

Estimating Dynamic Ideal Points for State Supreme Courts

Jason H. Windett

*Department of Political Science, Saint Louis University, 127 McGannon Hall, 3750 Lindell Blvd,
St Louis, MO 63108, USA
e-mail: jwindett@slu.edu (corresponding author)*

Jeffrey J. Harden

*Department of Political Science, University of Colorado Boulder, 416 Fleming, UCB 333, Boulder,
CO 80309, USA
e-mail: jeffrey.harden@colorado.edu*

Matthew E. K. Hall

*Department of Political Science, University of Notre Dame, 217 O'Shaughnessy Hall, Notre Dame,
IN 46556, USA
e-mail: matt.hall@nd.edu*

Edited by Justin Grimmer

Courts of last resort in the American states offer researchers considerable leverage to develop and test theories about how institutions influence judicial behavior. One measure critical to this research agenda is the individual judges' preferences, or ideal points, in policy space. Two main strategies for recovering this measure exist in the literature: Brace, Langer, and Hall's (2000, Measuring preferences of state supreme court judges, *Journal of Politics* 62(2):387–413) Party-Adjusted Judge Ideology and Bonica and Woodruff's (2014, A common-space measure of state supreme court ideology, *Journal of Law, Economics, & Organization*, doi: 10.1093/jleo/ewu016) judicial CFscores. Here, we introduce a third measurement strategy that combines CFscores with item response (IRT) estimates of judicial voting behavior in all fifty-two state courts of last resort from 1995 to 2010. We show that leveraging two distinct sources of information (votes and CFscores) yields a superior estimation strategy. Specifically, we highlight several key advantages of the combined measure: (1) it is estimated dynamically, allowing for the possibility that judges' ideological leanings change over time and (2) it maps judges into a common space. In a comparison against existing measurement strategies, we find that our measure offers superior performance in predicting judges' votes. We conclude that it is a valuable tool for advancing the study of judicial politics.

1 Introduction

The development of methods for the estimation of ideal points in policy space represents a critical advancement in the study of political institutions and elite behavior over the last several decades. In the American context, widely accepted ideal point measures are available for many institutions, including state legislatures, presidents, Congress, and the U.S. Supreme Court. Recent work seeks to create these measures for judges serving on state supreme courts, with the most prominent being Brace, Langer, and Hall's (2000) Party-Adjusted Judge Ideology (PAJID) and Bonica and Woodruff's (2014) judicial CFscores. While both of these measures are useful in some contexts, they also retain some shortcomings. In this letter, we discuss limitations with the current measures

Authors' note: The measures described here as well as complete replication materials are available at the *Political Analysis* Dataverse (Windett, Harden, and Hall 2015). This article is part of a larger research agenda by the authors on representation and state supreme courts. The ordering of names reflects a principle of rotation. We thank Mike Alvarez, Fred Boehmke, Jake Bowers, Paul Brace, David Nickerson, and Steve Rogers for helpful comments. We also thank Elizabeth Alberty, Eric Behna, Meaghan Gass, and Chad Williams for research assistance. Supplementary materials for this article are available on the *Political Analysis* Web site.

and introduce a third measurement strategy: the combination of a vote-based item response model with the CFscores' common space to produce the first vote-based dynamic indicator of state supreme court judges' ideal points.

Specifically, we employ a new judge-level data set of state supreme court cases and Martin and Quinn's (2002) Bayesian measurement model to generate dynamic ideal points for all judges serving on the fifty-two state courts of last resort from 1995 to 2010. We then utilize Bonica and Woodruff's (2014) CFscores as an instrument for mapping our ideal points into a common ideological space. Below, we present this measure and compare it to those currently in use. We show that the combination of two unique sources of information—the votes judges cast and their CFscores—produces a superior measurement strategy compared with those that rely on only one source of information. We highlight key advantages that our measure offers; most notably, it is a dynamic measure that maps judges into a common space. In a comparison with the alternatives, we show empirical evidence that our measure provides superior performance in predicting judges' votes. Overall, we conclude that our dynamic ideal points can be useful to researchers studying state judicial politics.

