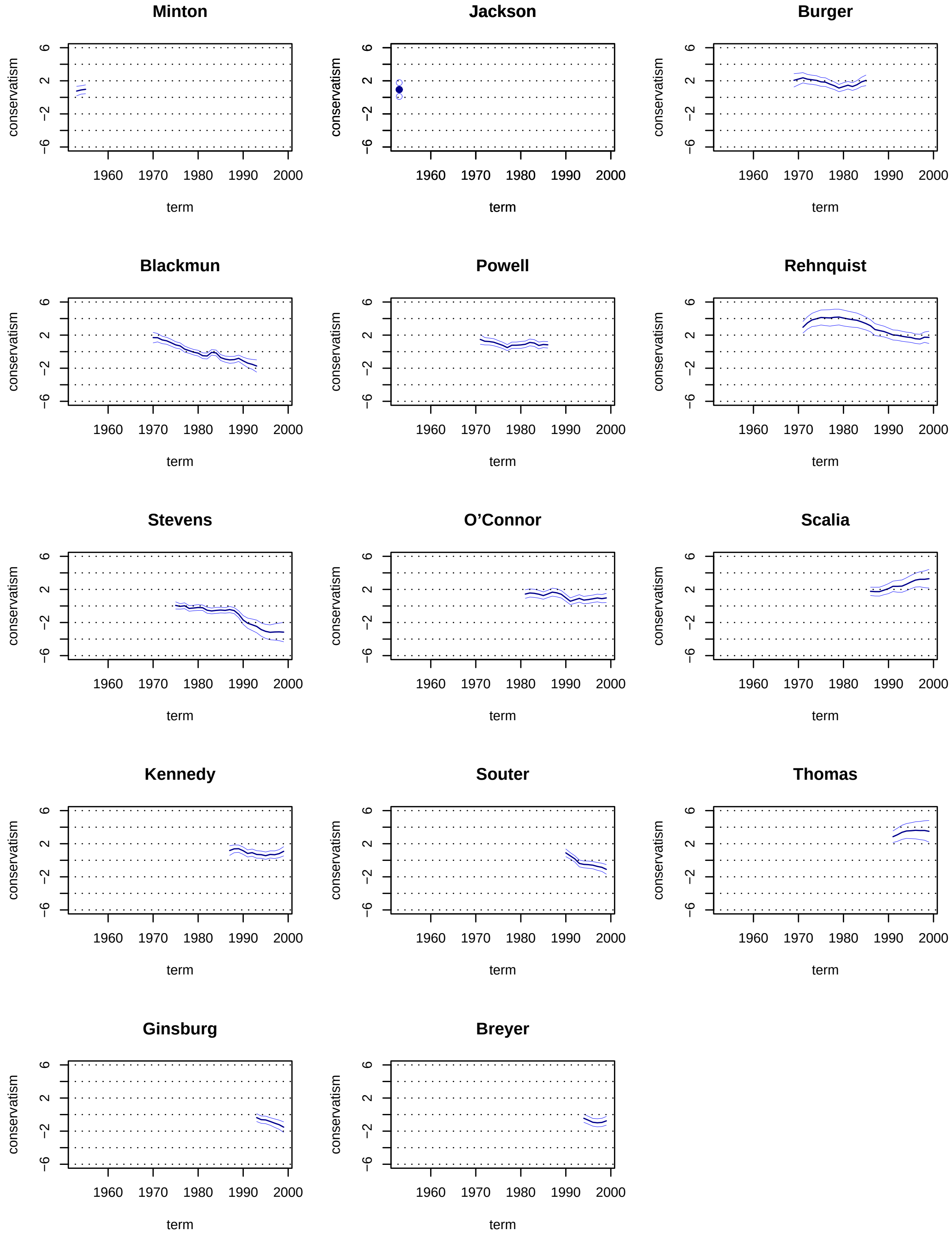


Figure 5: Posterior density summary of the ideal points of each justice for the terms in which they served for the dynamic ideal point model, Justices Harlan-Goldberg.



Minton Jackson Burger Blackmun Powell Rehnquist Stevens O'Connor Scalia Kennedy Souter Thomas Ginsburg Breyer

Figure 6: Posterior density summary of the ideal points of each justice for the terms in which they served for the dynamic ideal point model, Justices Minton-Breyer.

