Assessing Preference Change on the US Supreme Court

Andrew D. Martin[*]
Washington University School of Law

Kevin M. Quinn[**]
Harvard University

The foundation upon which accounts of policy-motivated behavior of Supreme Court justices are built consists of assumptions about the policy preferences of the justices. To date, most scholars have assumed that the policy positions of Supreme Court justices remain consistent throughout the course of their careers and most measures of judicial ideology—such as Segal and Cover scores—are time invariant. On its face, this assumption is reasonable; Supreme Court justices serve with life tenure and are typically appointed after serving in other political or judicial roles. However, it is also possible that the worldviews, and thus the policy positions, of justices evolve through the course of their careers. In this article we use a Bayesian dynamic ideal point model to investigate preference change on the US Supreme Court. The model allows for justices' ideal points to change over time in a smooth fashion. We focus our attention on the 16 justices who served for 10 or more terms and completed their service between the 1937 and 2003 terms. The results are striking—14 of these 16 justices exhibit significant preference change. This has profound implications for the use of time-invariant preference measures in applied work.

1. Introduction

Do the revealed preferences of Supreme Court justices change over time?[1] The answer to this question is of profound importance to both policymakers and academics. When nominating someone to the US Supreme Court, presidents typically want to appoint a like-minded individual who will hold his or her ideological course for the entirety of his or her life term. Similarly, before voting on a nominee, senators need to form an expectation of how that person will decide cases over as many as the next 20–30 years. If justices tend to exhibit temporally stable revealed preferences it is *relatively* easy to form such expectations of future behavior and we might expect the ideological makeup of the Court to be extremely reflective of the balance of power between the Democratic and Republican Parties at the times of nomination.

The temporal stability of revealed judicial preferences is also of great importance to scholars of judicial politics. If the assumption of preference stability does not hold, then the findings of all studies that rely on time-invariant measures of justice preferences may be called into question. We note that this includes a very large fraction of judicial politics work appearing in the top journals. But for the study of Epstein et al. (1998) no systematic empirical analysis has determined the extent to which judicial preferences change over time.

In this article we employ a Bayesian dynamic ideal point model developed by Martin and Quinn (2002) to estimate revealed preferences for all Supreme Court justices serving between 1937 and 2003. This model allows us to separate the effects of case content from the effects of justice-specific policy positions and estimate ideal points that are on a comparable scale over time. Furthermore, because these ideal point estimates are based on a statistical measurement model, we can gauge uncertainty of the estimates and other quantities of interest. We use the results from this model to demonstrate that the revealed preferences of Supreme Court justices are far from stable.

The current article differs from that of Martin and Quinn (2002) in that it is solely concerned with assessing whether or not the revealed preferences of Supreme Court justices change over time, whereas the article by Martin and Quinn (2002) focuses on the general modeling strategy that makes the present article possible. The current article takes the model developed in Martin and Quinn (2002) as a starting point and examines the issue of preference change in a more systematic and comprehensive fashion than was done in Martin and Quinn (2002). It also extends the model by including data back to the 1937 term.

---

1. Throughout this article we use the terms preferences and ideal points interchangeably to mean *revealed preferences*. By this we mean the preferred policy positions defined in an issue space revealed through the votes of the justices (Epstein and Mershon 1996). It is important to note that this notion of revealed policy preferences is conceptually very different from personal policy preferences or attitudes (Segal and Spaeth 1993). The revealed preferences that we estimate on a policy scale are likely caused by any number of factors, including personal attitudes, the decision context, and so forth. The purpose of this article is to document the change in revealed preferences.

We begin this article by reviewing the literature and arguing that the Epstein et al. (1998) study has methodological limitations that restrict its ability to uncover preference change. We then discuss the measurement model we employ, highlighting its ability to estimate preferences while controlling for changes in case stimuli. In Section 4, we discuss research design and present the evidence for preference change. The final section concludes.

2. A Methodological Critique of the Literature

Personal policy preferences, or attitudes, are key explanatory variables in attitudinal accounts of judicial behavior (Segal and Spaeth 1993). If the attitudinal model is true, then the revealed preferences, or ideal points, of the justices will correspond to their personal attitudes. Strategic accounts of Supreme Court decision making also make use of preferences for explaining interdependent behavior. As Epstein et al. (1998) note, the prevailing wisdom in the study of judicial behavior is that, "[t]he occasional anomaly notwithstanding, most jurists evince consistent voting behavior over the course of their careers" (801).

2.1 The Assumption of Constant Preferences

The attitudinal model (Schubert 1974; Rhode and Spaeth 1976; Segal and Spaeth 1993) asserts that justices have personal attitudes and that case material provides stimuli that trigger the justices' attitudes and consequently their decisions. The model does not *explicitly* assume that attitudes are fixed. However, nearly all empirical work related to the attitudinal model employs constant measures of attitudes. The most commonly used measures are Segal and Cover (1989) scores, which are based on newspaper editorials at the time of confirmation. Others use measures such as the party identification of the justice (George and Epstein 1992) or measures of social background (Tate 1981). In some areas, such as civil rights and civil liberties, these time-invariant measures are shown to be quite successful in accounting for votes, but in others, like economics and federalism cases, their performance is much less impressive (Epstein and Mershon 1996). An alternative explanation of behavior—the strategic model (Eskridge 1991; Epstein and Knight 1998)—asserts that justices have policy preferences and pursue their interests in an interdependent choice situation. Although, this model does not necessarily assert that preferences are constant, nearly all empirical work in this genre employs measures where this is the case, most notably Eskridge (1991) and Segal (1997) who use Segal and Cover (1989) scores, and Spiller and Gely (1992) who use the party of the appointing president.

The key point to take from the literature is that the assumption of constant preferences in not a theoretical one per se, but rather is chosen for empirical convenience. It is somewhat surprising, then, that little systematic research has been conducted to determine whether or not the assumption is consistent with the data. The anecdotal evidence suggests that preference change sometimes

occurs (see, e.g., Ulmer 1981; Atkins and Sloope 1986, as well as accounts in the law reviews and the popular press).

These anecdotal accounts are suggestive, but to draw definite conclusions it is necessary to systematically study the behavior of many justices over time. The first to do so was Baum (1988), who was primarily interested in *policy* change on the Court. Although he claims there may be some preference change, he concludes that case stimuli, not preference change, are what explain the observed dynamics. In the only study with the goal of assessing preference change, Epstein et al. (1998) look at all 16 justices who served 10 or more terms and served their entire career between 1937 and 1993.[2] They contend that it is vital to look at justices who have served for a long period of time and to only look at justices who have completed their entire service in the time period (otherwise, one could underestimate the number of justices who demonstrated significant preference change). To measure preferences, they argue that votes are the best place to look (Epstein and Mershon 1996) and use the Baum-corrected (1988) percentage of liberal votes on civil liberties as their measure of policy preferences. Given this measure, they fit linear, quadratic, and cubic regressions of preferences on time. They find that seven justices exhibit no significant preference change (Brennan, Burger, Burton, Harlan, Jackson, Marshall, and Stewart), four exhibit linear trends (Blackmun, Clark, Reed, and White), and five exhibit nonlinear change (Black, Douglas, Frankfurter, Powell, and Warren). Their conclusion is that preference change is significant and that it should be accounted for in future studies. However, these results depend on an ideal point estimator (a justice's Baum-corrected percentage of votes in the liberal direction on civil liberties cases) that is not well suited to the task at hand. As Baum (1988) notes, one of the three assumptions on which his method is based is "each justice's ideal point on the civil liberties dimension remains *constant throughout the justice's career*" (907, italics added).

