Chapter 2

The Measure of Law: Estimating Preferences across Institutions and Time

Our goal is to see if the court really is as unconstrained as the attitudinal model says it is. To do this we address several measurement challenges. The first is that we need to compare policy views across institutions. The good news on this front is that statisticians and political scientists have developed a number of fancy methods for measuring preferences within institutions. The bad news is that these approaches do not travel well across institutional boundaries.

The root of the problem is that existing measures are based on different scales and on
different agendas. Comparing them directly would be as problematic as comparing students’
test scores on different tests. Is a student who scores an 87 on one test more mathematically
inclined than a student who scores a 75 on a different test? Surely a student who gets
an 87 by adding three digit numbers is not more mathematically capable than a student
who got a 75 solving differential equations. Similarly, if Justice William Brennan voted
conservatively 19% of the time on the court and Representative Connie Morella (R, MD)
voted conservatively 20% of the time in the House of Representatives (as they did in the
samples of votes in our data set), would we have any confidence in saying that they have the
same preferences?

A second challenge arises from the need to compare preferences over time. Just as Justice
David Souter voted on different issues than did members of Congress, so too did Justice
Souter face a different set of cases than Justice Lewis Powell. If Souter voted in a liberal
direction more often than Powell, do we conclude that Souter was more liberal than Powell?
Or is it a reflection that the Court’s agenda during Souter’s tenure was different the Court’s
agenda during Powell’s tenure?

This chapter provides a framework for addressing these issues. First, we establish that
measurement matters. While no one doubts the technical basis for concerns about cross-
institutional and cross-time comparability, some may believe that simple measures provide
valid results without the headache and complication inherent in more technically sophis-
ticated approaches. In this context, however, simple does not suffice. In fact, even some
of the most prominent approaches are flawed. We show specific examples of how existing
approaches to inter-institutional and cross temporal measurement fail, sometimes rather
dramatically.

Second, we build an inter-institutional and inter-temporal measurement model.\(^1\) Key to
the approach is the use of “bridging” techniques that link actors across time and institutions.
To bridge across institutions, we incorporate data on presidents and members of Congress
taking positions on Supreme Court cases. To bridge across time, we incorporate observations
of individuals taking positions on cases and votes in the past and use information about the
substantive implications of cases and legislation. The statistical analysis uses flexible and
powerful Bayesian Markov Chain Monte Carlo methods (see, e.g., Jackman 2000).\(^2\)

This chapter uses positions taken by members of Congress and by executive branch actors
on Supreme Court cases to develop measures of preferences that are comparable over time
and across institutions. In subsequent chapters we demonstrate that the approach is valid for
more than just ideal point estimation, as we build on this approach to assess other influences
on the court.

\(^1\)The approach for measuring ideological preferences across institutions was originally detailed in Bailey (2007). The models
presented in chapters 3-5 employ a similar approach for measuring ideology. They also employ a dataset that is expanded
beyond Bailey (2007) to include recent court appointments of John Roberts and Samuel Alito.
\(^2\)Preference estimates and a methodological appendix will be available at www9.georgetown.edu/faculty/baileyma.
2.1 A Spatial Model of Supreme Court Decision-Making

Spatial models provide the foundation for measuring Supreme Court preferences. These models have a long pedigree and are standard in the literature (Downs 1957; Martin, Quinn and Epstein 2005). In them, individuals have “ideal points” in policy space that maximize their utility. Policy alternatives are points in that preference space, and individuals prefer spatially closer alternatives.

Figure 2.1 illustrates the classic spatial model. Each justice has an ideal point at $\theta_i$. Following standard spatial theory, a justice will vote for the defendant if his or her ideal point is on the same side of the midway point between the outcomes associated with voting for the defendant and plaintiff (which we refer to as the vote “cutpoint”). In the figure, justices one through five support the defendant and justices six through nine support the plaintiff. More concretely, suppose a liberal justice such as Justice Stevens has an ideal point of -1 while a conservative justice such as Justice Scalia has an ideal point of +1. If a case presented justices with the alternatives of voting to free a defendant (with a hypothetical policy outcome of -0.8) or to convict a defendant (with a hypothetical policy outcome of +0.6), Stevens would prefer to free the defendant and Scalia would prefer to convict.
2.2 Comparing preferences across institutions

Our first task is to estimate preferences in a manner that allows us to directly compare preferences of actors who operate in different institutions. This would be a simple task if justices voted on congressional roll calls and members of Congress voted on Supreme Court cases. In that case, we could directly apply any of the advanced measurement approaches in the literature (Clinton, Jackman and Rivers 2004; Poole and Rosenthal 1997). Unfortunately reality is not so accommodating.

One solution commonly employed assumes that preference estimates for members of one institution are directly comparable to preference estimates produced for another institution. For example, Segal (1997) assumed that preference measures for justices based upon an analysis newspaper editorials written while a nomination was pending (see Segal and Cover 1989 for a discussion of Segal-Cover scores) were directly comparable to scores for members
of Congress produced by Americans for Democratic Action based upon key roll-call votes (commonly known as ADA scores). Moraski and Shipan (1999) assumed that percent-liberal scores for justices were directly comparable to adjusted ADA scores from Groseclose, Levitt and Snyder (1999).³

Scholars have long recognized the problem with this approach. Indeed, Segal (1997, 36) employed a direct comparison and noted that such comparison was “obviously not an example of textbook scaling.” ⁴ The problem with the direct comparison approach is that only works if the distributions of vote cutpoints facing judicial and legislative actors are the same. Figure 2.2 demonstrates what can go wrong if this assumption is not accurate. At the top is a hypothetical court with judges with ideal points at \( J_1 \), \( J_2 \) and \( J_3 \). This court ruled on two cases. In the first, the cutpoint between the two alternatives was \( K_1 \); in the second, the cutpoint was \( K_2 \). The ideal point of the judge \( J_3 \) is greater than both cutpoints. If there is no randomness, this judge will vote conservatively 100 percent of the time. The ideal point of the median judge (\( J_2 \)) is greater than \( K_1 \), but less than \( K_2 \), leading this judge to vote liberally once and conservatively once. The ideal point of the most liberal judge is less than both cutpoints, yielding no conservative votes.