2 Existing Measures of State Supreme Court Ideal Points

Scholars have previously measured state judicial ideology in several different ways, with two strategies emerging as the most prominent. The first is Brace, Langer, and Hall's (2000) PAJID scores. This measure is based on the state's ideological context at the time the judge joined the court, as measured by Berry et al.'s (1998) indicators of state elite and citizen ideology. Elected judges are assigned the score of the state citizen ideology at the time of their election, and appointed judges are assigned the state elite ideology at the time of their appointment. These scores are then weighted based on the partisan affiliation of the judge. The PAJID scores represent an improvement over prior strategies. They offer greater variation and nuance than simply recording a judge's party affiliation (Ulmer 1962) and do not require news coverage of the selection process (Emmert and Traut 1994).

Bonica and Woodruff (2014) introduce an alternative approach to judicial ideal point estimation by employing campaign finance data (see also Bonica 2013, 2014). Their method scales a large data set of political donations over the period 1979–2012 into a common space. This approach provides ideal point estimates for a host of American political actors, including state supreme court judges, who may enter the data as a candidate, contributor, or appointee. This methodology represents a major advancement in ideal point estimation both generally and in the specific case of state supreme courts. It is a notable improvement over PAJID because it incorporates observed behavior rather than relying on indirect proxies. Furthermore, this approach scales actors in a common space and can generate ideal points for electoral winners and losers.

Nonetheless, both PAJID and CFscores may be problematic for estimating judicial ideology. Both methods assume that judges' ideal points are static. PAJID scores assume judges maintain consistent preferences from the time they join the court, and CFscores aggregate contribution data over time. This limitation is potentially problematic because past work finds evidence of temporal variation in judges' preferences on the U.S. Supreme Court (Epstein et al. 1998; Martin and Quinn 2002) and state supreme courts (Canes-Wrone, Clark, and Kelly 2014). The assumption of static ideal points in spite of evidence to the contrary leaves out a potentially important source of variation that scholars may wish to analyze.

Additionally, PAJID and CFscores both rely on only one source of information. PAJID scores employ the Berry et al. (1998) measures of state elite and citizen ideology, based on the assumption that “justices will reflect the ideology of their states at the time of their accession to the bench” (Brace, Langer, and Hall 2000, 393). CFscores utilize the donation behavior of judges and/or their appointing governors, which Bonica and Woodruff (2014) show is a useful source of information. Yet, while contribution behavior is undoubtedly informative, additional sources of information may offer improved estimates of the judges' ideology.

In what follows, we present a third alternative to measuring the ideology of state supreme court judges. Our measure does not assume that judges' ideal points are static and combines the information in judges' contribution behavior with their voting decisions. Specifically, we map estimates

from a model of judges' votes into the CFscores' common space to leverage the strengths of both measures in the estimation of judicial ideal points. We present the measure and validate it below. In the Online Appendix, we discuss more details and provide an application of the measure to the study of judicial responsiveness to public preferences.

3 Dynamic Ideal Points for State Supreme Courts

Utilizing the new data set constructed by Hall and Windett (2013), we estimate ideal points for state supreme court judges employing the same methodological approach that Martin and Quinn (2002) use for U.S. Supreme Court justices.¹ Following Martin and Quinn, we derive a dynamic item response (IRT) model for state supreme court judges from a spatial model of voting. This spatial model is based on the "attitudinal model" (Segal and Spaeth 2002).² We employ Markov chain Monte Carlo (MCMC) methods to fit a Bayesian measurement model of ideal points for all judges serving on every state supreme court from 1995 to 2010.

To identify the model, we consider all non-unanimous state supreme court cases with a full written opinion.³ With these data we estimate a single dimension of spatial voting behavior. Following Martin and Quinn (2002), we use semi-informative priors from PAJID scores and party affiliation to set a directional constraint within each court. These priors define a left-right directionality of preferences for our models. We then estimate the model for each court-year just as Martin and Quinn (2002) do for the U.S. Supreme Court. Our estimates are generated from 250,000 MCMC iterations with 25,000 burn-in exclusions and thinned to every 100th iteration to reduce sample autocorrelation.

3.1 Projecting to a Common Space

Because judges in different states do not vote on the same cases and there are no instances of overlapping membership, we are not able to use the IRT model to directly bridge these ideal points across states into a common space. As such, the ideal points outlined above are based on the unique dimension within each individual court. Instead, we utilize Bonica and Woodruff's (2015) CFscores as an instrument to map our IRT ideal points into a common space. We follow Shor and McCarty (2011) and produce this mapping using a linear projection within each state. Unlike Shor and McCarty (2011), who aggregate their ideal points, we retain the dynamic nature of our original measure. Moreover, although we utilize the CFscores' common space for scaling, our within-state ordering comes from the IRT estimates. As such, our estimates are based on the judges' voting behavior.