2.2 The Baum Correction

The Baum (1988) correction is the tool Epstein et al. (1998) employ to tie together estimates of ideal points throughout time. To formalize the problem of dynamic ideal point estimation, let $\theta_{t,j} \in \Re$ denote the ideal point or policy position of justice j in term t. Furthermore, without loss of generality, let $x_k^{(l)}, x_k^{(r)} \in \Re$ with $x_k^{(l)} \leq x_k^{(r)}$. If $x_k^{(l)} < x_k^{(r)}$ we can think of $x_k^{(l)}$ and $x_k^{(r)}$ as the locations of the liberal and conservative policy alternatives for case k, respectively. These two case parameters contain the information about the policy content of each case. The midpoint between these two policy locations $(x_k^{(l)} + x_k^{(r)})/2$ (also called the indifference point) determines the manner in which the justice votes.[3] If, for example, $x_k^{(r)}$ is strictly greater than $x_k^{(l)}$,

2. Bailey and Chang (2001) allow for preference change in their cross-institutional measurement model based on the findings of Epstein et al. (1998).

3. The midpoint is determinative under the assumption of symmetric utility functions, such as with the commonly used quadratic utility function (Enelow and Hinich 1984), or a Gaussian utility function (Poole and Rosenthal 1997).
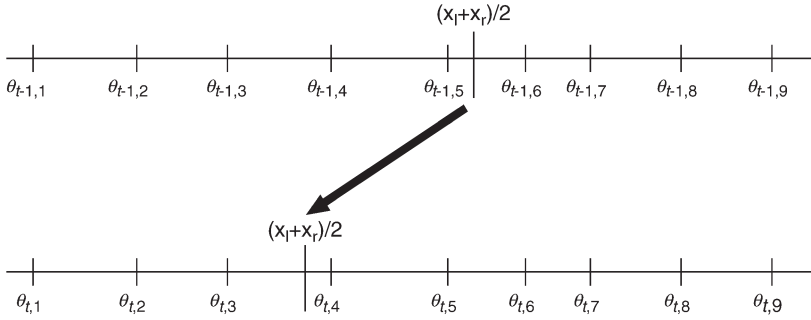
Figure 1. An illustration of policy change (Baum 1998) and the spatial voting model.

then those to the left of the midpoint will be more likely to vote for the liberal option and those to the right will be more likely to vote in the conservative direction.

One approach to estimating $\theta_{t,j}$ is to take the raw average of the number of liberal decisions made by justice j in term t. This approach is not without problems. Figure 1 contains an illustration of two hypothetical configurations of preferences in terms $t-1$ and t for nine justices. In the top line of the figure in term $t-1$ the midpoint falls between Justice 5 and 6. In the second line, the location of the midpoint has changed and now falls between Justice 3 and 4. Notice that in the figure the preferences remain the same, but the observed vote would change (it would be 5–4 in the first case, and 6–3 in the second). Thus, computing a raw average across a set of cases could be misleading unless the changes in case stimuli are controlled for.

Baum recognizes this fact and offers a correction to account for it. In his case, between two natural Courts (or, for Epstein et al. [1998], between terms), one computes the median change in the percentage of liberal votes made by each justice and then takes the median of these differences across all justices serving in those natural Courts (terms). The Baum correction constructs an ideal point estimate $\hat{\theta}_{t,j}^{\text{Baum}}$ by subtracting this median difference from each justice's percentage of liberal decisions. More formally,

$$\hat{\theta}_{t,j}^{\text{Baum}} = \hat{\theta}_{t,j}^{\text{raw}} - \underset{j' \in J_{(t-1):t}}{\text{median}}\{\hat{\theta}_{t,j'}^{\text{raw}} - \hat{\theta}_{t-1,j'}^{\text{raw}}\},$$

where $\hat{\theta}_{t,j}^{\text{raw}}$ is the percentage of liberal decisions made by justice j in time period t and $J_{(t-1):t}$ is the set of justices who served in both time period t and time period $t-1$. The Baum correction thus cleanses case content from justices' voting behavior by assuming that preferences are fixed and that any dynamics are solely in the case parameters.

As Baum notes, this correction is only appropriate if preferences are temporally constant. Indeed, by inspecting Figure 1, it is clear that if preferences were also allowed to move freely, it would be impossible to determine whether the derived correction was explained by changes in the case stimuli or changes

in preferences. Without additional modeling, the two are in fact conflated. Observed changes in Baum-corrected percent liberalism measures will tell us that *something* changed (either the ideal points or the case parameters), but it cannot tell us *which* changed.[4] Thus, since one of the major assumptions underlying the Epstein et al. (1998) study is inconsistent with the primary research goal of that study, one should view the results of Epstein et al. (1998) with some caution.

There are other limitations that call the Epstein et al. (1998) findings into further question. First, the authors treat their estimates as if they are known with certainty. However, they are estimates and thus have some estimation uncertainty attached. It is well known that failing to account for this uncertainty in the dependent variable will bias standard errors (SEs) downward. To test for significant preference change, Epstein et al. (1998) estimate linear, quadratic, and cubic models of ideal points regressed on time. From these regressions they draw the conclusions highlighted above. Not only does this assume that the variance of the regression disturbances are constant (which is unlikely to be the case given the underlying probability model) but it also rests heavily on the parametric assumption that ideal points follow low-order polynomials in time. If a justice's preferences change for only a small subset of terms, the slope estimates will be attenuated toward zero and thus biased against finding preference change. In short, even if the Baum correction was accurate, the Epstein et al. (1998) method of assessing preference change may be overly liberal by treating ideal point estimates as known quantities and may be overly conservative by estimating global models of preference change.

3. Bayesian Dynamic Ideal Point Estimation

From this review it is clear that the question of whether or not the preferences of Supreme Court justices change over time is still open. It is also clear that to answer the question one requires a statistical model that allows for ideal points to exhibit a wide range of dynamics, based on a parametric statistical model that *simultaneously* estimates case stimuli and ideal points. Measures of uncertainty should be reported, and accounted for in diagnosing preference change. Further, tools other than global regression models are needed to assess the amount and magnitude of preference change.