In the next two panels are hypothetical legislatures whose legislators have ideal points at \( L_1 \) through \( L_5 \). The legislature voted on legislation that had cutpoints at \( K_3 \) and \( K_4 \).

³ An exception is Epstein, Martin, Segal and Westerland (2007) which re-scales Poole-Rosenthal Common Space scores and Martin and Quinn (2002) scores. Because their approach uses both Poole-Rosenthal and Martin-Quinn scores, it inherits the concerns we discuss below about both methods.

⁴ Segal’s use of the approach is understandable given the fact his work appeared prior to more recent breakthroughs in ideal point estimation.
Figure 2.2: The challenge of making inter-institutional preference comparisons
Two conservative legislators had ideal points above both cutpoints, implying a 100 percent conservative rating. The median legislator’s ideal point was higher than the first cutpoint (implying one conservative vote) and lower than the second cutpoint (implying one liberal vote). Two liberal legislators’ ideal points were below both cutpoints, implying a 0 percent conservative rating.

If we measure preferences based on voting within each of the two institutions, we cannot know whether the depiction in the middle panel (where the legislative median is far to the left of the court median) or the bottom panel (where the legislative median is far to the right of the court median) is correct. Either depiction is logically possible, even as they differ dramatically from each other. More sophisticated within-institution preference estimation will not solve this fundamental inter-institutional problem.

This is not merely a technical problem. Figure 2.3 compares two plausible ways to implement direct comparability. The first treats percent liberal judicial scores as comparable to ADA scores. The second treats Poole and Rosenthal Common Space scores for senators and presidents as comparable to Martin and Quinn scores for justices.\(^5\) The figure presents the estimated preferences of the court median, Senate median and the president for these two approaches over time.\(^6\)

---

\(^5\)This approach to inter-institutional preference measurement is not endorsed by Poole, Rosenthal, Martin or Quinn; we present it simply as an example of assuming direct comparability across institutions.

\(^6\)The ADA and percent liberal measures have been subtracted from one in order to give them the same ideological polarity as the Poole-Rosenthal and Martin-Quinn measures.
"ADA Measure" Inter-institutional comparison assuming adjusted ADA scores are comparable to judicial percent liberal scores

"PRMQ Measure" Inter-institutional comparison assuming Poole & Rosenthal scores are comparable to Martin & Quinn scores

Figure 2.3: INTER-INSTITUTIONAL COMPARISONS ASSUMING DIRECT COMPARABILITY
ample, in the fifties, the ADA measure indicates the court was the most liberal and the Senate was the most conservative. The measures based on Poole and Rosenthal and Martin and Quinn scores, on the other hand, has the opposite, with the Senate the most liberal and the court the most conservative. The fact that these approaches differ this decisively illustrates the need to employ measurement theory while developing measurement strategies.

2.3 Comparing preferences over time

A related challenge is ensuring that preference measures are comparable over time. To see the issue, suppose we wish to compare the conservativism of two courts: one that voted seven to two in favor of a liberal outcome on *Roe v. Wade* (1973) and another that voted five to four in favor of a conservative outcome on *Webster v. Reproductive Health* (1989), a case in which the court upheld a Missouri law that banned abortions in public facilities and placed other limits on abortions.

Figure 2.4 presents a hypothetical depiction of preferences of the justices on each case. In the top panel, the seven justices who voted liberally on *Roe* are to the left of the cutpoint and the two who voted conservatively are on the right. The next panels depict hypothetical preferences of the justices who voted on *Webster*. Again, those who voted liberally are to the left of the cutpoint and those who voted conservatively are to the right. Here, though, we see the problem: Do we believe scenario 1 in which the vote cutpoint is similar to that
of *Roe*? Or do we believe scenario 2 in which the vote cutpoint has shifted to the left? Or is the correct view scenario 3 in which the cutpoint has shifted to the right? Based on only the vote information, we cannot say.

The stakes are high for estimating preference change over time: if we believe scenario 1, for example, there is little to indicate a significant change in court preferences. If we believe the scenario 4, however, there has been a significant rightward shift of the members of the bench. And given Baum’s (1988) results that the agenda has in fact changed over time, this problem could well be quite real.

In fact, widely used preference measures struggle with across time preference estimation. Consider first Martin and Quinn’s estimates of the median of the Supreme Court as presented in Figure 2.5 (see, e.g. Epstein, Friedman and Staudt 2008). Because the Supreme Court operates under more or less open rules and appears to mostly divide along one-dimensional lines, scholars frequently use the court median to summarize the court’s preferences for any given point in time. According to Figure 2.5 the court nearly reaches its post-war conservative peak in 1973. That means that according to these estimates, the court that deemed abortion a constitutional right in *Roe* and struck down the death penalty in *Furman v. Georgia* was actually one of the most *conservative* in the modern era.

The challenge of inter-temporal preference estimation extends to Congress. Poole and Rosenthal’s (1997) hugely influential work argues that congressional voting can be explained across long time periods with a one dimensional spatial model of preferences; only in limited
Figure 2.4: Difficulty in Identifying Preference Change or Cutpoint Change
Figure 2.5: Martin and Quinn Estimates of Supreme Court Median over Time
periods does a second dimension significantly help explain voting. One of the products of this research agenda has been “Common Space Scores” that “place the members of the House and Senate in the same space... [allowing] members to be compared across Chambers and across Congresses” (Poole 2005b; see also Poole 1998).

Figure 2.6 plots Common Space scores of selected members of Congress. The first dimension captures “party loyalty” and explains most votes in Congress. The second dimension - the so-called “race dimension” (Poole and Rosenthal 1997, 46-48) – was important in the fifties and sixties, but faded considerably with the Republican realignment in the South in the eighties. Average preferences are roughly (0, 0). Senators Helms and Kennedy are included as reference points, as these two often are treated as anchors on the conservative and liberal extremes on both dimensions.