The first step in this process is to create a mapping of our IRT ideal points for a given state into the CFscores' cross-state common space. To do so, we regress the CFscores for judges in that state on the yearly ideal point of each judge from our measure using ordinary least squares, as shown in equation (1):

$$\text{CFscore}_j = \alpha + \beta(\text{IRT Ideal Point})_{jt} + \varepsilon_{jt}. \quad (1)$$

In this equation, j subscripts individual judges and t subscripts each term a judge serves.

We then generate expected values from the parameter estimates of equation (1) to produce that state's scaled ideal point estimates in the cross-state common space, as shown in equation (2):

$$\widehat{\text{Scaled Dynamic IRT Score}}_{jt} = \hat{\alpha} + \hat{\beta}(\text{IRT Ideal Point})_{jt}. \quad (2)$$

Because this second step uses the regression in equation (1), we must assume that the dynamic IRT ideal points in that state are linearly related to the underlying dimension in the CFscores.

¹The measures described here as well as complete replication materials are available at the *Political Analysis* Dataverse (Windett, Harden, and Hall 2015).

²Of course, judges may vote strategically. Thus, our measure should be interpreted as the ideological preferences of judges as reflected in their voting patterns.

³As Martin and Quinn (2002, 137) note, unanimous cases contribute no information to the model's likelihood and make prior specification difficult. We use only cases with a full written opinion to distinguish judges in the majority and judges who dissent.

Additionally, we assume that the estimates from equation (1), which define the average relationship between the CFscores and the IRT ideal points in that state, can be applied to the year-specific dynamic IRT ideal points in order to impute dynamic common space scores. That is, we assume that we can make a projection into the common space for each judge-year.⁴

3.2 *Assessing the Measure Across States*

Because a judge's score generated by this method is unique in each year, the process results in a considerable amount of data. We summarize the scaled dynamic IRT (SDIRT) measure in Fig. 1. The graph plots the aggregated distributions of ideal points for each state court. The vertical black lines represent each court's median over the 16-year period. The boxes indicate the 25th and 75th percentiles of the court, and the whiskers indicate the minimum and maximum ideal points.

While we suppress temporal variation here, the graph does show differences across the states. Courts with relatively liberal medians include those in Hawaii, Vermont, and Oregon, while some conservative-leaning courts can be found in Alabama, Texas, and Iowa. The data also show differences in within-court variation. In some courts, such as those in New Mexico and Texas (Supreme Court), there is low variance within the court, indicating agreement and ideological homogeneity. Other courts, such as in Virginia and Wisconsin, evince much greater ideological spread.

3.3 *Assessing Dynamics*

As we discuss above, a key advantage of our measure is its dynamic nature. Figure 2 plots the individual ideal points from 1995 to 2010 for four states with different selection mechanisms: Delaware (gubernatorial appointment), Missouri (initial appointment by the governor, then retention elections), Ohio (nonpartisan elections), and West Virginia (partisan elections). In each graph the points and shading, respectively, represent posterior means and standard deviations generated from the IRT model for each judge in each year.

A major feature of these graphs is the change in ideal points for the judges over time. In Delaware and Missouri, the ideal points fluctuate in an incremental pattern, while in Ohio and West Virginia, judges' preferences change in a more dramatic, non-incremental fashion. In each instance (as well as in other states—see the Online Appendix), there is evidence that judges' ideal points do change over time. For 72% of the judges in our data, the posterior mean of the IRT estimate in at least 1 year moves outside the bounds of the posterior standard deviation in his or her first year. Thus, we find evidence supporting the need to allow for dynamics in measuring judges' ideal points (see also Martin and Quinn 2002; Canes-Wrone, Clark, and Kelly 2014).

