Online Appendix to "Estimating Dynamic Ideal Points for State Supreme Courts"

Jason H. Windett* Jeffrey J. Harden† Matthew E.K. Hall‡

Forthcoming, *Political Analysis*

Contents

A Data Collection Process **ii**

A.1 Reliability and Validity v

B Court Medians Over Time **vii**

C Dynamic IRT Ideal Points **vii**

D Additional Fit Statistics **vii**

D.1 Cross-Validated Fit Statistics with the Unscaled Dynamic IRT Measure vii

D.2 In-Sample Fit Statistics viii

D.3 Explaining the SDIRT Measure's "Losses" viii

D.4 Correlations Between Measures ix

E Application: Assessing Policy Responsiveness **x**

E.1 Testing the Determinants of Policy Responsiveness xii

E.2 Dynamic Representation in State Courts of Last Resort xiii

*Corresponding Author. Assistant Professor, Department of Political Science, Saint Louis University, 127 McGannon Hall, 3750 Lindell Blvd, St. Louis, MO 63108, jwindett@slu.edu.

†Assistant Professor, Department of Political Science, University of Colorado Boulder, 416 Fleming, UCB 333, Boulder, CO 80309, jeffrey.harden@colorado.edu.

‡Assistant Professor, Department of Political Science, University of Notre Dame, 217 O'Shaughnessy Hall, Notre Dame, IN 46556, matt.hall@nd.edu.

E.3 Research Design xv
E.4 Analysis xviii

A Data Collection Process

First, we outline our data collection process, which is also described in [citation omitted]. We collected our data via automated textual analysis through the Python programming language. This method allows us to quickly and reliably collect data on state supreme court decisions. Our computer program extracts critical pieces of information from text files of state supreme court opinions and converts the information into quantitative data. This data collection method produces reliable measures of state supreme court decision making and facilitates a wide range of empirical analyses.

We began by obtaining every state supreme court decision for each state from 1995 through 2010 through the Lexis-Nexis Academic online database. Lexis-Nexis includes both published and unpublished opinions for those state courts of last resort that allow unpublished decisions (see Serfass and Cranford (2001) for details on the variation in publishing rules across states). However, for two reasons, we do not utilize unpublished opinions in the creation of our ideal point estimates. First, and most importantly, these unpublished opinions do not report any disagreement between the justices. Accordingly, they all appear as unanimous decisions and are, therefore, omitted from our calculation of ideal point estimates. Second, unpublished opinions are generally used when a court does not "modify," "clarify," "explain," or even "call attention to" existing law (Serfass and Cranford 2001, 252). In some states, this practice is an explicit rule; in other states it is an informal norm. Therefore, these opinions are, by definition, unrelated to the creation of legal policy in which we are interested. Accordingly, unpublished opinions are comparable to voice votes in legislatures, which are always disregarded for the creation of ideal point estimates.

We then created automated textual analysis programs (or "scrapers") using the Python programming language to extract data from these text files. Because the formatting of published opinions varies between state supreme courts, it was necessary to create fifty-two unique scrapers, one for

each state supreme court.[1] We then utilized these scrapers to extract critical pieces of information on each state supreme court case and convert the text information into a spreadsheet with data for each state. Our data include the following pieces of information for each case in our dataset:

- The name of the case, case citation, and LexusNexis citation.
- The day, month, and year of the decision.
- The prior history, procedural posture, and disposition of the case in text format.
- The area of law at issue in the case as identified by the first Lexis-Nexis headnote.
- The opinion author and dissent author(s), if any.
- The vote of each judge in the case.

In order to create accurate Python code to extract this data, we began by hand-coding the dockets of a small sample of states across different years. Our hand-coding revealed that the format of the pertinent information in the dockets was relatively uniform for most of our variables; however, our initial hand-coding revealed that the dockets varied widely in the reporting of judges' votes both across and within states. For example, in 14 states, the dockets only listed the names of opinion-writing judges, dissenting justices, and those not participating. In these cases, we imputed the concurring votes of other sitting judges into the data. This process required a precise roster of service dates by all judges in these 14 states. To complete this task, we utilized State Court websites, Lexis-Nexis news searches, and direct contact with state courts to ensure that we accurately imputed the votes of judges who were actually on the court. Our Python scrapers code each judge in a specific year with a concurring vote unless he or she was noted in the decision as dissenting, dissenting in part, or not participating. When judges joined or left the court in the middle of a year, we manually deleted concurring votes by those judges in any case in which they did not sit based

[1] We have separate scrapers and data for the Courts of Criminal Appeals and State Supreme Courts in Oklahoma and Texas.

on the date of the oral argument. Consequently, our data identify the name of each judge sitting on each case and indicate whether that judge voted with the majority or dissented in the case.

We also corrected our data to account for changes in the structure, organization, and procedures of each state supreme court. For example, the Nevada judicial system lacks an intermediate court of appeals. Consequently, the Nevada Supreme Court hears most appeals in the state. To accommodate the court's large docket, the Nevada state legislature expanded the court from five members to seven in 1997. Even more importantly, the court hears most cases in panels of three judges. Accordingly, we adapted our scraper for the Nevada Supreme Court to detect and report only the specific panel of judges that heard each case.

Another coding problem arose when two or more judges on the same state supreme court shared a common surname. Most dockets typically list only the last name of sitting judges; however, in some states judges with the same surname are noted with a first initial. This practice was initially problematic for our scrapers because we used punctuation marks to demarcate judge names. To solve this problem, we revised our Python code to merge the first initial into the surname so that the program could distinguish between judges. For example, in Minnesota, Paul Anderson and Barry Anderson have served concurrently since 2004. Our scraper first changed their names in the dockets to "Panderson" and "Banderson" so that it could then easily distinguish between them to code their votes. We did the same type of change for individuals with two capital letters in their last name–for example, "McDonald" was changed to "Mcdonald".

Once we completed a scraper for specific state, we began a multi-step process to maximize the accuracy of our final data. Our Python scrapers produced four separate documents–(1) the data, (2) a list of cases and votes with any dissenting votes, (3) a list of cases with any judges listed as "not participating" or "recused", and (4) a list of cases in which any error was encountered while running the Python code. We then assigned a research assistant to review documents 2-4 and compare these reports to the data (document 1). If our data was missing any of the dissenting votes or non-participants, we went back to the dockets, looked for common errors, and adjusted the Python code accordingly. We then repeated these steps until the data correctly identified all

of the dissenting votes and non-participants. Lastly, we assigned a research assistant to compare about 25 cases in each year from the original dockets with our finalized data to ensure accuracy.

A.1 Reliability and Validity

We next turn our attention to reliability and validity checks of our data. Because the time span of our data collection effort overlaps with that of Brace and Hall, we are able to conduct a reliability check on our data by comparing it to the data they collected for the years 1995 through 1998. We randomly sampled 500 cases from the Brace and Hall data and compared their hand-coded data to the same 500 cases in our computer generated data. We then conducted a validity check of our data by assigning a research assistant to compare the same 500 cases in our data to the actual dockets from which we extracted the information.[2] We specifically examined between 14-18 pieces of information from each dataset: the case title, citation, majority opinion author, dissenting opinion author, and votes from 5-9 judges in each state depending on court size.[3] We code the votes of the judges as either joining or dissenting from the decision of the court. Results from both the reliability and validity checks are reported in Table A1.

The first important aspect of Table A-1 is the high level of agreement between our data and the Brace and Hall data. With the exception of the case citations, our data corresponds with the Brace and Hall data at over 93.9% for each variable. With regard to case citations, our data agrees with the Brace and Hall data in only 84.6% of cases. An examination of these cases indicates that almost every case of disagreement between our data and the Brace and Hall data is the result of human coder error in the Brace and Hall data. In many cases of disagreement, the Brace and Hall data has typos, inaccurate values, or missing alphabetic prefixes that some states include in their citations. When we match our data to the dockets, our citations match up 98.8% of the time. In fact, our validity check indicates that our automated scraper program was remarkably accurate at coding the data. Agreement between our data and the actual dockets was higher than 98% for each

[2]To ensure the accuracy of our validity check, we assigned a second research assistant to compare 100 of these cases with the actual dockets. These two research assists agreed in 100% of the cases they examined.

[3]The manner in which we scraped our data and Brace and Hall coded their data does not allow for a direct comparison across all pieces of information. For example, they provide numeric values of the individual steps in the appeals process, whereas we provide the direct text.

variable.

[Insert Table A-1 here]

Within the reliability checks, our largest concern is with the actual behavior of the judges. We match our data to the Brace and Hall data in 93.9% of all judge votes; however, our validity check suggests that our automated program outperforms the hand-coding in the Brace and Hall data. Most of the discrepancies between our data and the Brace and Hall data are the result of their human coders failing to include a judge in a case when we accurately pick up the vote. Overall, the Brace and Hall data missed 141 votes in these cases; our program missed only 46 votes.

This is not to say our data is without comparative disadvantages. The Brace and Hall data includes many nuances that our data collection process fails to detect. For example, we do not have substitute judges in our data. We only include the full-time judges. If a temporary judge becomes a full-time judge at some point in their career, we can pick up their votes when they were substitute judges. The Brace and Hall data codes the votes of all substitute judges; however, these votes comprise only a small minority of votes on these courts.

Our data collection process also encounters difficulties when judges retire after hearing a case, but before the decision is written. Our automated textual analysis software is not always able to detect the nuance required to code this information. For example, many states list the judges who heard oral arguments for the case; if the judge retires before the decision is written, this information is indicated in a footnote. Our data does not pick up these footnotes and, therefore, misrepresents these votes. This problem arises in fewer than 10 cases in our sample.[4]

Overall, our data has high levels of validity when compared to the actual dockets, meaning our scraping procedure yields data of remarkably high quality. Additionally, our data are quite reliable

[4]Below are two examples of cases with a footnote designating a special judge or a non-participating judge that we do not accurately pick up in our data.

JUDGES: Shearing, C.J., Rose, J., Sullivan, D.J.[4] SPRINGER, J., dissenting.

4 The Honorable Jerry V. Sullivan, Judge of the Sixth Judicial District Court, was designated by the Governor to sit in the place of the Honorable A. William Maupin, Justice. Nev. Const. art. 6.

JUDGES: Before Carson, Chief Justice, and Gillette, Van Hoomissen, Fadeley, Unis,** Graber, and Durham, Justices.

** Unis, J., did not participate in this opinion.

compared to the Brace and Hall data obtained using human coders. The most important aspect of the data, the judges' votes, is highly valid and reliable.

B Court Medians Over Time

Figure A-1 shows the dynamic element of our scaled ideal points. The graphs plot the dynamic common space median of each court for the time period 1995–2010. For ease of visibility, we separate the 52 dynamic medians into four groups of 13 state courts.

[Insert Figure A-1 here]

C Dynamic IRT Ideal Points

In Figures A-2–A-14 we graph our dynamic IRT ideal points of the justices not included in the main text. The graphs show considerable change over time in individual judges' ideal points.

[Insert Figures A-2–A-14 here]

D Additional Fit Statistics

Here we report additional results regarding the fit statistics presented in the main text. First, we present an additional out-of-sample predictive test generated and evaluated through cross-validation based solely on our unscaled dynamic IRT measure. Secondly, we report the in-sample fit statistics with our unscaled IRT measure. Finally, we report correlations between our unscaled measure, CFscores, and PAJID.

D.1 Cross-Validated Fit Statistics with the Unscaled Dynamic IRT Measure

Table A-2 replicates Table 1 from the main text, but with the unscaled version of our dynamic IRT measure.

[Insert Table A-2 here]

The results show that our unscaled dynamic IRT measure generally performs the best among the three alternatives. In this case the model with the IRT measure produces the largest CC and

APRE for 44 of the 52 state courts. The CFscores produce the best fit for seven of the eight remaining states according to both statistics. On average, the IRT-based measure's CC is 0.879, versus 0.812 for PAJID and 0.834 for CFscores. The average APRE values reflect the same pattern: 0.568 (IRT), 0.345 (PAJID), and 0.420 (CFscores).

D.2 In-Sample Fit Statistics

Table A-3 reports the fit statistics using the unscaled dynamic IRT measure and the full sample of cases (i.e., without cross-validation).

[Insert Table A-3 here]

The results again show that our unscaled dynamic IRT measure generally performs the best among the three alternatives; the model with that measure produces the largest fit statistics for 47 of the 52 state courts, with the CFscores producing the best fit in the other five states. On average, the IRT-based measure's CC is 0.893, versus 0.835 for PAJID and 0.856 for CFscores. The average APRE values reflect the same pattern: 0.594 (IRT), 0.385 (PAJID), and 0.458 (CFscores).

D.3 Explaining the SDIRT Measure's "Losses"

Table 1 in the main text shows a few instances in which PAJID or CFscores beat our measure with respect to the fit statistics. To the best of our knowledge there is nothing systematic across these states that would lead us to believe this is a major problem. What follows are the best explanations we can produce for each state. We do not consider Iowa in depth because the CFscores only beat our measure based on the proportion correctly classified, and even then it is by a very small amount (0.001).

In Michigan the CFscores benefit from a great deal of information compared to other states. When looking at the DIME database, on average, Michigan judges receive contributions from 825 individuals. This is far greater than the national mean of 276 contributors per judge. Furthermore, only Florida has a higher average number of contributors. In Florida, however, the data only cover four judges, while Michigan has 63 judges receiving contributions. Moreover, the maximum number of contributors in Michigan for a judge is 3,981. This is the second highest maximum number

of contributors in DIME. Thus, we suspect that our IRT estimates are contributing a smaller proportion of the total information compared to other states. In other words, adding the IRT information does not contribute much to what is already in the CFscores in that state.

In Alaska and Indiana, the left-right constraints set for the ideal point estimation may be causing higher levels of uncertainty in the ideal points. In each of those states, there is only a single judge who was appointed by a Republican governor, allowing us to only place one judge as the right constraint. In both cases there is not much data for this judge. In Indiana, this judge retired in 1996, and in Alaska this judge was not appointed until 2008. These issues could be contributing to the lower classification rate and APRE in the cross-validation tests. However, this is not to say that the scores in those states are necessarily wrong or skewed. As evident in the fit tests, they are still very close to the CFscores in predictive capacity.

Missouri is the only state in which the PAJID scores outperform both our measure and the CFScores. When looking at our scaled ideal points in Missouri, there are three distinct clusters of judges: a conservative group, a liberal group, and a moderate group. These judges were almost all appointed by governors ideologically similar to them: John Ashcroft (extreme conservative), Mel Carnahan, (extreme liberal), and Bob Holden (moderate). Thus, Missouri may simply be an example in which the theoretical framework of PAJID works well. These judges' voting choices very closely reflect the composition of the Missouri government at the time of their appointments.

It is also important to note that the magnitude of fit performance favors our measure. In the five state courts in which our measure gets beat on the proportion correctly classified, it is by an average of 0.01. In the four state courts in which our measure gets beat on APRE, it is by about 0.03, on average. In the other 47 and 48 courts, respectively, our measure outperforms the second best measure by an average of 0.048 and 0.163.

D.4 Correlations Between Measures

In Table A-4, we report the correlation coefficients between the PAJID scores and our unscaled dynamic IRT ideal points and between the CFscores and our unscaled dynamic IRT ideal points. Our measure has a moderately strong correlation with the CFscores, with an overall correlation

coefficient of 0.4494. The state variation ranges from -0.1989 in California to a high of 0.8752 in Wisconsin. In contrast, the PAJID scores have a weak correlation with our IRT ideal points, with an overall coefficient of -0.2492. In a number of states, the correlation coefficients are very close to 0, indicating no relationship, while in a few states (e.g., Oklahoma, Delaware, and Missouri) there is a negative correlation stronger than -0.8.

[Insert Table A-4 here]

E Application: Assessing Policy Responsiveness

Numerous studies find that elite policy decisions often tend to reflect popular opinion (e.g., Erikson, MacKuen, and Stimson 2002; Erikson, Wright, and McIver 1993; Fiorina 1974; Froman 1963). In fact, a variety of policymakers appear to follow public sentiment, including legislators (Stimson, MacKuen, and Erikson 1995), executives (Canes-Wrone and Shotts 2004), bureaucrats (Rabinovich 2008), and judges (McGuire and Stimson 2004). Yet, scholars continue to debate which mechanisms drive this process—why does public policy mirror popular opinion?

Undoubtedly, this mass-elite policy connection is partially driven by the changing composition of governing institutions. Policy decisions by public officials tend to reflect popular sentiment as those officials are replaced over time, whether through electoral turnover (Hakhverdian 2010; Stimson, MacKuen, and Erikson 1995), judicial appointment (Norpoth and Segal 1994), or the selection of bureaucratic agents (Brehm and Gates 1997). After securing office, elite decisions may correlate with shifts in popular opinion due to the independent influence of social forces (Collins, Norton, Manning, and Carp 2008; Giles, Blackstone, and Vining 2008) or interest group pressure (Carpenter 2002; Collins 2007). It is also possible that public officials actually influence popular opinion (Franklin and Kosaki 1989; Kernell 1997; Sigelman 1980; Tulis 1988). However, individual policymakers may also directly respond to shifts in popular preferences for several different reasons.

First, the replacement of some officials may pressure others to comply with changing popular preferences. The system of checks and balances deliberately empowers separate institutions

to exert influence over one another: executives can veto legislative enactments, legislatures can thwart executive policies, and both can sanction judges. Consequently, even unelected officials, such as bureaucrats and judges, may respond to popular preferences due to pressure from other officials (Stimson, MacKuen, and Erikson 1995; Weingast 1984). For example, U.S. Supreme Court justices may react to public opinion because they fear legislative override of their decisions (Ferejohn and Shipan 1990; Gely and Spiller 1990), retributive sanctions against their institutional power (Casillas, Enns, and Wohlfarth 2011; Epstein and Knight 1998), or nonimplementation of their decisions (Hall 2014; McGuire and Stimson 2004). Thus, even without a personal reelection incentive, appointed judges may have good reason to heed popular sentiment. However, evidence of the High Court's responsiveness is tentative (Flemming and Wood 1997; Mishler and Sheehan 1993), disputed (Giles, Blackstone, and Vining 2008; Norpoth and Segal 1994; Segal and Spaeth 2002), and possibly restricted to a small subset of cases (Hall 2014).

Alternatively, electoral accountability may be the critical element inducing policy responsiveness. The goal of reelection is a powerful motivation; indeed, "[elections are] where representative form[s] of government begin and end" (Fenno 1996, 9; see also Mayhew 1974; Miller and Stokes 1963). Accordingly, policymakers may adjust their decisions to reflect shifts in public opinion in order to improve their odds of reelection (Erikson, MacKuen, and Stimson 2002; Gordon and Huber 2002; Kousser, Lewis, and Masket 2007). In fact, the threat of electoral defeat may be so powerful that it prompts responsiveness irrespective of a legislator's electoral experiences (Fiorina 1974). For example, Bartels finds that "representatives who win with 100% of the vote appear to be about as responsive to constituency opinion as those who win with 51% of the vote" (1991, 468; see also Gulati 2004). This behavior may arise because larger vote margins do not necessarily indicate greater incumbent security (Jacobson 1987). Consequently, "incumbents are subject to random terror" (Ansolabehere, Brady, and Fiorina 1992, 36) and, therefore, respond to popular preferences regardless of their personal electoral experiences.

E.1 Testing the Determinants of Policy Responsiveness

Several practical limitations hinder the study of how elections and competition influence elite behavior. First, it is difficult to compare the responsiveness of public officials who do and do not have to stand for reelection because most major political offices do not vary in selection mechanism. Some studies overcome this problem by testing the influence of electoral rules and institutional design in cross-national studies (e.g., Hobolt and Klemmemsen 2007; Roberts and Kim 2011); however, this research examines only a few countries that vary on a wide range of social and political dimensions. The complex nature of cross-national research makes it difficult to isolate institutional effects. Even more importantly, these studies compare different types of electoral systems rather than elected versus unelected offices. Accordingly, few previous studies have directly assessed the influence of elections on policy responsiveness (cf. Besley and Coate 2003).