Most striking about the figure is that Southern segregationist senators are indistinguishable from modern moderate Democrats on both dimensions. Senator Hollings (D-SC) – who, among other things, voted to override President Bush’s 1988 veto of the Civil Rights Restoration Act and voted for the 1991 Civil Rights Act – is measured to be at least as conservative as Senator Harry Byrd (D-VA), who advocated “massive resistance” to civil rights rulings by the court at a time when African-Americans were routinely denied voting rights and segregated in public. Byrd (not to be confused with Senator Robert Byrd of West Virginia) stated that that “we should exclude the Negro population” from voting and that “non-segregation would be most unwise and I will certainly oppose it in every way I can”
Figure 2.6: Common Space Scores of Selected Members of Congress

Pro-Segregation Senators
1) Ellender (D-LA, 1937-73)
2) Russell (D-GA, 1933-73)
3) Eastland (D-MS, 1943-79)
4) H. Byrd (D-VA, 1933-65)
5) Sparkman (D-AL, 1946-79)

Modern Moderate Democrats
6) Hollings (D-SC, 1966-04)
7) Stenholm (D-TX, 1979-04)
8) Zorinsky (D-NE, 1976-89)
9) W. Jones (D-NC, 1965-92)
10) Exon (D-NE, 1979-96)
Hollings also has a similar score to Senator Eastland (D-MS) who had stated “I assert that the Negro race is an inferior race. ... I know that the white race is a superior race. ... It is responsible for all the progress on earth.” (DeParle 2004, 32, citing *Congressional Record* 79th Congress (June 29, 1945) p. s7000).7

Why do these immensely influential preference estimates contradict our understandings of the ideology of Supreme Court justices and members of Congress? Why do these measures lack face validity? The problem is that it is very hard to locate preference when the underlying policy agenda is changing. These approaches use only vote data, implicitly assuming that the distributions of vote cutpoints over time do not vary.8 However, without information on the relative location of vote cutpoints, the estimation has no reason to favor any one of the bottom three scenarios in Figure 2.4 and will on average, assume that the cutpoint has not changed. This can cause a problem if there is a rapid increase in the number of conservative votes (as occurred in the early 1970s). The cutpoint distribution is by default assumed to be the same or to the right, meaning that the estimation procedure will lean towards the alignment in scenario 3 even if, in fact, scenario 1 or 2 reflects reality.

For Congress, the anomalous preference estimates are due to two factors. The first is the same as for Martin and Quinn’s estimates: the measures are based only on vote data

---

7 It is also interesting to note that Hollings’ views changed over time. On *Brown v. Board of Education*, Hollings said: “It certainly is the most significant judicial decision of that century ...There is no question in my mind that was for the good. I had my doubts at that particular time... [The plaintiffs] understood the Constitution in America better than this particular Senator” (*Congressional Record* 108th Congress (May 13, 2004), p. S5457).

8 Although Martin and Quinn allow individual judicial preferences to vary from term to term, they constrain preferences from moving too much from one term to the next by constraining justice’s ideal points not to move too much from one term to the next.
so may not be well equipped to handle situations in which preferences and agendas shift together. The second factor is that the dimensionality of the Poole and Rosenthal measures vary over time. Their “race dimension” exerted an independent effect on congressional voting from roughly 1940 to 1966 and then was gradually absorbed into the first dimension by the mid-1980s. This means that one cannot use just one of these scores for studies that span the post-war era. If one were to use only first dimension scores, one would have a poor measure for issues addressed by the court such as race, busing, school prayer and internal security as these loaded heavily on the second dimension. For example, the Common Space scores of Senators Baucus (D-MT), Evan Bayh (D-IN), Reid (D-NV), and Brooke (R-MA) are more conservative on the first dimension than the indisputably racist Senator Bilbo (D-MS) who served from 1935 to 1947 and infamously wrote an anti-civil rights tract entitled *Take Your Choice: Separation or Mongrelization*. In fact, the Common Space scores of most moderate Republicans are more conservative on the first dimension than most signers of the segregationist Southern Manifesto of 1956. One may consider using second dimension scores, but this dimension faded to near-irrelevance by the late eighties, making these scores useless for analysis of what is happening at the start of the twenty-first century.\footnote{Poole and Rosenthal conceive of their NOMINATE scores as reflecting underlying latent preference dimensions that typically do not change. The mapping of policies onto these dimensions does change, however and the important changes that have occurred in the mapping of policy to the one (and sometimes two) dimensional policy space in the post-war era (Poole and Rosenthal, 1997, 6 and Chapter 5) makes interpretation quite complex.}

The point of these examples is that measurement matters. As scholars we need to reflect carefully on if and how the preference measures we use are consistent with the political
concepts we are trying to measure. In many cases, the easiest to acquire measures will not be appropriate for the question at hand. For example, in our efforts to isolate political preferences on cases before the court, we cannot use preference measures that imply that the modern Southern Democrats are the same as their segregationist predecessors. Or, in the same spirit, as we try to assess political constraints on the court, we may be reluctant to use preference measures that place the Supreme Court near its conservative peak in a time when it was handing down decidedly liberal opinions.

2.4 Generating Comparable Preferences

Our goal is to build a measurement model that can span institutions and time. Key to our approach will be “bridge” observations of judges and elected officials taking positions on cases/votes in other institutions and other periods of time. These observations help us pin down preferences of a diverse set of actors in a common space.

Educational testers face an analogous and mathematically isomorphic problem when trying to create scores that are comparable across students who take exams with different test questions (see e.g. Baker 1992). Educational testers model each student as having some “ability” and each question as having some “difficulty.” If every student took the same test, then it would be straightforward to compare student’s abilities as they were faced with questions of the same difficulty. However, it is not wise for standardized tests to ask the same
questions year in and year out, so they must ask different questions to different cohorts. The challenge is to make these scores comparable. The solution is to have a modest number of questions that are asked across the cohorts. For these overlapping questions, the difficulty is the same, facilitating estimation of ability across test-taking cohorts; this in turn facilitates comparable estimation of the difficulty of different questions.

We pursue an analogous strategy. We model each justice as having some “ideology” and each case as having (as above) some cutpoint. Our “overlapping questions” are of two kinds: Supreme Court cases on which members of Congress take public positions such as the 2008 District of Columbia v. Heller gun control case on which 322 members of Congress joined amicus briefs on both sides and Supreme Court cases on which non-contemporaneous justices take positions such as occurred in 2002 Atkins v. Virginia death penalty case in which justices took positions on previous 1989 Penry v. Lynaugh, with some affirming the earlier decisions and others rejecting it.