4 *Evaluating the Dynamic Ideal Point Measure*

Our final task is to evaluate the SDIRT measure—which combines information from the IRT model and the CFscores—by comparing it with PAJID and CFscores alone. We utilize the most recent publicly available PAJID scores, which include judges in all states placed on the bench from 1995 to 2008, and CFscores from all states except South Carolina due to sparse data availability in that state.⁵ We evaluate these measures by comparing their performance in predicting individual votes. Following Bonica and Woodruff (2014), we estimate a logistic regression model of the votes for each non-unanimous case on each ideal point measure.⁶ We then use those models to predict the direction of each vote based on whether each judge's fitted value is above 0.50 (concur) or below 0.50 (dissent). We then compute two fit statistics: the proportion of votes correctly classified (CC) and the aggregate proportional reduction in error (APRE; see Armstrong et al. 2014).

⁴Thus, these ideal points should be used with appropriate caution because, just like any other estimates from a model, they are measured with error. If the measure is used as a dependent variable, this is an efficiency issue. If it is used as a covariate, researchers may need to employ an errors-in-variables approach in model estimation.

⁵We use the unscaled version of our measure in South Carolina.

⁶The data occasionally produce separability problems due to small within-case sample sizes. We employ a penalized likelihood approach in estimation to address this issue (Heinze et al. 2013).