Second, assessing the effects of electoral competition on public officials is difficult because competition (e.g., in the form of a challenger or a viable challenger) is endogenous to incumbent behavior (Gordon and Huber 2007). That is, incumbents may respond to public pressure when they face competition, but if they respond to public pressure they should avoid competition. Consequently, testing the effect of actual competition in the next election on incumbent behavior is fraught with problems. An alternative approach is to use the vote margin of another race (such as presidential vote share) as a proxy for two party competition (Griffin 2006; Gulati 2004). This approach avoids the endogeneity concern, but sacrifices some theoretical fit. Our interest here is in state supreme court judges' perceived electoral threats, not generic evaluations of whether the two major parties are competitive among the electorate.[5]

To overcome these limitations, we follow several groundbreaking studies that focus on state courts of last resort (e.g., Brace and Hall 1995; Hall 1992). These institutions offer several advantages for the study of electoral competition and policy responsiveness. First, state supreme court elections and decisions attract very little public attention (Cann and Wilhelm 2011; Mathias 1990).

[5]Previous electoral experiences may also suffer from problems of endogeneity (Gulati 2004; Sullivan and Uslaner 1978).

Accordingly, these institutions pose a conservative test for responsiveness and the potential influence of competition (Gordon and Huber 2007, 110). Second, despite their relatively low public profile, these courts possess important policymaking responsibilities. Therefore, unpopular rulings may attract negative attention from other officials. Third, these courts share numerous similarities in operation, organization, and social context to facilitate meaningful comparisons, yet, they vary significantly in selection mechanism. Fourth, state supreme court justices usually represent the same constituency. This institutional feature enables us to disentangle a state's level of two-party competition from a judge's personal electoral experiences.

Judicial institutions also pose a special normative dilemma for policy responsiveness. In the legislative and executive branches, responsiveness is a hallmark of democratic legitimacy; however, in courts of law, responsiveness lies in tension with principles of justice, precedent, and judicial independence. Accordingly, judges may tend to be less responsive than other public officials, and a judge's responsiveness may be moderated by her personal *legal* experiences in addition to her electoral experiences. Do legal principles win out over popular pressure and electoral incentives? Or do state supreme court justices tend to behave like other politicians? And, perhaps most importantly, what institutional structures and individual experiences moderate judicial responsiveness? To examine these questions, we turn to research on policy responsiveness in state courts.

E.2 Dynamic Representation in State Courts of Last Resort

Numerous studies argue that elections prompt state judges to follow popular preferences. Most of this literature focuses on the manner in which judges reach the bench and retain their jobs. The American states vary widely in judicial selection mechanisms. Seven states elect justices through partisan elections, 15 hold nonpartisan elections, and 16 appoint members initially, then hold retention elections to remain on the court. Finally, 12 states use gubernatorial or legislative appointment, with many of these states requiring reappointment after varying term lengths or a mandatory retirement age (American Judicature Society 2013).[6] These selection differences influ-

[6]Reappointment mechanisms vary across these 12 states. Massachusetts and New Hampshire allow judges to serve until they turn 70 and then must retire. Connecticut, Delaware, Maine, New Jersey, New York, and Rhode Island have gubernatorial reappointment processes that include a confirmation vote by one or both legislative chambers. Vermont

ence the choices judges make (Brace and Hall 1990; Hall and Brace 1989; Brace and Hall 1993; Langer 2002). Elected trial court judges (Gordon and Huber 2007; see also Huber and Gordon 2004) and state supreme court justices (Brace and Hall 1995, 1997; Brace, Hall, and Langer 1999, 2001; Brace and Boyea 2008) are both more responsive to public preferences than judges selected through appointment.

Other studies emphasize electoral competition; i.e., state judges may respond to public opinion only when they rationally anticipate future competition. These studies test the effects of competition in different ways. For example, judges tend to follow popular preferences when they near the end of their terms (Hall 1992; Caldarone, Canes-Wrone, and Clark 2009) and when general partisan competition in the state is high (Brace and Hall 1995, 1997). These findings suggest that judges respond to public opinion because they anticipate future competition; however, they do not capture rational expectations at the individual level. State-wide competition and the proximity of a judge's reelection contest are only indirectly related to that judge's rational expectations regarding future competition.

Secondly, mandatory retirements may influence the behavior of both appointed and elected judges. In thirty-three states, judges may not serve past a certain age (typically 70 or 72). Extant literature on behavior of state courts often overlooks mandatory retirement as an influence on judicial behavior. In other institutions, however, there is mixed evidence on the influence of retirement on roll call behavior (Rothenberg and Sanders 2000; Clark and Williams 2013; Wright 2007). The influence of public opinion in state courts may be conditional based on mandatory retirement. The electoral and re-appointment constrain is no longer influencing the behavior of judges.

This rich, extensive literature offers valuable insights into policy responsiveness; yet, this groundbreaking work has been hampered by a lack of available data. For example, some studies examine judicial behavior over only a few years (e.g., Brace and Boyea 2008; Cann and Wilhelm 2011). Because these studies essentially employ cross-sectional rather than longitudinal data, they are unable to test whether judges actually change their voting behavior in response to shifts in pub-

calls for a vote of the general assembly to reappoint a judge. Hawaii requires reappointment by a commission. Finally, Virginia and South Carolina hold legislative elections for both initial appointments and retention.

lic opinion over time. Others include only a small number of states (Hall 1987, 1992, 1995; Brace and Hall 1997) or only states with certain selection mechanisms (Caldarone, Canes-Wrone, and Clark 2009). Finally, many studies focus exclusively on rare case types—namely abortion (Brace, Hall, and Langer 1999, 2001; Caldarone, Canes-Wrone, and Clark 2009) and capital punishment (Brace and Boyea 2008; Brace and Hall 1995, 1997; Canes-Wrone, Clark, and Kelly 2014). As an example, Brace and Boyea's (2008) analysis of capital punishment decisions uses only 889 out of the 15,000 cases in the State Supreme Court Data Project (SSCDP). More recently, Cann and Wilhelm (2011) use the SSCDP to examine judicial responsiveness across a wide range of issue areas. They find that only judges facing contestable elections respond to the public and then only in "highly visible" cases, which are just 1.3% of the cases in their data (2011, 569).

In short, there is some evidence of policy responsiveness in state supreme courts. However, the research to date often focuses on a small subset of salient cases, employs cross-sectional analyses for a brief time period, examines only a few states, and/or does not incorporate individual electoral experiences. Our study improves on previous work in several important ways. First, we test responsiveness in every state supreme court by generating aggregate measures of decision making based on every non-unanimous decision in these courts. Second, we test whether justices' voting behavior responds to, rather than simply correlates with, popular preferences by modeling the dynamic effects of public opinion over time. Finally, we test whether state supreme court justices anticipate electoral competition or change their behavior when they are faced with mandatory retirement from the court.

E.3 Research Design

Our goal is to test the effects of citizen preferences on state supreme court voting behavior. Because our ideal point estimates cover a period of 16 years, our measure of judicial voting facilitates an assessment of dynamic effects across time. This approach provides better leverage on the question of whether judges are indeed *responding* to citizen opinion rather than simply voting in a manner that is correlated with it. Indeed, a problem for many cross-sectional studies of representation is determining whether a political official's behavior is consistent with constituency

preferences because he or she is responding to their views or because he or she shares those views already and was elected to office because of that agreement. As Kuklinski and Stanga state, "[t]he use of data collected at a single point in time... precludes separating agreement resulting from simple elite-mass sharing of policy attitudes and that due to the actual response of officeholders" (1979, 1091). Our data and estimation strategy avoid this problem by modeling *changes* in judges' voting behavior over time as a function of *changes* in citizen preferences over time.

To capture these dynamic effects, we use single-equation error correction models (ECM) with judge-level random effects.[7] The dependent variable in our models is the change in the scaled ideal point for each individual judge in a given year (the SDIRT measure from the main text). Positive values of the ideal point estimates indicate a judge moving in a conservative direction, and negative values indicate a move in a liberal direction. To test the moderating role of judicial selection mechanisms, we estimate four separate ECMs: one for states with partisan elections, one for states with non-partisan elections, one for states with initial appointment and unchallenged retention elections, and one for states with lifetime appointments.

Our use of ECMs provide a strong link between our theoretical expectations and our empirical model. The ECM estimates a coefficient for both the differenced and lagged values of each predictor.[8] This allows us to examine both short-term effects, in which a change in public preferences produces an immediate change in judges' voting behavior, and long-term effects, in which the past value of public preferences influences current and future voting behavior through an equilibrium relationship (De Boef and Keele 2008). Allowing for separate estimation of short- and long-term effects accounts for the fact that the judiciary is a reactive branch of government. While a court can have substantial influence on policy, it cannot exert such influence until a relevant case is argued before it. Thus, while immediate responses to public preferences are possible, our strategy also allows for the possibility that a change in the public's preferences may take several years to

[7] Regardless of selection type, each of our proposed models contains cointegrated data, necessitating the use of the error correction approach (Persyn and Westerlund 2008).

[8] A lagged independent variable in an ECM is a re-parameterized estimate of the error term from regressing Y on X. Consequently, the ECM models the extent to which the disequilibrium between X and Y is adjusted or corrected through changes in Y.

influence judges' behavior.

Our primary independent variable is the measure of *state mood* as computed by Enns and Koch (2013). This measure relies on over 500 different surveys with more than 700,000 respondents to calculate yearly estimates of state mood through multilevel regression and poststratification (MRP, see also Lax and Phillips 2009; Park, Gelman, and Bafumi 2004). To date, this is the only measure of yearly state-level mood that relies on public opinion polls rather than proxies for the electorate's preferences or pooled multi-year estimates (e.g., Erikson, Wright, and McIver 1993; Berry, Ringquist, Fording, and Hanson 1998; Carsey and Harden 2010).[9]

We hypothesize that responsiveness to public opinion is not only a function of selection method, but also a judge's personal electoral experiences. Utilizing data from *The Judicial Elections Data Initiative* (JEDI), we include a variable indicating whether the judge is up for election in the next year. We then interact the state mood and election variables to measure how changes in public opinion exert different influences when a judge will face an election in the next year

Furthermore, we include an indicator variable if a judge is in the year prior to mandatory retirement. We collected the data of birth of each judge over this 16 year time period, as well as the mandatory retirement age of each state if applicable. We then interact the state mood and mandatory retirement variable to examine if judges respond different to public mood when they are forced to leave the bench.

Finally, we include fixed effects to account for state-level variation; that is, we include dichotomous indicator variables for each state taking on the value of 1 for observations within that state and 0 otherwise. These fixed effects set different intercepts for each state in order to account for the possibility that decision making in certain states tends to be more or less liberal than decision making in other states. Thus, our models test the effects of our predictor variables *within* states. This is an important choice because there are many differences between states that might affect the decisions presented to judges, and thus the estimated ideal points. For example, some state

[9]We reversed the original coding of this variable to match the ideological direction of our judge-level indicator. This variable now ranges from 38.194 to 74.35, with lower values indicating a more liberal state and higher values representing a more conservative state. The mean state mood is 62.02 with a standard deviation of 5.56.

courts have discretionary dockets while others do not. Our modeling strategy accounts for these differences because all between-state variance is explained by the fixed effects.[10]

E.4 Analysis

We present the coefficient estimates and standard errors from our four models in Table A-5. The columns report the results from states with partisan elections, non-partisan elections, retention elections, and judicial appointment, respectively. Overall, the results in Table A-5 comport with the expectations that retiring judges will behave differently, but we do not find evidence indicating that judges are responsive to public preferences.

[Insert Table A-5 here]

More specifically, we do not see a change in judicial behavior due to an upcoming election. In each of our models, the interaction between public mood and an election year is insignificant. In partisan election states, the coefficient is in the right direction, however this relationship fails to reach traditionally accepted levels of significance.

Mandatory retirement, however, negatively influences policy responsiveness in both partisan election and appointment states, but not in non-partisan or merit selection states. As evident in Table A-5, Columns 1 and 4, judges who face mandatory retirement change their voting behavior in the opposite direction of public mood. This relationship reaches traditional levels of statistical significance.

We plot the marginal effects of this interaction in Figure A-15 and A-16. In both of these figures, judges in non-retirement years see no significant movement in their judicial behavior, as evident by the relatively flat line. Judges in retirement years, however, in both graphics exhibit a sharp countermajoritarian behavior when faced with retirement. This lends evidence to retirement lifting the potential pull of the public on judges behavior.

[Insert Figure A-15 here]

[10] We estimated alternative models controlling for term length and the number of seats on the court. Both of these variables are statistically insignificant in each of our models. Furthermore, the overall model fit declines in the over-specified models. We therefore only report results from the analyses outlined above.

[Insert Figure A-16 here]

Overall, this substantive application highlights a key benefit of our new SDIRT measure of judicial ideal points: its dynamic nature. A common problem in studies on representation is that static data cannot provide evidence for or against responsiveness. For example, previous work shows that state public opinion is positively correlated with judges' voting behavior in elected states but not non-elected states (e.g., Brace and Boyea 2008). However, these cross-sectional associations do not provide evidence that judges' voting behavior actually moves with, or responds to, changes in public opinion. In this application we examined actual responsiveness to public opinion through an ECM. The result is a closer connection between the theoretical framework of most representation studies and the empirical test. In this case, we find no evidence that judges are responsive across several different types of selection mechanism. In fact, we find that in some circumstances—namely mandatory retirement—judges' voting behavior actually moves in a countermajoritarian direction.

References

American Judicature Society. 2013. "Methods of Judicial Selection." http://www.judicialselection.us/judicial_selection/methods/selection_of_judges.cfm?state=. Accessed January, 8, 2014.

Ansolabehere, Stephen, Davis Brady, and Morris P. Fiorina. 1992. "The Vanishing Marginals and Electoral Responsiveness." *British Journal of Political Science* 22(1): 21–38.

Bartels, Larry M. 1991. "Constituency Opinion and Congressional Policy Making: The Reagan Defense Buildup." *American Political Science Review* 85(2): 457–474.

Berry, William D., Evan Ringquist, Richard C. Fording, and Russell L. Hanson. 1998. "Measuring Citizen and Government Ideology in the American States, 1960–93." *American Journal of Political Science* 41(1): 327–348.

Besley, Timothy, and Stephen Coate. 2003. "Elected Versus Appointed Regulators: Theory and Evidence." *Journal of the European Economic Association* 1(5): 1176–1206.

Bonica, Adam, and Michael J. Woodruff. 2015. "A Common-Space Measure of State Supreme Court Ideology." Forthcoming, *Journal of Law, Economics, & Organization*.

Brace, Paul, and Brent D. Boyea. 2008. "State Public Opinion, the Death Penalty, and the Practice of Electing Judges." *American Journal of Political Science* 52(2): 360–372.

Brace, Paul, and Melinda Gann Hall. 1990. "Neo-Institutionalism and Dissent in State Supreme Courts." *Journal of Politics* 52(1): 54–70.

Brace, Paul, and Melinda Gann Hall. 1993. "Integrated Models of Judicial Dissent." *Journal of Politics* 55(4): 914–935.

Brace, Paul, and Melinda Gann Hall. 1995. "Studying Courts Comparatively: The View from the American States." *Political Research Quarterly* 48(1): 5–29.

Brace, Paul, and Melinda Gann Hall. 1997. "The Interplay of Preferences, Case Facts, Context, and Rules in the Politics of Judicial Choice." *Journal of Politics* 59(4): 1206–1231.

Brace, Paul, Melinda Gann Hall, and Laura Langer. 1999. "Judicial Choice and the Politics of Abortion: Institutions, Context, and the Autonomy of Courts." *Albany Law Review* 62(4): 1265–1303.

Brace, Paul, Melinda Gann Hall, and Laura Langer. 2001. "Placing State Supreme Courts in State Politics." *State Politics & Policy Quarterly* 1(1): 81–108.

Brehm, John, and Scott Gates, eds. 1997. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor, MI: University of Michigan Press.

Caldarone, Richard P., Brandice Canes-Wrone, and Tom S. Clark. 2009. "Partisan Labels and Democratic Accountability: An Analysis of State Supreme Court Abortion Decisions." *Journal of Politics* 71(3): 560–573.

Canes-Wrone, Brandice, and Kenneth W. Shotts. 2004. "The Conditional Nature of Presidential Responsiveness to Public Opinion." *American Journal of Political Science* 48(4): 690–706.

Canes-Wrone, Brandice, Tom S. Clark, and Jason P. Kelly. 2014. "Judicial Selection and Death Penalty Decisions." *American Political Science Review* 108(1): 23–39.

Cann, Damon M., and Teena Wilhelm. 2011. "Case Visibility and the Electoral Connection in State Supreme Courts." *American Politics Research* 39(3): 557–581.

Carpenter, Daniel P. 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval." *American Journal of Political Science* 46(3): 490–505.

Carsey, Thomas M., and Jeffrey J. Harden. 2010. "New Measures of Partisanship, Ideology, and Policy Mood in the American States." *State Politics & Policy Quarterly* 10(2): 136–156.

Casillas, Christopher, Peter K. Enns, and Patrick C. Wohlfarth. 2011. "How Public Opinion Constrains the U.S. Supreme Court." *American Journal of Political Science* 55(1): 74–88.

Clark, Jennifer H., and R. Lucas Williams. 2013. "Parties, Term Limits, and Representation in the U.S. States." *American Politics Research* 42(1).

Collins, Paul M., Jr. 2007. "Lobbyists Before the U.S. Supreme Court: Investigating the Influence of Amicus Curiae Briefs." *Political Research Quarterly* 60(1): 55–70.

Collins, Paul M., Jr., Daniel A. Norton, Kenneth L. Manning, and Robert A. Carp. 2008. "Unified Government, Divided Government, and Party Responsiveness." *Justice System Journal* 29(2): 121–144.

De Boef, Suzanne, and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52(1): 184–200.

Enns, Peter K., and Julianna Koch. 2013. "Public Opinion in the U.S. States: 1956 to 2010." *State Politics & Policy Quarterly* 13(3): 349–372.

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

Erikson, Robert S., Gerald C. Wright, and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States*. New York: Cambridge University Press.

Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2002. *The Macro Polity*. New York: Cambridge University Press.

Fenno, Richard F. 1996. *Senators on the Campaign Trail: The Politics of Representation*. Norman, OK: University of Oklahoma Press.

Ferejohn, John, and Charles Shipan. 1990. "Congressional Influence on Bureaucracy." *Journal of Law, Economics, & Organization* 6: 1–20.

Fiorina, Morris P. 1974. *Representatives, Roll Call, and Constituencies*. Lexington, MA: Lexington Books.

Flemming, Roy B., and B. Dan Wood. 1997. "The Public and the Supreme Court: Individual Justice Responsiveness to American Policy Moods." *American Journal of Political Science* 41(2): 468–498.

Franklin, Charles H., and Liane C. Kosaki. 1989. "Republican Schoolmaster: The U.S. Supreme Court, Public Opinion and Abortion." *American Political Science Review* 83(3): 751–771.

Froman, Lewis A. 1963. *Congressmen and their Constituencies*. Chicago: McNally.

Gely, Rafael, and Pablo T. Spiller. 1990. "A Rational Choice Theory of Supreme Court Statutory Decisions with Applications to the *State Farm* and *Grove City* Cases." *Journal of Law, Economics, & Organization* 6(2): 263–300.

Giles, Micheal W., Bethany Blackstone, and Richard L. Vining. 2008. "The Supreme Court in American Democracy: Unraveling the Linkages between Public Opinion and Judicial Decision Making." *Journal of Politics* 70(2): 293–306.

Gordon, Sanford C., and Gregory A. Huber. 2002. "Citizen Oversight and the Electoral Incentives of Criminal Prosecutors." *American Journal of Political Science* 46(2): 334–351.