The model we employ begins with two assumptions. First, justices vote in accordance with the spatial model outlined in the previous section. That is, all votes can be explained solely by considering the ideal points of the justices and the case stimuli. Second, we assume that a single issue dimension structures all

---

4. Moreover, there is no theoretical justification for using the median change as the correction. The justice who exhibits the median change in percent liberalism will typically not be the median justice on the Court. Even if that were the case, it is clear that the median justice does not always prevail on the Supreme Court, as we observe many votes of 6–3, 7–2, and so forth.

decision making from 1937 to the present.[5] With these two assumptions and a further assumption that disturbances to the latent utility of voting are independent Gaussian random variables, one can show that a standard two-parameter item response model can be used to estimate both case parameters and the ideal points from voting data (Clinton et al. 2004). This model, which we call a constant ideal point model, assumes that ideal points are time invariant.

Martin and Quinn (2002) extend this model and propose a dynamic ideal point model that allows for preferences to change over time. Conceptually, this model estimates all the case parameters from a distribution common to all terms and an ideal point for each justice in each term on a comparable scale across terms. The model is formalized as follows. Let $K_t \subset \{1, 2, \ldots, K\}$ denote the set of cases heard by the Supreme Court in term t, and $J_k \subset \{1, 2, \ldots, J\}$ denote the set of justices who heard case k. The cardinality $|J_k|$ denotes the number of justices sitting on a case k, which is typically nine, but fewer in certain cases. 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 unidimensional issue space. We code all votes in term t on case k by justice j as either being in favor of the conservative option ($v_{t,k,j} = 1$) or the liberal option ($v_{t,k,j} = 0$). The observed data matrix $\mathbf{V}$ is thus a $(K \times J)$ matrix of votes and missing values. We note that, for reasons that will become apparent below, it matters not whether we code votes as liberal/conservative, majority/minority, affirm/reverse, and so forth.

The spatial model suggests that the ideal points of the justices $\theta_{t,j}$ for the jth justice in term t, and the case stimuli $x_k^{(l)}$ and $x_k^{(r)}$, determine votes on the merits. These are the quantities we wish to make inferences about from the data. To do so, let $z_{t,k,j}$ denote the difference between the latent random utility of voting for the conservative policy and the latent random utility of voting for the liberal policy. We expect that this latent random utility explains the votes on the merits in the following fashion:

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

_____

5. Both of these are strong assumptions, but ones that can be tested with the data. If the statistical model we propose fits well, then the assumptions are reasonable. Moreover, there are a handful of reasons to suspect that the judicial policy space from 1937 to 2003 is unidimensional. Using multidimensional scaling techniques, Grofman and Brazill (2002) show that between 1953 and 1991 a single dimension best explains Supreme Court decision making. To assess dimensionality throughout the time period of our study, we have computed the eigenvalues of the double-centered agreement score matrix for each term. Poole and Rosenthal (1997) suggest that the number of eigenvalues greater than 1 is suggestive of the true underlying dimensionality. For every term, the second eigenvalue is less than 1, which suggests that a unidimensional model is appropriate. Moreover, the policy dimension we estimate is highly correlated with the many (nonorthogonal) scales uncovered by Schubert (1974) and Rohde and Spaeth (1976). This evidence, along with the model fit discussed below, justifies these simplifying assumptions.

The spatial model suggests that (Clinton et al. 2004; Martin and Quinn 2002):

$$z_{t,k,j} = \alpha_k + \beta_k \theta_{t,j} + \varepsilon_{t,k,j} \tag{2}$$

where $\alpha_k = [x_k^{(l)} x_k^{(l)} - x_k^{(r)} x_k^{(r)}]$, $\beta_k = 2[x_k^{(r)} - x_k^{(l)}]$, and $\varepsilon_{t,k,j}$ is a random error term which we assume is homoscedastic with known variance.[6] The two case parameters α_k and β_k characterize the case characteristics. More specifically, the ratio $-\alpha_k/\beta_k$ is the midpoint between $x_k^{(l)}$ and $x_k^{(r)}$. The policy position for justice j in term t is $\theta_{t,j}$. This model differs from that standard two-parameter item response model in that these ideal points are allowed to change over time.

To complete the model it is necessary to assign prior distributions to all parameters. We begin by assuming standard Gaussian prior distributions for the case parameters[7]:

$$\begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} \sim N_2(\mathbf{0}, \mathbf{I_2}) \quad \forall\, k \in \{1, 2, \ldots, K\}. \tag{3}$$

The Martin and Quinn (2002) model departs from the standard approach in the prior distribution over the ideal points by assuming that the idea points follow a random walk process. This allows for, but does not force, a wide range of smooth dynamics for the ideal points. We model these dynamics with a separate random walk prior for each justice:

$$\theta_{t,j} \sim N\left(\theta_{t-1,j}, \Delta_{\theta_{t,j}}\right) \quad \text{for} \quad t = \underline{T}_j, \ldots, \overline{T}_j \text{ and justice } j \text{ on the Court}, \tag{4}$$

where $\underline{T}_j$ is the first term justice j served and $\overline{T}_j$ is the last term j served. We do not estimate ideal points for terms in which a justice did not serve. $\Delta_{\theta_{t,j}}$ is an evolution variance parameter which is fixed a priori by the researcher. Its magnitude determines how much borrowing of strength (or smoothing) takes place from one time period to the next. Note that as $\Delta_{\theta_{t,j}} \to 0$, we approach a model with temporally constant ideal points. At the other extreme, as $\Delta_{\theta_{t,j}} \to \infty$, 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 zero. Here, in a slight abuse of notation, we let 0 denote time period $\underline{T}_j - 1$ for all j. We assume that:

$$\theta_{0,j} \sim N(m_{0,j}, C_{0,j}). \tag{5}$$

This random walk prior is what makes comparisons over time possible. The model is identified by fixing three justices in their first term of service (see below), thus defining a metric for measurement. The prior ties together all model estimates through time, including the case parameters, which control for possible changes in the docket.

---

6. Note that since α_k and β_k are specific to case k and only depend on the differences of policy outcomes and their squares, votes can be coded in any way that is consistent across justices within a case and the likelihood will not change.

7. We have relaxed the assumption of standard Normal priors, and assigned prior variances of $3 \cdot \mathbf{I_2}$, $5 \cdot \mathbf{I_2}$, and $10 \cdot \mathbf{I_2}$; the substantive results remain the same. This suggests that the data contain a reasonable amount of information about these case parameters.

This approach is similar to that of Berry et al. (1999).

To estimate this model, we adopt the strategy of Martin and Quinn (2002), which is based on standard item response theory (Bock and Liberman 1970; Hambleton and Swaminathan 1985; Albert 1992; Bradlow et al. 1999; Johnson and Albert 1999) and Bayesian dynamic linear models (West and Harrison 1997). The strategy uses Markov chain Monte Carlo (MCMC) methods (Jackman 2000; Gill 2002) to simulate from the posterior distribution $f(\theta, \alpha, \beta|\mathbf{V}) \propto f(\mathbf{V}|\theta, \alpha, \beta)\pi(\theta)\pi(\alpha, \beta)$. These methods allow one to simulate from a distribution that is otherwise computationally intractable.[8] There are many advantages to using Bayesian methods in the context of ideal point estimation; see Jackman (2001) and Bafumi et al. (2005) for a review. Due to the large number of parameters, maximum likelihood estimation for our dynamic ideal point model would be intractable.