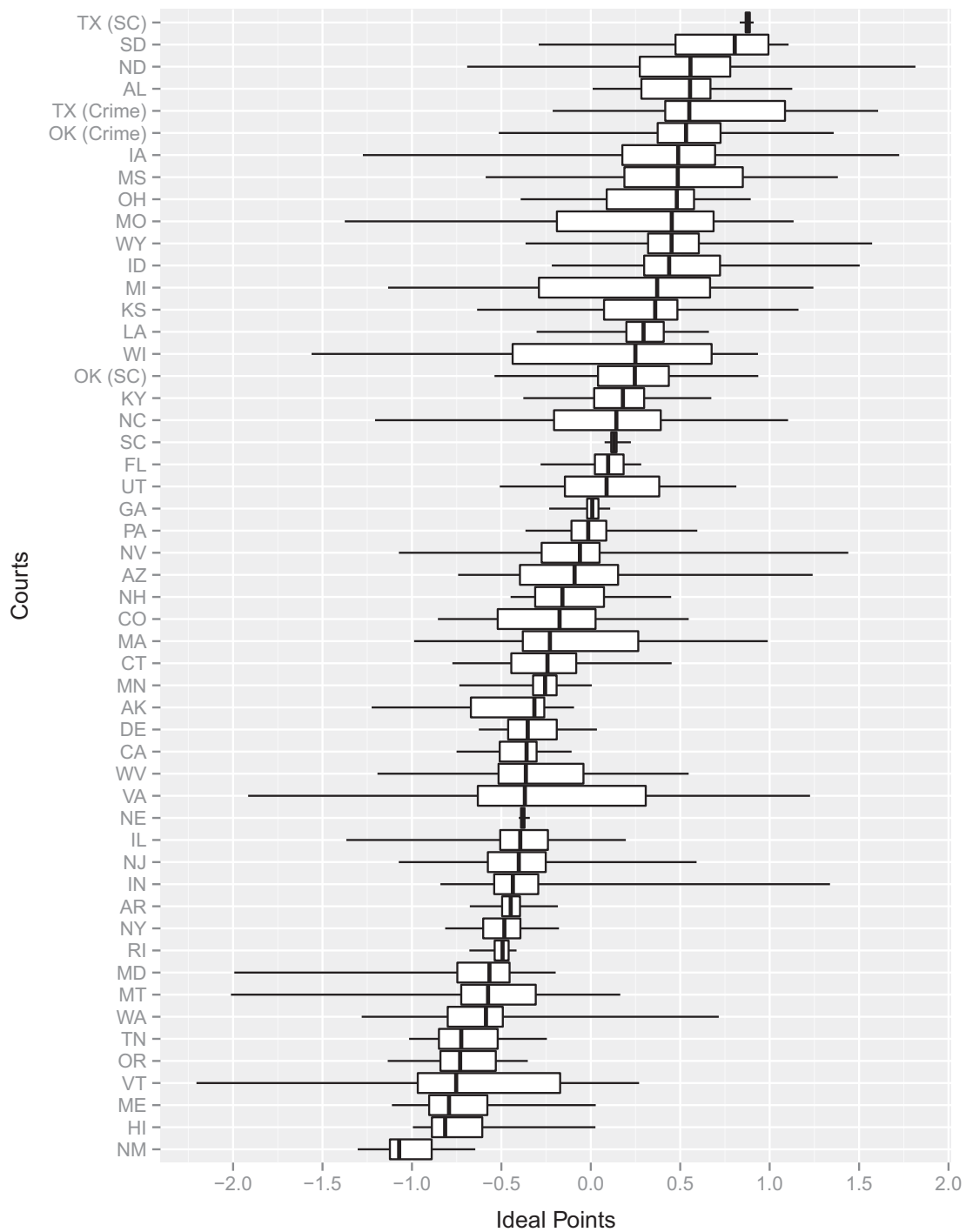


Fig. 1 Distributions of scaled dynamic IRT ideal points for state courts of last resort, aggregated over 1995–2010. The graph plots the distribution of ideal points by court, aggregated over time. The vertical black lines represent each court’s median over the 16-year period. The boxes indicate the 25th and 75th percentiles of the court, and the whiskers indicate the minimum and maximum ideal points.