Our approach therefore uses the following pieces of information: justices’ votes on questions presented to the court, members of Congress’s votes the questions (roll calls) presented to Congress and presidents and some members of Congress positions on some questions presented to the Supreme Court. The conservatism of political actors is analogous to test-taking ability and the cases on which non-judicial actors have taken positions are analogous to the overlapping questions (Bailey and Chang 2001).

These bridge observations provide fixed references against which the preferences of actors
Figure 2.7: Identifying Preferences in an Inter-Institutional Context

The Constrained Court
across institutional boundaries can be judged. Suppose we are interested in comparing preferences of a three-person court (with justices $J_1$, $J_2$ and $J_3$) to a five-person senate (with senators $S_1$, $S_2$, $S_3$, $S_4$ and $S_5$). These hypothetical individuals are aligned from liberal to conservative in each institutional context in Figure 2.7(a). We also include the percentage of the time each of our hypothetical senators and justices votes in a conservative manner as a tool for illustrating again the weakness of assuming direct comparability of ideology scores. As discussed earlier, we cannot calibrate preferences across the two contexts based only on votes within the respective contexts. However, if we observe the position of the justices and senators on “Case X” we have information that is very helpful. In the example, two justices were liberal and one was conservative on Case X while two senators were liberal and three were conservative on the same case. Using the cutpoint of Case X as a fixed reference, we can align the preferences across the two institutions as in Figure 2.7(b). Using this basic insight, the statistical model described below incorporates such information in a large-scale fully-specified dichotomous choice statistical model.

Our approach is to find fixed reference points that allow preferences to be estimated even when preferences change over time. It follows a similar reasoning as for the inter-institutional bridges. To see the logic, first suppose that instead of having two separate institutions, we have the same institution at two separate points in time. There may be some overlap of membership, but if we allow preferences to change over time, we will not be able to align preferences across institutions without additional information or assumptions.
One very useful source of information is the existence of cases (such as “Case X”) on which individuals at both points in time took positions. This produces comparability just as in Figure 2.7. For example, when Justice Thomas wrote in *Planned Parenthood v. Casey* (1992) that *Roe* was “wrongly decided,” he provided an indication his preferences in 1992 relative to *Roe*, a case decided well before he came to the Court by justices with whom he did not generally overlap.

Incorporating these bridge observations has the additional salutary effect of increasing information about case parameters. As Londregan (1999) emphasized, cutpoint estimates for institutions with a small number of actors – such as the Supreme Court – will be poorly estimated. The best way to mitigate the problem is to add “votes” whenever possible and to incorporate other sources of information about vote parameters (see also Clinton and Meirowitz 2001).

We also use information based on the relative position of vote cutpoints. Doing this allows us to incorporate substantive information about agenda changes when it is available. We know, for example, that the 1964 Civil Rights Act was more liberal than the 1960 Civil Rights Act which in turn was more liberal than the 1957 Civil Rights Act. Using this information allows us to more precisely measure preferences across time in a way that is comparable across time.

Figure 2.8 illustrates how this works with a specific example involving two death penalty cases. In 1988 the court assessed in *Thompson v. Oklahoma* whether execution of people un-
der 16 was permissible; in 1989 the court assessed in *Stanford v. Kentucky* whether execution of people between 16 and 18 years old was permissible. The key point for our purposes is that allowing execution of minors under 16 logically implies execution of individuals over 16 is acceptable. If we do not incorporate this relative cutpoint information the estimation could, for the reasons laid out in the discussion of Figure 2.3 estimate the cutpoint for *Thompson* to be to the left of the cutpoint for *Stanford*. When we do incorporate the constraint on cutpoints, however, we ensure that on this vote the implied ideal point for justices seven through nine are to the right of the cutpoint on *Stanford* and the implied ideal points for justices 10 through 13 are to the left of the cutpoint in *Thompson*.10

To identify the relative position of cutpoints does not mean that we can characterize all justices’ implicit positions on the other case. In our example, the cutpoint information does

---

10 *Stanford* was overturned by *Roper v. Simmons* (2005), providing additional information for cross-temporal preference comparisons.
not determine positions for justices 1 through 6 and justices 14 through 18 on the other case in the example. For example, our knowing that the cutpoint on *Stanford* does not rule out the possibility that all liberals on Thompson would have been conservative on *Stanford* (unlike what is in the figure).

We have many inter-institutional examples. The abortion issue provides several cases for which cutpoints have clear relations. In many of the abortion cases which followed *Roe* (including, for example, *Webster*) the court was asked to rule on legislation that regulated, but did not outlaw abortion. If one thought that states could outlaw abortion as did conservatives on *Roe*, then logically, states must be able to regulate it in a manner that stops short of outlawing it, as was true in all these cases. Hence, someone to the right of *Roe* must be to the right on these cases that do less than outlaw abortion, implying the cutpoints of these cases must be to the left of *Roe*. Sometimes the cutpoint information spans institutions. For example, senators voted on April 28, 1976 on a “right to life” amendment that deemed fetuses people with legal rights. Voting in favor of this amendment implied that a senators not only thought that states could regulate abortion (which was at issue in *Roe v. Wade*) but that states should ban it. Hence, the cutpoint on the right to life vote is to the right of the cutpoint in *Roe*.

A few other issues account for a large proportion of the substantive cutpoint linkages. As referenced earlier, civil rights legislation over the years got progressively more liberal; if you were liberal on the 1991 Civil Rights Act which expanded the ability of minorities to
succeed with discrimination claims, you would certainly have been liberal on the 1964 Civil Rights Act which established certain basic principles of non-discrimination. There were also many court cases and congressional votes on busing; if one was liberal on busing, that meant one had to be liberal on Brown v. Board of Education; in our terms, this means the cutpoint on busing cases and votes was to the left of the cutpoint in Brown. There are also many death penalty votes and cases with cutpoint constraints. The cutpoint on a vote in Congress expanding the death penalty to “drug kingpins” will be to the right of the cutpoint in Gregg v. Georgia. Liberals on Gregg were willing to strike all death penalties even after post-Furman reforms; they would certainly be liberal on efforts to expand the death penalty.

