

Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999

Andrew D. Martin

*Department of Political Science, Washington University,
Campus Box 1063, One Brookings Drive,
St. Louis, MO 63130-4899
e-mail: admartin@artsci.wustl.edu*

Kevin M. Quinn

*Department of Political Science and
Center for Statistics in the Social Sciences, Box 354320,
University of Washington, Seattle, WA 98195-4322
e-mail: quinn@stat.washington.edu*

At the heart of attitudinal and strategic explanations of judicial behavior is the assumption that justices have policy preferences. In this paper we employ Markov chain Monte Carlo methods to fit a Bayesian measurement model of ideal points for all justices serving on the U.S. Supreme Court from 1953 through 1999. We are particularly interested in determining to what extent ideal points of justices change throughout their tenure on the Court. This is important because judicial politics scholars oftentimes invoke preference measures that are time invariant. To investigate preference change, we posit a dynamic item response model that allows ideal points to change systematically over time. Additionally, we introduce Bayesian methods for fitting multivariate dynamic linear models to political scientists. Our results suggest that many justices do not have temporally constant ideal points. Moreover, our ideal point estimates outperform existing measures and explain judicial behavior quite well across civil rights, civil liberties, economics, and federalism cases.

1 Introduction

Whether one adopts an attitudinal (Segal and Spaeth 1993) or a strategic (Eskridge 1991; Epstein and Knight 1998) approach to explain the behavior of Supreme Court justices, most

Authors' note: An early version of this paper was awarded the 2001 Harold Gosnell Prize by the Political Methodology Section of the American Political Science Association. We would like to thank Michael Bailey, Sara Benesh, 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 Weidenbaum Center at Washington University as well as the Center for Statistics and the Social Sciences with funds from the University Initiatives Fund at the University of Washington. Data for replication can be obtained from the Supreme Court Database Web site at Michigan State University (<http://www.ssc.msu.edu/~pls/pljp/sctdata.html>). Documented C++ code for these models and supplementary results are available in a Web appendix at both the *Political Analysis* and the authors' Web sites.

scholars invoke preferred policy positions of the justices as the key explanatory variables.¹ For both of these approaches, an important substantive question is, Do ideal points of U.S. Supreme Court justices change over time? If it is the case that preferences do vary over time, the conclusions drawn from past work with static preferences measures are suspect, and the implications for future work are clear (Epstein et al. 1998). The purpose of this paper is to present a modeling strategy that can be used to estimate ideal points from longitudinal data.

The measurement of latent preferences of political actors 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 (see, e.g., Schubert 1965, 1974; Rohde and Spaeth 1976; Segal and Cover 1989; Poole and Rosenthal 1997; Clinton et al. 2000; Londregan 2000; Jackman 2001; Bailey and Chang 2001). Nonetheless, this paper offers several important methodological innovations.

First, the model that we derive is a *dynamic* item response model. To investigate preference change, we introduce a dynamic hierarchical prior on the latent ideal points to test for preference change. This is similar in spirit to the D-NOMINATE model proposed by Poole and Rosenthal (1997). However, our approach is quite different in terms of specification and estimation strategy. The models we present are unique, as they provide the most flexible structure employed to date to model the dynamics of ideal points on the Supreme Court.

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). Bailey and Chang (2001) have estimated models of interinstitutional preferences using the EM algorithm. But, to our knowledge, no one has developed Bayesian models for multiple cross sections of votes that flexibly model the temporal dependence that one might expect to find. The Bayesian methods we employ are well suited to the small number of cases heard in each term and the small number of justices on the Court. Moreover, taking a Bayesian approach facilitates moving from a static cross-sectional model to a dynamic one.

Third, we introduce Bayesian inference for dynamic linear models (DLMs) to political scientists. The method we employ makes use of the DLM machinery of West and Harrison (1997) and is similar in spirit to standard state-space modeling (Beck 1989). These methods are useful in many settings beyond ideal point estimation. They are particularly useful in borrowing strength (smoothing) across time or when dealing with unevenly spaced time series. The applications to standard time series data and time-series cross-section data (Stimson 1985; Beck and Katz 1995) are fairly apparent. The methods discussed in this paper for modeling ideal point dynamics carry over directly to binary time-series cross-section (BTSCS) data; one can think of the data set used in this paper as BTSCS data with no measured covariates.² The search for good methods to model such data is currently one of the most active research areas in political methodology.

¹The 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 (Enelow and Hinich 1984), where each actor has an ideal point and acts to achieve the policy outcome closest to that point. For expositional economy, we employ the parlance of the spatial voting model throughout this paper.

²The dynamic model proposed below could be used for BTSCS with measured covariates by replacing α_k and β_k with measured covariates. The model would smooth the structural parameters $\theta_{t,j}$ over time. This could be done for any number of covariates using a multidimensional generalization of our model.

This paper proceeds as follows. In the following section we briefly review other approaches scholars have employed to measure judicial preferences and present the data. This is followed, in Section 3, by a presentation of a dynamic item response model. We present our MCMC estimation strategy and an introduction to Bayesian DLMS in Section 4. In Section 5 we present our ideal point estimates and other quantities of interest for the Supreme Court from 1953 through 1999. The final section concludes with a discussion of directions for future research.

2 Previous Dynamic Approaches and the Data

Many generations of judicial politics scholars have estimated the ideal points of Supreme Court justices. The seminal work is *The Judicial Mind*, in which Schubert (1965) presents a comprehensive look at the ideologies of all Supreme Court justices serving from the end of the Roosevelt era to the beginning of the Johnson administration. Using multidimensional scaling techniques, he finds two primary dimensions—political and economic liberalism—that explain votes on the merits for the Court from the late 1940s to the early 1960s. He also finds evidence of three “minor scales” that are salient in only a subset of terms. In a later work, Schubert (1974) finds that these two dimensions structure behavior through the Vinson and Warren Courts. The use of latent variable models to measure ideal points is also seen in the work of Rohde and Spaeth (1976), who find three dimensions that structure the Warren Court and the first five terms of the Burger Court: freedom, equality, and “New Dealism.”

Others have taken novel approaches to measuring judicial preferences: Segal and Cover (1989) content analyze newspaper editorials at the time of appointment to the Court, Epstein et al. (1989) look at the percentage liberal votes in a single policy area, and Tate and Handberg (1991) record background characteristics of the justices. While these approaches are highly creative, we feel that a better approach to explaining votes as a function of policy preferences is to posit a theoretical model of decision making and then directly operationalize it into a statistical model. Such a structural model has the property that if the theoretical model is correct (implying that the posited statistical assumptions hold), then it can explain votes in terms of preferences even though preferences are treated as latent variables. For a similar unified structural model of voting and preference estimation in Congress, see Clinton et al. (2000).

An important question is whether the preferences of Supreme Court justices are constant throughout time. In reviewing the literature, Epstein et al. (1998, p. 801) summarize the conventional wisdom in the judicial politics field: “The occasional anomaly notwithstanding, most jurists evince consistent voting behavior over the course of their careers.” On the other hand, it is plausible that as justices learn by deciding new and different cases, their worldview—and thus their ideal points—changes over time. Epstein et al. (1998) challenge the notion that judicial preferences are constant and seek to discover whether or not ideal points 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 the scores of all justices and is geared to control for agenda changes. Of the 16 justices under study, Epstein et al. find nine justices that exhibit changes over the course of their careers.