Gordon, Sanford C., and Gregory A. Huber. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2(2): 107–138.

Griffin, John D. 2006. "Electoral Competition and Democratic Responsiveness: A Defense of the Marginality Hypothesis." *Journal of Politics* 68(4): 909–919.

Gulati, Girish J. 2004. "Revisiting the Link Between Electoral Competition and Policy Extremism in the U.S. Congress." *American Politics Research* 32(5): 495–520.

Hakhverdian, Armen. 2010. "Political Representation and its Mechanisms: A Dynamic Left-Right Approach for the United Kingdom, 1976–2006." *British Journal of Political Science* 40(4): 835–856.

Hall, Matthew E.K. 2014. "The Semi-Constraind Court: Public Opinion, the Separation of Pow-

ers, and the U.S. Supreme Court's Fear of Nonimplementation." *American Journal of Political Science* 58(2): 352–366.

Hall, Melinda Gann. 1987. "Constituent Influence in State Supreme Courts: Conceptual Notes and a Case Study." *Journal of Politics* 49(4): 1117–1124.

Hall, Melinda Gann. 1992. "Electoral Politics and Strategic Voting in State Supreme Courts." *Journal of Politics* 54(2): 427–446.

Hall, Melinda Gann. 1995. "Justices as Representatives: Elections and Judicial Politics in the American States." *American Politics Quarterly* 23(4): 485–503.

Hall, Melinda Gann, and Paul Brace. 1989. "Order in the Courts: A Neo-Institutional Approach to Judicial Consensus." *Western Political Quarterly* 42(3): 391–407.

Hobolt, Sara Binzer, and Robert Klemmemsen. 2007. "Responsive Government? Public Opinion and Government Policy Preferences in Britain and Denmark." *Comparative Political Studies* 41(3): 309–337.

Huber, Gregory A., and Sanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind Whn it Runs for Office?" *American Journal of Political Science* 48(2): 247–263.

Jacobson, Gary C. 1987. *The Politics of Congressional Elections*. Boston: Little, Brown.

Kernell, Samuel. 1997. *Going Public: New Strategies of Presidential Leadership*. 2nd ed. Washington, DC: Congressional Quarterly Press.

Kousser, Thad, Jeffrey B. Lewis, and Seth E. Masket. 2007. "Ideological Adaptation? The Survival Instinct of Threatened Legislators." *Journal of Politics* 69(3): 828–843.

Kuklinski, James H., and John E. Stanga. 1979. "Political Participation and Government Responsiveness: The Behavior of California Superior Courts." *American Political Science Review* 73(4): 1090–1099.

Langer, Laura. 2002. *Judicial Review in State Supreme Courts: A Comparative Study*. Albany, NY: State University of New York Press.

Lax, Jeffrey R., and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107–121.

Mathias, Sara. 1990. *Electing Justice: A Handbook of Judicial Election Reforms*. Chicago: American Judicature Society.

Mayhew, David R. 1974. *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

McGuire, Kevin T., and James A. Stimson. 2004. "The Least Dangerous Branch Revisited: New Evidence on Supreme Court Responsiveness to Public Preferences." *Journal of Politics* 66(4): 1018–1035.

Miller, Warren E., and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1): 45–56.

Mishler, William, and Reginald S. Sheehan. 1993. "The Supreme Court as a Countermajoritarian

Institution? The Impact of Public Opinion on Supreme Court Decisions." *American Political Science Review* 87(1): 87–101.

Norpoth, Helmut, and Jeffrey A. Segal. 1994. "Popular Influence on Supreme Court Decisions." *American Political Science Review* 88(3): 711–24.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4): 375–385.

Persyn, Damiaan, and Joakim Westerlund. 2008. "Error-Correction–Based Cointegration Tests for Panel Data." *The Stata Journal* 8(2): 232–241.

Rabinovich, Julia. 2008. "Administrative Responsiveness to Public Opinion." Ph.D. Dissertation, Northwestern University.

Roberts, Andrew, and Byung-Yeon Kim. 2011. "Policy Responsiveness in Postcommunist Europe: Public Preferences and Economic Reforms." *British Journal of Political Science* 41(4): 819–839.

Rothenberg, Lawrence S., and Mitchell S. Sanders. 2000. "Severing the Electoral Connection: Shirking in the Contemporary Congress." *American Journal of Political Science* 44(2): 316–325.

Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. New York: Cambridge University Press.

Serfass, Melissa M., and Jessie L. Cranford. 2001. "Federal and State Court Rules Governing Publication and Citation of Opinions." *Journal of Appellate Practice and Process* 3(1): 251–285.

Sigelman, Lee. 1980. "Gauging the Public Response to Presidential Leadership." *Presidential Studies Quarterly* 10(3): 427–433.

Stimson, James A., Michael B. MacKuen, and Robert S. Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89(3): 543–565.

Sullivan, John L., and Eric M. Uslaner. 1978. "Congressional Behavior and Electoral Marginality." *American Journal of Political Science* 22(3): 536–563.

Tulis, Jeffrey K. 1988. *The Rhetorical Presidency*. Princeton, NJ: Princeton University Press.

Weingast, Barry R. 1984. "The Congressional-Bureaucratic System: A Principal Agent Perspective (with applications to the SEC)." *Public Choice* 44(1): 147–191.

Wright, Gerald C. 2007. "Do Term Limits Affect Legislative Roll Call Voting? Representation, Polarization, and Participation." *State Politics & Policy Quarterly* 7(3): 256–280.

Table A-1: Reliability and Validity Check, 1995-1998

Variable	Total Possible Matches	Matches with Brace and Hall (%)	Matches with Dockets (%)
Case Title	500	483 (96.6)%	499 (99.8%)
Citation	500	426 (84.6%)	494 (98.8%)
Opinion Author	500	474 (94.8%)	496 (99.2%)
Dissent Author	500	471 (94.2%)	494 (98.8%)
Judge Votes	3,075	2,888 (93.9%)	3,029 (98.5%)

Table A-2: Cross-Validated Vote Prediction Model Fit Statistics with Unscaled Dynamic IRT

State Court	Proportion Correctly Classified				Aggregate Proportional Reduction in Error				Court Cases (Testing Set)
	IRT	PAJID	CFscores	Best Fit	IRT	PAJID	CFscores	Best Fit	
AK	0.804	0.805	0.843	CFscores	0.407	0.407	0.694	CFscores	54
AL	0.909	0.884	0.888	IRT	0.575	0.448	0.458	IRT	297
AR	0.830	0.812	0.801	IRT	0.437	0.391	0.355	IRT	154
AZ	0.902	0.838	0.878	IRT	0.600	0.333	0.500	IRT	25
CA	0.876	0.812	0.808	IRT	0.547	0.348	0.328	IRT	114
CO	0.898	0.835	0.856	IRT	0.579	0.341	0.429	IRT	105
CT	0.899	0.801	0.812	IRT	0.735	0.492	0.517	IRT	78
DE	0.950	0.883	0.833	IRT	0.833	0.667	0.250	IRT	5
FL	0.932	0.780	0.814	IRT	0.680	0.140	0.214	IRT	255
GA	0.810	0.767	0.795	IRT	0.335	0.185	0.282	IRT	388
HI	0.928	0.902	0.950	CFscores	0.618	0.441	0.735	CFscores	52
IA	0.754	0.718	0.755	CFscores	0.253	0.138	0.230	IRT	40
ID	0.899	0.833	0.864	IRT	0.590	0.333	0.462	IRT	34
IL	0.838	0.795	0.817	IRT	0.420	0.327	0.444	CFscores	84
IN	0.871	0.809	0.908	CFscores	0.643	0.480	0.751	CFscores	108
KS	0.890	0.794	0.809	IRT	0.527	0.182	0.255	IRT	35
KY	0.828	0.776	0.922	CFscores	0.505	0.368	0.737	CFscores	133
LA	0.888	0.840	0.845	IRT	0.529	0.308	0.346	IRT	146
MA	0.925	0.843	0.911	IRT	0.567	0.200	0.533	IRT	33
MD	0.865	0.822	0.804	IRT	0.551	0.386	0.400	IRT	97
ME	0.847	0.802	0.789	IRT	0.453	0.314	0.314	IRT	51
MI	0.910	0.795	0.925	CFscores	0.772	0.470	0.776	CFscores	269
MN	0.846	0.811	0.824	IRT	0.426	0.344	0.361	IRT	79
MO	0.851	0.860	0.833	PAJID	0.464	0.524	0.405	PAJID	49
MS	0.909	0.810	0.860	IRT	0.604	0.215	0.371	IRT	268
MT	0.909	0.828	0.869	IRT	0.655	0.300	0.489	IRT	172
NC	0.828	0.770	0.776	IRT	0.310	0.207	0.207	IRT	18
ND	0.867	0.813	0.816	IRT	0.632	0.453	0.462	IRT	74
NE	0.827	0.792	0.769	IRT	0.488	0.388	0.257	IRT	37
NH	0.938	0.847	0.835	IRT	0.773	0.455	0.409	IRT	17
NJ	0.918	0.828	0.814	IRT	0.564	0.164	0.109	IRT	46
NM	0.877	0.759	0.841	IRT	0.571	0.143	0.429	IRT	11
NV	0.868	0.851	0.796	IRT	0.630	0.568	0.420	IRT	54
NY	0.874	0.808	0.850	IRT	0.425	0.151	0.288	IRT	56
OH	0.852	0.805	0.836	IRT	0.476	0.285	0.410	IRT	212
OK (Crime)	0.947	0.928	0.836	IRT	0.807	0.737	0.509	IRT	43
OK (SC)	0.826	0.772	0.806	IRT	0.332	0.120	0.298	IRT	190
OR	0.845	0.798	0.774	IRT	0.455	0.303	0.212	IRT	21
PA	0.807	0.804	0.805	IRT	0.369	0.347	0.341	IRT	210
RI	0.964	0.741	0.782	IRT	0.875	0.250	0.313	IRT	11
SC	0.917	0.856	—	IRT	0.667	0.427	—	IRT	72
SD	0.926	0.803	0.789	IRT	0.741	0.353	0.360	IRT	103
TN	0.891	0.815	0.829	IRT	0.571	0.343	0.371	IRT	32
TX (Crime)	0.919	0.854	0.878	IRT	0.724	0.412	0.604	IRT	189
TX (SC)	0.876	0.784	0.796	IRT	0.575	0.327	0.345	IRT	58
UT	0.807	0.794	0.752	IRT	0.420	0.360	0.240	IRT	33
VA	0.857	0.811	0.870	CFscores	0.481	0.342	0.577	CFscores	48
VT	0.886	0.824	0.789	IRT	0.565	0.362	0.232	IRT	58
WA	0.854	0.766	0.822	IRT	0.545	0.294	0.422	IRT	188
WI	0.920	0.742	0.909	IRT	0.791	0.391	0.765	IRT	109
WV	0.955	0.815	0.870	IRT	0.864	0.460	0.621	IRT	153
WY	0.890	0.767	0.883	IRT	0.588	0.235	0.559	IRT	26
Means	0.879	0.812	0.834		0.568	0.345	0.420		100

Note: Cell entries report cross-validated model fit statistics for our unscaled dynamic IRT ideal point measure, PAJID, and CFscores. The first four columns give the proportion of votes correctly classified by the model with each measure and the best fitting-measure. The next four columns give the aggregate proportional reduction in error by the model with each measure and the best-fitting measure. The final column gives the number of court cases used in model fitting for each state court. South Carolina is omitted from the CFscores computations due to sparse data.

Table A-3: In-Sample Vote Prediction Model Fit Statistics with Unscaled Dynamic IRT

	Proportion Correctly Classified				Aggregate Proportional Reduction in Error				
State Court	IRT	PAJID	CFscores	Best Fit	IRT	PAJID	CFscores	Best Fit	Court Cases
AK	0.811	0.790	0.862	CFscores	0.428	0.364	0.676	CFscores	275
AL	0.930	0.902	0.907	IRT	0.616	0.439	0.476	IRT	1,743
AR	0.852	0.817	0.817	IRT	0.503	0.376	0.386	IRT	805
AZ	0.911	0.836	0.889	IRT	0.653	0.359	0.565	IRT	140
CA	0.879	0.803	0.793	IRT	0.574	0.334	0.323	IRT	569
CO	0.909	0.835	0.842	IRT	0.655	0.389	0.419	IRT	526
CT	0.938	0.827	0.826	IRT	0.776	0.471	0.464	IRT	444
DE	0.941	0.928	0.927	IRT	0.759	0.704	0.714	IRT	50
FL	0.939	0.791	0.817	IRT	0.720	0.169	0.221	IRT	1,309
GA	0.819	0.768	0.791	IRT	0.361	0.186	0.263	IRT	1,940
HI	0.950	0.920	0.947	IRT	0.726	0.543	0.644	IRT	275
IA	0.827	0.782	0.757	IRT	0.391	0.243	0.213	IRT	220
ID	0.944	0.889	0.904	IRT	0.758	0.534	0.596	IRT	218
IL	0.905	0.858	0.884	IRT	0.520	0.345	0.439	IRT	614
IN	0.865	0.822	0.896	CFscores	0.607	0.479	0.693	CFscores	541
KS	0.887	0.847	0.858	IRT	0.496	0.308	0.365	IRT	191
KY	0.849	0.798	0.932	CFscores	0.513	0.353	0.723	CFscores	706
LA	0.882	0.831	0.838	IRT	0.517	0.308	0.350	IRT	729
MA	0.924	0.822	0.909	IRT	0.624	0.218	0.515	IRT	197
MD	0.858	0.819	0.802	IRT	0.507	0.362	0.353	IRT	484
ME	0.844	0.775	0.774	IRT	0.467	0.274	0.337	IRT	269
MI	0.932	0.796	0.915	IRT	0.816	0.479	0.753	IRT	1,346
MN	0.879	0.836	0.861	IRT	0.495	0.347	0.431	IRT	434
MO	0.891	0.876	0.879	IRT	0.655	0.606	0.602	IRT	272
MS	0.941	0.869	0.899	IRT	0.642	0.257	0.382	IRT	1,940
MT	0.907	0.840	0.877	IRT	0.635	0.337	0.499	IRT	1,018
NC	0.921	0.869	0.901	IRT	0.468	0.205	0.314	IRT	177
ND	0.867	0.820	0.809	IRT	0.620	0.480	0.450	IRT	371
NE	0.843	0.800	0.802	IRT	0.528	0.400	0.335	IRT	183
NH	0.955	0.905	0.913	IRT	0.773	0.532	0.571	IRT	174
NJ	0.926	0.896	0.869	IRT	0.524	0.323	0.202	IRT	363
NM	0.905	0.863	0.893	IRT	0.647	0.491	0.603	IRT	95
NV	0.907	0.882	0.827	IRT	0.707	0.627	0.461	IRT	353
NY	0.927	0.866	0.915	IRT	0.542	0.238	0.463	IRT	394
OH	0.798	0.804	0.835	CFscores	0.310	0.301	0.436	CFscores	1,058
OK (Crime)	0.947	0.935	0.838	IRT	0.796	0.756	0.505	IRT	217
OK (SC)	0.839	0.801	0.815	IRT	0.364	0.206	0.282	IRT	949
OR	0.859	0.844	0.825	IRT	0.346	0.314	0.212	IRT	132
PA	0.858	0.825	0.815	IRT	0.505	0.383	0.336	IRT	1,150
RI	0.953	0.802	0.868	IRT	0.770	0.204	0.363	IRT	118
SC	0.918	0.867	—	IRT	0.674	0.475	—	IRT	358
SD	0.915	0.805	0.798	IRT	0.727	0.395	0.406	IRT	515
TN	0.881	0.847	0.859	IRT	0.511	0.391	0.420	IRT	176
TX (Crime)	0.911	0.864	0.871	IRT	0.676	0.408	0.587	IRT	946
TX (SC)	0.907	0.858	0.872	IRT	0.528	0.312	0.366	IRT	432
UT	0.845	0.834	0.820	IRT	0.557	0.516	0.475	IRT	209
VA	0.867	0.757	0.899	CFscores	0.593	0.311	0.665	CFscores	239
VT	0.830	0.821	0.768	IRT	0.441	0.394	0.227	IRT	291
WA	0.839	0.764	0.806	IRT	0.538	0.312	0.406	IRT	942
WI	0.924	0.753	0.883	IRT	0.770	0.292	0.636	IRT	543
WV	0.945	0.820	0.873	IRT	0.834	0.475	0.632	IRT	766
WY	0.923	0.855	0.896	IRT	0.723	0.483	0.609	IRT	201
Means	0.893	0.835	0.856		0.594	0.385	0.458		550

Note: Cell entries report in-sample model fit statistics for our unscaled dynamic IRT ideal point measure, PAJID, and CFscores. The first four columns give the proportion of votes correctly classified by the model with each measure and the best fitting-measure. The next four columns give the aggregate proportional reduction in error by the model with each measure and the best-fitting measure. The final column gives the number of court cases used in model fitting for each state court. South Carolina is omitted from the CFscores computations due to sparse data.

Table A-4: Correlation of Ideal Point Scores

State Court	IRT and PAJID	IRT and CFscores
AK	−0.2596	0.5114
AL	−0.4669	0.7022
AR	−0.2761	0.2571
AZ	0.1521	0.7522
CA	0.3227	−0.1989
CO	0.3178	0.6285
CT	−0.7523	0.5888
DE	−0.9114	0.3401
FL	−0.082	0.2742
GA	−0.2665	0.4192
HI	0.6132	0.1235
IA	−0.0206	0.6162
ID	−0.4859	0.6402
IL	−0.3436	0.6331
IN	−0.4028	0.7802
KS	−0.3834	0.5075
KY	−0.1415	0.4034
LA	−0.5945	0.7699
MA	0.3699	0.725
MD	−0.1822	0.0322
ME	−0.552	0.567
MI	−0.2228	0.8448
MN	−0.7183	0.2649
MO	−0.8521	0.7899
MS	−0.5795	0.6444
MT	−0.0707	0.7106
NC	−0.2637	0.7981
ND	−0.6345	0.5452
NE	−0.2617	0.0432
NH	0.4387	0.2628
NJ	−0.2668	−0.0717
NM	0.3572	0.4674
NV	−0.5319	0.1607
NY	−0.2517	0.316
OH	−0.3572	0.5272
OK (Crime)	−0.8611	0.8454
OK (SC)	−0.2274	0.5428
OR	−0.1193	0.364
PA	−0.2971	0.4711
RI	−0.0653	0.2453
SC	0.3349	—
SD	0.1415	−0.0117
TN	−0.3404	−0.4653
TX (Crime)	−0.3431	0.6566
TX (SC)	0.3352	−0.082
UT	0.0396	0.454
VA	−0.2272	0.849
VT	0.0844	0.679
WA	−0.0578	0.5013
WI	−0.4143	0.8752
WV	−0.2793	0.7573
WY	−0.2833	0.5325
Total	−0.2492	0.4494

Note: Cell entries report Pearson's correlation coefficients for the dynamic IRT ideal point measure and both of the other ideal point measures discussed in the main text.