Before we turn to our specific application, it is important to recognize some additional properties of this model. First, this is a fully parametric statistical model, which not only solves the fundamental problem of dynamic ideal point estimation but also allows us to report measures of uncertainty for all quantities of interest. Second, this approach does not conflate possible changes in case stimuli and ideal points. Both are estimated separately in the model: the case parameters α_k and β_k are the estimates of the case stimuli, and the $\theta_{t,j}$ are the ideal point estimates in each term. Third, this model allows for ideal points to change over time. The use of the random walk prior allows for change to take an extremely wide range of smooth forms and is *much more flexible* than assuming linear or polynomial change, such as the D-NOMINATE model of Poole and Rosenthal (1997).

4. The Evidence for Preference Change, 1937–2003

To make our results comparable to those of Epstein et al. (1998) and due to data availability we focus our attention on the Supreme Court from the 1937 to the 2003 terms. During this time period, 41 justices served ($J = 41$). We obtain data from three sources: (1) data for the 1953–2003 terms comes from the *Original United States Supreme Court Database* (Spaeth 2004); (2) data for the 1946–1952 terms comes from the *Vinson–Warren Court Database* (Spaeth 2001); (3) data for the 1937–1945 terms comes from an unpublished data set collected and used by Epstein et al. (1998).[9] This selection results in

---

8. This application is extremely computationally expensive due to the size of the problem. On a dedicated MacOS X G5 2.0 gHz workstation it took 1 week to complete 200,000 scans after 20,000 burn-in scans.

9. The 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. As discussed by Martin and Quinn (2002), dropping unanimous cases does not impact estimation of ideal points or case parameters because these cases do not contribute to the likelihood. For comparability with Epstein et al. (1998) and across the three data sources, we code all votes as either liberal or conservative. As noted above, the coding protocol is arbitrary; one would obtain identical ideal point estimates if all votes were coded affirm or reverse, minority or majority, and so forth.

$K = 4741$ total cases, the most heard in the 1972 term (108) and the fewest heard in the 2003 term (41).

To ensure a common scale for the ideal points across time, it is necessary to assign informative priors for justices that span the entire length of the study. In our case, we set the prior mean for the ideal points $m_{0,j}$ to zero for all justices except Black, Stewart, and Rehnquist, with prior means -2.0, 1.0, and 3.0, respectively. The prior variances $C_{j,0}$were set to 1 for all justices but for these three; the prior variances for these three are set to 0.1. Note that this prior is only on the first term in which the justice served. For all other terms, the ideal point in the previous term serves as the prior mean. To complete the prior we set the evolution variance $\Delta_{\theta_{t,j}} = 0.1$for all justices in all terms after their first.[10] After specifying the priors, we employ the Martin and Quinn (2002) MCMC algorithm to simulate from the posterior distribution. With the posterior sample in hand, we first performed standard convergence tests. All suggested that the chain has reached the stationary distribution.[11]

Our main quantity of interest is the ideal points of the justices. Due to space considerations, we only report the ideal point estimates of the 16 justices Epstein et al. (1998) considered. These justices are chosen because they served for 10 or more terms and because they completed their entire terms of service between the 1937 and 2003 terms inclusive.[12] The ideal point estimates, for each justice in each term in which they served, are presented in Figure 2. The large point in the middle is the posterior mean of the ideal points, and the error bars represent plus or minus two posterior standard deviations (SDs). One can think of the posterior mean as a point estimate and the posterior SD as a SE. The amount of uncertainty of the estimates depends primarily on two factors; *ceteris paribus* more extreme justices are estimated with less certainty than centrist justices, and terms with more less cases are estimated with less certainty than those with more cases.

The scale we estimate is a conservatism scale—higher values represent greater conservatism. The results in Figure 2 are striking. Many justices seem to trend over the course of their careers. Black begins his career as a liberal, but gets more conservative over time. Frankfurter and Reed also trend toward conservatism; Reed ends his career as a moderate, whereas Frankfurter ends his career as a conservative. Other justices get more liberal over time. The classic example is Blackmun, a Nixon appointee who was actually quite conservative in his first few terms. Yet at the time of his retirement in the mid-1990s, Blackmun became quite liberal and in fact was one of the most liberal

---

10. Our findings are quite robust to other prior specifications. The results look quite similar for other evolution variances. We have fit models with $\Delta_{\theta_{t,j}} = 0.01$, $\Delta_{\theta_{t,j}} = 0.25$, $\Delta_{\theta_{t,j}} = 0.5$, and $\Delta_{\theta_{t,j}} = 3.0$ and find nearly identical results. This implies that regardless of the amount of smoothing, many justices exhibit preference change.

11. We utilized the diagnostic tests of Geweke (1992) and Heidelberger and Welch (1981).

12. Ideal points are estimated for all justices, and similar comparisons can be made for those who are remained on the bench at the end of the study. Those estimates suggest that a number of other justices, including Rehnquist and Stevens, also exhibit significant preference change.
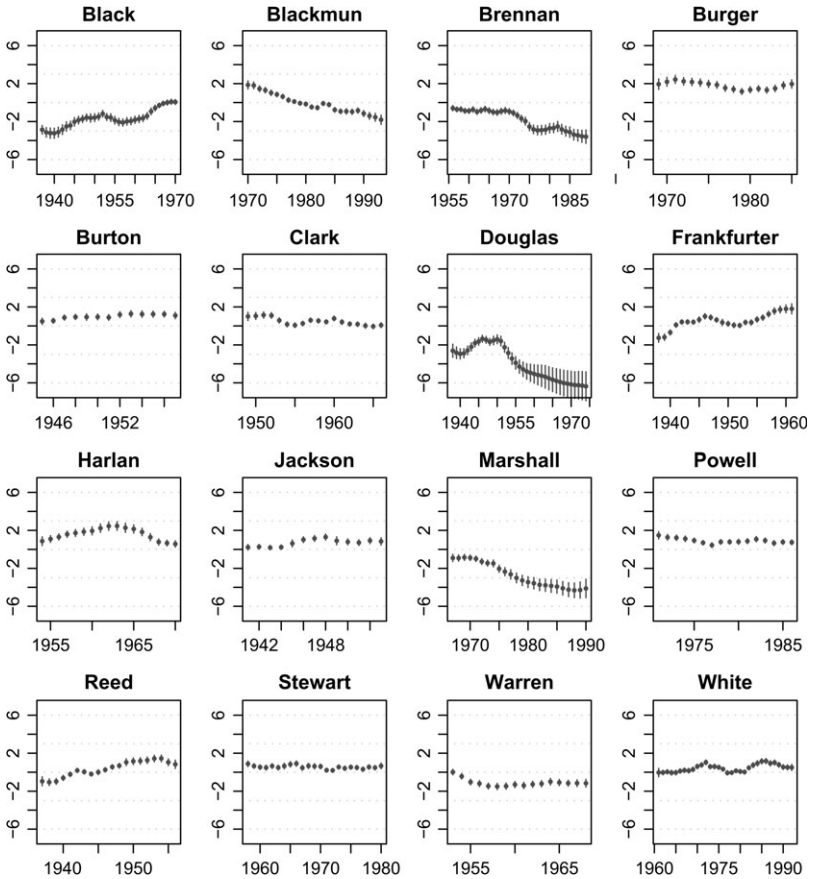
Figure 2. Estimated posterior distribution of the ideal points of selected justices from the dynamic ideal point model, 1937–2003. The *y* axis in all plots is the estimate on the ideal point conservatism scale, and the *x* axis denotes the terms in which the justice served. The centerline points the posterior mean, and the error bars are plus or minus two posterior SDs.