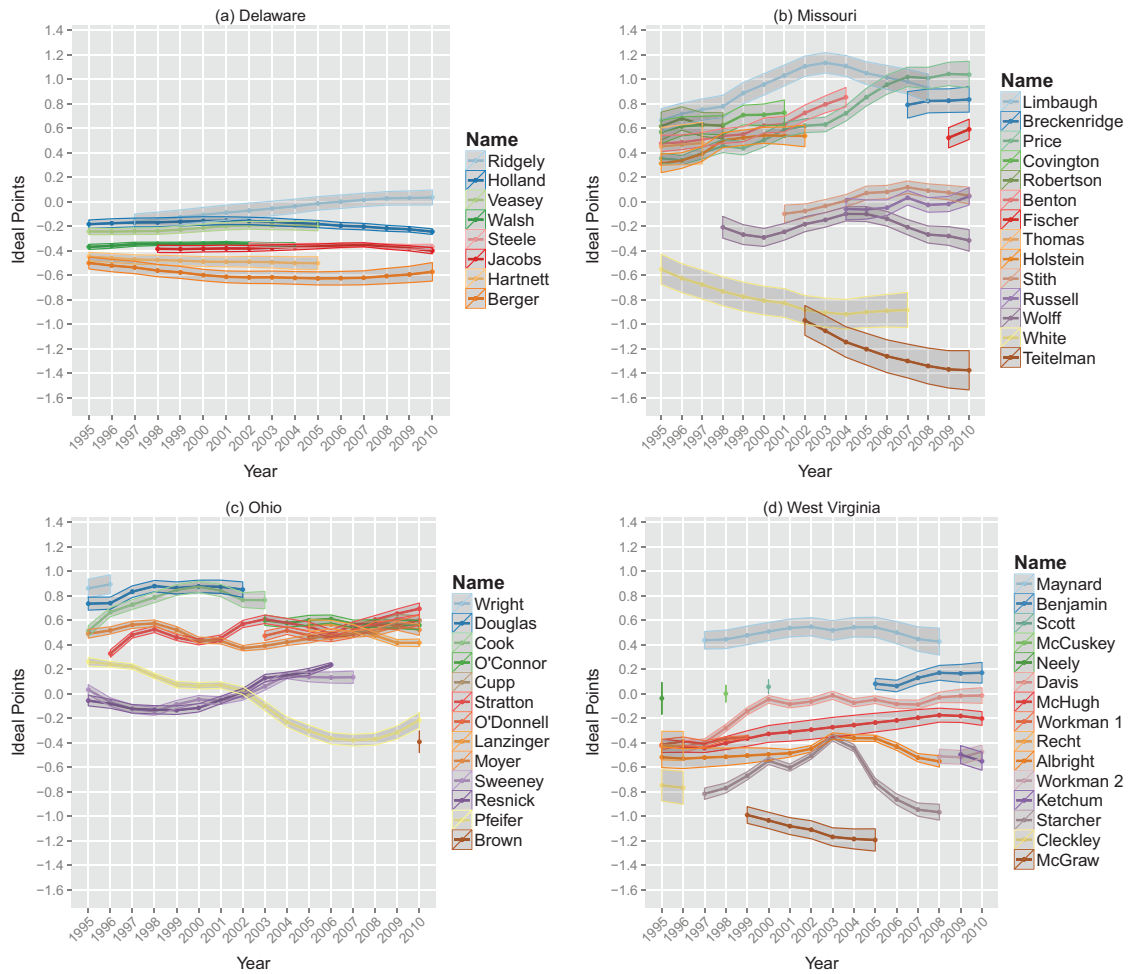


Fig. 2 Scaled dynamic IRT ideal point posterior means and standard deviations. The graphs present the ideal point estimates for all judges serving in Delaware, Missouri, Ohio, and West Virginia from 1995 to 2010. Points represent posterior means and shading represents posterior standard deviations from Martin and Quinn’s (2002) dynamic IRT model after projection into the CFscores’ common space.

Importantly, we perform this assessment using a cross-validation routine with our 1995–2010 vote data. Cross-validation is necessary to guard against the in-sample bias that would come with evaluating the SDIRT estimates on the same data used for estimation (PAJID and CFscores are already out-of-sample measures). Specifically, we randomly split the cases in each court into a training set (80% of the cases) and testing set (20%). We then fit the IRT models using only the cases in the training set. Finally, we estimate the two fit statistics using the SDIRT estimates from the training set on the cases in the testing set.

Table 1 reports the results from this exercise for the SDIRT measure, PAJID, and the CFscores alone. The cell entries report the CC and APRE statistics for each measure, the best-fitting measure by each statistic, and the number of court cases used in the fitting for each state court. Note that the number of usable cases is small in some courts.⁷ In the Online Appendix, we report the fit statistics using the full sample of data and find similar results.

⁷Dropping 20% of the cases from the training set leads to some judges being omitted from the estimation (if the cases on which they voted did not go into the training set). This creates (random) missingness on the cross-validated IRT estimates, which yields too few observations to fit the logit model to some cases in the testing set. Thus, the number of usable cases in the testing set is slightly less than 20% in most states.

Table 1 Cross-validated vote prediction model fit statistics

State court	<i>Proportion correctly classified</i>				<i>Aggregate proportional reduction in error</i>				Court cases (Testing set)
	SDIRT	PAJID	CFscores	Best fit	SDIRT	PAJID	CFscores	Best fit	
AK	0.829	0.805	0.843	CFscores	0.676	0.407	0.694	CFscores	54
AL	0.909	0.884	0.888	SDIRT	0.575	0.448	0.458	SDIRT	297
AR	0.830	0.812	0.801	SDIRT	0.437	0.391	0.355	SDIRT	154
AZ	0.902	0.838	0.878	SDIRT	0.600	0.333	0.500	SDIRT	25
CA	0.905	0.812	0.808	SDIRT	0.667	0.348	0.328	SDIRT	114
CO	0.898	0.835	0.856	SDIRT	0.579	0.341	0.429	SDIRT	105
CT	0.895	0.801	0.812	SDIRT	0.717	0.492	0.517	SDIRT	78
DE	1.00	0.883	0.833	SDIRT	1.00	0.667	0.250	SDIRT	5
FL	0.932	0.780	0.814	SDIRT	0.680	0.140	0.214	SDIRT	255
GA	0.810	0.767	0.795	SDIRT	0.335	0.185	0.282	SDIRT	388
HI	0.990	0.902	0.950	SDIRT	0.941	0.441	0.735	SDIRT	52
IA	0.754	0.718	0.755	CFscores	0.253	0.138	0.230	SDIRT	40
ID	0.899	0.833	0.864	SDIRT	0.590	0.333	0.462	SDIRT	34
IL	0.840	0.795	0.817	SDIRT	0.517	0.327	0.444	SDIRT	84
IN	0.888	0.809	0.908	CFscores	0.698	0.480	0.751	CFscores	108
KS	0.890	0.794	0.809	SDIRT	0.527	0.182	0.255	SDIRT	35
KY	0.923	0.776	0.922	SDIRT	0.745	0.368	0.737	SDIRT	133
LA	0.888	0.840	0.845	SDIRT	0.529	0.308	0.346	SDIRT	146
MA	0.925	0.843	0.911	SDIRT	0.567	0.200	0.533	SDIRT	33
MD	0.872	0.822	0.804	SDIRT	0.618	0.386	0.400	SDIRT	97
ME	0.851	0.802	0.789	SDIRT	0.488	0.314	0.314	SDIRT	51
MI	0.910	0.795	0.925	CFscores	0.772	0.47	0.776	CFscores	269
MN	0.846	0.811	0.824	SDIRT	0.426	0.344	0.361	SDIRT	79
MO	0.851	0.860	0.833	PAJID	0.464	0.524	0.405	PAJID	49
MS	0.909	0.810	0.860	SDIRT	0.604	0.215	0.371	SDIRT	268
MT	0.909	0.828	0.869	SDIRT	0.655	0.300	0.489	SDIRT	172
NC	0.828	0.770	0.776	SDIRT	0.310	0.207	0.207	SDIRT	18
ND	0.867	0.813	0.816	SDIRT	0.632	0.453	0.462	SDIRT	74
NE	0.826	0.792	0.769	SDIRT	0.429	0.388	0.257	SDIRT	37
NH	0.938	0.847	0.835	SDIRT	0.773	0.455	0.409	SDIRT	17
NJ	0.918	0.828	0.814	SDIRT	0.564	0.164	0.109	SDIRT	46
NM	0.877	0.759	0.841	SDIRT	0.571	0.143	0.429	SDIRT	11
NV	0.868	0.851	0.796	SDIRT	0.630	0.568	0.420	SDIRT	54
NY	0.874	0.808	0.850	SDIRT	0.425	0.151	0.288	SDIRT	56
OH	0.854	0.805	0.836	SDIRT	0.479	0.285	0.410	SDIRT	212
OK (Crime)	0.947	0.928	0.836	SDIRT	0.807	0.737	0.509	SDIRT	43
OK (SC)	0.822	0.772	0.806	SDIRT	0.348	0.120	0.298	SDIRT	190
OR	0.845	0.798	0.774	SDIRT	0.455	0.303	0.212	SDIRT	21
PA	0.807	0.804	0.805	SDIRT	0.369	0.347	0.341	SDIRT	210
RI	0.964	0.741	0.782	SDIRT	0.875	0.250	0.313	SDIRT	11
SC	0.917	0.856	—	IRT	0.667	0.427	—	IRT	72
SD	0.927	0.803	0.789	SDIRT	0.748	0.353	0.360	SDIRT	103
TN	0.891	0.815	0.829	SDIRT	0.571	0.343	0.371	SDIRT	32
TX (Crime)	0.930	0.854	0.878	SDIRT	0.769	0.412	0.604	SDIRT	189
TX (SC)	0.876	0.784	0.796	SDIRT	0.575	0.327	0.345	SDIRT	58
UT	0.807	0.794	0.752	SDIRT	0.420	0.360	0.240	SDIRT	33
VA	0.891	0.811	0.870	SDIRT	0.635	0.342	0.577	SDIRT	48
VT	0.886	0.824	0.789	SDIRT	0.565	0.362	0.232	SDIRT	58
WA	0.854	0.766	0.822	SDIRT	0.545	0.294	0.422	SDIRT	188
WI	0.920	0.742	0.909	SDIRT	0.791	0.391	0.765	SDIRT	109
WV	0.955	0.815	0.870	SDIRT	0.864	0.460	0.621	SDIRT	153

(continued)

Table 1 Continued

State court	<i>Proportion correctly classified</i>				<i>Aggregate proportional reduction in error</i>				Court cases (Testing set)
	<i>SDIRT</i>	<i>PAJID</i>	<i>CFscores</i>	<i>Best fit</i>	<i>SDIRT</i>	<i>PAJID</i>	<i>CFscores</i>	<i>Best fit</i>	
WY	0.890	0.767	0.883	SDIRT	0.588	0.235	0.559	SDIRT	26
Means	0.885	0.812	0.834		0.597	0.345	0.420		100

Notes: Cell entries report cross-validated model fit statistics for our scaled dynamic IRT ideal point measure (SDIRT), PAJID, and CFscores. The first four columns give the proportion of votes CC by the model with each measure and the best-fitting measure. The next four columns give the APRE 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. We use the unscaled IRT measure for South Carolina due to sparse CFscores data.

The results show that our SDIRT measure generally performs the best among the three alternatives; the model with that measure produces the largest fit statistics for forty-seven (CC) and forty-eight (APRE) of the fifty-two state courts. PAJID is the best performer in Missouri, and the CFscores produce the best fit in the other courts. On average, the SDIRT measure's CC is 0.885, versus 0.812 for PAJID and 0.834 for CFscores. The average APRE values reflect the same pattern: 0.597 (SDIRT), 0.345 (PAJID), and 0.420 (CFscores).

In the Online Appendix, we show that our measure's superior performance emerges when the unscaled version is used instead of the scaled version and when the IRT estimates are fit and evaluated on the full sample of court cases.⁸ However, the SDIRT measure's performance shown in Table 1 is the strongest. If we use the unscaled version, our measure performs best in forty-four state courts (CC and APRE). Thus, we conclude that the scaled version of our measure, which combines our dynamic IRT estimates with the CFscores' common space, provides the best measure of ideal points on state courts of last resort.

5 Conclusions

The main contribution of this letter is to introduce and evaluate a new measure of ideal points for judges on state courts of last resort. We highlight key advantages that this measure offers over existing measurement strategies. By combining two sources of information—votes and CFscores—our approach allows for a dynamic rather than static measure that places judges in a common space. We contend that this measure can be useful to applied researchers seeking a dynamic measure of state supreme court ideal points over the period 1995–2010.

While there are several advantages to our new measure, we do not suggest that it renders PAJID or CFscores obsolete. PAJID may be useful if the key theoretical concept researchers wish to measure is the ideological tenor of each state at the time a judge joined the bench. CFscores are very useful because they provide measures for electoral winners and losers, all in a common ideological space. Moreover, they are necessary to create the scaled version of our measure. However, our cross-validated fit statistics show that, on their own, CFscores are not an optimal choice. Instead, we conclude that the combination of our IRT estimates and the CFscores gives our scaled measure the broadest applicability to state courts research.

Funding

The National Science Foundation (SES-1424013) and a President's Research Fund Award at Saint Louis University.

Conflict of interest statement. None declared.

⁸The Online Appendix also contains explanations for the few instances in which PAJID or CFscores outperform our measure.

References

- Armstrong, David A., Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal. 2014. *Analyzing spatial models of choice and judgment with R*. Boca Raton, FL: CRC Press.
- Berry, William D., Evan J. 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* 42(1):327–48.
- Bonica, Adam. 2013. Ideology and interests in the political marketplace. *American Journal of Political Science* 57(2):294–311.
- . 2014. Mapping the ideological marketplace. *American Journal of Political Science* 58(2):367–86.
- Bonica, Adam, and Michael J. Woodruff. 2015. A common-space measure of state supreme court ideology. *Journal of Law, Economics, & Organization*, doi: 10.1093/jleo/ewu016.
- Brace, Paul, Laura Langer, and Melinda Gann Hall. 2000. Measuring preferences of state supreme court judges. *Journal of Politics* 62(2):387–413.
- 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.
- Emmert, Craig F., and Carol Ann Traut. 1994. The California supreme court and the death penalty. *American Politics Quarterly* 22(1):41–61.
- Epstein, Lee, Valerie Hoekstra, Jeffrey A. Segal, and Harold J. Spaeth. 1998. Do political preferences change? A longitudinal study of U.S. Supreme Court justices. *Journal of Politics* 60(3):801–18.
- Hall, Matthew E. K., and Jason H. Windett. 2013. New data on state supreme court cases. *State Politics & Policy Quarterly* 13(4):427–45.
- Heinze, Georg, Meinhard Ploner, Daniela Dunkler, and Harry Southworth. 2013. logistf: Firth's Bias Reduced Logistic Regression. R package version 1.21. <http://CRAN.R-project.org/package=logistf>.
- Martin, Andrew D., and Kevin M. Quinn. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* 10(2):134–53.
- Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the attitudinal model revisited*. New York: Cambridge University Press.
- Shor, Boris, and Nolan McCarty. 2011. The ideological mapping of American legislatures. *American Political Science Review* 105(3):530–51.
- Ulmer, S. Sidney. 1962. The political party variable in the Michigan supreme court. *Journal of Public Law* 11:352–62.
- Windett, Jason H., Jeffrey J. Harden, and Matthew E. K. Hall. 2015. Replication data for: Estimating dynamic ideal points for state supreme courts, *Political Analysis Dataverse*. <http://dx.doi.org/10.7910/DVN/PPPKMF>.