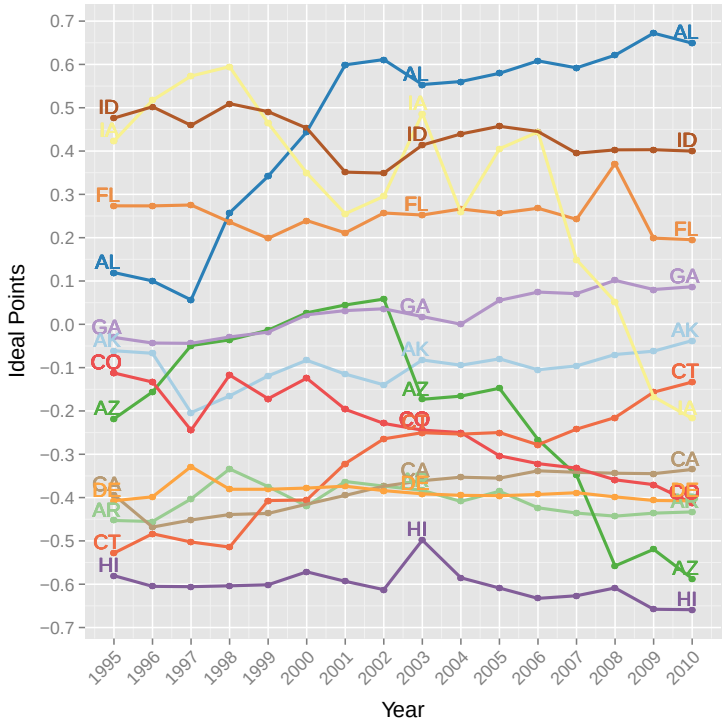
Table A-5: Dynamic Representation on State Supreme Courts

	Partisan Elections	Non-Partisan Elections	Merit Selection	Appointment
Ideal Point$_{t-1}$	0.0125*	0.0115*	0.0272*	0.0177*
	(0.0051)	(0.0042)	(0.0032)	(0.0047)
Δ Mood	-0.0005	-0.0001	-0.0001	-0.0003
	(0.0007)	(0.0005)	(0.0004)	(0.0004)
Mood$_{t-1}$	-0.0001	-0.0001	0.0002	-0.0005†
	(0.0005)	(0.0004)	(0.0003)	(0.0003)
Year Before Election	-0.0583	0.0348	0.0263	
	(0.0603)	(0.0499)	(0.0357)	
Year Before Election $\times$ Mood$_{t-1}$	0.0009	-0.0005	-0.0004	
	(0.0009)	(0.0008)	(0.0006)	
Year Before Mandatory Retirement=1	0.2611*	-0.0706	0.0085	0.1393*
	(0.1202)	(0.1355)	(0.0879)	(0.0656)
Year Before Mandatory Retirement $\times$ Mood$_{t-1}$	-0.0043*	0.0013	-0.0001	-0.0026*
	(0.0019)	(0.0022)	(0.0014)	(0.0011)
R^2	.1796	.1812	.4237	.2749
N	783	1484	1518	1040

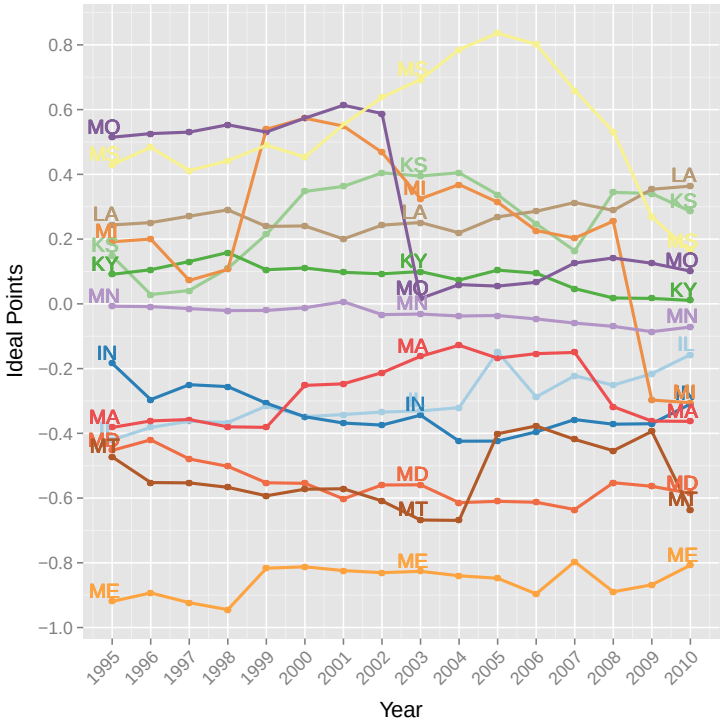
Estimated coefficients from Error Correction Models are reported along with standard errors in parentheses. * = $p \leq 0.05$ (two tailed) † = $p \leq 0.1$ (two tailed).

Figure A-1: State Court Medians of the Dynamic Common Space Ideal Points

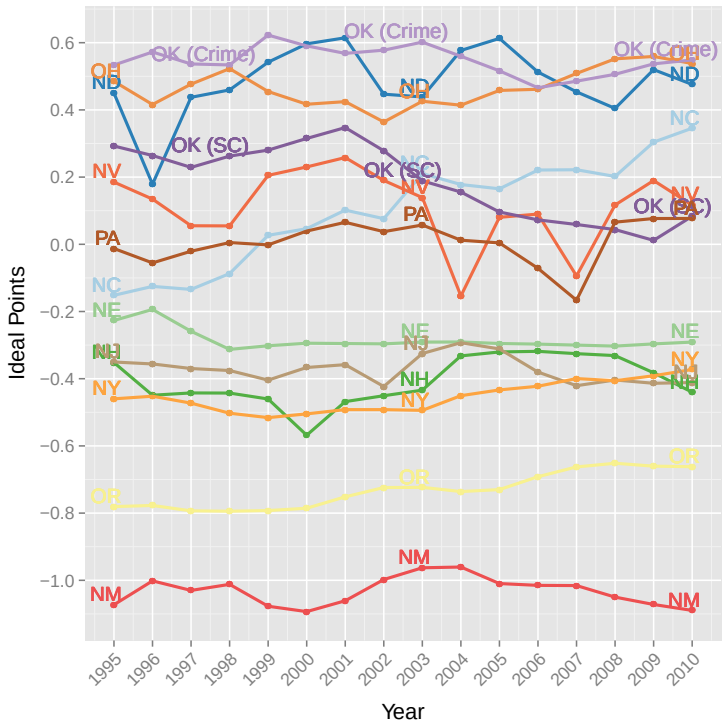
(a) Alaska–Idaho



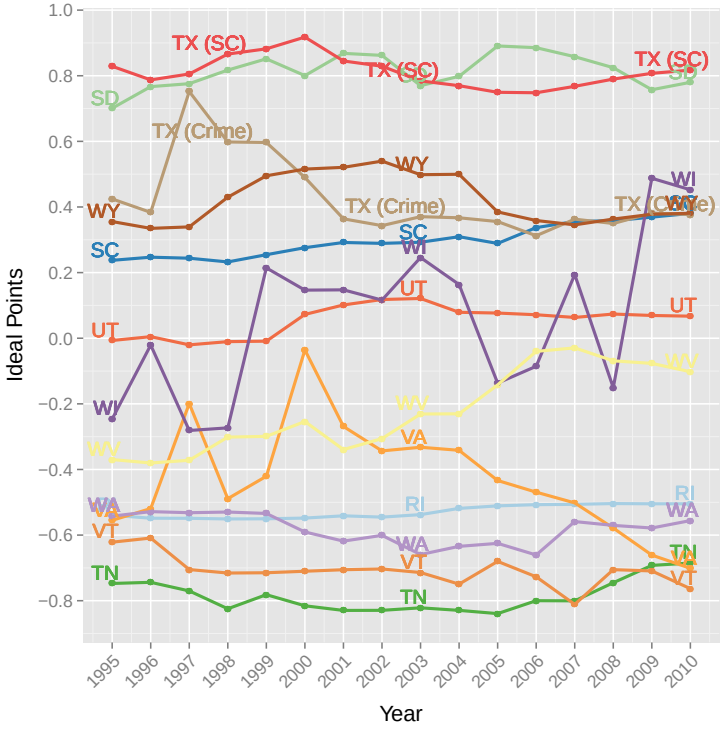
(b) Illinois–Montana



(c) North Carolina–Pennsylvania



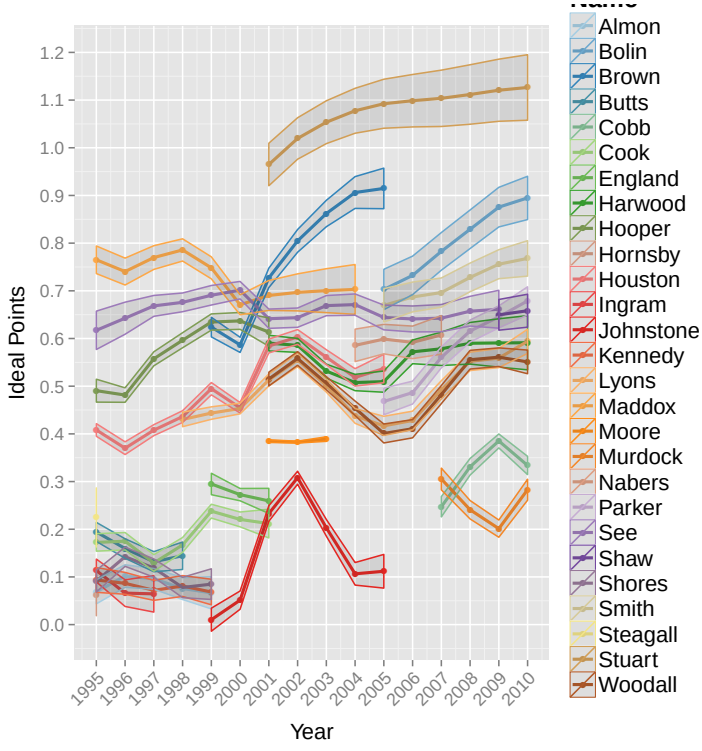
(d) Rhode Island–Wyoming



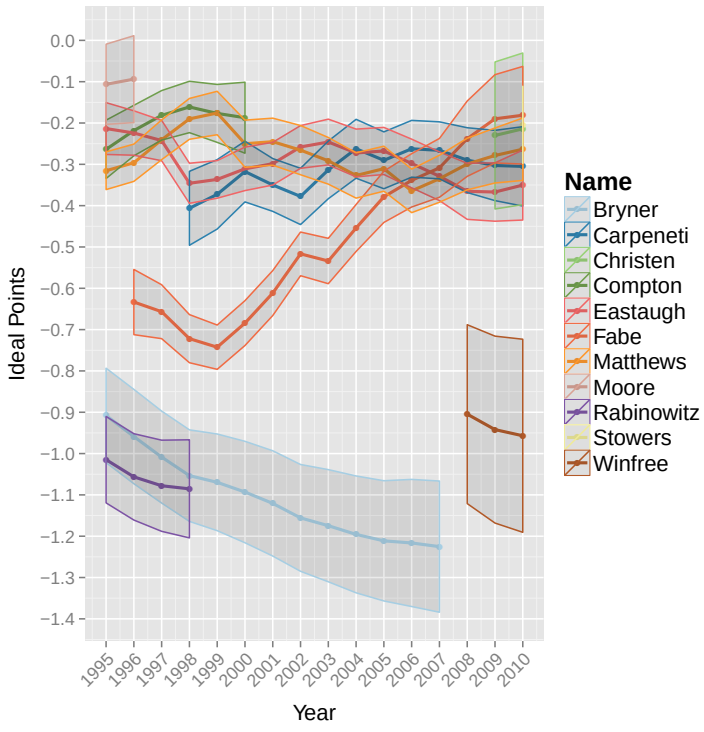
Note: The graphs plot each court's median over the 16 year period in the dynamic common space mapped from the CFscores (Bonica and Woodruff 2015).

Figure A-2: Scaled Dynamic IRT Ideal Point Posterior Means and Standard Deviations

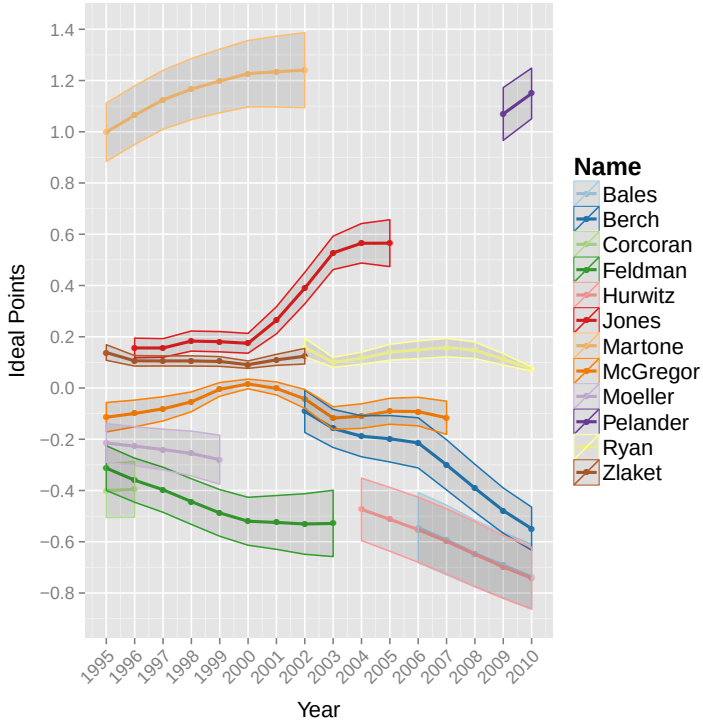
(a) Alabama



(b) Alaska



(c) Arizona



(d) Arkansas

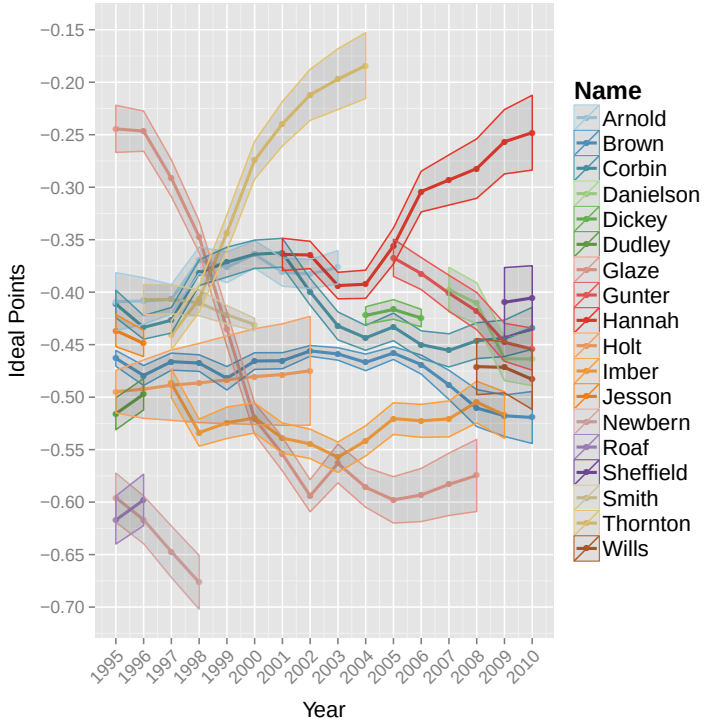
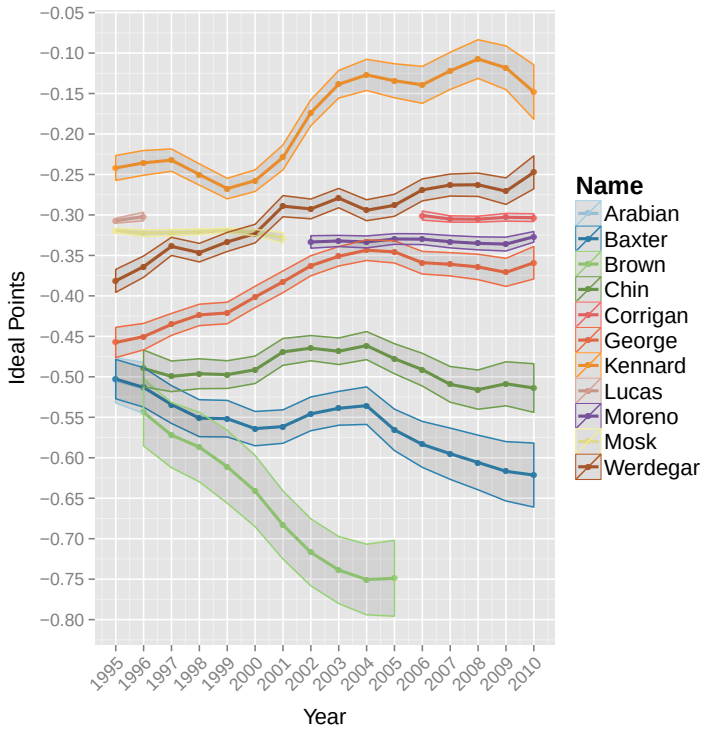
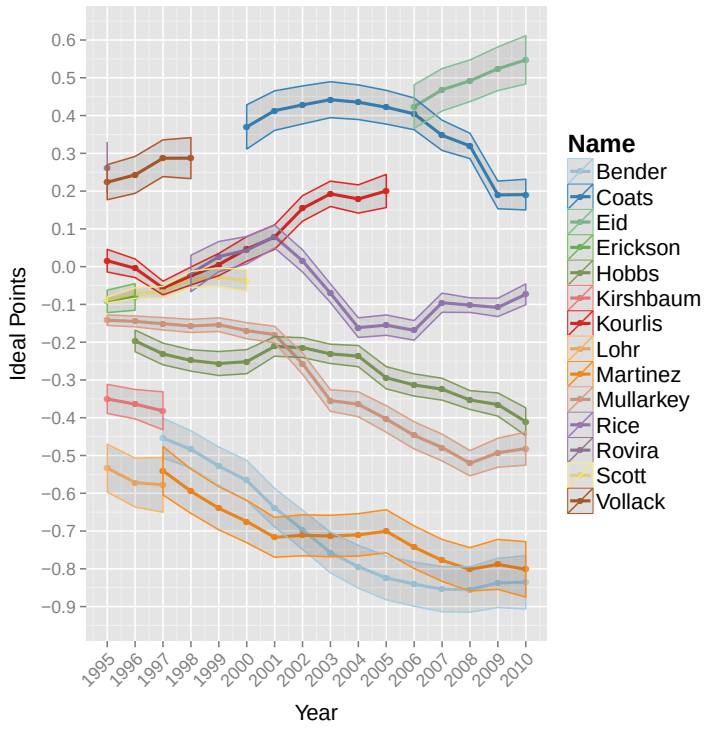