justices on the Court. Clark, Powell, and Warren also seem to become slightly more liberal over the course of their careers.

Changes in ideal points are not limited to directional trends. Some justices remain somewhat constant throughout the course of their careers, such as Stewart and White. Others exhibit more exotic patters. Douglas begins his career as liberal, becomes more moderate through the late 1940s and early 1950s, and then becomes increasingly more liberal through the remainder of his career. Harlan too has a parabolic shape, although the amount of change is far less dramatic than that of Douglas.

How well does the model fit? In short, quite well. In particular, these results correlate highly with percent liberalism in civil rights, civil liberties, economics, and federalism cases (Martin and Quinn 2002). This is surprising, as most
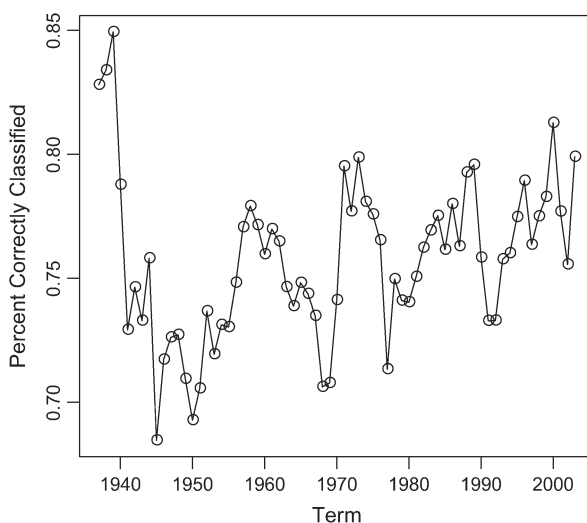
Figure 3. Term-by-term percent cases correctly classified, 1937–2002.

existing measures only fare well for civil rights and civil liberties cases (Epstein and Mershon 1996). Additionally, the model has solid explanatory power. Overall, the model correctly classifies 75.7% of the votes.[13] In Figure 3 we plot the term-by-term percent correct classification for our model. The model performed worst in the 1945 term (68.5% correctly classified) and best in the 1939 term (84.9% correctly classified). There are clearly dynamics in the classification rate, but compared to a baseline of 50% classification, the model does well. Not surprisingly, the model appears to do increasingly better in terms with stable membership; the big dips correspond to the appointment of new justices.

From Figure 2 it appears as if the preferences of Supreme Court justices change a great deal over time. However, these are quantities that are measured with uncertainty. It is important to account for that uncertainty when making claims about whether or not there is significant change in ideal points for individual justices. To assess preference change, it is also important *not* to rely on global models of change, such as linear regression models. In fact, the probability of interest is the posterior probability that a particular justice is more conservative in subsequent terms than in a baseline term. As Hagle (1993) notes, justices learn a great deal during their first term of service. Maltzman and Wahlbeck (1996) also show that justices are amenable to persuasion early in their careers. This implies that the first term of service is not a terribly

---

13. These are computed by averaging over the parameter uncertainty of the ideal points and the case parameters. Compared to other measures of model fit, such as the rank-order measures used by Poole and Rosenthal (1997) or using only the point estimates, this is a very conservative measure. Although this measure of fit can say nothing about external validity, it does show that our model is not inconsistent with the observed data.

reliable baseline. For our first comparison, we take the mean ideal point of each justice's second, third, and fourth terms of service as the baseline for comparison θ_j^*. Then, for all subsequent terms, we compute the posterior probability that the justice is more conservative than the baseline. Formally, we compute:

$$\Pr(\theta_{r,j} > \theta_j^*) \text{ for } r = 5, \ldots, \overline{T_j}, \quad \text{where } \theta_j^* = \frac{1}{3}(\theta_{2,j} + \theta_{3,j} + \theta_{4,j}), \qquad (6)$$

using a Monte Carlo estimation strategy detailed in Appendix. By using all draws from the posterior distribution, we account for all parameter uncertainty when estimating these quantities of interest.

For each of the 16 justices, we plot these posterior probabilities in Figure 4. Each cell of the figure contains dotted horizontal lines at 0.025 and 0.975. If the estimated probability is greater than 0.975, then we can conclude that the justice was significantly more conservative in that term. If the estimated probability is less than 0.025 percentile, we can conclude that the justice was significantly more liberal in that term than in the baseline term.

The results in Figure 4 are striking. Justices Black, Douglas, Frankfurter, Harlan, Jackson, Reed, and White are significantly more conservative in some subsequent terms than the baseline. Justices Blackmum, Brennan, Burger, Clark, Douglas, Harlan, Marshall, Powell, and Warren are significantly more liberal in some subsequent terms.

Justices Douglas and Harlan are significantly more conservative than the baseline in some subsequent terms and significantly more liberal than the baseline in other subsequent terms. This is consistent with the parabolic trajectories in Figure 2. But the patterns are quite different, as Harlan is only significantly more liberal in his final term than the baseline, whereas Douglas is significantly more liberal for well over his final decade. Another interesting pattern is White, who is significantly more conservative than the baseline for two periods. Only two justices—Burton and Stewart—demonstrate no significant change in their ideal points, *even after controlling for changes in case stimuli*.[14] The implication is clear—Epstein et al. (1998) *underestimate* the amount of preference change on the Supreme Court. And, their conclusion that Brennan, Burger, Harlan, Jackson, and Marshall exhibit no significant change is incorrect. This is likely due to their measurement strategy and using only a global test of preference change. Indeed, our results for Harlan, Douglas, and White show that assuming a particular functional form, either linear or parabolic, for ideal points trajectories is an inappropriate assumption. The findings for these justices would be masked (as with Harlan) or attenuated (as with Douglas and White) when using a global measure.