There are some weaknesses of the Epstein et al. (1998) study that 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 (percentage liberal on civil liberties cases) is treated as if it were a known quantity. However, with the

small and varying number of cases heard in each term, it is important to account for the inherent uncertainty in the measure. Failing to do so will result in incorrect standard errors. 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 justice's outlook. Finally, the analysis 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 may dictate the manner in which future scholars should measure preferences and will call into question the findings of previous research. Our focus in this paper is the $T = 47$ terms of the Court from 1953 through 1999. We obtain data from *The United States Supreme Court Judicial Database* (Spaeth 2001) and select all nonunanimous cases that are either formally decided cases with written opinions after full oral argument or cases decided by an equally divided vote.³ 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 $K = 3450$ cases was decided, with an average of 70 cases per term. The most active Courts sat during the 1972 and 1986 terms, deciding 102 and 104 cases, respectively. The least active was the 1996 term, which decided only 40 cases.

The observed data matrix \mathbf{V} is thus a $(K \times J)$ matrix of votes and missing values, which can be partitioned by term t . We code all votes in term t on case k by justice j as either being in favor of reversing ($v_{t,k,j} = 1$) or affirming ($v_{t,k,j} = 0$) the decision of a lower court.⁴ A total of 2189 lower court cases was reversed, while 1261 were affirmed. Using the disposition variable in the *Supreme Court Judicial Database*, we are able to recode 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).

3 The Dynamic Ideal Point Model

In this section we derive a dynamic item response model from a simple spatial model of voting. Judicial politics scholars refer to this model as the “attitudinal model,” which simply states that justices vote their true attitudes (or preferences) when voting on the merits (Segal and Spaeth 1993). The dynamic model we propose can perhaps be used to uncover the ideal points of Supreme Court justices. Our work parts from that of Clinton et al. (2000) and Jackman (2001) in the specification of a prior distribution—that explicitly models the temporal dependence we might expect to find—for the latent ideal points. To estimate the

³The unit of analysis is the case citation (ANALU = 0). We select cases where the type of decision (DEC_TYPE) equals 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) and drop all unanimous cases. We exclude unanimous cases because they contribute no information to the likelihood. Including unanimous cases also makes it quite difficult to specify reasonable prior distributions for the case parameters since one would expect a number of the α_k values (those from unanimously decided cases) to approach $-\infty$ or $+\infty$, while the remainder would be fairly close to 0. Using a Gaussian prior for the case parameters would force the α_k values associated with unanimously decided cases toward 0, which would in turn influence our ideal point estimates. Additionally, we suspect that unanimity results from motivations rooted in institutional legitimacy that cannot be explained by our spatial model.

⁴We use the disposition variable (DIS) in the Supreme Court Date Base to code affirmances and reversals. An affirmance (DIS = 1) and a petition denied or appeal dismissed (DIS = 8) 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), or vacated (DIS = 7) are coded as reversals. Cases when the Court granted a stay, petition, or motion (DIS = 0), affirmed in part and reversed in part (DIS = 5), or certified to a lower court (DIS = 9) were dropped from the analysis.

model, we rely on standard Gibbs sampling theory and the dynamic linear model (DLM) machinery as discussed in Section 4.

3.1 Formalizing the Spatial Model

Let $K_t \subset \{1, 2, \dots, K\}$ denote the set of cases heard in term t . Note that K_t partitions the set of cases, and its cardinality $|K_t|$ denotes the number of cases heard in a term. Similarly, let $J_k \subset \{1, 2, \dots, J\}$ denote the set of justices who heard case k . The cardinality of J_k denotes the number of justices sitting on a case k . This number is typically nine, although in certain circumstances fewer justices heard a particular case. We are interested in modeling the decisions made in terms $t = 1, \dots, T$ on cases $k \in K_t$ by justices $j \in J_k$ in a unidimensional issue space.⁵ 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 $u_{t,k,j}^{(a)}$ be the utility to justice $j \in J_k$ of voting to affirm on case $k \in K_t$ in term t , and $u_{t,k,j}^{(r)}$ be the utility to justice $j \in J_k$ of voting to reverse on case $k \in K_t$ in term t . We assume that justice j will vote to reverse on case k when $z_{t,k,j} = u_{t,k,j}^{(r)} - u_{t,k,j}^{(a)} > 0$. We can write and simplify this utility difference as follows (see Clinton et al. 2000):

$$\begin{aligned} z_{t,k,j} &= u_{t,k,j}^{(r)} - u_{t,k,j}^{(a)} = -|\theta_{t,j} - x_k^{(r)}|^2 + \xi_{t,k,j}^{(r)} + |\theta_{t,j} - x_k^{(a)}|^2 - \xi_{t,k,j}^{(a)} \\ &= [x_k^{(a)} x_k^{(a)} - x_k^{(r)} x_k^{(r)}] + 2[x_k^{(r)} - x_k^{(a)}]\theta_{t,j} + [\xi_{t,k,j}^{(r)} - \xi_{t,k,j}^{(a)}] \\ &= \alpha_k + \beta_k \theta_{t,j} + \varepsilon_{t,k,j}. \end{aligned} \quad (1)$$

$\theta_{t,j} \in \mathbb{R}$ is justice j 's ideal point in the uni-dimensional issue space for term t , $x_k^{(a)} \in \mathbb{R}$ is the location of the policy under an affirmance vote, $x_k^{(r)} \in \mathbb{R}$ is the location of the policy under a reversal, and $\xi_{t,k,j}^{(a)}$ and $\xi_{t,k,j}^{(r)}$ are Gaussian disturbances with zero means and fixed variances.⁶ Given this model, justice j will vote to reverse on case k when $z_{t,k,j} > 0$.

3.2 A Dynamic Item Response Model

This model is easily translated into a statistical model by noting the relationship between the theoretical model and our observed data. For observed votes, we assume that

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

where

$$z_{t,k,j} = \alpha_k + \beta_k \theta_{t,j} + \varepsilon_{t,k,j} \quad \varepsilon_{t,k,j} \stackrel{iid}{\sim} \mathcal{N}(0, 1). \quad (3)$$

⁵While the model employed in this paper is unidimensional, it is easy to extend into multiple dimensions using the same technology. For examples of multidimensional models, see Jackman (2001) and Martin and Quinn (2001a).

⁶It is important to note that these measures are summarizing observed behavior in a manner consistent with the spatial model. It might be the case that justices are behaving in a nonsincere, strategic fashion. Viewed in this light, these measures could be interpreted as revealed preferences (which are likely to change over time due to strategic considerations). If one believes the attitudinal model, our estimates could be viewed as true ideal points. This distinction, however, is beyond the scope of this paper.

Note that we have fixed the variance of $\varepsilon_{t,k,j}$ to 1 since this variance and the other model parameters are not separately identified in the likelihood.⁷ This results in a standard two-parameter item response model; the only difference is that the latent traits $\theta_{t,j}$ vary across time.