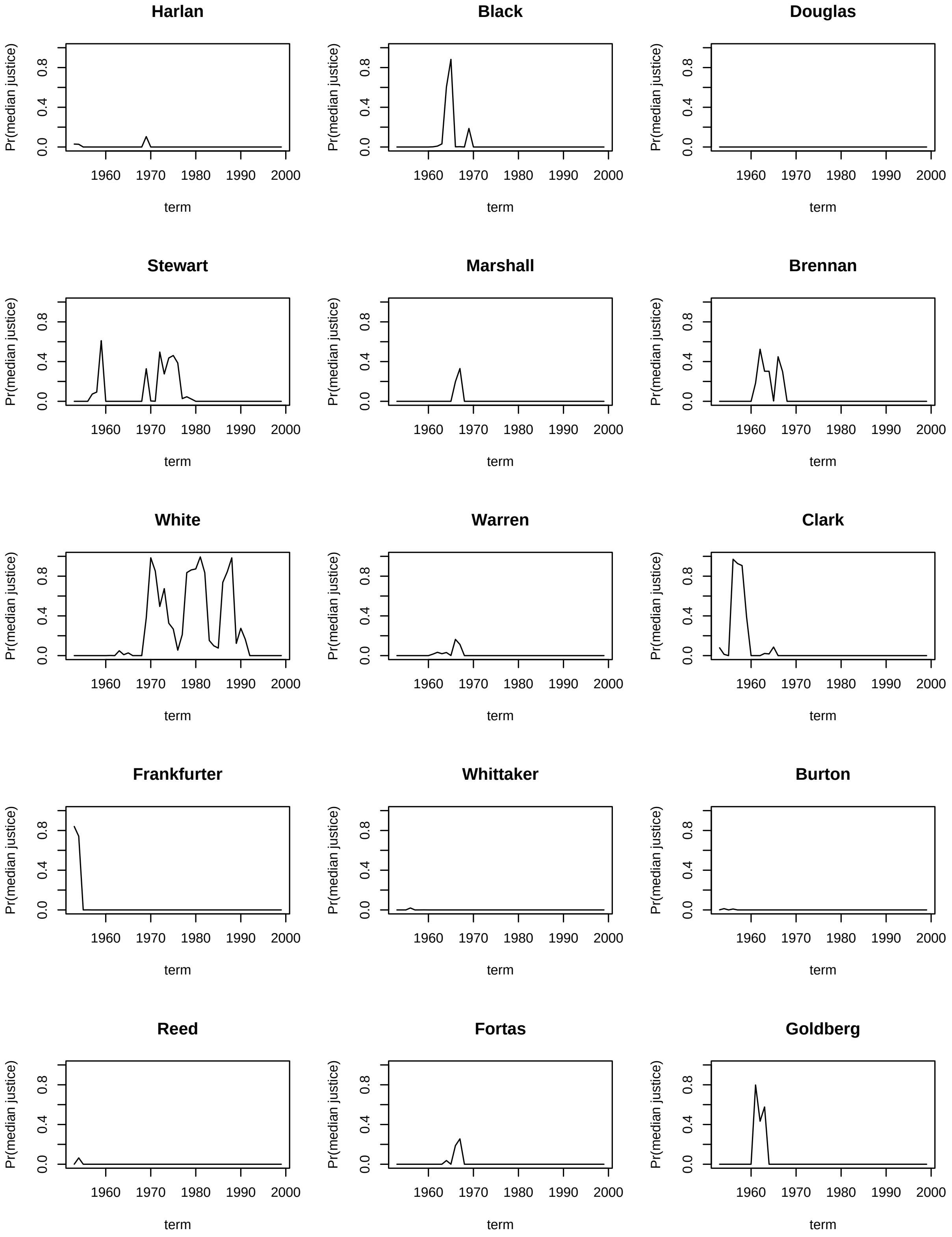


Figure 7: Estimated posterior probabilities that each justice in a given term is the median justice on the Court for the dynamic ideal point model, Justices Harlan-Goldberg.