Instead of imposing a particular baseline for comparison, we present further evidence of preference change in Figure 5. To construct this figure, we compute the posterior probability that a given justice is more conservative in term

---

14. The choice of the baseline category is arbitrary. To test for robustness, we have created these same plots for more or less terms in the baseline, and the results are substantively the same. These results are available in the Web appendix.
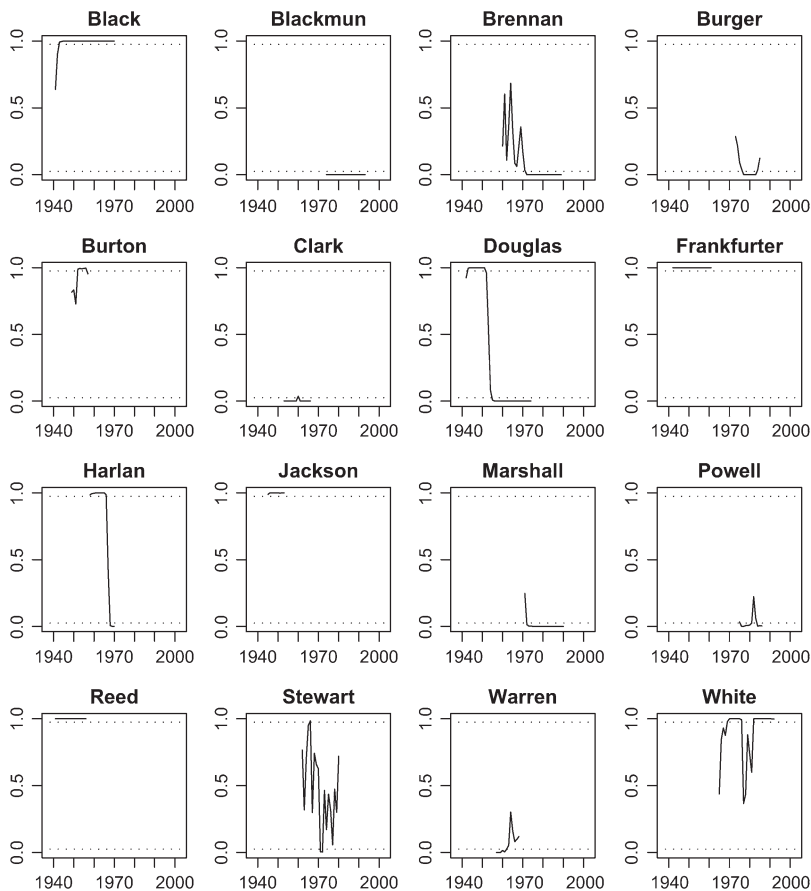
Figure 4. Estimated posterior probabilities that the justice is more conservative in subsequent terms than their mean location in their second through fourth terms. These plots are for selected justices from the dynamic ideal point model, 1937–2003. The *y* axis denotes the estimated probability, and the *x* axis is the term in which the justice served.

r than term s for all possible combinations where $r > s$. The specific algorithm used to compute this quantity is discussed in Appendix. We summarize these posterior probability profiles for four justices of interest in Figure 5. The baseline term is on the y axis, and the comparison term is on the x axis. For example, if one is interested in determining whether or not Justice Black is more conservative in subsequent terms than the 1950 term, one would read across from left to right at the 1950 tick on the y axis. The legend shows how the probabilities are encoded in the figure. The bright red color implies that the justice is significantly more liberal, and the bright blue color implies that the justice is significantly more conservative. For the sake of space, we only present these profiles for Justice Black, Harlan, Marshall, and Stewart. The profiles for the other 12 justices are available in the Web appendix.
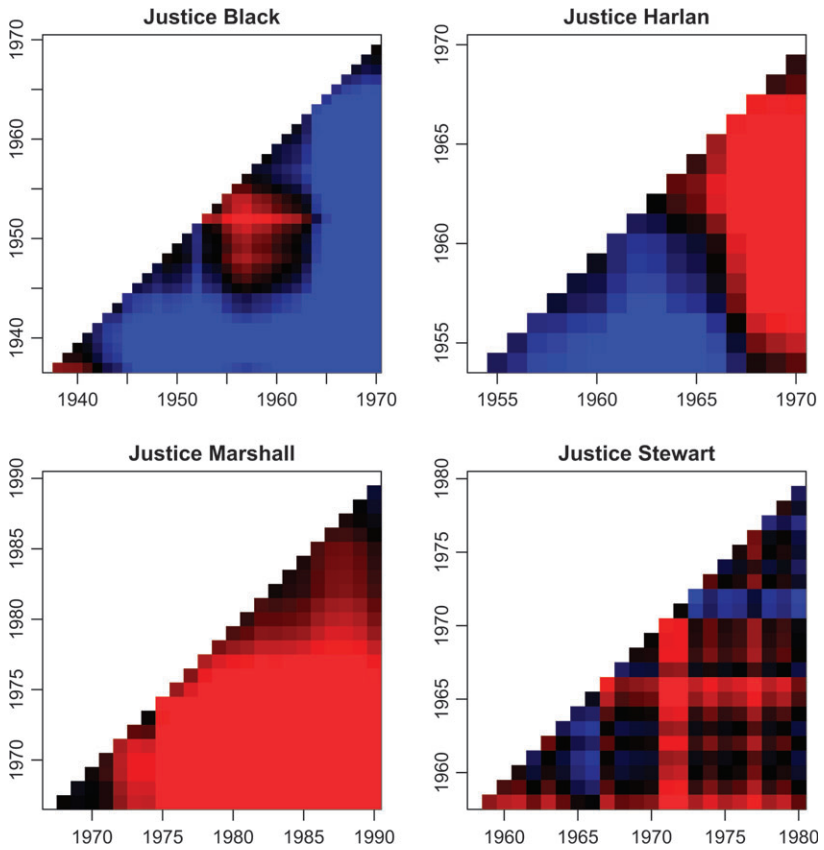
Figure 5. Estimated posterior preference change profiles for Justices Black, Harlan, Marshall, and Stewart from the dynamic ideal point model.