Let α denote the stacked α_k for $k = 1, \dots, K$ and β denote the stacked β_k for $k = 1, \dots, K$. Further, let θ denote the stacked ideal points of all justices $j = 1, \dots, J$ in all terms. Bayesian inference for the justices' ideal points θ and the case parameters α and β proceeds by summarizing the posterior density given by:

$$p(\alpha, \beta, \theta | \mathbf{V}) \propto p(\mathbf{V} | \alpha, \beta, \theta)p(\alpha, \beta, \theta), \tag{4}$$

where $p(\alpha, \beta, \theta)$ represents our prior beliefs about these parameters, and

$$p(\mathbf{V} | \alpha, \beta, \theta) \propto \prod_{t=1}^T \prod_{k \in K_t} \prod_{j \in J_k} \Phi(\alpha_k + \beta_k \theta_{t,j})^{v_{t,k,j}} (1 - \Phi(\alpha_k + \beta_k \theta_{t,j}))^{1-v_{t,k,j}} \tag{5}$$

is the sampling density. Here $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

It is well known that item response models suffer from identification problems (see, e.g., Albert 1992; Johnson and Albert 1999). The first problem is called scale invariance. The parameters of interest α , β , and θ are identified only up to an arbitrary scale factor. Thus, one must anchor the scale on which ideal points are measured, just as one would calibrate a scale to measure any quantity. An additional identification problem is called rotational invariance. For the one-dimensional case, multiplying all of the model parameters by -1 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? In the Bayesian context, both of these identification problems are resolved through the use of semiinformative prior distributions. The prior variances on θ , α and β define the metric on which the ideal points and case parameters are measured, and the informative priors on the ideal points for certain justices ensure that only one of the posterior modes has nonnegligible mass.

3.3 Prior Distributions

We begin by assuming standard prior distributions for the case parameters α and β . The standard approach is to assume that these are drawn from multivariate normal distributions. In other words,

$$\begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} \sim \mathcal{N}_2(\mathbf{b}_0, \mathbf{B}_0) \quad \forall k \in \{1, \dots, K\}. \tag{6}$$

Our dynamic ideal point model departs from the literature in the assumed prior distribution on the ideal points. Item response models were developed in the educational testing field, where the standard assumption is that the latent quantity of interest—intelligence or

⁷This is a typical assumption; for example, one makes this assumption for the standard probit model for a dichotomous dependent variable. In principle, it would be possible to fix the variance of one justice's latent disturbances and to estimate separate disturbance variances for the other justices relative to this reference justice. We leave this computationally expensive exercise for future research.

aptitude—is constant throughout the course of the study. For this application, we would need to assume that Supreme Court justices have policy preferences that do not change over their careers. An alternative assumption is to estimate separate models for term. This is equivalent to assuming that each justice’s ideal point at time t is independent of her ideal point at $t - 1$. This does not seem plausible to us, as surely justices demonstrate at least some stability in their attitudes over time.

The approach we take is the middle ground between these extremes. Our assumption is that the ideal points follow a random walk process. This strikes a balance between the overly strong assumption that ideal points are constant over time and the assumption that a justice’s ideal points in a given time period are completely independent of that justice’s ideal points in other time periods. We model the dynamics of the ideal points with a separate random walk prior for each justice:

$$\theta_{t,j} \sim \mathcal{N}(\theta_{t-1,j}, \Delta_{\theta_{t,j}}) \quad \text{for } t = \underline{T}_j, \dots, \overline{T}_j \quad \text{and justice } j \text{ on the Court, } (7)$$

where \underline{T}_j is the first term justice j served, and \overline{T}_j is the last term justice j served.⁸ We do not estimate ideal points for terms in which a justice did not serve. $\Delta_{\theta_{t,j}}$ is an evolution variance parameter which is fixed *a priori* by the researcher. Its magnitude parameterizes how much borrowing of strength (or smoothing) takes place from one time period to the next. Note that if $\Delta_{\theta_{t,j}} = 0$, this would be the same as assuming fixed ideal points. At the other extreme, as $\Delta_{\theta_{t,j}} \rightarrow \infty$ we get a model in which the ideal points are temporally independent. To complete the prior, we must anchor each time series at the unobserved time period 0. We assume that

$$\theta_{0,j} \sim \mathcal{N}(m_{0,j}, C_{0,j}). \quad (8)$$

4 MCMC for the Dynamic Ideal Point Model

Our inferential goal is to summarize the joint posterior distribution of the model parameters given the observed data: $p(\alpha, \beta, \theta | \mathbf{V})$. We accomplish this by using MCMC methods to sample from this distribution and then summarizing the resulting draws. More specifically, we use a version of the data augmentation algorithm of Albert and Chib (1993) to sample the latent matrix of utility differences \mathbf{Z} . This makes sampling from the other conditional distributions, $p(\alpha, \beta | \mathbf{V}, \mathbf{Z}, \theta)$ and $p(\theta | \mathbf{V}, \mathbf{Z}, \alpha, \beta)$, much easier.

The MCMC algorithm for this dynamic item response model contains three steps over which we iterate a large number of times. The algorithm proceeds as follows.

1. For $t = 1, \dots, T, k \in K_t$ and $j \in J_k$: Simulate $z_{t,k,j}$ from Eq. (9).
2. For $t = 1, \dots, T, k \in K_t$: Simulate $(\alpha_k \beta_k)'$ from Eq. (10).
3. For $j = 1, \dots, J$: Simulate $\{\theta_{t,j}\}_{t=\underline{T}_j}^{\overline{T}_j}$ using the forward-filtering, backward sampling algorithm for the DLM specified in Eqs. (26) and (27).

In the remainder of this section, we detail these three steps and present a general introduction to estimation of DLMs in the Bayesian context.

⁸This notation is a bit clumsy because justices serve a variable number of terms with different starting and ending dates. Nonetheless, we use 0 to denote the term before the justice began serving, \underline{T}_j to denote the first term served, and \overline{T}_j to denote the last term served. This results in series that, for example, look like 0, 1963, 1964, ..., 1976 in the data.

4.1 Simulating the Latent Utilities

In the first step, we sample the latent utilities from the conditional distribution $p(\mathbf{Z} | \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. For terms $t = 1, 2, \dots, T$, cases $k \in K_t$, and justices $j \in J_k$, we simulate the latent utilities from the following distribution:

$$p(z_{t,k,j} | v_{t,k,j}, \theta_{t,j}, \alpha_k, \beta_k) = \begin{cases} \mathcal{N}_{[0,\infty)}(\alpha_k + \beta_k \theta_{t,j}, 1) & \text{if } v_{t,k,j} = 1 \\ \mathcal{N}_{(-\infty,0]}(\alpha_k + \beta_k \theta_{t,j}, 1) & \text{if } v_{t,k,j} = 0 \\ \mathcal{N}(\alpha_k + \beta_k \theta_{t,j}, 1) & \text{if } v_{t,k,j} \text{ is missing and } t \in \{\underline{T}_j, \dots, \bar{T}_j\}, \end{cases} \quad (9)$$

where $\mathcal{N}_{[a,b]}$ denotes the Gaussian distribution truncated on the interval $[a, b]$. This is the standard data augmentation step for a binary response model with missing data (Albert and Chib 1993; see also Johnson and Albert 1999).

4.2 Sampling the Case Parameters

After sampling the latent utilities we sample the case parameters α_k and β_k in a single, case-specific block for all k . Let $\boldsymbol{\eta}_k = (\alpha_k \beta_k)'$. We draw the case parameters from the conditional distribution $p(\boldsymbol{\eta} | \mathbf{V}, \mathbf{Z}, \boldsymbol{\theta})$. To aid in exposition, we need to define two additional quantities. First, let $\boldsymbol{\theta}_{t,j}^* = (1 \theta_{t,j})'$ denote a 2×1 -column vector that contains a constant followed by the ideal point estimate for justice j in term t . Second, we define $\boldsymbol{\theta}_{t,\cdot}^*$ to be the $|J_k| \times 2$ matrix formed by stacking the transpose of these elements for justices $j \in J_k$. For all cases $k \in K_t$ in all terms t , we simulate the case-specific parameters $\boldsymbol{\eta}_k$ from a bivariate normal distribution:

$$p(\boldsymbol{\eta}_k | \mathbf{V}, \mathbf{Z}, \boldsymbol{\theta}) = \mathcal{N}_2(\mathbf{e}, \mathbf{E}), \quad (10)$$

where $\mathbf{e} = \mathbf{E}[\boldsymbol{\theta}_{t,\cdot}^* \mathbf{z}_{t,k,\cdot} + \mathbf{B}_0^{-1} \mathbf{b}_0]$, $\mathbf{E} = [\boldsymbol{\theta}_{t,\cdot}^* \boldsymbol{\theta}_{t,\cdot}^* + \mathbf{B}_0^{-1}]^{-1}$, and $\mathbf{z}_{t,k,\cdot}$ is the column vector of latent utilities for all justices on case k . This is the standard update for a multivariate regression model.