Figure A-3: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

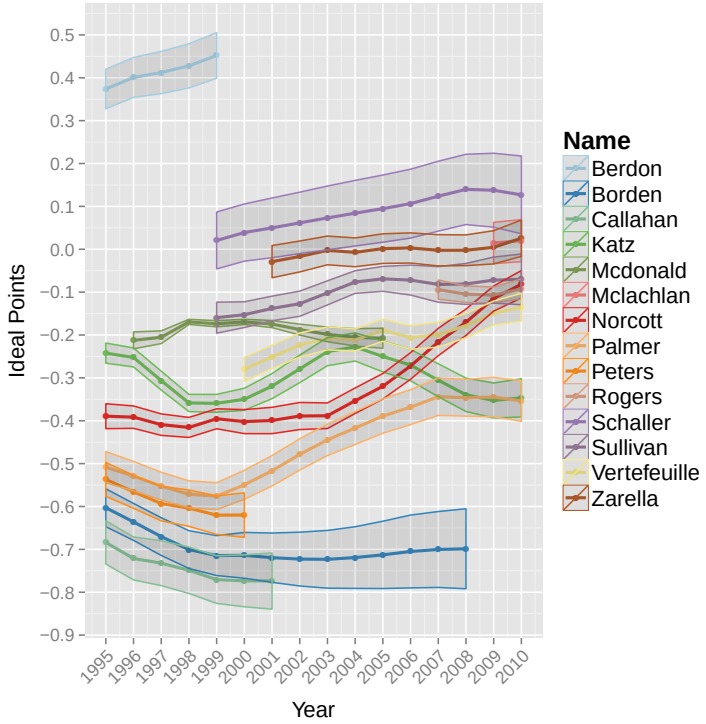
(a) California



(b) Colorado



(c) Connecticut



(d) Delaware

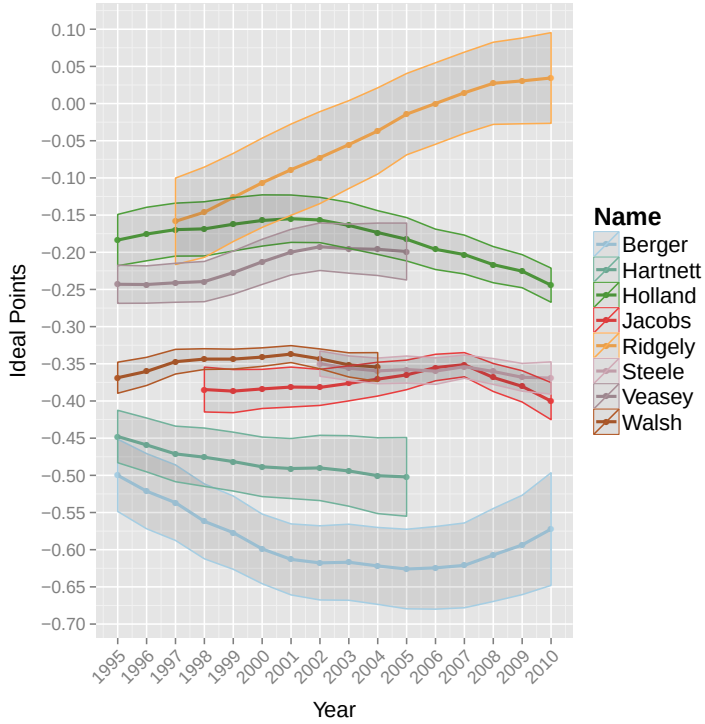
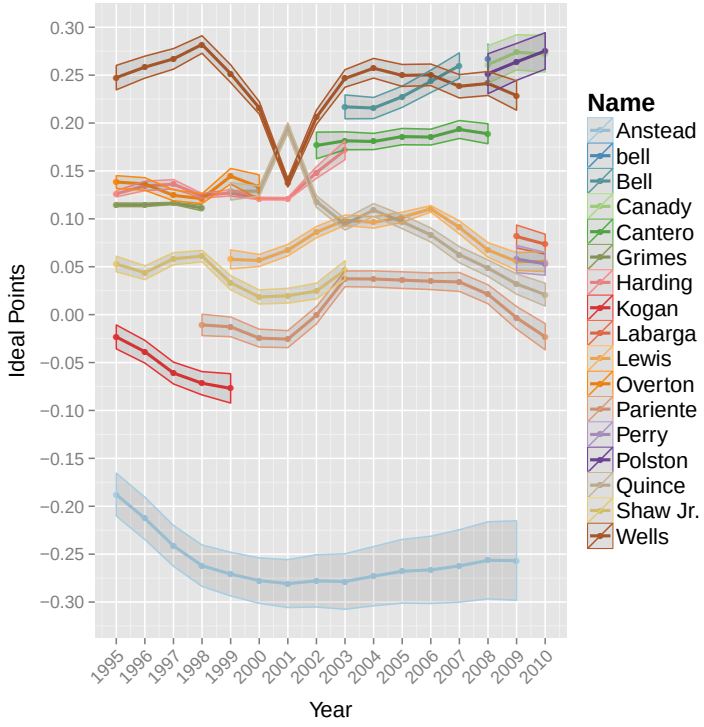
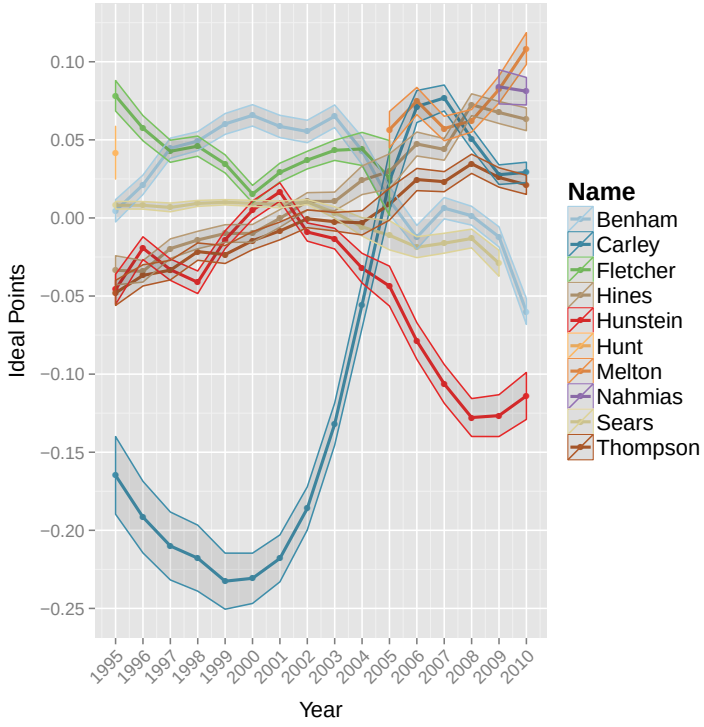


Figure A-4: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

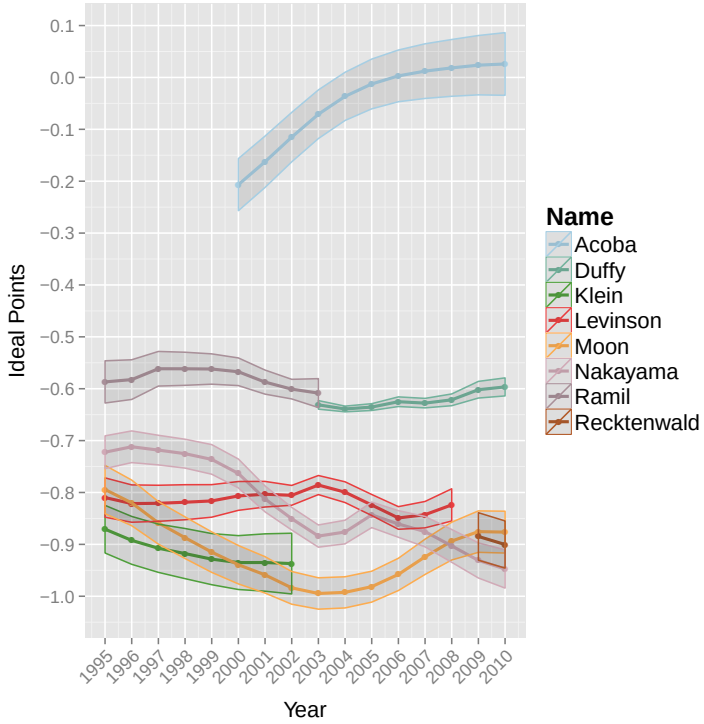
(a) Florida



(b) Georgia



(c) Hawaii



(d) Idaho

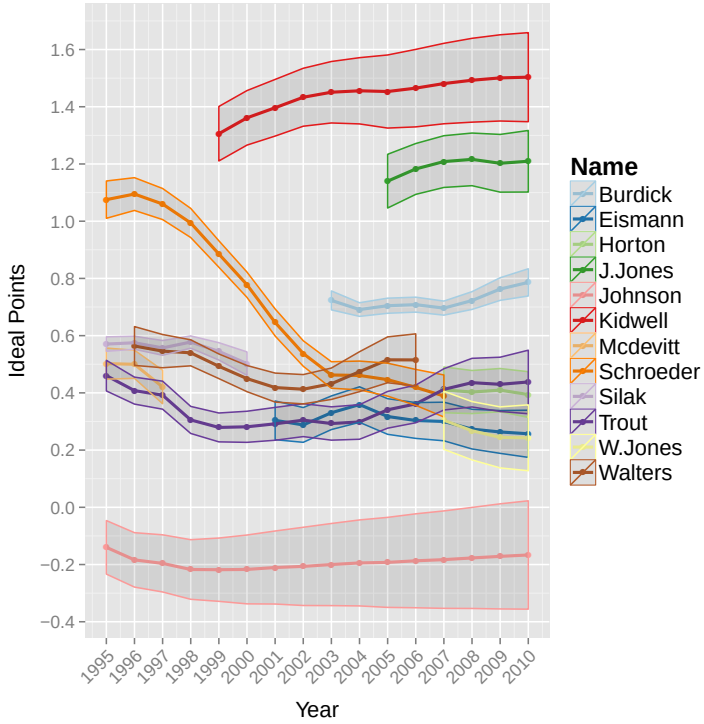
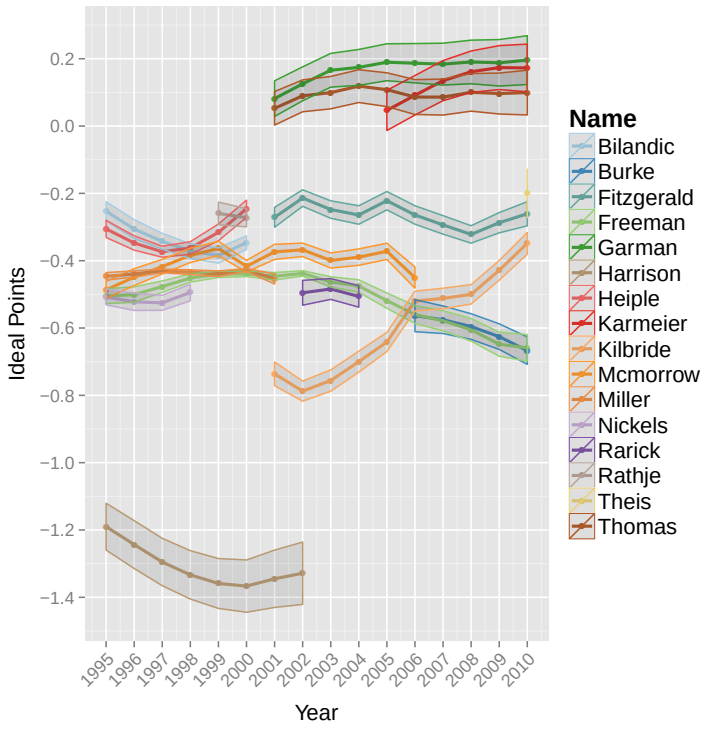
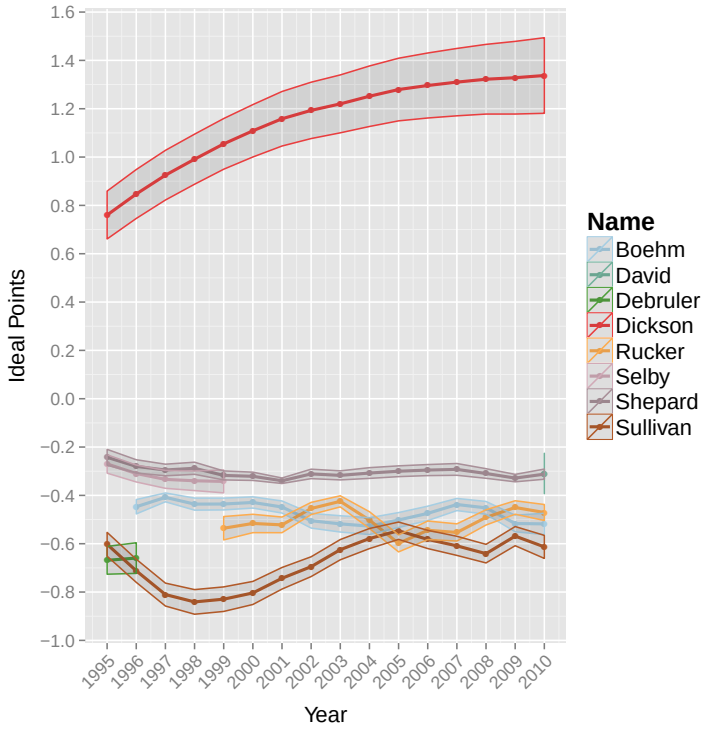


Figure A-5: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

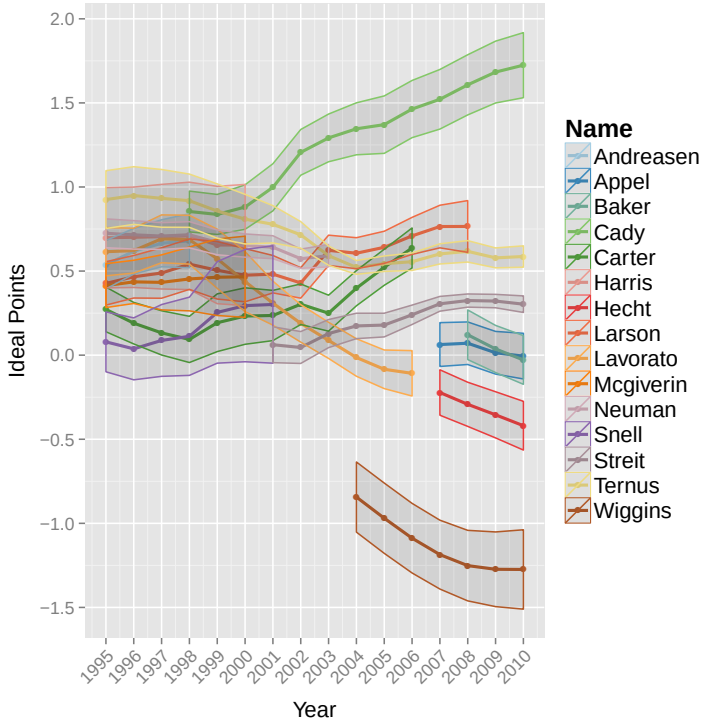
(a) Illinois



(b) Indiana



(c) Iowa



(d) Kansas

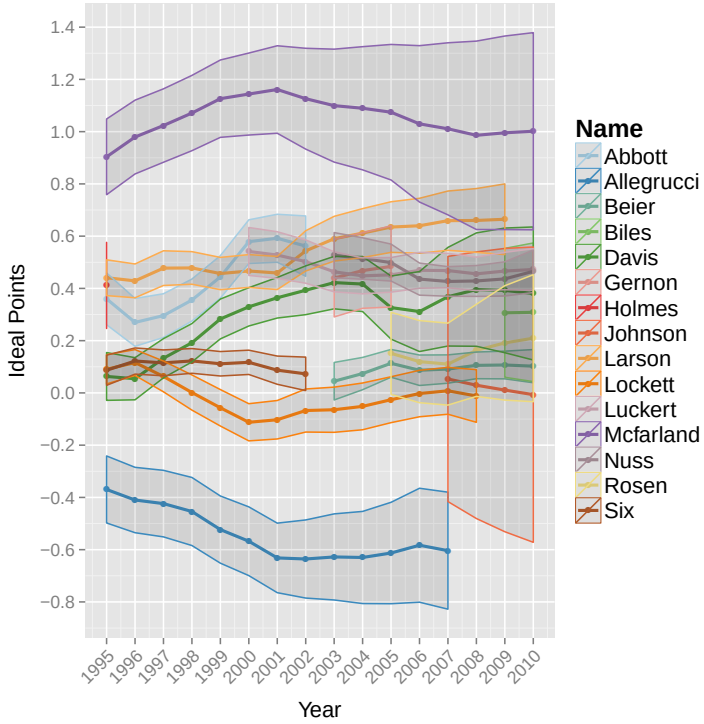
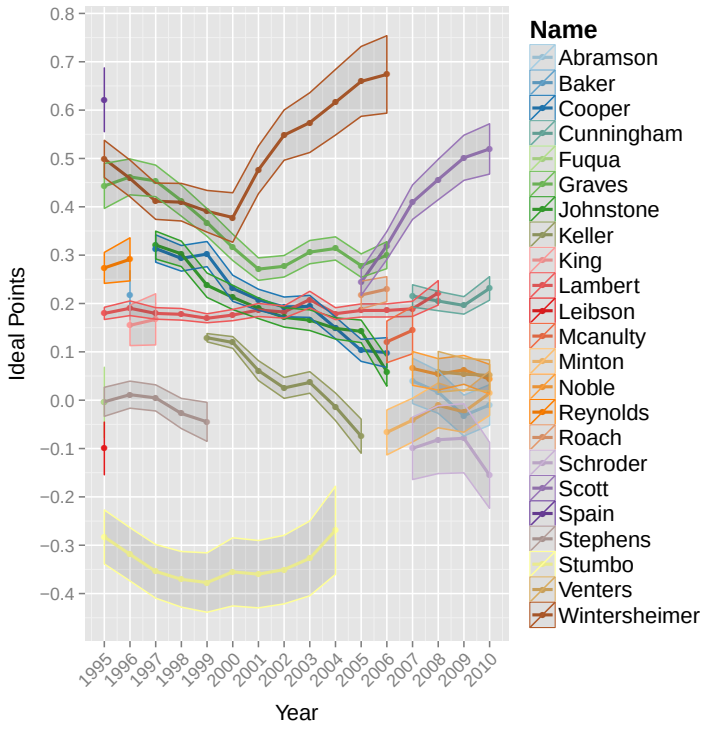
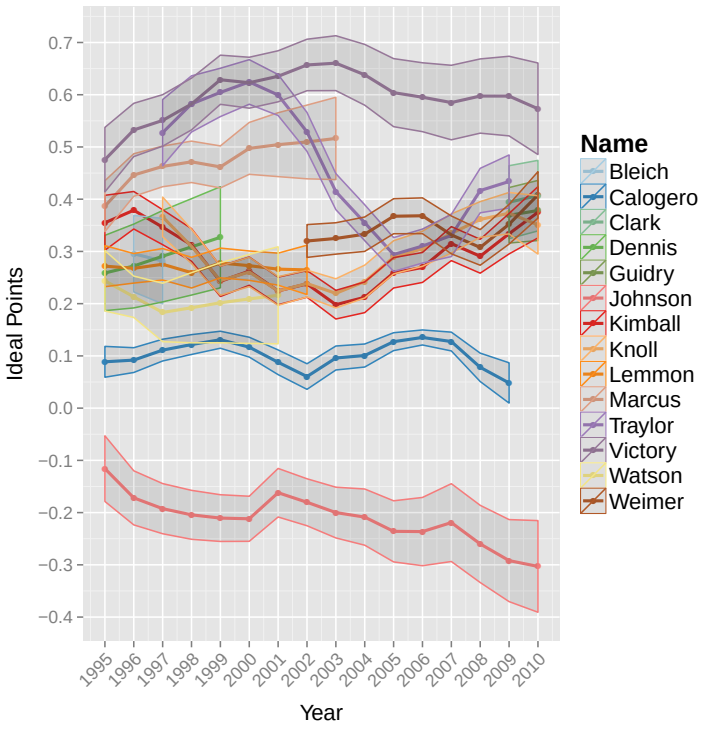