The results in Figure 5 confirm the conclusions drawn from Figure 4. Compared to his first terms, Justice Black is significantly more conservative in nearly every subsequent term. Also, if we chose terms in the mid-1950s as the baseline, we also see that Justice Black was also significantly more conservative in the late 1960s. Similarly, we see that Justice Marshall became significantly more liberal throughout the term of his service. But, by the early 1980s, he remains a stable liberal. Justice Harlan, with the parabolic trajectory, is another interesting case. The estimated posterior probability profile shows that depending on the baseline category, Harlan became significantly more conservative or significantly more liberal. Finally, the cell for Justice Stewart shows no significant preference change regardless of the baseline. These profiles show that global tests of preference change are inappropriate; rather, one should use local estimates of the probabilities of interest as we have done here. The findings from these results are striking—preference change is a common phenomenon that occurs quite often in the Supreme Court.
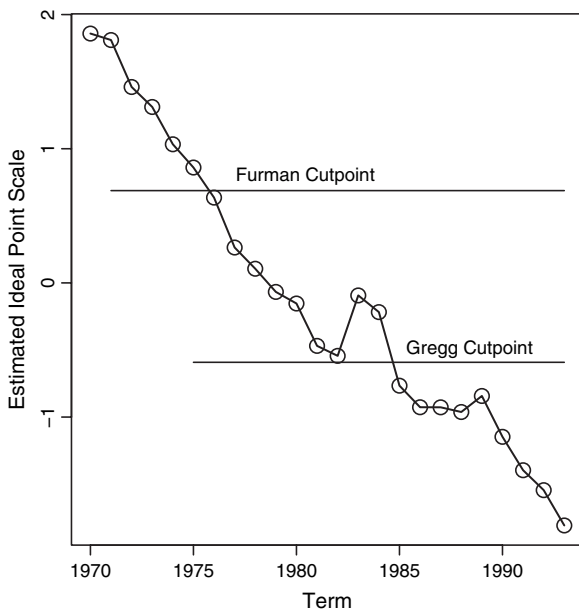
Figure 6. Estimation of ideal points of Justice Blackmun throughout his career, with estimated cut-points from Furman v. Georgia, 408 U.S. 238 (1972), and Gregg v. Georgia, 428 U.S. 153 (1976).

To get a sense of how our results fit with more qualitative accounts of attitudinal change we examine the path of Justice Blackmun's ideal points together with the cut-points from two important death penalty cases. Justice Blackmun was the justice with perhaps the most dramatic change in preferences over the course of his career (Greenhouse 2005). When he was nominated by President Nixon, the model suggests that Justice Blackmun was the second most conservative member of the Court (second only to Chief Justice Burger); when he retired, he was the second most liberal member of the Court (Justice Stevens was more liberal).

Figure 6 plots the ideal point trajectory of Justice Blackmun along with the cut-points for two major death penalty cases. In 1972, the Supreme Court decided Furman v. Georgia, 408 U.S. 238 (1972), and held that the death penalty as currently employed in the states was a cruel and an unusual punishment. The Court ruled that the arbitrariness of the use of the death penalty violated the Constitution. Blackmun dissented from this decision, thus taking the conservative side. Figure 6 shows the cut-point in the Furman case. The model predicts that justices above the cutline should vote in the conservative direction. In the 1971 term Blackmun was clearly on the conservative side of the cut-point. A second important death penalty case was decided in 1976; Gregg v. Georgia, 428 U.S. 153 (1976). In Gregg the Court reversed course and upheld the constitutionality of death penalty legislation spawned in response to Furman. The Court held that the new death penalty statutes in the states were consistent with

the protections of the Eighth and Fourteenth Amendments. Blackmun concurred with this conservative decision, which is also consistent with the model; in the 1975 term he falls above the Gregg cut-point (on the conservative side).

Looking at Blackmun's votes on these cases in isolation one would have no reason to believe that his views were becoming more liberal. This would be hard to square with his later decisions and writings, which clearly show that he had moved to the left (Greenhouse 2005). However, because the model used here looks at all Blackmun's votes over time together with the votes of the justices he served with, it points to a gradual leftward shift throughout most of Blackmun's tenure on the Court. The substantive content of Blackmun's decisions and other writings never enter the model, yet the model's results are remarkably consistent with what we know of Blackmun from more qualitative accounts.

The model also allows us to entertain counterfactuals about how justices would have decided cases at different points in time. Although, for obvious reasons, such an exercise should not be taken too seriously, it does allow us to get a sense of how *important* the changes are that we identify. For instance, the results from our model suggest that if the Court heard Furman in 1976 there is less than a 50% chance that Blackmun would have upheld the existing Georgia death penalty statute. Similarly, had Gregg been decided by the Court in 1985 or later, our model suggests that Blackmun would have voted to strike down the new death penalty laws. The changes in Blackmun's revealed that preferences are not only statistically significant but also they are substantively important.

5. Implications and Conclusion

The results presented above strongly suggest that the policy positions of Supreme Court justices do not remain constant throughout the course of their careers. This finding goes against much of the prevailing wisdom in judicial politics research and calls into question the results from a large body of research that explicitly assumes temporal stability of preferences. When scholars employ preference measures that are constant across time, such as Segal and Cover (1989) scores, the independent variable capturing preferences will be measured with *systematic error*. It is well known that this can lead to bias in the estimation of structural parameters of interest and can lead to incorrect substantive conclusions. Oftentimes, when one is analyzing a single cross section of data, this measurement error will be of little consequence. But when studying Supreme Court behavior over time using time-invariant preference measures can lead to incorrect conclusions about the effect of preferences or other variables on the outcome of interest. Since most statistical studies of judicial behavior look at behavior over time (e.g., Segal and Spaeth 1993; Maltzman and Wahlbeck 1996), this is a serious concern.

The most commonly used measure of judicial preferences, Segal and Cover (1989) scores, are time invariant. These measures have the advantage of being truly exogenous from behavior, but the results presented here demonstrate that

the assumption of constancy is unwarranted. Epstein and Mershon (1996) demonstrate that the explanatory power of Segal and Cover (1989) scores is limited to civil rights and liberties issues and that the scores should only be used to study aggregated votes for those issue areas. Our results go a step further. Based on our results, Segal and Cover (1989) scores should generally not be used to study judicial behavior that occurs over time. Other measures, such as party identification of the justices or social background characteristics suffer the same ills.

What is the solution to this problem? One important by-product of our research is that we estimate an ideal point for every justice in every term for all justices serving between 1937 and 2003. These measures are available in electronic form on the Web appendix. Our measures are time varying and thus do not suffer from the deficiency of other commonly used measures. When studying phenomena other than votes on the merits, these measures can be employed as independent variables to explain behavior. However, in a strict sense, the measures should not be used directly in probit or logit models to study votes on the merits because these same votes were used to construct the measures.

One important question that we leave for future research is: what *explains* preference change? Although the results clearly show that justices exhibit change in revealed preferences over time, we do not offer an exhaustive theory for why this is so. While it is tempting to speculate about what such a grand theory might look like,[15] we suspect that personal, idiosyncratic reasons operating at the justice level are at least, if not more, important than broad system-level mechanisms. Moreover, with only a small number of justices, attacking this problem with a *statistical* research design is implausible. A far better approach would be to perform a historical or doctrinal analysis of the behavior of individual justices. We look forward to future work that pursues this research agenda.

At the end of the day, this article contributes to the judicial politics literature in a number of ways. Our dynamic ideal point model provides an extremely powerful and flexible method to simultaneously learn about the effects of justice policy positions and case content on decision making. The data suggest that nearly every justice exhibits statistically significant preference change over the course of his or her career. The patterns uncovered by the model are substantively interesting and facially valid. Just as important as what we have learned from this article is what we still do not fully understand. Although there is a great deal of evidence that revealed preferences of justices change over time, more research needs to be done to more fully understand the mechanisms that cause such change.