4.3 Sampling the Ideal Points Using DLMs

We next use dynamic linear models (DLMs) to estimate the ideal points of the justices over time. More generally, these models are extremely flexible and can be employed to model a wide variety of temporal dependencies. We begin by laying out the structure of the most general case, which is followed by a discussion of estimation via MCMC and how we employ this technology to estimate ideal points.

4.3.1 The General Multivariate DLM

We begin by introducing another notation that can be used for any multivariate time series. Let $\{\mathbf{y}_t\}_{t=1}^T$ denote a multivariate time series observed over $t = 1, \dots, T$.⁹ Each \mathbf{y}_t is an $(N \times 1)$ vector. Note that the time points need not be equally spaced. Further, let \mathbf{X}_t be an

⁹For 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.

$(N \times P)$ matrix of known covariates observed at time t . The observation equation can be written

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_N(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (11)$$

where $\boldsymbol{\gamma}_t$ is a $(P \times 1)$ column vector of state parameters to be estimated. The system or evolution equation can be written

$$\boldsymbol{\gamma}_t = \mathbf{W}_t \boldsymbol{\gamma}_{t-1} + \boldsymbol{\delta}_t \quad \boldsymbol{\delta}_t \sim \mathcal{N}_P(\mathbf{0}, \boldsymbol{\Delta}_t), \quad (12)$$

where \mathbf{W}_t is a $(P \times P)$ matrix of covariates observed at time t . The observational error variance–covariance matrix $\boldsymbol{\Sigma}_t$ and the evolution error variance–covariance matrix $\boldsymbol{\Delta}_t$ are assumed to be independent and known.¹⁰ This is a very general model that can be used to estimate a wide range of standard time series models. Note that one can think of the system equation as specifying a prior density on the $\boldsymbol{\gamma}_t$ that acts to smooth them over time. As the variance terms in $\boldsymbol{\Delta}_t$ decrease, smoothing increases.

4.3.2 MCMC for the General Multivariate DLM

We are interested in the joint posterior distribution of $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_t\}_{t=1}^T$. We assume a multivariate normal prior on the initial unobserved $\boldsymbol{\gamma}_0$:

$$\boldsymbol{\gamma}_0 \sim \mathcal{N}_P(\mathbf{m}_0, \mathbf{C}_0). \quad (13)$$

Let D_t denote all information available up to time t , i.e., $D_t = \{\mathbf{y}_t, D_{t-1}\}$, and as a convention let

$$D_0 = \{\mathbf{m}_0, \mathbf{C}_0, \{\mathbf{X}_t\}_{t=1}^T, \{\mathbf{W}_t\}_{t=1}^T, \{\boldsymbol{\Sigma}_t\}_{t=1}^T, \{\boldsymbol{\Delta}_t\}_{t=1}^T\}. \quad (14)$$

We can exploit the structure of the system equation to factor the joint posterior density as

$$\begin{aligned} p(\boldsymbol{\gamma} | D_T) &= p(\boldsymbol{\gamma}_T | D_T) p(\boldsymbol{\gamma}_{T-1} | \boldsymbol{\gamma}_T, D_{T-1}) p(\boldsymbol{\gamma}_{T-2} | \boldsymbol{\gamma}_{T-1}, D_{T-2}) \\ &\quad \cdots p(\boldsymbol{\gamma}_1 | \boldsymbol{\gamma}_2, D_1) p(\boldsymbol{\gamma}_0 | \boldsymbol{\gamma}_1, D_0). \end{aligned} \quad (15)$$

We can thus obtain a sample from the joint distribution $p(\boldsymbol{\gamma} | D_T)$ by

1. Sampling $\boldsymbol{\gamma}_T$ from $p(\boldsymbol{\gamma}_T | D_T)$.
2. Sampling $\boldsymbol{\gamma}_t$ from $p(\boldsymbol{\gamma}_t | \boldsymbol{\gamma}_{t+1}, D_t)$ for $t = (T - 1), (T - 2), \dots, 1, 0$.

From the factorization in Eq. (15) this sequence of draws is exactly equivalent to a direct multivariate draw from $p(\boldsymbol{\gamma} | D_T)$. The benefit of this algorithm over more direct Gibbs sampling approaches is that it allows us to sample $\boldsymbol{\gamma}$ (the set of all of the state parameters) directly in one block rather than sampling from the component by component full

¹⁰The assumption that the variance parameters are known is easily relaxed. It is assumed here only to facilitate exposition. Additionally, in our application, the observation variances must be fixed to identify the model, and the evolution variances represent our prior beliefs of how evolution takes place.

conditionals. Given the generally high correlations between the components of γ (which is why one would want to fit a dynamic model to begin with), approaches that sample from the component-by-component full conditionals mix very slowly.

What remains is to determine what form each of these conditional distributions take. Frühwirth-Schnatter (1994) and Carter and Kohn (1994) show that these conditional distributions are all multivariate normal, with means and variances that can be computed using quantities from the Kalman filter and the future values of the γ_t . That is, one moves from $t = 1, \dots, T$ and computes the quantities defined below, which is the “forward filtering” part of the algorithm. At $t = T$ one samples from $p(\gamma_t | D_T)$. Then one samples backward through time $t = T, \dots, 0$, updating the values of \mathbf{h}_t and \mathbf{H}_t (also defined below) at each iteration. For each time point, we then sample from

$$p(\gamma_t | \gamma_{t+1}, D_t) = \mathcal{N}_p(\mathbf{h}_t, \mathbf{H}_t). \quad (16)$$

This is the “backward sampling” part of the algorithm.

To complete the algorithm, it is necessary to write down the quantities used to sample from the conditional distributions above. We begin by filtering forward from $t = 1, \dots, T$. At each time period, we compute the following quantities:

- The prior mean of γ_t given D_{t-1} :

$$\mathbf{a}_t = \mathbf{W}_t \mathbf{m}_{t-1}. \quad (17)$$

Note that for the first iteration, this can be computed directly from the prior on γ_0 and the known covariates. For subsequent iterations, we have already computed \mathbf{m}_{t-1} (see below).

- The prior variance of γ_t given D_{t-1} :

$$\mathbf{R}_t = \mathbf{W}_t \mathbf{C}_{t-1} \mathbf{W}_t' + \mathbf{\Delta}_t. \quad (18)$$

Again, note that for the first iteration, this can be computed directly from the prior variance of γ_0 and the known covariates and evolution variance. For subsequent iterations, we have already computed \mathbf{C}_{t-1} (see below).

- The mean of the one-step ahead forecast of \mathbf{y}_t given D_{t-1} :

$$\mathbf{f}_t = \mathbf{X}_t \mathbf{a}_t. \quad (19)$$

This is easily computed using the \mathbf{a}_t computed above.