Figure A-6: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

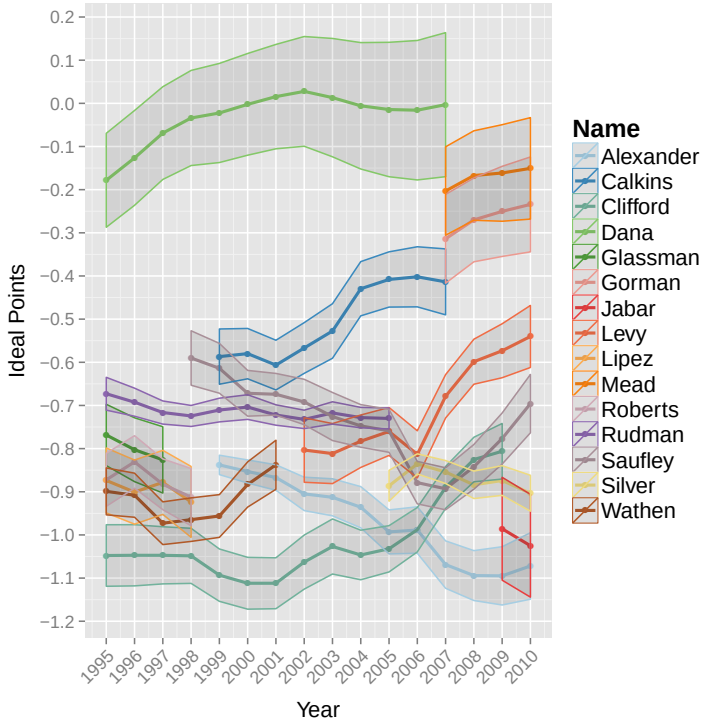
(a) Kentucky



(b) Louisiana



(c) Maine



(d) Maryland

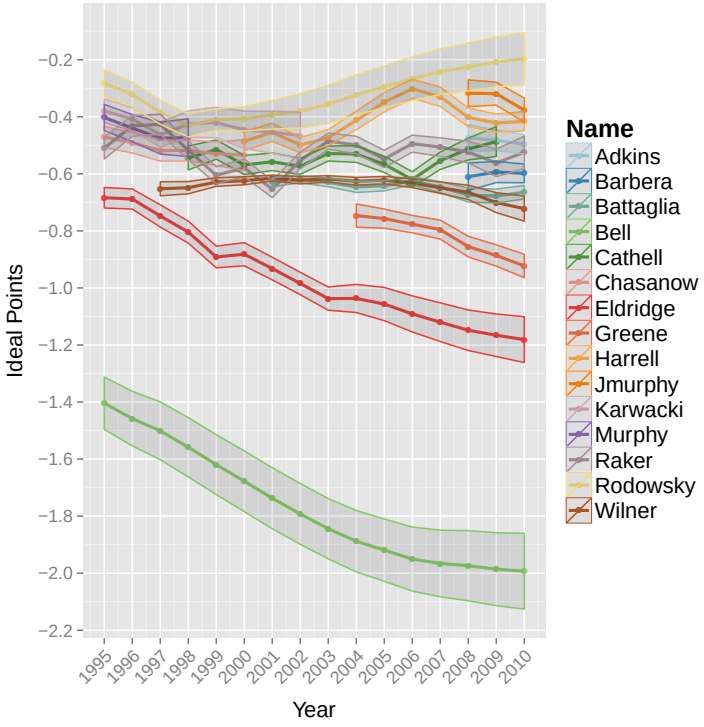
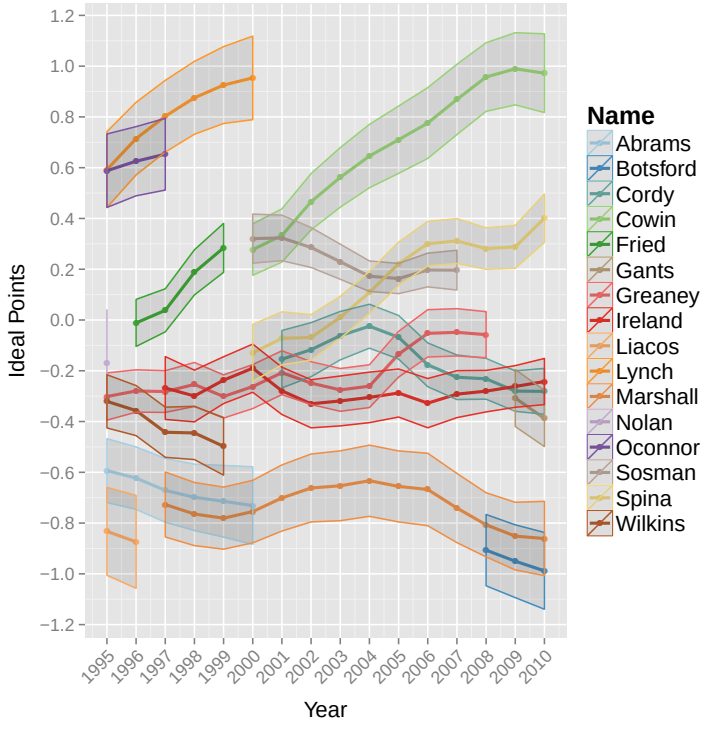
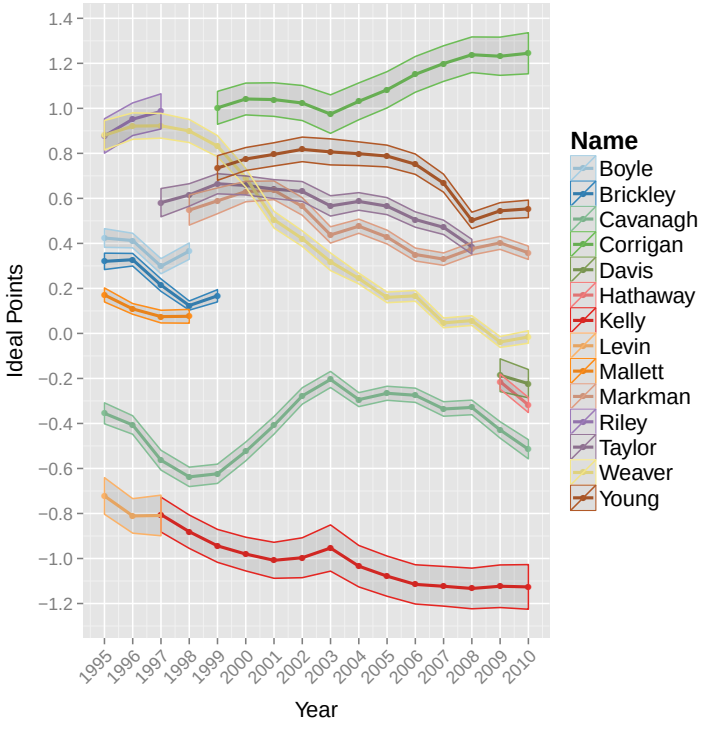


Figure A-7: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

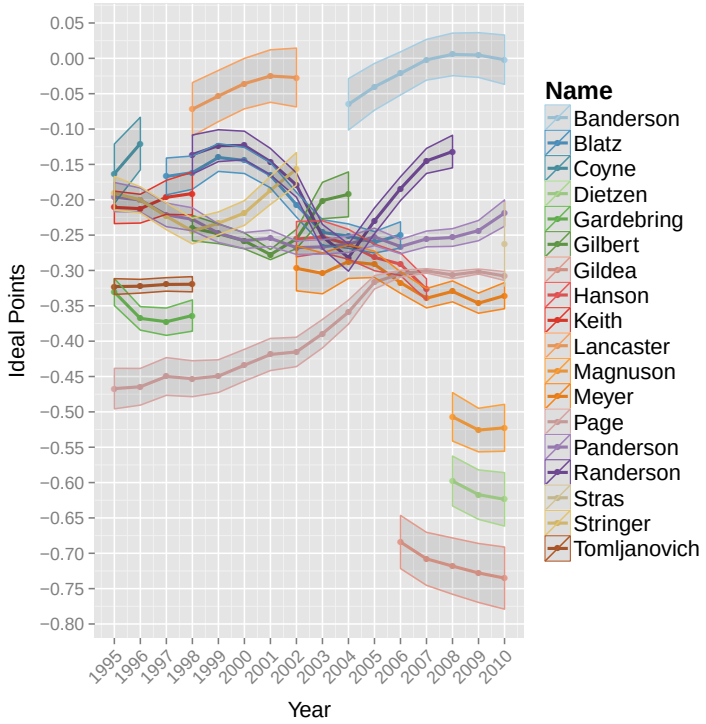
(a) Massachusetts



(b) Michigan



(c) Minnesota



(d) Mississippi

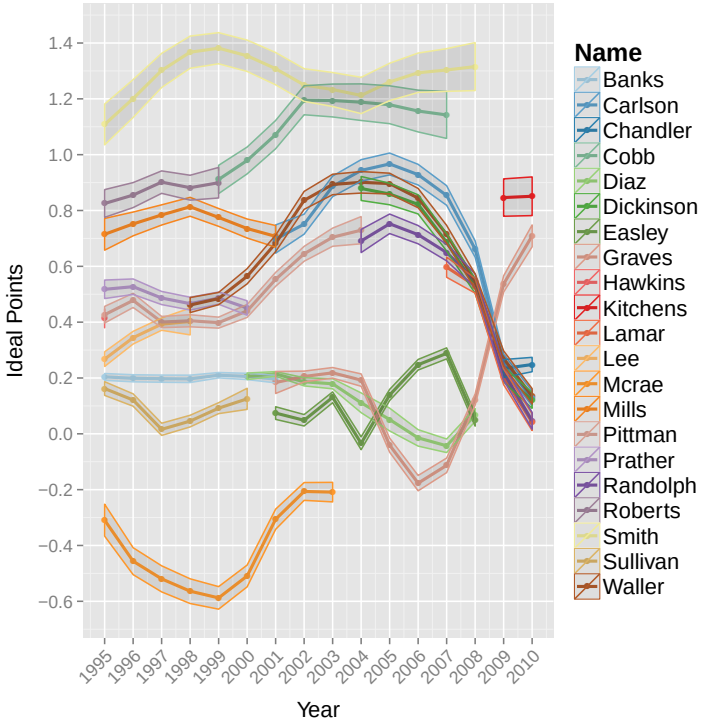
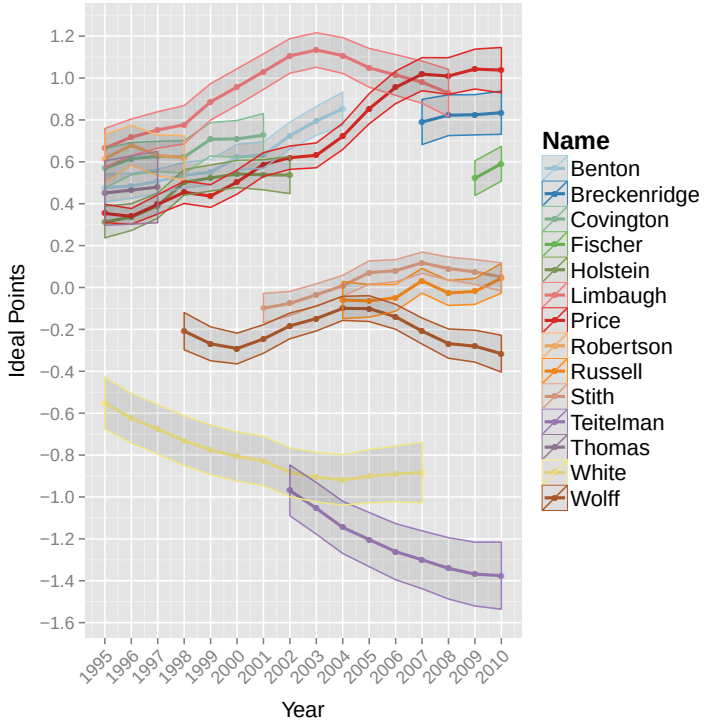
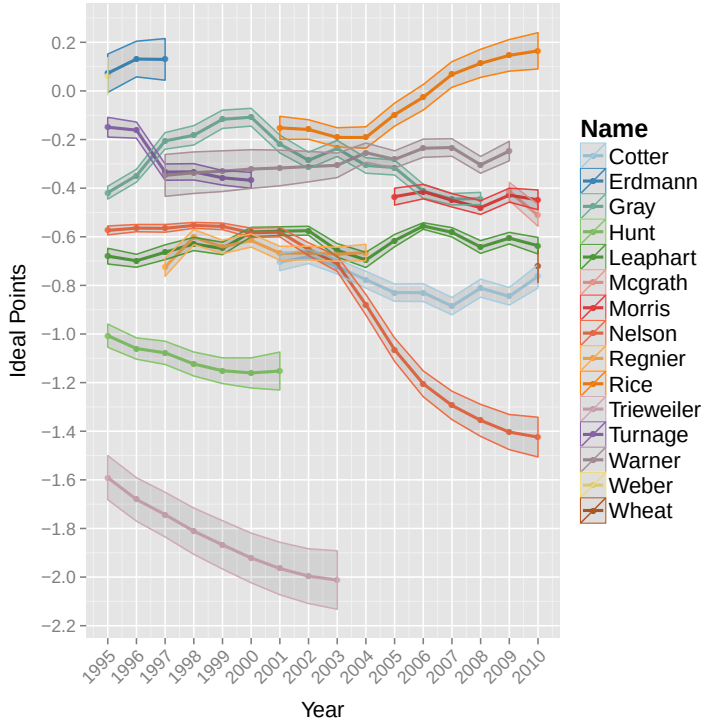


Figure A-8: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

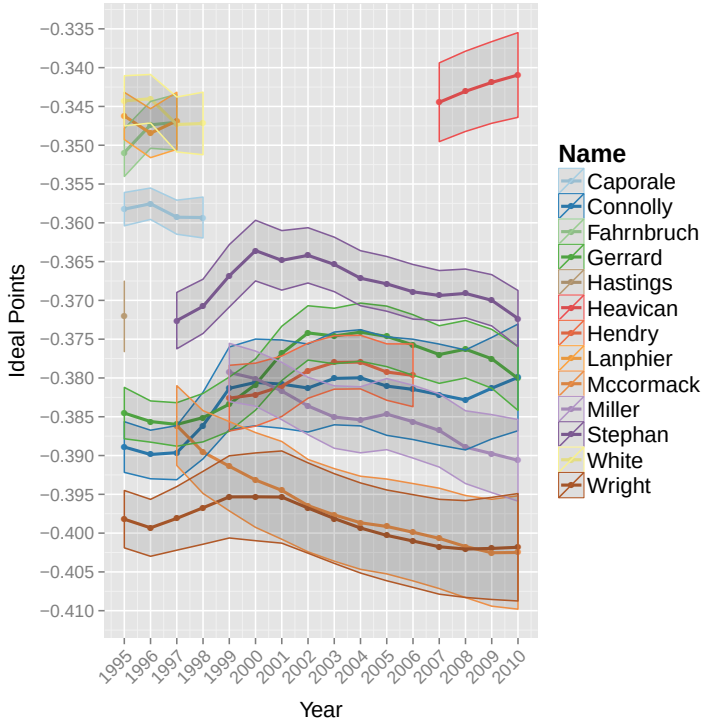
(a) Missouri



(b) Montana



(c) Nebraska



(d) Nevada

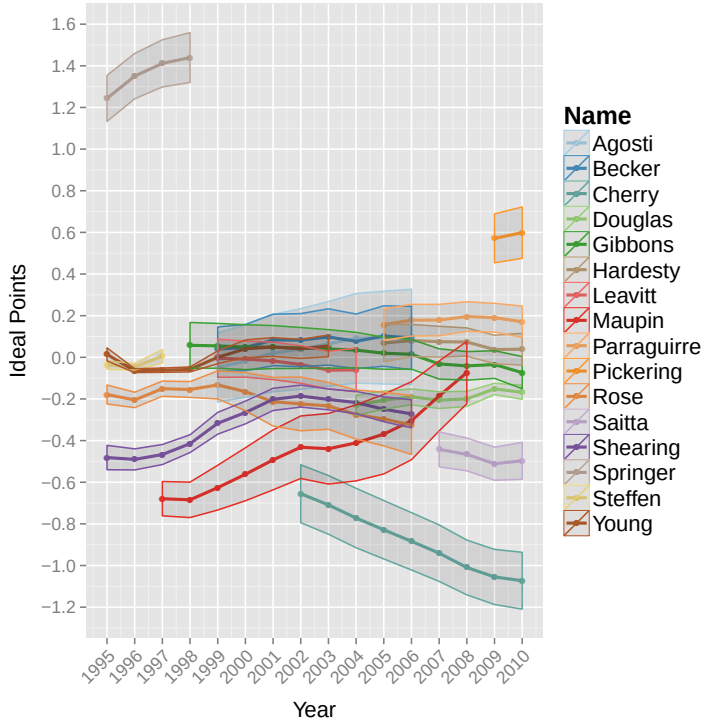
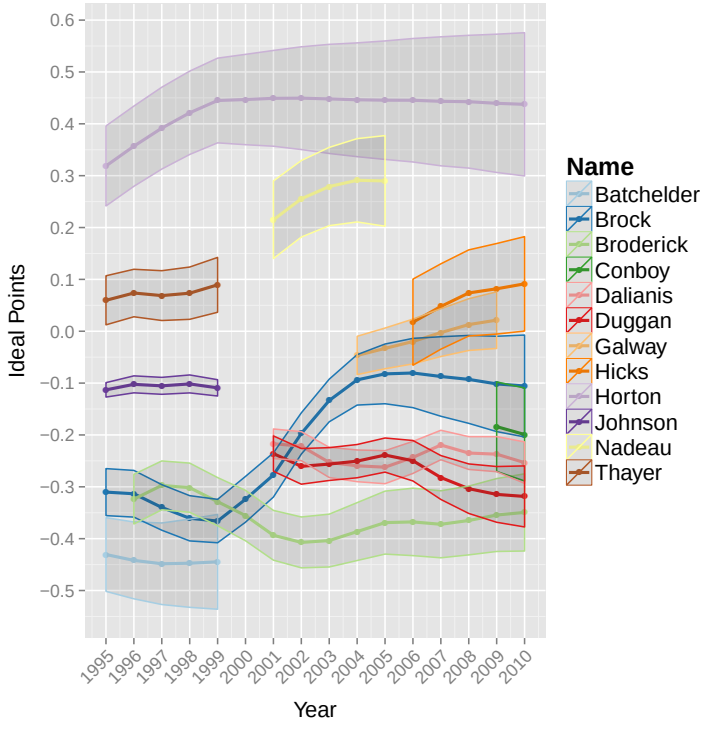
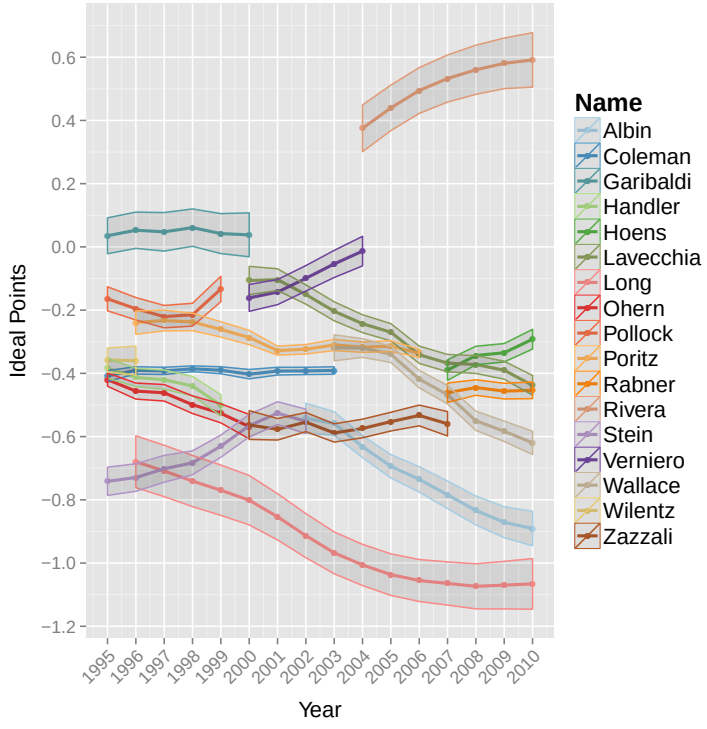