---

15. Some factors that might influence the revealed preferences of individual justices include social psychology (Eagly and Chaiken 1993), strategic considerations (Maltzman and Wahlbeck 1996; Epstein and Knight 1998), or the macropolitical context (McGuire and Stimson 2004).

Appendix

Monte Carlo Estimates of Quantities of Interest

To estimate the quantities of interest regarding preference change, we use the following Monte Carlo algorithms. To estimate the quantity in equation (6), the following pseudocode illustrates how this quantity can be calculated for term r and justice j using the draws from the posterior distribution of $\boldsymbol{\theta}$:

Algorithm A1. PROBMORECONS1 $(\boldsymbol{\theta}, r, j)$

$P_{r,j} \leftarrow 0$

for $g \leftarrow 1$ **to** G

$\text{\textbf{do}} \begin{cases} \theta_j^* = \frac{1}{3}(\theta_{2,j}^{(g)} + \theta_{3,j}^{(g)} + \theta_{4,j}^{(g)}) \\ \text{\textbf{if} } (\theta_{r,j}^{(g)} > \theta_j^*) \\ \text{\textbf{then} } P_{r,j} \leftarrow P_{r,j} + 1 \end{cases}$

$P_{r,j} \leftarrow G^{-1} P_{r,j}$

return $(P_{r,j})$.

Here G is the number of MCMC draws and $P_{r,j}$ is the probability that justice j is more conservative than the baseline in term r.

For the quantity of interest plotted in Figure 3:

$$\Pr(\theta_{r,j} > \theta_{s,j}) \text{ for } s = 1, \ldots, \overline{T_j} - 1, \quad r = 2, \ldots, \overline{T_j}, \text{ and } r > s. \qquad \text{(A1)}$$

We employ the following algorithm:

Algorithm A2. PROBMORECONS2 $(\boldsymbol{\theta}, r, s, j)$

$P_{r,s,j} \leftarrow 0$

for $g \leftarrow 1$ **to** G

$\text{\textbf{do}} \begin{cases} \text{\textbf{if} } (\theta_{r,j}^{(g)} > \theta_{s,j}^{(g)}) \\ \text{\textbf{then} } P_{r,s,j} \leftarrow P_{r,s,j} + 1 \end{cases}$

$P_{r,s,j} \leftarrow G^{-1} P_{r,s,j}$

return $(P_{r,s,j})$.

Here $P_{r,s,j}$ is the probability that justice j is more conservative in term r than in term s.

References

Albert, James H. 1992. "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling," 17 *Journal of Educational Statistics* 251–69.

Atkins, Burton M., and Terry Sloope. 1986. "The 'New' Hugo Black and the Warren Court," 18 *Polity* 621–37.

Bafumi, Joseph, Andrew Gelman, David Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation," 13 *Political Analysis* 171–87.

Bailey, Michael, and Kelly Chang. 2001. "Comparing Presidents, Senators, and Justices: Inter-institutional Preference Estimation," 17 *Journal of Law, Economics, and Organization* 477–506.

Baum, Lawrence. 1988. "Measuring Policy Change in the United States Supreme Court," 82 *American Political Science Review* 905–12.

Berry, S. M., C. S. Reese, and P. M. Larkey. 1999. "Bridging Different Eras in Sports," 94 *Journal of the American Statistical Association* 668–79.

Bock, R. Darrell, and Marcus Lieberman. 1970. "Fitting a Reponse Model for n Dichotomously Scored Items," 35 *Psychometrika* 179–97.

Bradlow, Eric T., H. Wainer, and X. H. Wang. 1999. "A Bayesian Random Effects Model for Testlets," 64 *Psychometrika* 153–68.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data," 98 *American Political Science Review* 355–70.

Eagly, Alice H., and Shelly Chaiken. 1993. *The Psychology of Attitudes.* Fort Worth, Tex.: Harcourt Brace Jovanovich.

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

Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences," 40 *American Journal of Political Science* 261–94.

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make.* Washington, D.C.: CQ 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," 60 *Journal of Politics* 801–18.

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

George, Tracey, and Lee Epstein. 1992. "On the Nature of Supreme Court Decision Making," 86 *American Political Science Review* 323–37.

Geweke, John. 1992. "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments," in J.M. Bernardo, J.O. Berger, and A.F.M Smith, eds., *Bayesian Statistics 4.* Oxford, UK: Clarendon Press.

Gill, Jeff. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Boca Raton, Fla.: Chapman & Hall.

Greenhouse, Linda. 2005. *Becoming Justice Blackmun: Harry Blackmun's Supreme Court Journey.* New York: Times Books.

Grofman, Bernard, and Timothy J. Brazill. 2002. "Identifying the Median Justice on the Supreme Court Through Multidimensional Scaling: Analysis of 'Natural Courts' 1953–1991," 112 *Public Choice* 55–79.

Hagle, Timothy M. 1993. "Freshman Effects for Supreme Court Justices," 37 *American Journal of Political Science* 1142–57.

Hambleton, Ronald K., and Hariharan Swaminathan. 1985. *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Heidelberger, P., and P. D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulation," 24 *Communications of the ACM* 233–45.

Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo," 44 *American Journal of Political Science* 369–98.

———. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking," 9 *Political Analysis* 227–41.

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

Maltzman, Forrest, and Paul J. Wahlbeck. 1996. "Strategic Policy Considerations and Voting Fluidity on the Burger Court," 90 *American Political Science Review* 581–592.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999," 10 *Political Analysis* 134–53.
———. 2007. "Scythe Statistical Library, Release 0.3," http://scythe.wustl.edu/.

McGuire, Kevin T., and James A. Stimson. 2004. "The Least Dangerous Branch Revisited: New Evidence on Supreme Court Responsiveness to Public Preferences," 66 *Journal of Politics* 1018–1035.

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

Segal, Jeffrey A. 1997. "Separation-of-Powers Games in the Positive Theory of Law and Courts," 91 *American Political Science Review* 28–44.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices," 83 *American Political Science Review* 557–65.

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

Spaeth, Harold J. 2001. *The Vinson–Warren Supreme Court Judicial Database, 1946–1968 Terms* (computer file). Ann Arbor, Mich.: Inter-University Consortium for Political and Social Research, first edition.
———. 2004. *United States Supreme Court Judicial Database, 1953–2003 Terms* (computer file), http://www.as.uky.edu/polisci/ulmerproject/.

Spiller, Pablo T., and Rafael Gely. 1992. "Congressional Control or Judicial Independence: The Determinants of U.S. Supreme Court Labor-Relations Decisions, 1949–1988," 23 *RAND Journal of Economics* 463–92.

Tate, C. Neal. 1981. "Personal Attribute Models of Voting Behavior of U.S. Supreme Court Justices: Liberalism in Civil Liberties and Economics Decisions," 75 *American Political Science Review* 355–67.

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

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