- The variance of the one-step ahead forecast of \mathbf{y}_t given D_{t-1} :

$$\mathbf{Q}_t = \mathbf{X}_t \mathbf{R}_t \mathbf{X}_t' + \mathbf{\Sigma}_t. \quad (20)$$

This too is easily computed using the known covariates and observational error and the prior variance \mathbf{R}_t computed above.

- The posterior mean of γ_t given D_t :

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t. \quad (21)$$

$\mathbf{A}_t = \mathbf{R}_t \mathbf{X}_t \mathbf{Q}_t^{-1}$ is the weight given to the likelihood relative to the prior (i.e., the Kalman gain), and $\mathbf{e}_t = \mathbf{y}_t - \mathbf{f}_t$ is the forecast error. All of these quantities are known.

- The posterior variance of γ_t given D_t :

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t' \quad (22)$$

All of these quantities are known. Note that as we filter forward through time, the past values of \mathbf{m}_t and \mathbf{C}_t are used in the computation of the conditional prior means and variances above.

After forward filtering is complete and these quantities have been saved, we can begin the sampling of the γ . For time $t = T$, we sample γ_T from a multivariate Normal distribution with mean \mathbf{m}_T and variance \mathbf{C}_T , which are the last values computed in the filtering step. We then sample backward from $t = (T - 1)$ to $t = 0$, at each time point drawing from

$$p(\gamma_t | \gamma_{t+1}, D_t) = \mathcal{N}(\mathbf{h}_t, \mathbf{H}_t). \quad (23)$$

For backward sampling, the quantities needed at each step are as follows.

- The mean of the conditional distribution of γ_t given γ_{t+1} and D_t :

$$\mathbf{h}_t = \mathbf{m}_t + \mathbf{B}_t(\gamma_{t+1} - \mathbf{a}_{t+1}), \quad (24)$$

where $\mathbf{B}_t = \mathbf{C}_t \mathbf{W}'_{t+1} \mathbf{R}_{t+1}^{-1}$, which can be computed from known values from the filtering step. \mathbf{h}_t is computed from known values from the filtering step. The future value of γ_{t+1} is also available because we are sampling backward.

- The variance of the conditional distribution of γ_t given γ_{t+1} and D_t :

$$\mathbf{H}_t = \mathbf{C}_t - \mathbf{B}_t \mathbf{R}_{t+1} \mathbf{B}_t' \quad (25)$$

These are all known from the filtering step.

One repeats this forward-filtering, backward sampling process many times to characterize the joint posterior distribution $p(\gamma | D_T)$. For derivations of the above formulas and proofs, see West and Harrison (1997).

4.3.3 DLMS for Dynamic Ideal Point Estimation

Given the latent \mathbf{Z} and the case parameters α and β , it is easy to show that the time series of ideal points for each justice is a multivariate dynamic linear model (West and Harrison 1997). We can rewrite Eq. (3) as

$$\mathbf{z}_{t,\cdot,j} - \alpha_t = \beta_t \theta_{t,j} + \varepsilon_{t,\cdot,j} \quad \varepsilon_{t,\cdot,j} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{|K_t|}), \quad (26)$$

where $\mathbf{z}_{t,\cdot,j}$ and $\varepsilon_{t,\cdot,j}$ are the $|K_t| \times 1$ vectors of latent utility differences and disturbances, respectively, for all cases heard by justice j in term t . In a slight abuse of notation, α_t is a $|K_t| \times 1$ vector of the stacked α_k for all $k \in K_t$, and β_t is a $|K_t| \times 1$ vector of the stacked β_k for all $k \in K_t$. Equation (26) acts as the observation equation in the DLM. We can write the evolution equation

$$\theta_{t,j} = \theta_{t-1,j} + \delta_t \quad \delta_t \sim \mathcal{N}(0, \Delta_{\theta_{t,j}}). \quad (27)$$

It is easy to see that this is equivalent to the prior specification in Eq. (7). For each iteration of the Gibbs sampler for the ideal point model, for each justice, we invoke the DLM algorithm

outlined above to simulate $\{\theta_{t,j}\}_{t=1}^{\bar{T}_j}$. The correspondence between the DLM notation and the dynamic ideal point model is as follows: $\mathbf{y}_t \Leftrightarrow \mathbf{z}_{t,\cdot,j} - \boldsymbol{\alpha}_t$; $\mathbf{X}_t \Leftrightarrow \boldsymbol{\beta}_t$; $\gamma_t \Leftrightarrow \theta_{t,j}$; $\boldsymbol{\Sigma}_t \Leftrightarrow \mathbf{I}_{|k_t|}$; $\mathbf{Z}_t \Leftrightarrow \mathbf{I}_{|k_t|}$; $\boldsymbol{\Delta}_t \Leftrightarrow \Delta_{\theta_{t,j}}$; $\mathbf{m}_0 \Leftrightarrow m_{0,j}$; $\mathbf{C}_0 \Leftrightarrow C_{0,j}$. Recall that $|K_t|$ represents the number of cases in a given term, which changes over time. This changing dimensionality of the DLM does not influence the estimation of the model because of the assumed independence of the observational and evolution variances across cases.

Interestingly, these computations can be done in parallel. We fit the models below using a serial processor machine, but the time necessary to fit this model could be reduced dramatically (approaching a factor of $1/J$) by using a parallel computer. It is important to note that this MCMC scheme is very computationally expensive. For each iteration of the algorithm, we have to simulate the latent utilities for each justice on each case, the case parameters for each case, and fit a separate DLM for each justice. For this reason, the models below are estimated in C++ using the Scythe Statistical Library (Martin and Quinn 2001b).

5 Results

We now turn to the data from the Supreme Court from 1953 through 1999 and present ideal point estimates and other quantities of interest from two models. The modeling strategy discussed above is quite general and can be used in many settings, including other courts, legislative bodies, and committees. In principle, it also can be used for an issue space of any dimensionality. For our application, however, we restrict our attention to the unidimensional case. This is an assumption made in nearly all statistical analyses of Supreme Court behavior. Additionally, Martin and Quinn (2001a) demonstrate that for the Burger Court from 1981 to 1985, approximately 93% of all cases fall on a single dimension. In this application, we employ the following priors on the case parameters: $\mathbf{b}_0 = \mathbf{0}$ and $\mathbf{B}_0 = \mathbf{I}_2$ for all cases k .

5.1 The Constant Ideal Point Model

We begin by summarizing the posterior distribution for a model when we assume the ideal points are constant, that is, $\Delta_{\theta_{t,j}} = 0$. This serves as the baseline for the dynamic ideal point model that we present next. Estimation of this model is straightforward and can be done using the Gibbs sampling algorithm of Albert (1992) and Clinton et al. (2000). This does not require using the DLM approach, which speeds computation time considerably.

We summarize the posterior densities of the ideal points in Table 1.¹¹ Columns 2 and 3 in the table contain the posterior means and posterior medians for the justices. It is worth noting that the scale ranges from Douglas—a well-known liberal—on the far left (-6.656) to Clarence Thomas on the far right (3.884). These results are *prima facie* 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 fourth column, can be thought of as standard errors. For the most part, these standard deviations are 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 the estimates are somewhat less precise for extremists. This is intuitive, as there are more nonunanimous cases that divide justices at the middle of the ideological spectrum and, hence, provide more information regarding the location of the latent ideal points. The final two columns in Table 1 contain

¹¹These models were run for 200,000 Gibbs scans after 5000 burn-in scans. On a dedicated Pentium 933 Linux workstation this took 11 hr to estimate. Standard diagnostics performed on the posterior samples suggests that the chain has reached steady state.