Figure A-9: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

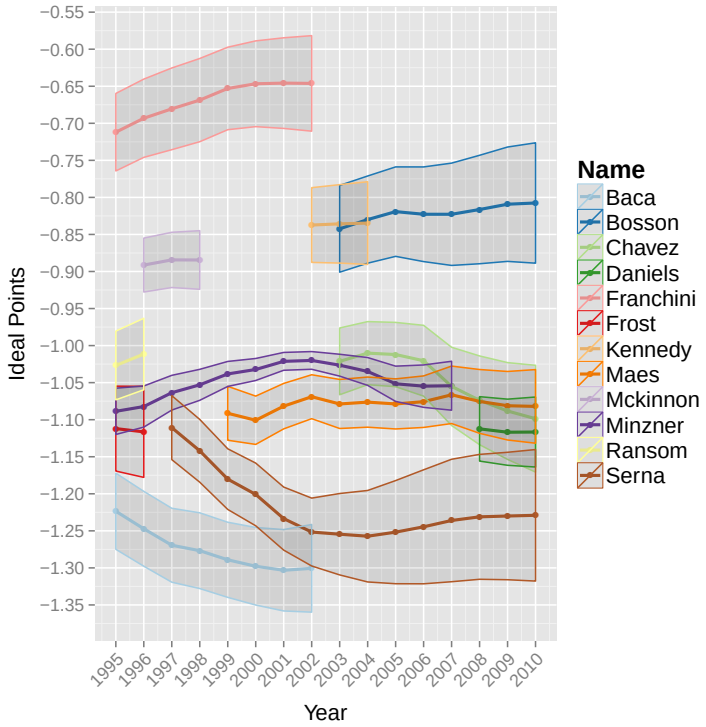
(a) New Hampshire



(b) New Jersey



(c) New Mexico



(d) New York

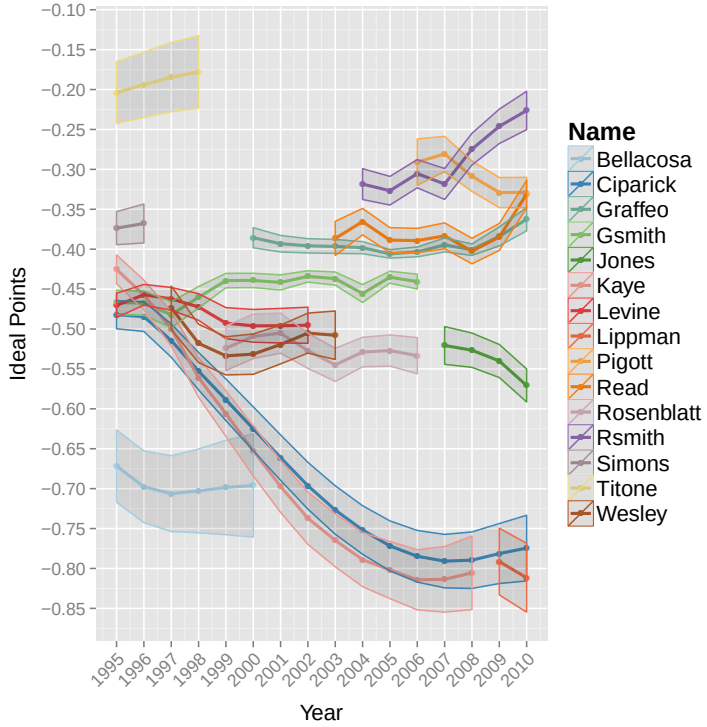
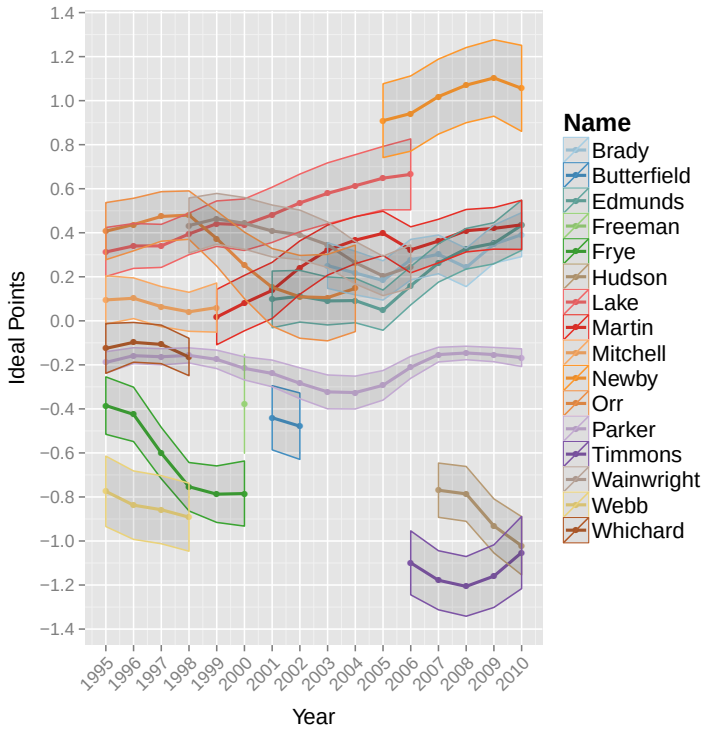
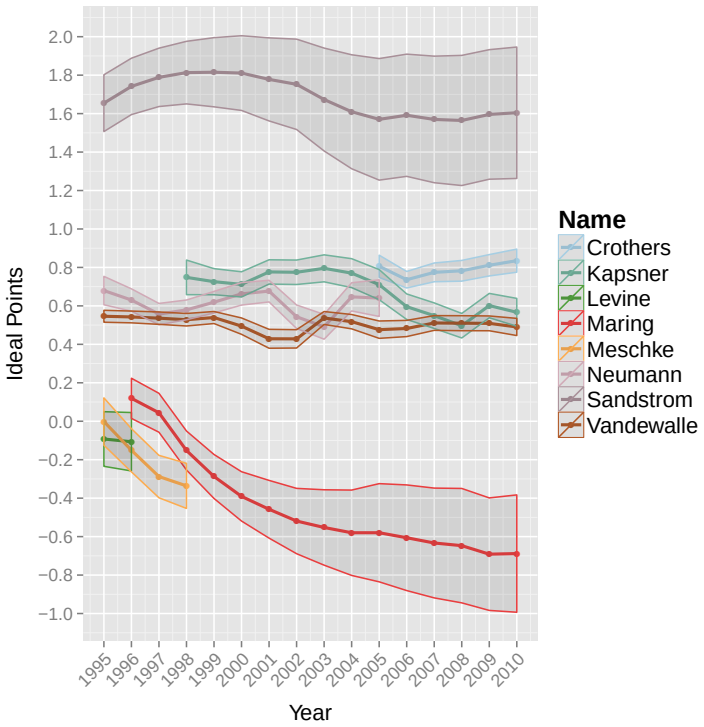


Figure A-10: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

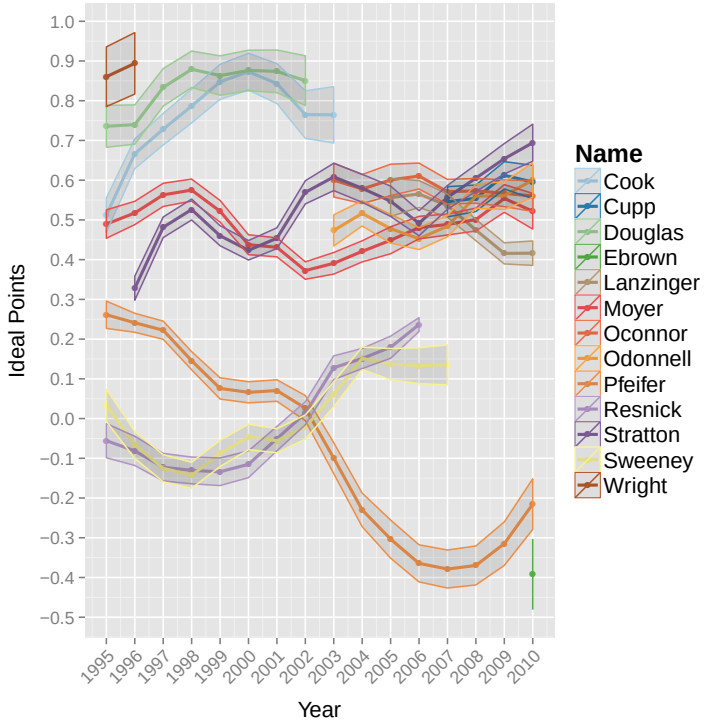
(a) North Carolina



(b) North Dakota



(c) Ohio



(d) Oklahoma (Crime)

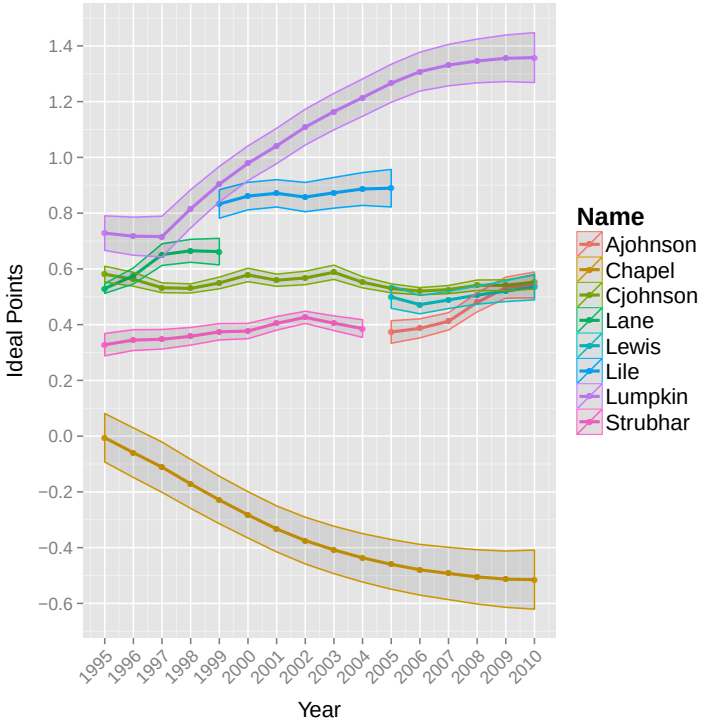
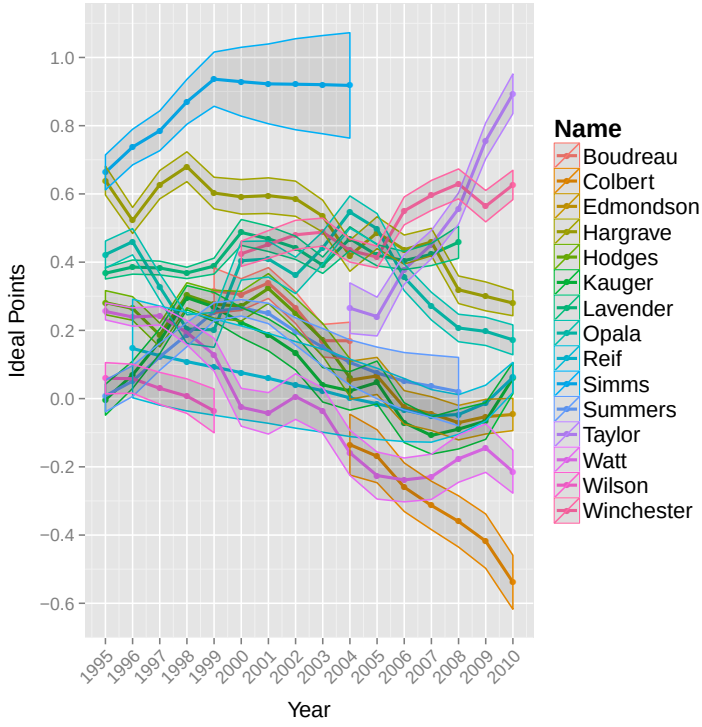
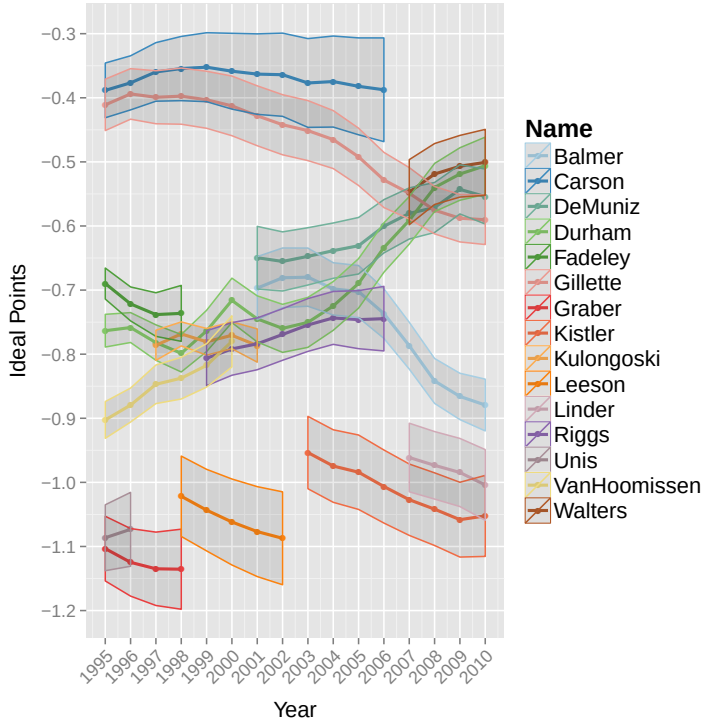


Figure A-11: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

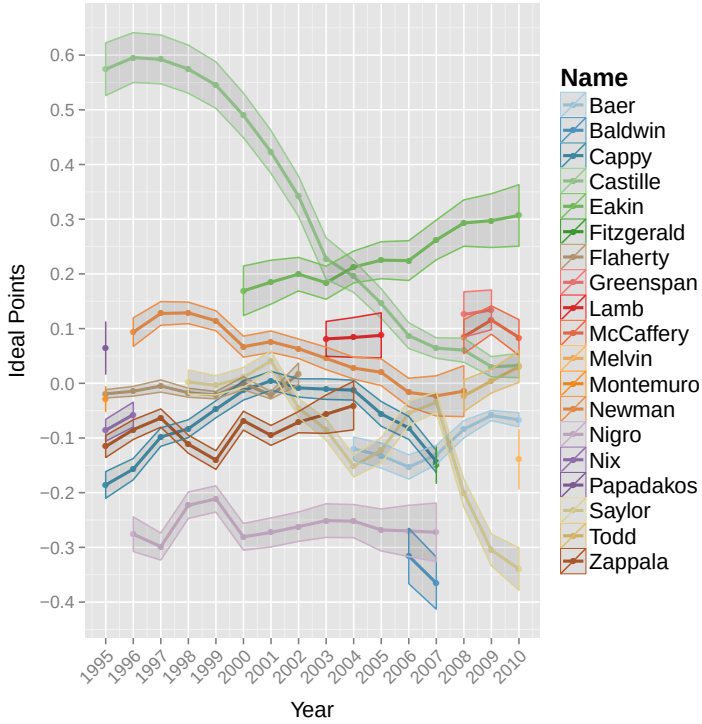
(a) Oklahoma (SC)



(b) Oregon



(c) Pennsylvania



(d) Rhode Island

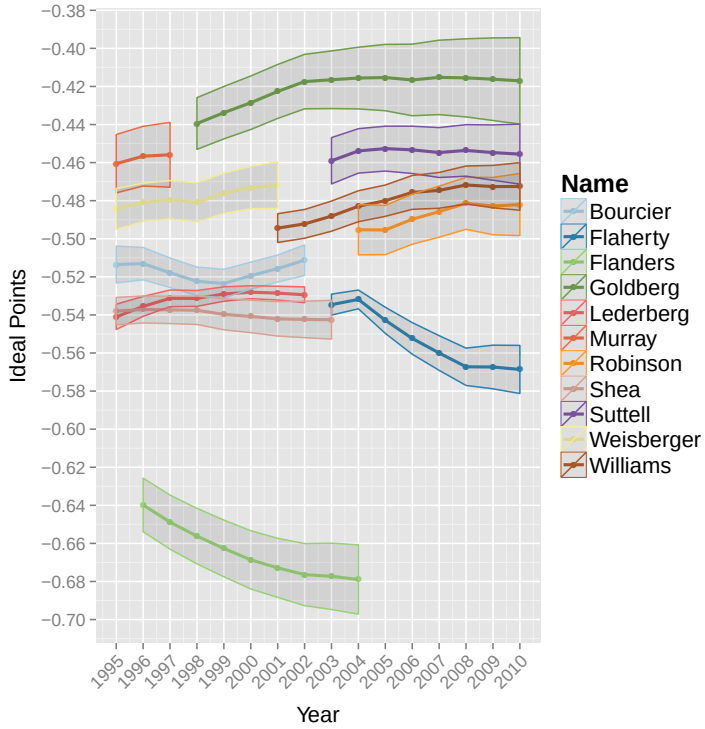
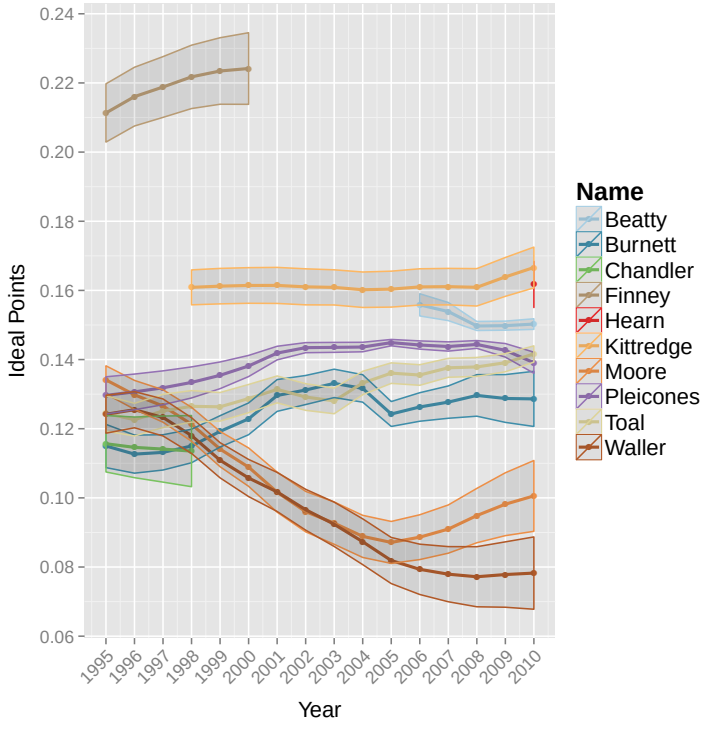
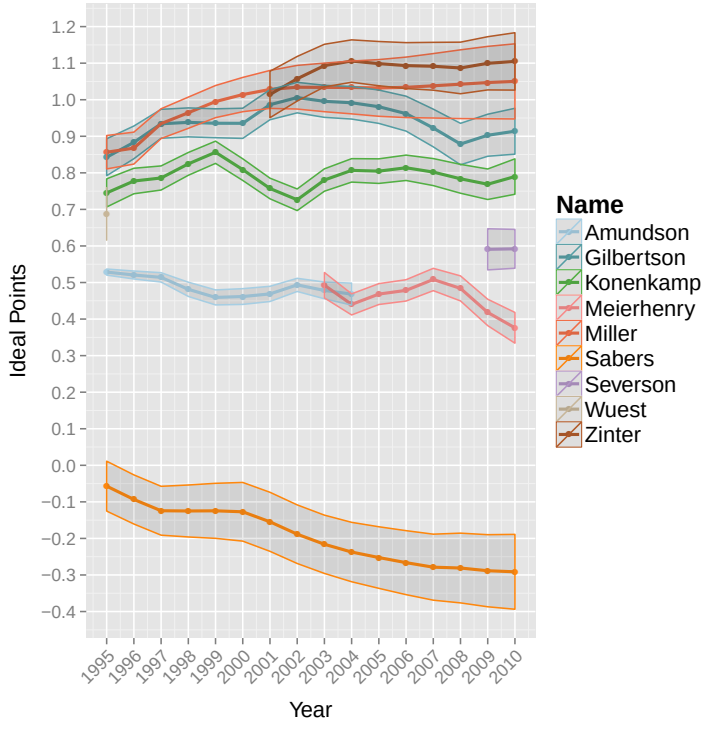