Information on relative position of case cutpoints was drawn from a variety of sources. Legal reference books such as O’Brien (2003) were used to follow the case evolution of specific issue areas. Congressional votes that had potential relations either to court cases or to congressional votes in previous Congresses were investigated, typically using the relevant edition of Congressional Quarterly’s Congressional Almanac. A full listing of the cutpoint constraints is available in the appendix.

**Statistical model**  In this section we formalize our model; readers not interested in statistical details may wish to skip to the data description section. Readers interested in more details can refer to the book’s appendix.

The model builds on the canonical formulation of latent preferences in the ideal point
estimation literature (see, e.g., Bailey 2001, Clinton, Jackman and Rivers 2004). Let \( i = 1, \ldots, N \) index individuals and \( v = 1, \ldots, V \) index votes. The utility of actor \( i \) of voting for the conservative alternative is

\[
\begin{align*}
  u_i(\lambda^C_v) &= - (\theta_{it} - \lambda^C_v)^2 + \epsilon^C_{iv} \\
\end{align*}
\]

(2.1)

where \( \lambda^C_v \) is the spatial location of the conservative alternative, \( \theta_{it} \) is the ideal point of the actor at the time of proposal \( t \) and \( \epsilon^C_{iv} \) is a random shock. The utility of voting for the liberal alternative with spatial location of \( \lambda^L_v \) is analogous.

Let \( y^*_itv \) be the utility difference between the conservative and liberal alternatives. It is

\[
\begin{align*}
  y^*_itv &= -(\theta_{it} - \lambda^C_v)^2 + \epsilon^C_{iv} + (\theta_{it} - \lambda^L_v)^2 - \epsilon^L_{iv} \\
  &= 2\theta_{it}(\lambda^C_v - \lambda^L_v) + \lambda^L_v^2 - \lambda^C_v^2 + \epsilon^C_{iv} - \epsilon^L_{iv} \\
  &= (\lambda^C_v - \lambda^L_v)(2\theta_{it} - (\lambda^L_v + \lambda^C_v)) + \epsilon^C_{iv} - \epsilon^L_{iv} \\
\end{align*}
\]

(2.2)

At this point we treat ideology as a single dimensional characteristic of individual justices that best predicts voting. That is, we are measuring revealed preferences, preferences that may be affected by more than the justices’ personal ideological preferences. For example, a justice may be affected by \textit{stare decisis}, by public opinion, by strategic considerations with regard to Congress and so forth. The goal of subsequent chapters is to add such considerations to the framework developed here.

Let \( \kappa_v = \frac{\lambda^L_v + \lambda^C_v}{2} \) be the vote “cutpoint,” (the midway point between the spatial location
of the yea and nay alternatives), $\alpha_v = 2(\lambda_v^C - \lambda_v^L)$ be the vote “discrimination parameter,” and $\epsilon_{iv} = \epsilon_{iv}^C - \epsilon_{iv}^L$ be a N(0, 1) random variable that is uncorrelated with other errors; then

$$y_{itv}^* = \alpha_v(\theta_{it} - \kappa_v) + \epsilon_{iv}$$ \hspace{1cm} (2.3)

which implies that

$$\text{Prob}(y_{itv} = 1) = \Phi(\alpha_v(\theta_{it} - \kappa_v))$$ \hspace{1cm} (2.4)

Assuming independence across individuals and votes, the joint posterior probability of the observed data is

$$g(\theta, \alpha, \kappa|Y) \propto L(\theta, \alpha, \kappa|Y)g(\theta, \alpha, \kappa)$$ \hspace{1cm} (2.5)

where

$$L(\theta, \alpha, \kappa|Y) = \prod_{i=1}^{N}\prod_{v=1}^{V}\Phi(\alpha_v(\theta_{it} - \kappa_v))^{y_{iv}} \times (1 - \Phi(\alpha_v(\theta_{it} - \kappa_v)))^{1-y_{iv}}$$ \hspace{1cm} (2.6)

and $g(\theta, \alpha, \kappa)$ is the prior distribution over the parameters to be estimated.

For cases and votes that are identical across voting bodies (mostly votes on conference legislation taken in the House and Senate) we constrain the cutpoints to be the same by relabeling the votes with a common label. For cases and votes for which we have information on the relative locations of the cutpoints, we constrain the cutpoints to satisfy the inequality constraint implied by the information. This is implemented in the Bayesian sampling process via rejection sampling.

---

11 This parameter is standard in ideal point estimation theory and its precursor, item response theory (Baker 1992). Votes for which the alternatives are relatively close (meaning $(\lambda_v^C - \lambda_v^L)$ is relatively small) will have a low discrimination parameter as the non-spatial error term will be more likely to induce actors with preferences higher than the cutpoint to vote liberally and vice versa.
The ideal points of individuals vary over time in order to account for the ideological evolution of justices and long-serving senators and representatives. For justices, there is a broad consensus that at least some individuals exhibited substantial preference evolution over the course of their service (see, e.g., Bailey and Chang 2001; Martin and Quinn 2002; Epstein, Hoekstra, Segal and Spaeth 1998); for members of Congress, this is more debatable, as Poole and Rosenthal (1997) find little preference change. As we shall see, the results here provide evidence of substantial preference evolution of members of Congress (which is consistent with the anecdotal evidence on Senator Hollings in footnote 7).

There is a broad consensus that the preferences of at least some justices change over time (Epstein, Hoekstra, Segal and Spaeth 1988). Therefore for long-serving individuals (individuals who served more than 20 years), we assume that the ideology of individual i at time t is

$$\theta_{it} = \gamma_{0i} + \gamma_{1i}X_{it} + \gamma_{2i}X_{it}^2 + \gamma_{3i}X_{it}^3$$  \hspace{1cm} (2.7)

where the $\gamma$ parameters are preference parameters to be estimated, and $X_{it}$ is the years the individual has been in office.\(^{12}\) For Justices and members of Congress who serve between 16 and 23 years, we estimate their preferences with a quadratic equation (meaning we estimate $\gamma_{0i}$, $\gamma_{1i}$, and $\gamma_{2i}$). For Justices and members of Congress who serve between nine and 15 years, we estimate their preferences with a linear equation (meaning we estimate $\gamma_{0i}$ and $\gamma_{1i}$). We assume individuals who served eight or fewer years have fixed preferences (meaning

\(^{12}\)The years of service data are expressed in terms of deviations from mean years of service for computational convenience.
we estimate only $\gamma_{0i}$).