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

<i>Justice</i>	<i>Post. mean</i>	<i>Post. median</i>	<i>Post. SD</i>	<i>BCI^a</i>	
				<i>2.5%</i>	<i>97.5%</i>
Harlan	1.528	1.523	0.116	1.315	1.761
Black	-1.272	-1.270	0.093	-1.460	-1.097
Douglas	-6.656	-6.648	0.607	-7.884	-5.506
Stewart	0.372	0.372	0.050	0.274	0.469
Marshall	-2.019	-2.015	0.105	-2.236	-1.823
Brennan	-1.618	-1.616	0.073	-1.764	-1.481
White	0.420	0.419	0.042	0.340	0.504
Warren	-1.407	-1.404	0.101	-1.611	-1.211
Clark	0.194	0.195	0.064	0.068	0.320
Frankfurter	1.047	1.044	0.109	0.847	1.270
Whittaker	1.040	1.035	0.135	0.790	1.322
Burton	1.179	1.175	0.151	0.894	1.484
Reed	1.106	1.095	0.198	0.740	1.509
Fortas	-1.580	-1.569	0.228	-2.044	-1.160
Goldberg	-1.024	-1.023	0.187	-1.403	-0.675
Minton	0.802	0.798	0.191	0.442	1.197
Jackson	0.835	0.814	0.380	0.139	1.676
Burger	1.558	1.556	0.097	1.372	1.755
Blackmun	-0.039	-0.038	0.042	-0.122	0.044
Powell	0.818	0.817	0.064	0.696	0.948
Rehnquist	2.996	2.987	0.180	2.664	3.367
Stevens	-0.555	-0.556	0.050	-0.654	-0.461
O'Connor	1.325	1.323	0.082	1.172	1.491
Scalia	2.446	2.436	0.196	2.089	2.862
Kennedy	1.345	1.345	0.105	1.146	1.553
Souter	0.245	0.244	0.082	0.085	0.407
Thomas	3.884	3.835	0.445	3.106	4.851
Ginsburg	-0.179	-0.179	0.108	-0.398	0.026
Breyer	-0.156	-0.155	0.114	-0.386	0.067

^aBayesian credible interval.

the 95% Bayesian credible intervals for the ideal points, which again illustrates the precision with which they are estimated.

How does this measure compare to existing measures of judicial preferences? We first correlate our measure with the percentage conservative voting on civil rights, civil liberties, economics, and federalism cases across a justice's 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 , -0.83 , -0.70 , and -0.60 , respectively ($J = 29$). These clearly outperform the scores of Segal and Cover (1989), which correlate at 0.63 , 0.60 , 0.43 , and 0.33 ($J = 29$). This justifies the unidimensionality assumption because one dimension explains a good deal of variance across many issue areas. The measure also correlates highly with Schubert's (1974) C-scale (0.77) and E-scale (0.56) ($J = 17$) and Rohde and Spaeth's (1976) freedom (-0.79), equality (-0.75), and New Deal (-0.85) scales ($J = 18$). 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 thus turn to our dynamic ideal point model.

5.2 *The Dynamic Ideal Point Model*

For our dynamic ideal point model it is necessary to posit the specific prior distributions employed. We use the same prior distribution for the case parameters as in the constant ideal point model. For the ideal points, we set the prior mean for the ideal points $m_{0,j}$ to 0 for all justices except Harlan, Douglas, Marshall, Brennan, Frankfurter, Fortas, Rehnquist, Scalia, and Thomas. Their prior means at time 0 were set to 1.0, -3.0, -2.0, -2.0, 1.0, -1.0, 2.0, 2.5, and 2.5, respectively. The prior variances $C_{j,0}$ were set to 1 for all justices except the aforementioned nine, which were set to 0.1. These informative priors resolve the rotational invariance problems. Nonetheless, they are not overly restrictive, as they are priors for only the first term in which a justice served. The more important parameters are the evolution variances $\Delta_{\theta_{t,j}}$ of the ideal points. The magnitude of the evolution variances determines the amount of smoothing that takes place. As $\Delta_{\theta_{t,j}}$ gets smaller relative to the variance of the observation equation, the amount of smoothing increases. As noted below, we have tested our substantive conclusions for other values of the smoothing parameter. We set these equal to 0.1 for all justices except Douglas. Because of the small number of cases that Douglas heard toward the end of his career and his ideological extremity, it was necessary to use a more informative value of $\Delta_{\theta_{t,j}} = 0.001$ to help identify his sequence of ideal points.

One may wonder to what extent our results depend on our choice of prior distributions. In particular, one may wonder under what range of values for $\Delta_{\theta_{t,j}}$ evidence of temporal change in ideal points still remains. The short answer is that the ideal point series of several justices still display significant trending even when $\Delta_{\theta_{t,j}}$ is set as small as 0.01. This is 1/100 of the observation equation variance, which indicates that a large amount of smoothing is occurring. We take this to indicate that there is strong evidence that some justices' ideal points are changing over time. As the values of $\Delta_{\theta_{t,j}}$ get close to 0.001 (1/1000 the observation equation variance) the ideal point series tend to become fairly flat. At the other extreme, we have fit an independent ideal point model where no smoothing takes place, and many justices exhibit temporal change in ideal points.

In Fig. 1 we summarize the posterior densities of the justices' ideal points throughout their careers.¹² The thick, dark line represents the posterior mean, and the light lines are ± 2 posterior standard deviations away from the posterior mean. From this figure, it is clear that many justices trend. Black, Frankfurter, and Thomas trend toward conservatism, while Marshall, Brennan, Blackmun, Stevens, and Souter trend toward liberalism. White was more conservative toward the end of his career than he was at the beginning of his career. Blackmun, a Nixon appointee who by the time of his retirement was one of the most liberal justices on the Court, was 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 move over time, although our results are substantively quite different.¹³ They find that neither Marshall nor Brennan trend, while we find that they do.

¹²This model was run for 200,000 scans after 20,000 burn-in scans. On a dedicated Pentium 933 Linux workstation this took over 6 days to estimate. Standard convergence tests suggest that the chain has reached steady state. Term-by-term posterior means and standard deviations for each justice are available in the Web appendix.

¹³We can make no claims about the percentage of justices who trend over time, because our data truncates justices on the left who served before 1953 and truncates those on the right in 1999. We will thus never overestimate the amount of trending, but we might underestimate it. Without data that extends back to the Founding, this problem is intractable.