Figure A-12: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

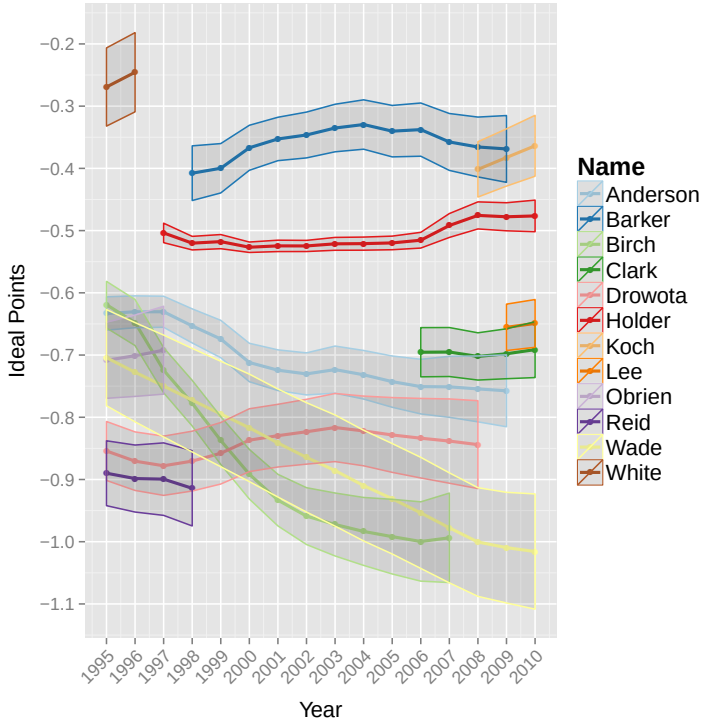
(a) South Carolina



(b) South Dakota



(c) Tennessee



(d) Texas (Crime)

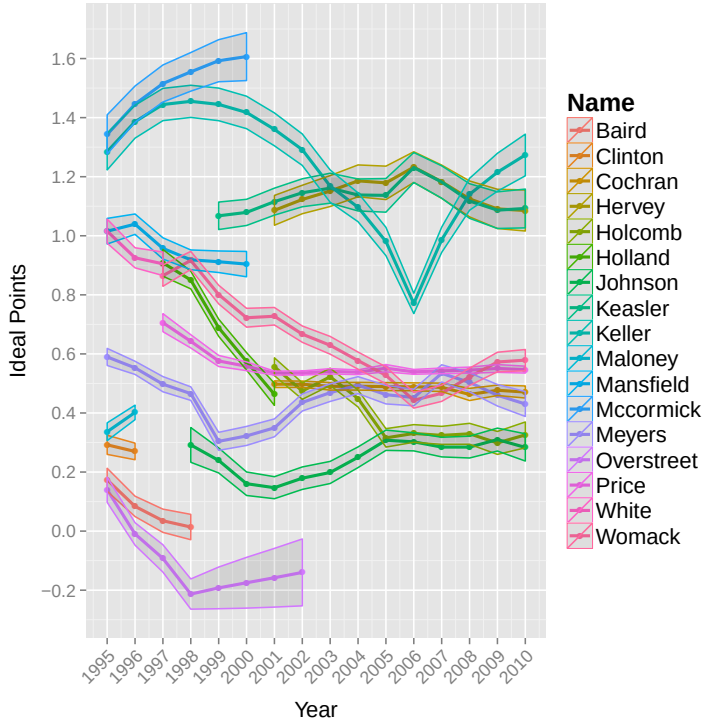
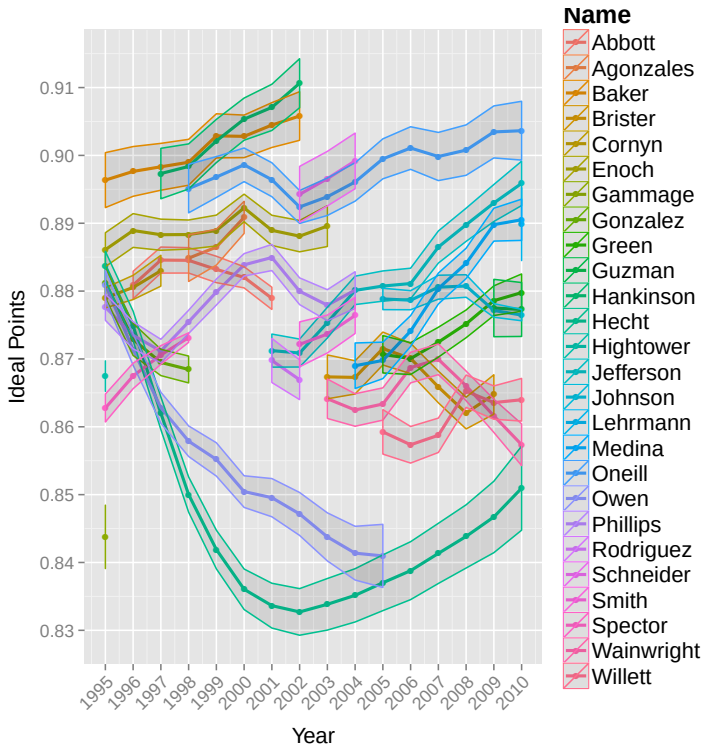
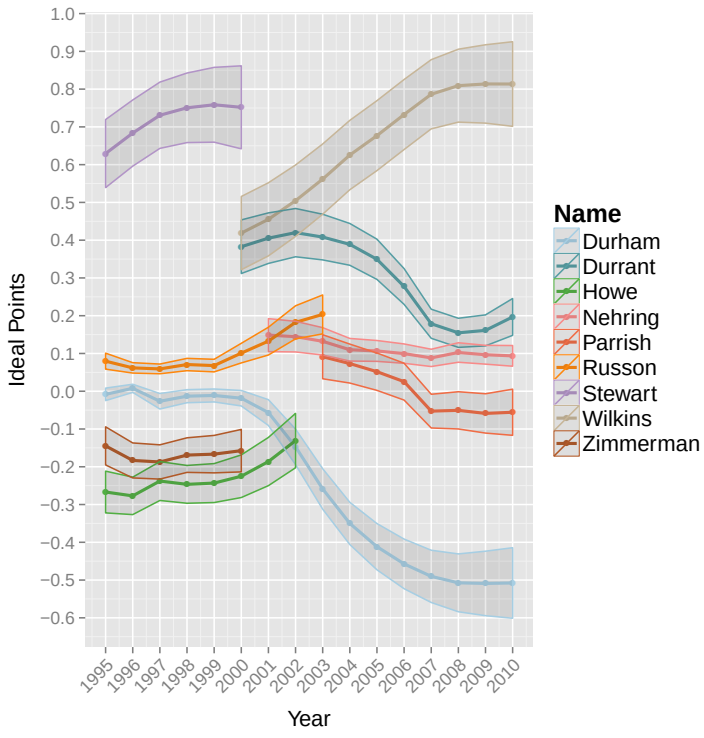


Figure A-13: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

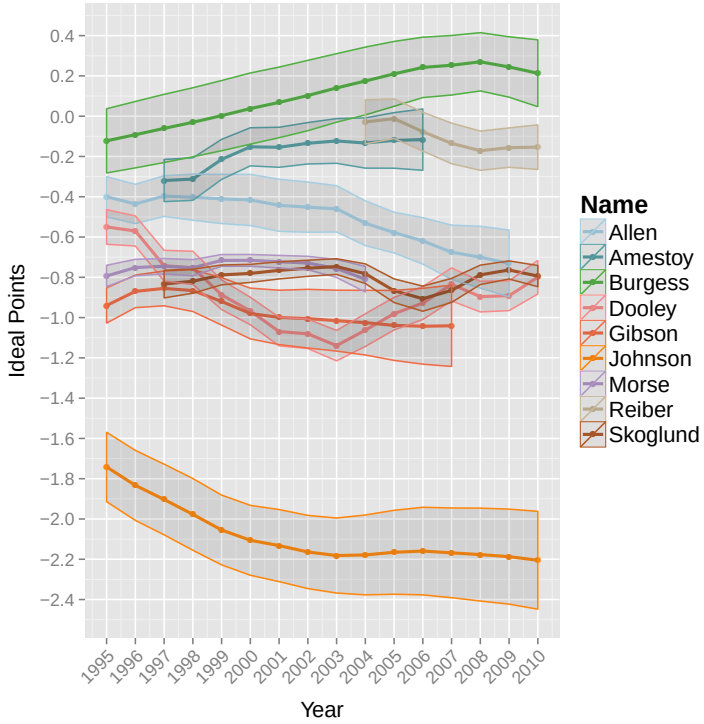
(a) Texas (SC)



(b) Utah



(c) Vermont



(d) Virginia

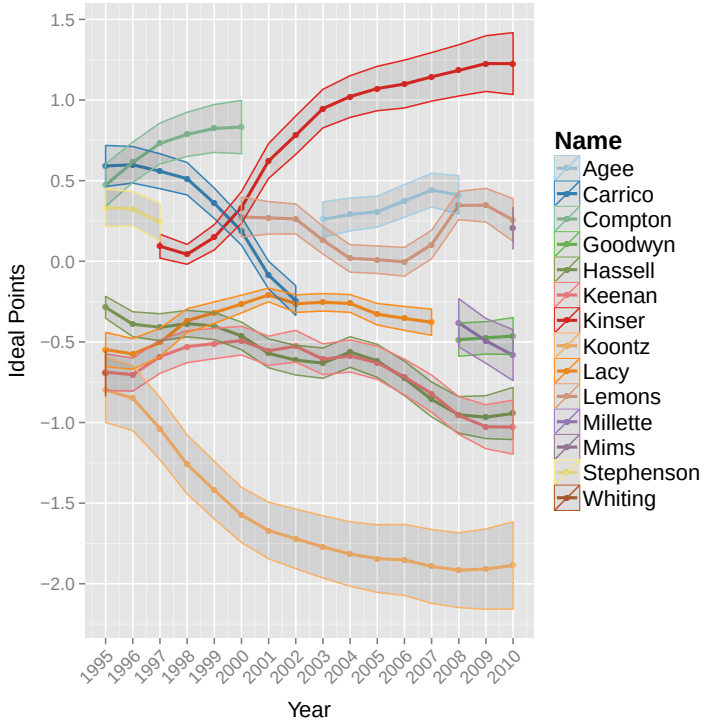
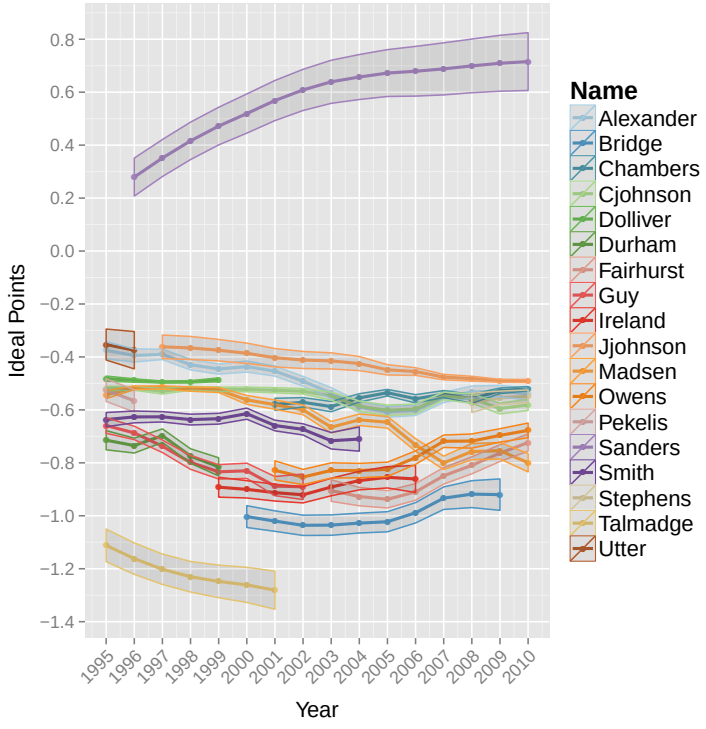
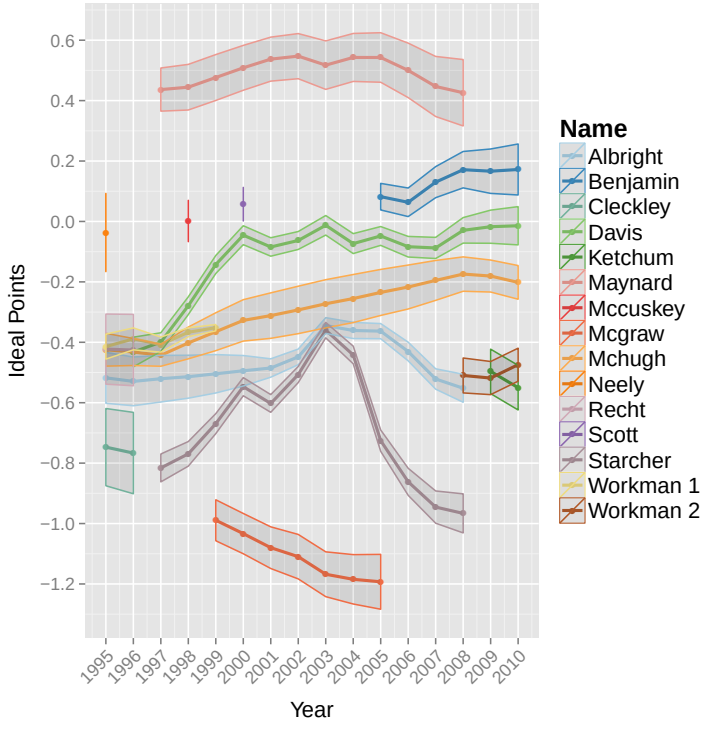


Figure A-14: Dynamic IRT Ideal Point Posterior Means and Standard Deviations

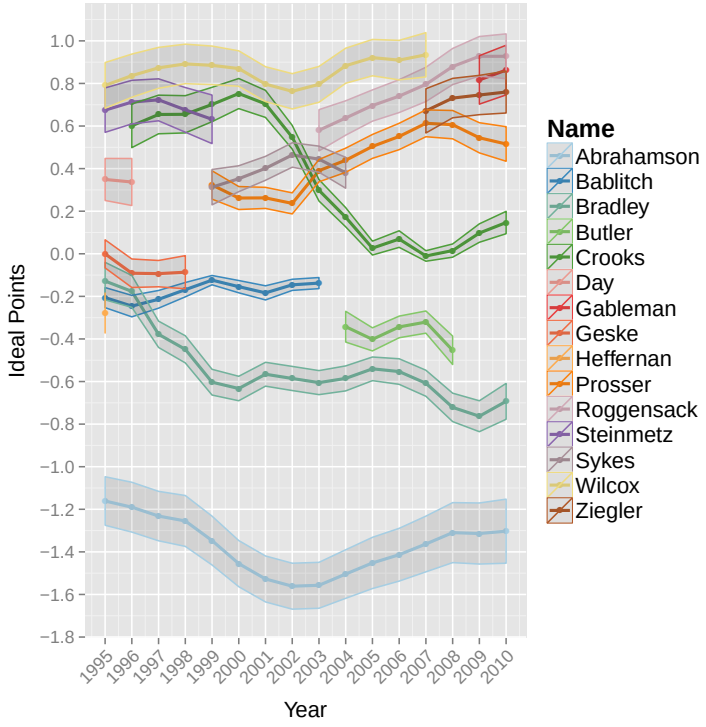
(a) Washington



(b) West Virginia



(c) Wisconsin



(d) Wyoming

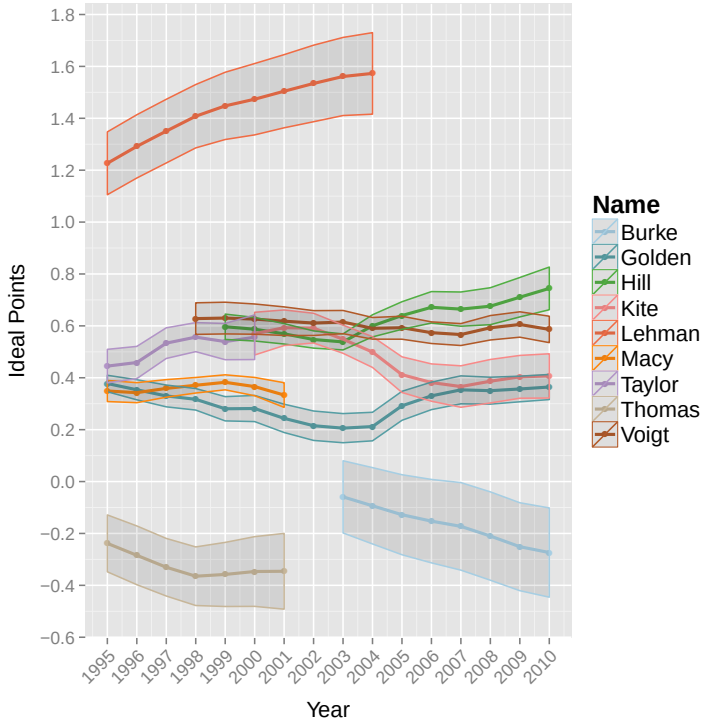
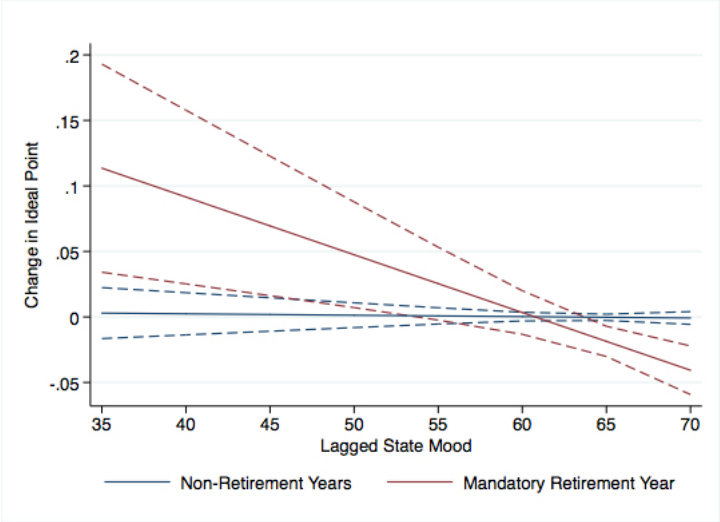
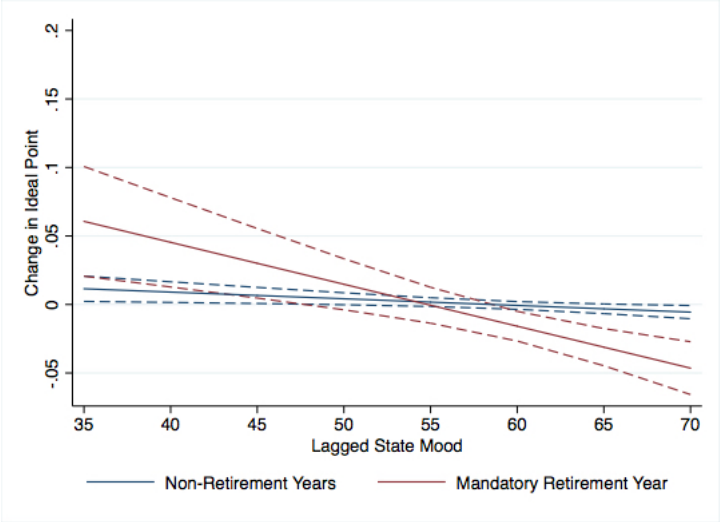


Figure A-15: The Marginal Effects of Mandatory Retirement and State Mood on Judicial Behavior in Partisan Election States



Note: The graph plots the marginal effects of mandatory retirement and lagged state mood on the change in the ideal points of state court judges in partisan election states.

Figure A-16: The Marginal Effects of Mandatory Retirement and State Mood on Judicial Behavior in Appointment States



Note: The graph plots the marginal effects of mandatory retirement and lagged state mood on the change in the ideal points of state court judges in appointment states.