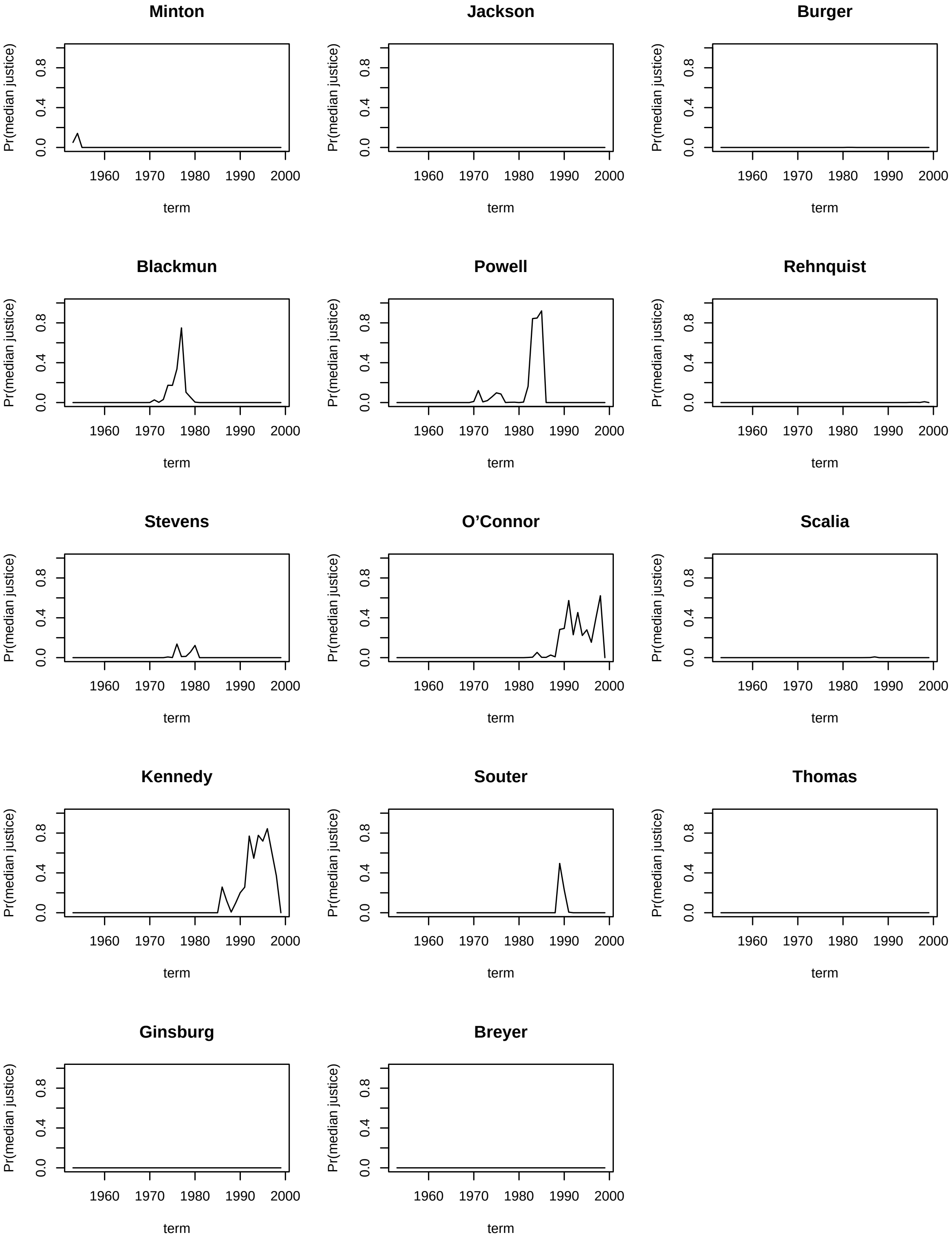


Figure 8: Estimated posterior probabilities that each justice in a given term is the median justice on the Court for the dynamic ideal point model, Justices Minton-Breyer.

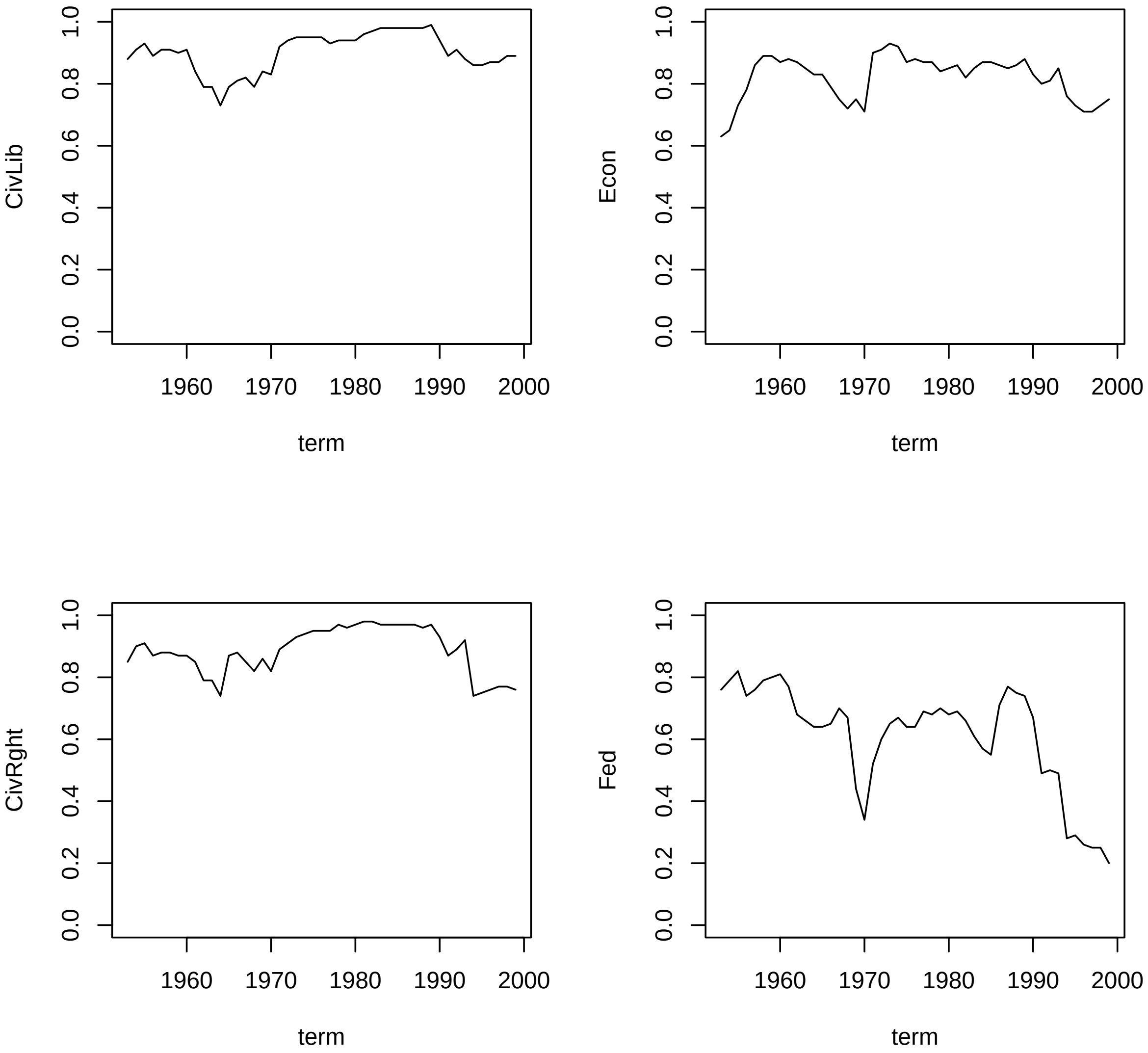


Figure 9: Term-by-term correlations of the ideal point posterior mean for the dynamic ideal point model with the percent conservative votes in civil liberties, civil rights, economics, and federalism cases, 1953-1999.

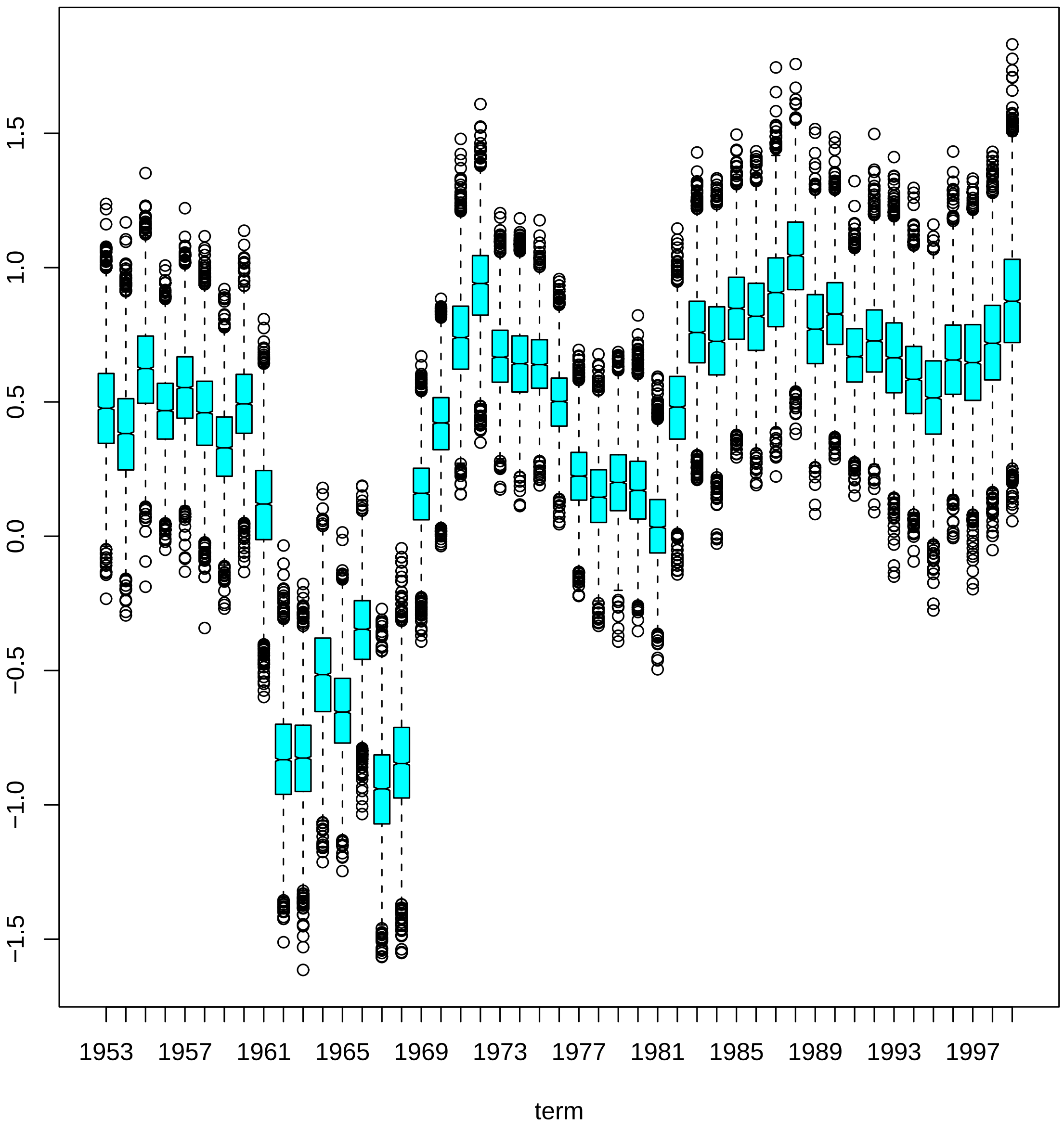


Figure 10: Estimated posterior distribution of the location of the median justice for the dynamic ideal point model.

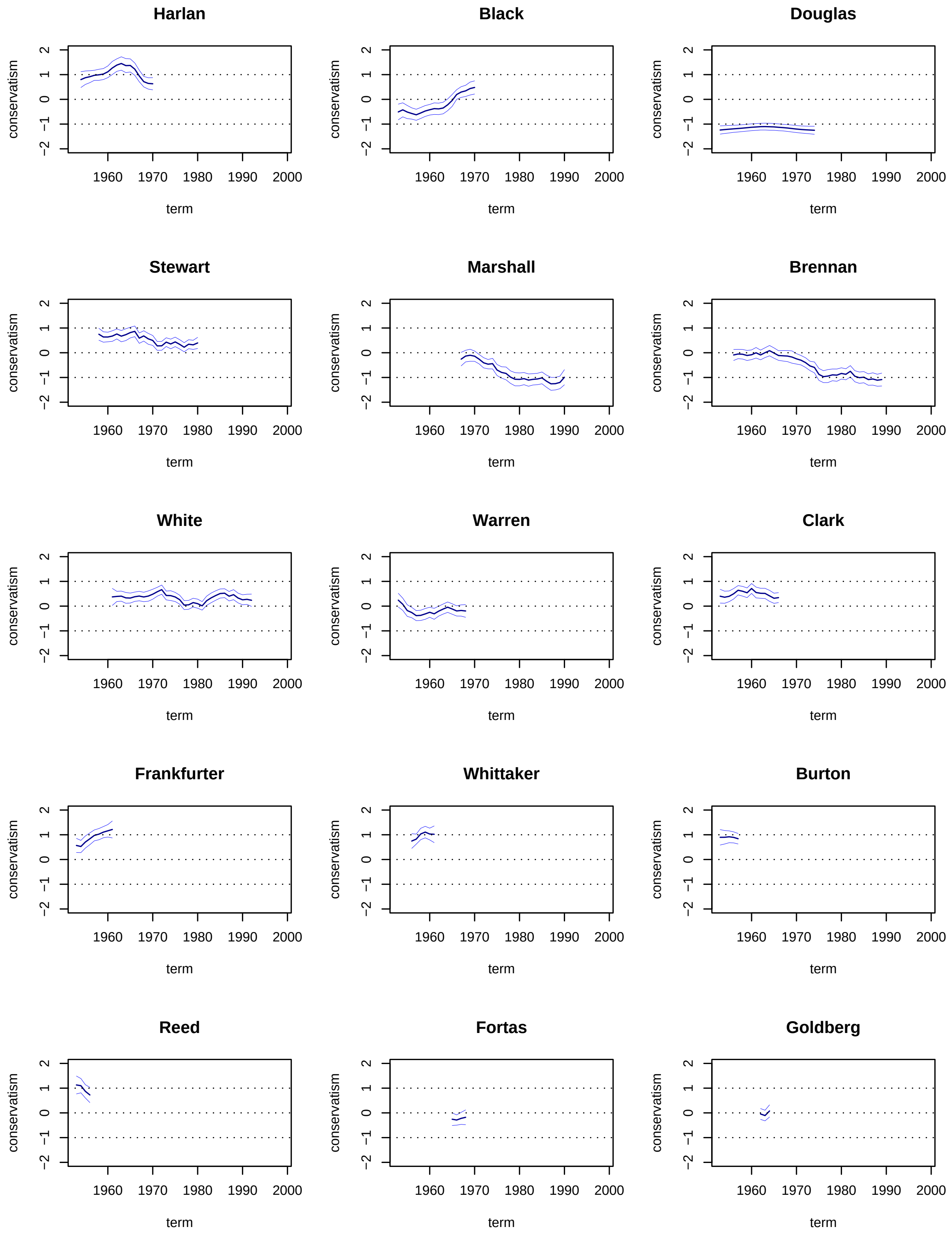


Figure 11: Posterior density summary of the ideal points of each justice for the terms in which they served for the dynamic ideal point and case parameter model, Justices Harlan-Goldberg.

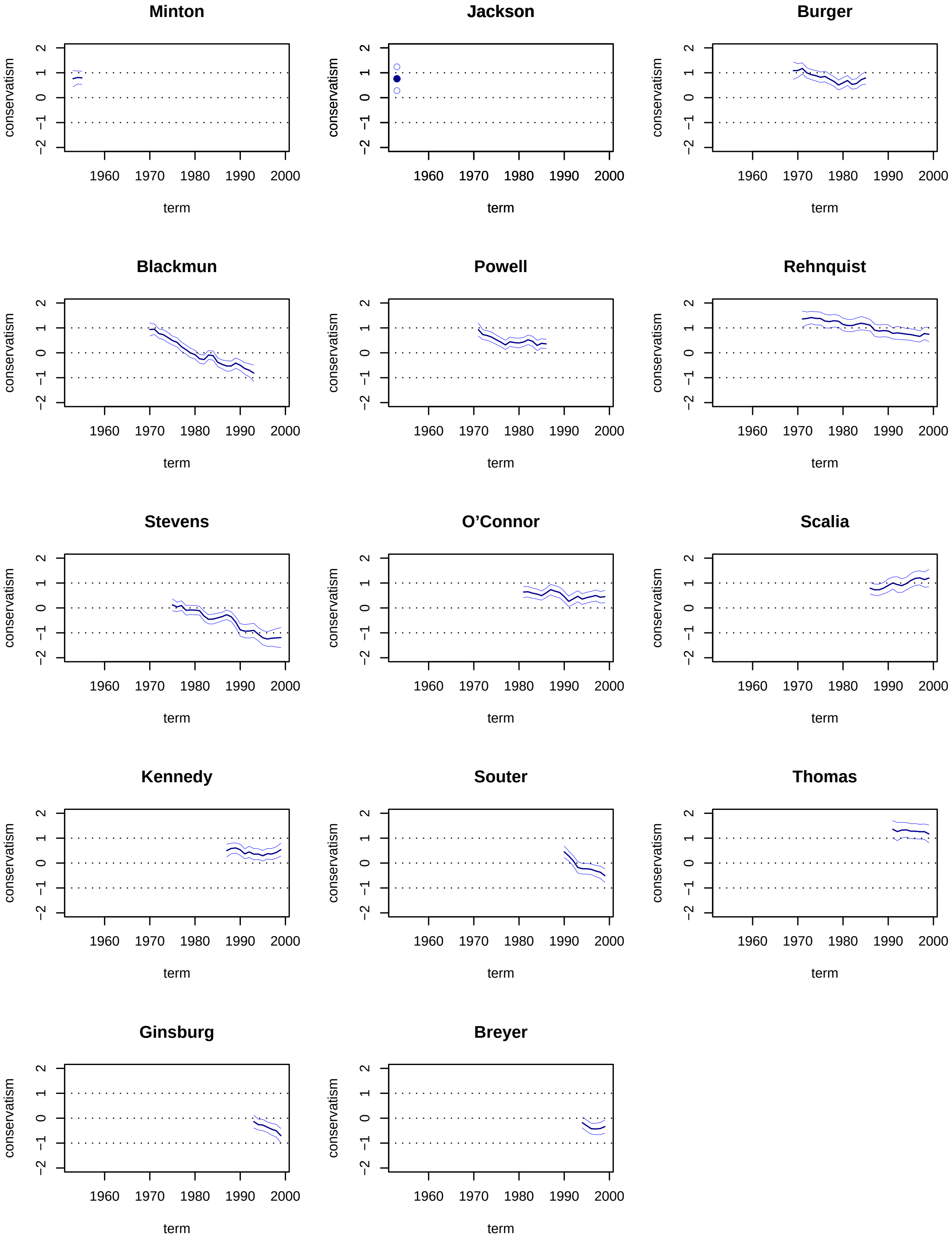


Figure 12: Posterior density summary of the ideal points of each justice for the terms in which they served for the dynamic ideal point and case parameter model, Justices Minton-Breyer.

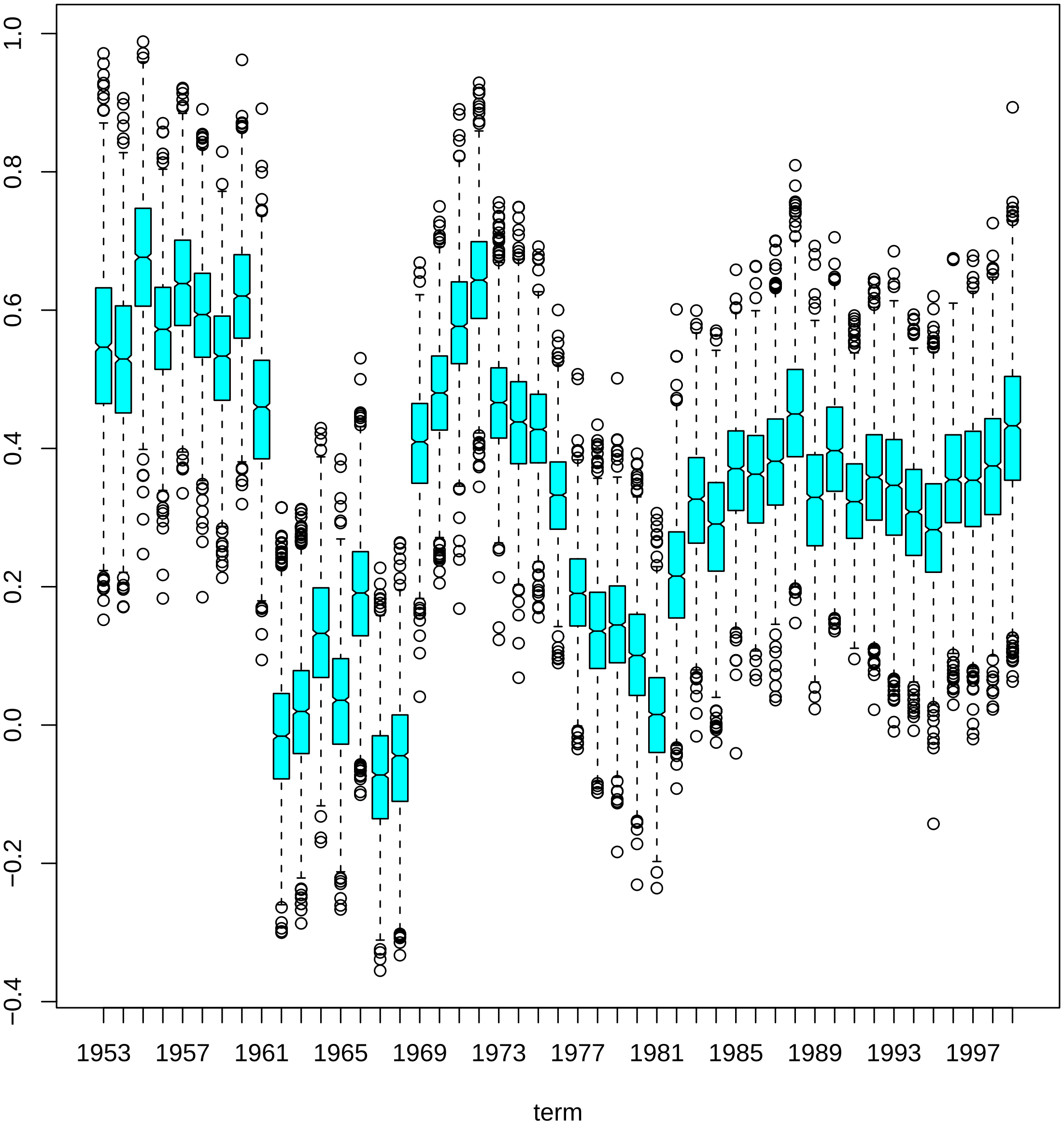


Figure 13: Estimated posterior distribution of the location of the median justice for the dynamic ideal point and case parameter model.

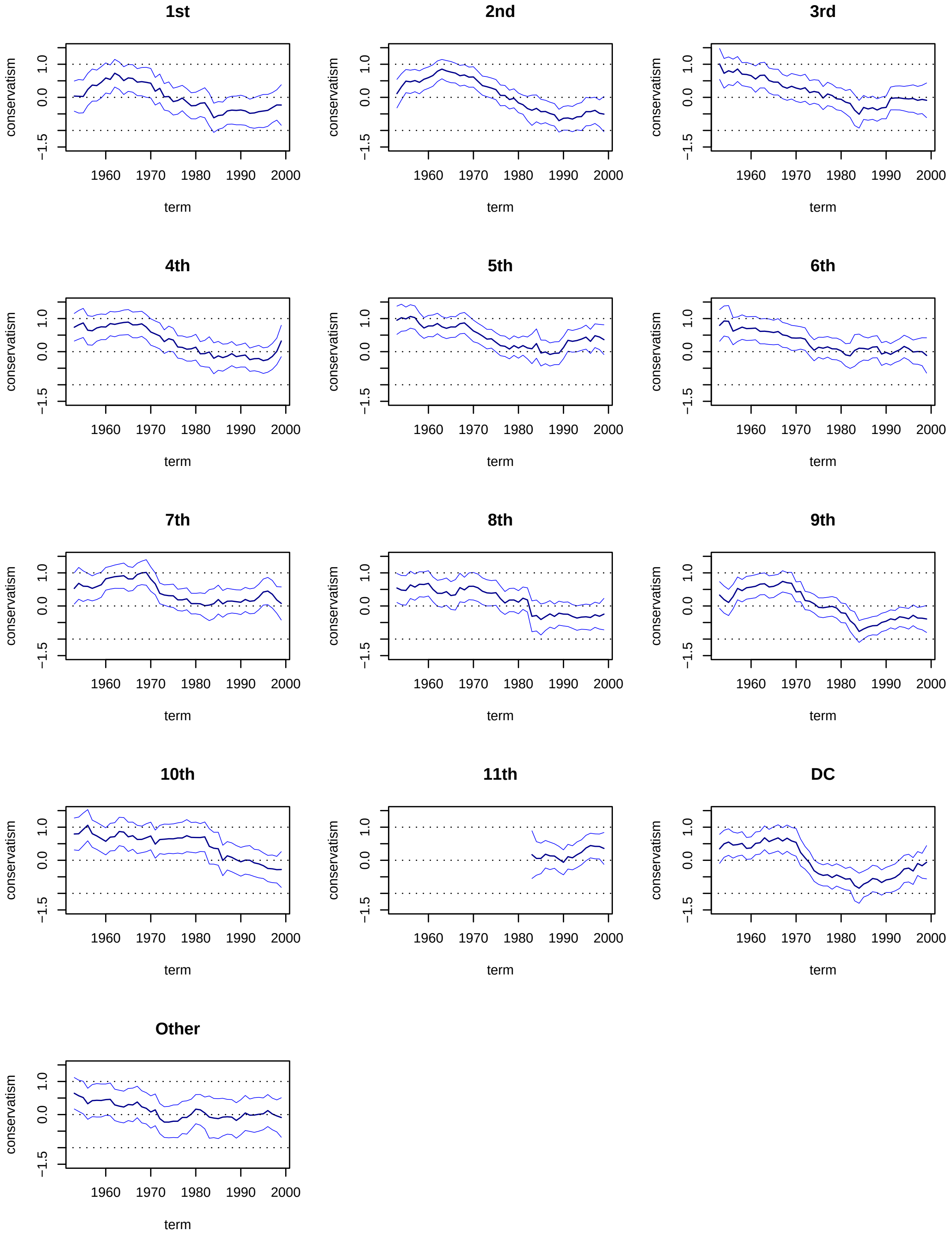


Figure 14: Estimated position of the prior mean ($\gamma_{c,t}$) of affirmance points [$\mathbf{x}_t^{(a)}$] from the lower court of origin c for the dynamic ideal point and case parameter model.