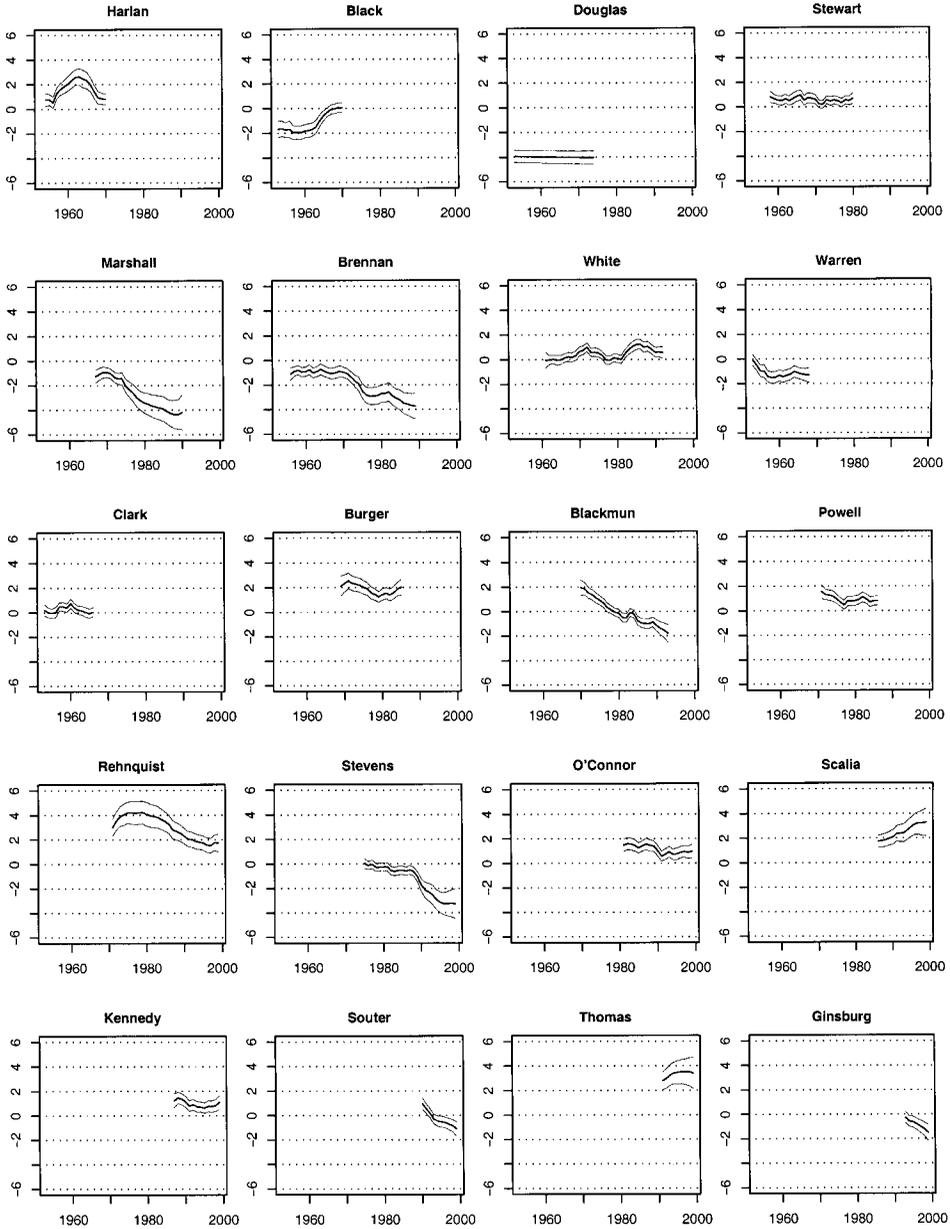


Fig. 1 Posterior density summary of the ideal points of selected justices for the terms in which they served for the dynamic ideal point model.

Similarly, they find trends for Warren, Clark, and Powell that we do not. By controlling for case stimuli and accounting for estimation uncertainty, we reach substantively very different conclusions.

We might also be interested in the posterior probability that a given justice's ideal point in one term is greater than that justice's ideal point in another term. This quantity is easily calculated from the MCMC output. As an illustration, we compute the posterior probability

that a justice's ideal point at time t is greater than his ideal point in his first term. The probability that Blackmun's ideal point in 1980 is farther to the right than his ideal point in his first term is 0.00. The posterior probability that Stewart's ideal point in 1979 is greater than his ideal point in 1958 is 0.14. This demonstrates that Blackmun did in fact become significantly more liberal over the course of his career, while Stewart did not. Another interesting example is Justice Harlan. The probability that his 1964 ideal point is farther to the right than in his first term (1954) is 1.00, while the probability that his 1970 ideal point is greater than his 1954 ideal point is 0.55.

How does this measure fare against the Epstein and Mershon (1996) validity test? In short, extremely well. We present the term-by-term correlations between the posterior mean ideal points from the dynamic model with percent conservative decisions in Fig. 2. To make the findings consistent with Epstein and Mershon (1996), the term-by-term measures are correlated with the percent conservative decisions over each justice's entire 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 cases. It also does quite well in economics cases, which is the major weakness of the Segal and Cover scores (1989). It does well in federalism cases except for a few terms around 1970 and throughout

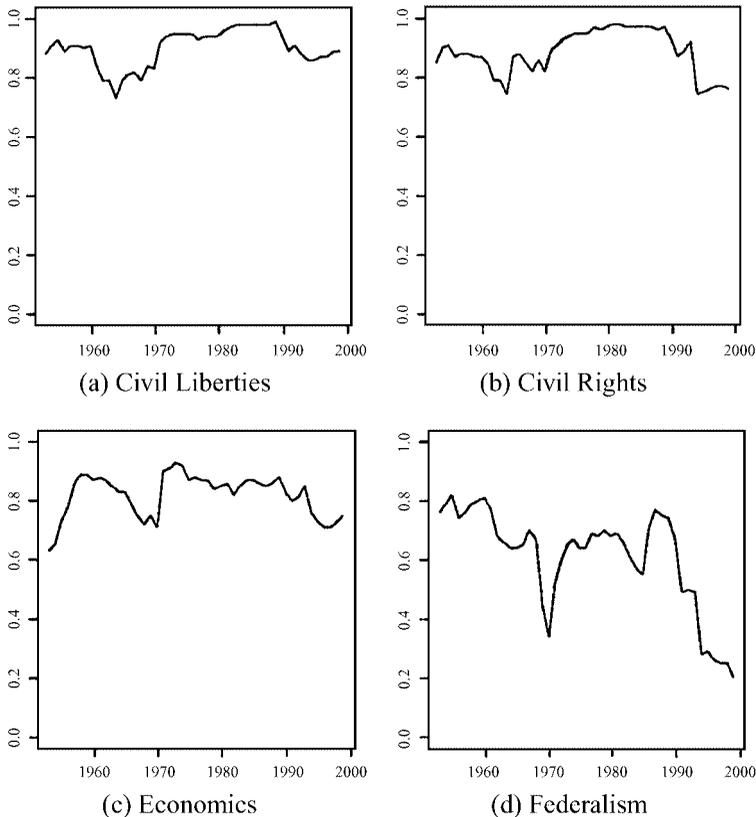


Fig. 2 Term-by-term correlations of the ideal point posterior mean for the dynamic ideal point model with the percentage conservative votes in civil liberties, civil rights, economics, and federalism cases, 1953–1999.

the 1990s. This may be suggestive of a second issue dimension that structures decision making in federalism cases during this time period or, alternatively, may be due to the large amount of turnover on the Court in 1970 and 1971. When comparing these correlations with those from the constant model, it is clear that the dynamic model does a superior job. This measure not only outperforms the Segal and Cover (1989) scores but also does better than the constant ideal point model, based on the Epstein and Mershon (1996) criterion.

To get a sense of the adequacy of this model we examine the mean percentage of cases correctly classified by the model. Given the posterior distribution, these percentages are computed by comparing the predicted votes from the model with the actual votes. The dynamic ideal point model correctly classifies 76% of the decisions. This is an improvement over the observed marginal percentage of reversals (63%). As a point of comparison, the model in which ideal points are assumed constant over time classifies 74% of the decisions correctly. While the improvement in model fit is modest when moving from the constant ideal point model to the dynamic ideal point model, we feel that the ability to test for temporal changes in the ideal points compensates for the increased computational cost of fitting the dynamic ideal point model. Moreover, the dynamic model also provides an efficiency gain over a model in which the ideal points are assumed to be temporally independent. Figure 3 depicts this increase in efficiency. This figure plots the posterior standard deviations of the ideal points from our dynamic model against the posterior standard deviations of the ideal points from a model that assumes independent ideal points. Note that the precision of the estimated ideal points is almost always greater in the dynamic model.