This functional form represents a tradeoff between flexibility and computation. The third order polynomial estimates four parameters to represent non-linear patterns of ideal point evolution. In contrast, Martin and Quinn estimate preferences for each term, with a Bayesian prior that preferences are similar to those in the previous term. This provides more flexibility, but increases complexity and computational time. In addition, the patterns of preference evolution that they do find with their method seem generally explicable in terms of a quadratic equation or, what we have here, a third order polynomial.

The model is estimated with Markov Chain Monte Carlo methods. These methods repeatedly sample from the posterior density of the parameter distribution. The mode, mean and standard error of the distribution of the parameters can then easily be derived from the mode, mean and standard error of the sampled observations. The appendix and references provide additional explanation.\textsuperscript{13}

**Data** The data consist of canonical data sets on court cases and congressional roll calls merged with an originally collected data set of bridge observations. There are 17,882 bridge observations, including 1,213 judicial comments on previous cases. The comments reflect the preference of the actor at the time the statement is made.

For both the Congress and the Supreme Court, we look only at votes and cases related to

\textsuperscript{13}While Bailey and Chang (2001) used an EM algorithm, the sampling Bayes approach is more flexible and readily estimates standard errors (Clinton, Jackman and Rivers 2004).
the major topics addressed by the courts in the post-war area, including crime, civil rights, free speech, religion, abortion and privacy. Focusing on these issues allows us to focus on the most relevant areas of political-judicial exchange and to minimize chances that our results are affected by behavior on secondary issues that did not necessarily have the same structure of preferences (Murphy 1962, 75; Dean 2001, 41). Nonetheless, it seems that in recent years positions on the issues we focus on appear to correlate highly with positions on economic and other matters (see, e.g., Martin and Quinn 2001).

We begin in 1950. Even as more data becomes available, we believe this remains an appropriate starting point as the ideological splits of the New Deal era revolved around economic legislation associated with the New Deal, topics that were quite different than those of the Vinson Court (1949-1953) and beyond when “social issues” such as race, civil rights and privacy dominated the court’s agenda. Given our assumption of a single policy dimension, we do not want to push too far into the past when this assumption becomes increasingly difficult to defend.

**Presidents**  Presidential positions on Supreme Court cases are drawn from two sources. One is a set of all statements by presidents on Supreme Court cases. We collected these data from presidential public papers, presidential library web sites and other sources. For example, George H.W. Bush on June 24, 1992 stated he was “very disappointed” by the Supreme Court’s ruling in *Lee v. Weisman* that a religious figure could not deliver an invocation at a
public school graduation ceremonies.

The second source of presidential positions on Supreme Court cases are Solicitors General amicus filings.\textsuperscript{14} Given the influence of the president in the selection of the Solicitor General and the power of the president to overrule or remove him or her, these can be treated as administration positions. Others make a similar assumption (Segal 1989 and Stimson, MacKuen and Erikson 1995).\textsuperscript{15}

Presidential positions on Senate and House votes are based on Congressional Quarterly data provided by Keith Poole. McCarty and Poole (1995) and Poole (1998) led the way in estimating presidential preferences simultaneously with members of Congress by including presidents’ on roll call votes.

\textsuperscript{14}We do not use Solicitor General positions on cases where the United States is a party. When the United States is a party to a case, precedent, the stakes of winning, the sometimes non-voluntary participation, or other non-ideological actors may be behind the position taken by the Solicitor General.

\textsuperscript{15}Several institutional and historical factors support the use of these filings for this purpose. The overt sources of presidential influence on the Solicitor General are clear: “the clearest and most important institutional linkage is with the President. It is the President who, by statute, nominates the Solicitor General and at whose pleasure he serves. Should he care to, the President has the coercive language to direct the activities of even a reticent Solicitor General” (Cooper 1990, 7).

These institutional powers may lead to influence even if we seldom see the president coercing the Solicitor General. First, Solicitor General appointees often share the preferences of the appointing president. As one of Reagan’s former Solicitors General, Charles Fried, said, “I have no trouble saying what the Attorney General and his crew want me to, because I’m more conservative than they are” (Cooper 1990, 7). Reagan surely anticipated Fried’s preferences before appointing him. Second, even when the preferences of the Solicitor General and the President diverge, the Solicitor General may choose to do the President’s bidding, out of deference or out of a desire to avert explicit intervention by the President.

In general, the evidence suggests that Solicitors General follow the desires of presidents. Meinhold and Shull (1998) found that presidential policy statements predicted Solicitor General amicus curiae briefs. In addition, examples in which presidents guide Solicitor General activities sometimes make it to the public sphere. Presidents Clinton and Bush both ordered their Solicitors General to change positions on cases (Fraley 1996, 22). President Kennedy had frequent contact with Solicitor General Cox (Segal 1989, 142), and President Eisenhower personally added several sentences to the government’s brief in \textit{Brown v. Board of Education} (Days 1995, 5). President Bush (at the urging of White House Counsel Alberto Gonzales) had the administration brief on \textit{Grutter v. Bollinger} re-written to be more accepting of affirmative action (Novak 2003).

In addition, when Solicitors General stray, presidents push them out. Reagan essentially fired Solicitor General Lee when he expressed reluctance in pursuing Reagan’s agenda (Norman-Major 1994). Nixon forced out Solicitor General Griswold in 1972 due to a perception that Griswold was too liberal (Salokar 1992, 41).
Senators and Representatives  Congressional positions on Supreme Court cases are based on amicus filings by members of Congress, statements in support of or in opposition to specific decisions by the Supreme Court, sponsorship data for legislation that explicitly or implicitly took a position on Supreme Court cases, and roll call votes that explicitly took a position on specific Supreme Court cases.\textsuperscript{16} A good example of a Congressional bridge observation would be an October 1999 vote in the Senate on an amendment stating that \textit{Roe} was “an appropriate decision and secures an important constitutional right.”

Justices  Data on Supreme Court voting from 1951 to 2008 are available from Spaeth (2009). We include all cases with bridge or linkage information and was “important.” \textsuperscript{17}

The observations of justices taking positions on cases that arose prior to the justices’ arrival on the bench were taken from written opinions. Opinions were identified by (1) searching for the phrases such as “wrongly decided” or “correctly decided” (2) examining every case that overturned precedent and (3) working through issue-specific discussions in legal reference books. An example is Justice Thomas’s position on \textit{Roe} discussed above.

When a case clearly and directly overturned a precedent, a vote in favor of overturning

\textsuperscript{16}The amicus filings were identified either in the \textit{Supreme Court Compendium} or with Lexis/Nexis searches we performed. The statements are primarily from an extensive search of the \textit{Congressional Record}. For example, Sen. Paul Douglas (D, IL) characterized \textit{Brown} as a “correct and noble decision” (\textit{Congressional Record} 110: 20910) while almost 100 Southern Democrats signed the Southern Manifesto stating the decision was “a clear abuse of judicial power” (\textit{Congressional Record} 102: 4459).

\textsuperscript{17}A case was deemed “important” if the case received news coverage in the New York Times (cite), Congressional Quarterly (cite), bridge observation. There was a twofold reason for limiting the cases in this manner. The effects we are interested in only truly matter if they affect important cases, broadly defined. In addition, the MCMC estimation process is very time consuming and limiting cases reduces the time for each estimation. Among the various robustness analyses reported below are estimates in which the case selection was not limited to important cases; the results do not appear to differ. In order to ensure an adequate number of observations for every year, we include randomly selected cases that are not important for years that would otherwise have a small number of cases. We also do not use cases that cannot be coded on a liberal and conservative scale (Harvey 2008).
the precedent was also coded as a vote against the original decision. For example, in *Roper v. Simmons* (2005), the Court explicitly overruled *Stanford v. Kentucky* (1989). A vote in favor of *Roper* was coded as a vote against *Stanford*. In some instances, such as this one, a justice changed his or her position on a case. In this case, Kennedy was conservative on the original case and liberal on the later case. Such observations are useful in gauging the ideological evolution of individual justices.

**Data Validity**  Even though using position taking that crosses institutional boundaries helps us solve some problems, doing so raises new questions. First, is non-voting position-taking less consequential than court and congressional voting and thus less valid? This is a good question that cuts to the heart of the approach here. Three factors lead us to believe that this is not a fundamental problem. First, these observations tend to reflect commitment to the positions stated. They are, in one way or another, based on official acts (ranging from amicus filings to bill co-sponsorship to statements on the floor of Congress). In addition, the member publicly stated his or her position more than one time for more than 20 percent of the observations (although we do not use or count repeat observations in the analysis unless they are separated by more than five years). Second, public position-taking on Supreme Court cases has clear electoral and political consequences. No contemporary politician would treat his or her position on *Roe* as a trivial act, nor would politicians in the fifties and sixties treat their public pronouncements on *Brown* or busing cases as inconsequential. Even comments
on less prominent cases can be politically relevant, as happened when Sen. Santorum (R, PA) created a controversy with comments on *Texas v. Lawrence* (2003) (Loughlin 2003).\(^{18}\) Indeed, it is the importance of such statements that have made the use of non-voting data for the purpose of preference measurement more common. For example, presidential NOMINATE and ADA scores are partially based on presidential position taking. Likewise, Ansolabehere, Snyder and Stewart (2001) use comment data from candidate surveys to identify legislators’ preferences. Third, we must not overstate the consequences of most roll call votes. Because most roll call votes are decided by more than one vote, legislators have considerable leeway to vote based on position-taking rather than substance. In addition, Poole and Rosenthal (1997, 69) provide evidence that “roll call voting is concerned with position-taking rather than policy-making.”

A second question is whether non-vote data are fundamentally different because of their more optional nature. Members may virtually be forced to take positions via roll call votes, but generally may avoid making public statements about Supreme Court cases. One consequence is that a non-random selection of legislators may take public positions on Supreme Court cases, something that in fact comes to pass in our data. This, however, will not bias the estimation because for the results to be contaminated by selection bias, the error in the selection equation must be correlated with the error in the preference equation (Greene 2000, \(^{18}\)In the process of making clear his opposition to the decision, then Senator Santorum stated “Every society in the history of man has upheld the institution of marriage as a bond between a man and a woman. Why? Because society is based on one thing: that society is based on the future of the society. And that’s what? Children. Monogamous relationships. In every society, the definition of marriage has not ever to my knowledge included homosexuality. That’s not to pick on homosexuality. It’s not, you know, man on child, man on dog, or whatever the case may be.”)
Selection bias is not induced simply if relatively extreme members are more likely to take positions.

Another possibility that is statistically more troubling is that the preferences of individuals are somehow different when they choose to take positions than when they are forced to vote on cases or roll calls. In this case, the error in the selection equation would be correlated with error in the outcome and selection bias would be possible. This could occur if, for example, a member of Congress only takes positions on court cases in order to look conservative (and does not do that on roll calls, the above notwithstanding). While it would appear reasonable to assume that the public persona politicians would like to exhibit would be similar whether acting on roll calls or other public acts, we cannot know for certain that this is the case. To test for the possibility of selection bias, we compare preferences expressed via roll call votes and preferences expressed via non-vote public positions. To do so, we generate two sets of preference estimates: one based only on Senate roll call votes, and the other based on court data and voluntary senate data such as public statements and amicus filings. Clear differences in preference ordering across these two estimation procedures would indicate that the ordinal ranking of senators based on voluntary observations was markedly different from senatorial behavior on roll-call observations. This does not appear to be the case, however, as the correlation between the two preference estimates is 0.89.

---

19 Of course, justices and members of Congress are not forced to vote, but it appears that abstention is costly. A justice who fails to vote without a good reason will probably lose the respect of his or her peers. A member of Congress who abstains may lose respect of other members and may suffer campaign attacks based on low-voting rates.

20 Note that selection bias can occur even if there were no difference in preference distributions between commentators and all members of Congress. That is, a perfectly representative sample of legislators who made comments only when they were on
2.5 Preference Estimates

The results from our data collection and estimation procedure are plotted in Figures 2.9 and 2.10. The scale in each plot is the same, allowing us to see that the results accord with intuition. Rehnquist and Thomas are at the conservative end of the spectrum, O’Connor is toward the middle and Stevens and Souter are at the liberal end. Souter began his career on the Supreme Court close to O’Connor, but moved left over time. Harlan’s estimated ideal points were generally between those estimated for O’Connor and Souter, while Brennan’s were consistently to the left of Stevens and the rest.

Figure 2.11 plots the estimated preferences of the Senate median, the president and the court median. In the fifties, the court median started between the President and the Senate and then moved to the left of even the president. After intense court-curbing activities of the mid-fifties, the court median moved back to the middle. In the sixties, the court moved firmly in the liberal direction, especially after Goldberg replaced Frankfurter in 1962. After a big rightward shift during the early Burger Court, the court median trended conservative in the seventies and eighties. In the nineties, the court median was more conservative than President Clinton and congressional medians. The court remained to the right of the congressional medians in George W. Bush’s first term; the congressional medians then briefly converged to where the court was in 2005 and 2006, but then moved clearly to the left after

---

21One needs to be careful in ascribing a single median to the court in a year. In years with an appointment, the median may shift considerably. In later chapters in this book when we need court medians for analytical purposes, we calculate them on a case by case basis depending who was on the court.
Figure 2.9: Estimated ideal points of Supreme Court Justices, page 1
Figure 2.10: Estimated ideal points of Supreme Court Justices, page 2
the Democratic victories in the 2006 elections.

One of our interests is whether the elected branches of government constrain the Supreme Court. Our full analysis is in Chapter 4, but for now note that the revealed one-dimensional preferences of the median of the Supreme Court have generally been within the pareto set defined by the president, House median and Senate median. The figure does not answer the extent to which this has been driven by the appointment process or by politically imposed discipline or perhaps another mechanism, but it does indicate that the institutional environment of the court in the post-war era has generally kept it from moving outside of the
bounds defined by the elected branches of government (Dahl 1957).

We can also explore whether apparent anomalies found in other preference estimates recur when using our approach. In contrast to Martin and Quinn’s estimates which imply that the court median neared its conservative peak in 1972, our estimates indicate that the court moved consistently to the right in the late sixties and eighties, but was more liberal in the early seventies than it is today (compare Figure 2.11 and Figure 2.5).

Our congressional estimates do not have the anomalies found in Common Space scores. Consider the preferences of southern Democrats over time. For example, Figure 2.12 depicts the preferences over time of a southern segregationist Senator (Eastland) and a moderate modern Democrat (Hollings). (Bayesian 90 percent confidence intervals are indicated with light colored lines around the estimates.) In contrast to Common Space preference estimates, our results show clear differences between these two; our results also are consistent with evi-
idence that at least some members of Congress changed preferences over time. Bailey (2007) shows that ideal points based on the bridging approach are better able to explain Senate voting on Supreme Court nominations than a Epstein, Lindstadt, Segal and Westerland’s (2006) estimation approach that builds on Common Space scores.

There are two factors behind the differences in our results from the Common Space scores. First and substantively important here and even more so in other projects, our data is limited to “social issues” broadly construed (e.g. civil rights, civil liberties, speech, crime, abortion). Poole and Rosenthal scores are based on all votes across years in which dimensionality changes. To see to what extent this explains the seemingly better fit with reality for our measures, we estimated Nominate scores based only on the roll calls in the sample.\(^{22}\) In these estimates, the modern and segregationist southern Democrats are distinguishable; for example, Eastland and Ellender each have estimated ideal points of 0.23 while Exon and Hollings have estimated ideal points of 0.04 and 0.06, respectively (with the Nominate-produced conditional standard errors in the range of 0.01 to 0.03). This implies that the similarity of southern segregationist and modern moderate Democrats in both dimensions of Common Space scores was due to shifting mapping of issues onto the underlying latent space, a complication we avoid in this book by limiting the analysis to a more focused set of issues.

But sample selection is far from the whole story. The Nominate scores from the restricted

\(^{22}\)These scores were estimated with W-Nominate from Poole’s Voteview website and using only congressional and presidential data. I fixed Sen. Edward Kennedy (D-MA) as the “left” anchor and Rep. Charles Stenholm (D-TX) as the “up” anchor. Results differed with different dimensionality or anchoring assumptions, but the general pattern discussed here was stable.
sample produce results that imply, for example, that the cutpoint for the 1991 Civil Rights Act was to the right of the 1964 Civil Rights Act (meaning someone who was conservative in the 1991 Civil Rights Act would have been conservative in the 1964 act). The restricted sample Nominate measures also imply, for example, that the Senate median was more or less constant from 1961 to 1980 and that the Senate median was not affected by the 1958 elections. The methods and data described in this chapter are specifically designed to facilitate careful analysis of whether we believe such results and, as it happens, produce quite different results (note the evolution of Senate preferences in Figure 2.11).

2.6 Conclusion

Accurate measurement is essential for quantitative theory testing. If we cannot characterize preferences with confidence, we cannot determine the forces that shape them or how they in turn affect outcomes. In the case of research crossing institutional boundaries and spanning time, it has been particularly challenging to generate comparable preference estimates, a fact that has left several research agendas waiting on development of valid preference measures that are comparable over time and across institutions.

In this chapter, we offer two contributions. First, we show that measurement matters. We highlight theoretical challenges and show the relevance of these issues for active research. Ad hoc and equally plausible approaches to inter-institutional comparisons yield starkly
different conclusions about the relative policy preferences of the president and congressional medians. Martin and Quinn’s and Poole and Rosenthal’s widely-used preference estimates imply temporal preference relations that are, in some respects, hard to believe. The examples and analysis here not only set the stage for the rest of the book, but also point scholars to be more critically reflective about the preference measures they use.

Second, we provide a method and data for producing preference estimates that are comparable across time and institutions. We use three types of information: “bridge” observations of actors taking positions on cases or votes in another institution; “bridge” observations of actors taking positions on cases or votes in a previous time period; and substantive information about the relationship of vote cutpoints over time. These data are incorporated into a spatial ideal point model estimated via Bayesian Markov chain simulation methods. The payoff is that the method produces preference estimates that avoid the anomalies found with other widely-used measures, measures that can be used to address a broad array of questions in the literature.