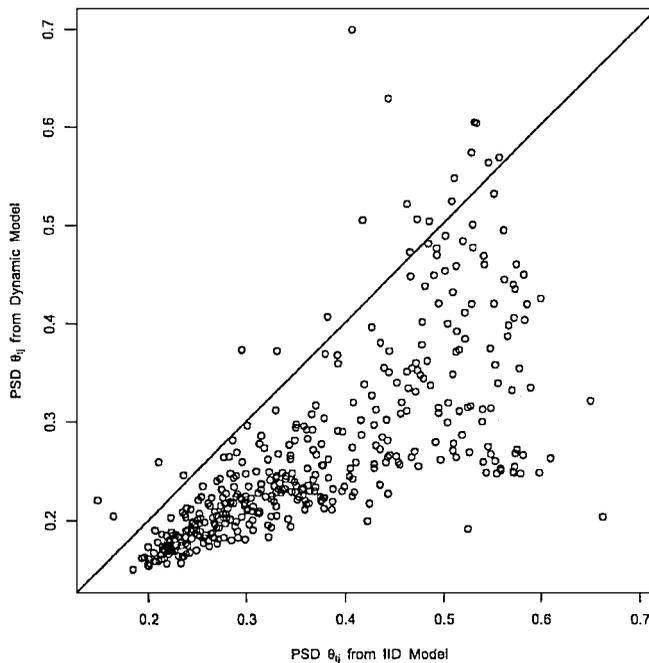


Fig. 3 The efficiency gains of the dynamic model over a model in which the ideal points are assumed independent and identically distributed. PSD, posterior standard deviation.

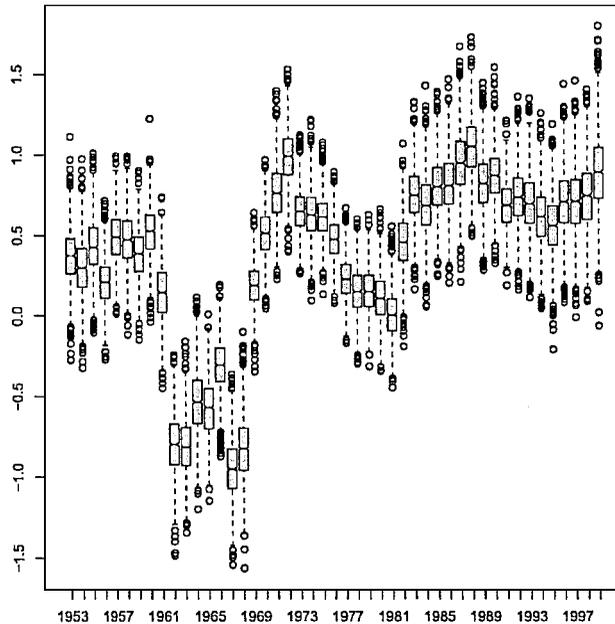


Fig. 4 Estimated posterior distribution of the location of the median justice for the dynamic ideal point model.

5.3 Other Quantities of Interest

One advantage of performing Bayesian inference is that it is straightforward to summarize the posterior density of quantities of interest other than the model parameters. One quantity of great substantive importance is the location of the median justice. This is valuable to study not only 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. We first compute the probabilities that a given justice is the median in each term.¹⁴ This analysis suggests that Frankfurter was 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, Powell, and Souter are medians in certain terms. Finally, in the 1990s, O'Connor and Kennedy are the median justices in different terms.

The identity of the median justice is interesting, but so too is the location of the median on the estimated issue dimension. In Fig. 4 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, with the addition of 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 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

¹⁴These complete results are available in the Web appendix.

moderate during the Bush administration, and then became somewhat more conservative during the Clinton years. Our model suggests that the most conservative Court sat in the 1988 term, and the most liberal sat in the 1967 term.

6 Discussion

The results from these models tell us a great deal about the Supreme Court from 1953 to the present. Not only have we estimated superior measures of judicial preferences, but we provide strong evidence that the ideal points of many 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. For most justices, this change is monotonic. Our results imply 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. In addition to presenting posterior standard deviations for all measures, we also identify and locate the median justice on the Court during each term.

The paper also highlights a useful modeling strategy for dynamic social systems. Using DLMS in a Bayesian framework makes this relatively easy; substantive assumptions can be directly translated into our models. DLMS are an important methodological tool that other political scientists can use to model dynamic phenomena of interest. For nearly any statistical model with the usual conditional independence assumptions, DLMS can be used to smooth parameters over time. This is a happy medium between complete pooling of parameters estimated over time by assuming parameter homogeneity and estimation of separate models for each time point. They are particularly useful for cases when the number of observations in each time period is small or for when observations from some time periods are missing. MCMC methods make estimating these models tractable.

This research is only the first step of a broader 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 actors under study*. Another important step is explaining the causes of the dynamics we observe. Why do justices' ideal points change over time? Is this a psychological process, or do revealed preferences change with different compositions of the Court or, perhaps, because of separation of powers concerns? It may also be the case that we can get better estimates of the case-specific parameters and the ideal points by using additional information about the cases, such as the lower court of origin or the issue area of the case. Including these types of covariates in the above models is a logical extension. Finally, there are 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.

References

- Albert, James H. 1992. "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling." *Journal of Educational Statistics* 17:251–269.
- Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88:669–679.
- Bailey, Michael, and Kelly H. Chang. 2001. "Comparing Presidents, Senators, and Justices: Interinstitutional Preference Estimation." *Journal of Law, Economics, & Organization* 17:477–506.

- Baum, Lawrence. 1988. "Measure Policy Change in the United States Supreme Court." *American Political Science Review* 82:905–912.
- Beck, Nathaniel. 1989. "Estimating Dynamic Models Using Kalman Filtering." *Political Analysis* 1:121–156.
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.
- Carter, C. K., and R. Kohn. 1994. "One Gibbs Sampling for State-Space Models." *Biometrika* 81:541–533.
- 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.
- 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* 60:801–818.
- 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* 40:261–294.
- 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* 33:825–841.
- Epstein, Lee, Jeffrey A. Segal, Harold J. Spaeth, and Thomas G. Walker. 2001. *The Supreme Court Compendium: Data, Decisions, and Developments*, 2nd ed. Washington, DC: CQ Press.
- Eskridge, William N. 1991. "Overriding Supreme Court Statutory Interpretation Decisions." *Yale Law Journal* 101:331–417.
- 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* 9:227–241.
- Johnson, Valen E., and James H. Albert. 1999. *Ordinal Data Modeling*. New York: Springer.
- Londregan, John. 2000. "Estimating Legislators' Preferred Points." *Political Analysis* 8:35–56.
- 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* 24:146–178.
- Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll-Call Voting*. Oxford: Oxford University Press.
- Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco: W. H. Freeman.
- Schubert, Glendon. 1965. *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946–1963*. Evanston, IL: 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* 83:557–565.
- Segal, Jeffrey A., and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge: Cambridge University Press.
- Spaeth, Harold J. 2001. *United States Supreme Court Judicial Database, 1953–2000 Terms* [Computer File], 16th ed. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.
- Stimson, James A. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29:914–947.
- 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* 35:460–480.
- West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer.