Judges as Candidates: Responsiveness to Judicial Selection Methods

Mason Reece

Massachusetts Institute of Technology

Most Recent Draft: May 17, 2024

Abstract

Members of Congress, presidents, and bureaucrats all strategically operate in relation to their selection for office – indeed much of their work is intentionally focused on this selection date. Do state level judges, who are selected in a panoply of manners, also respond strategically to selection or do they uphold the normative apolitical nature of their office. I use the text of over 26,000 judicial opinions in all 50 states in combination with a semi-supervised topic model to estimate the degree to which ideology structures judicial opinions. In turn, I use those estimates to identify how judges respond to selection methods in their state. Although I do not find that ideology is a dominating latent dimension of judicial opinions, I suggest several future areas of improvement for the data and models that could be used to advance our knowledge of judges as candidates.

This paper was written for MIT's Quantitative Research Methods IV and American Political Economy courses. I am indebted to feedback on this project from Kathleen Thelen and In Song Kim. I acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing HPC resources that have made the research results reported within this paper possible.

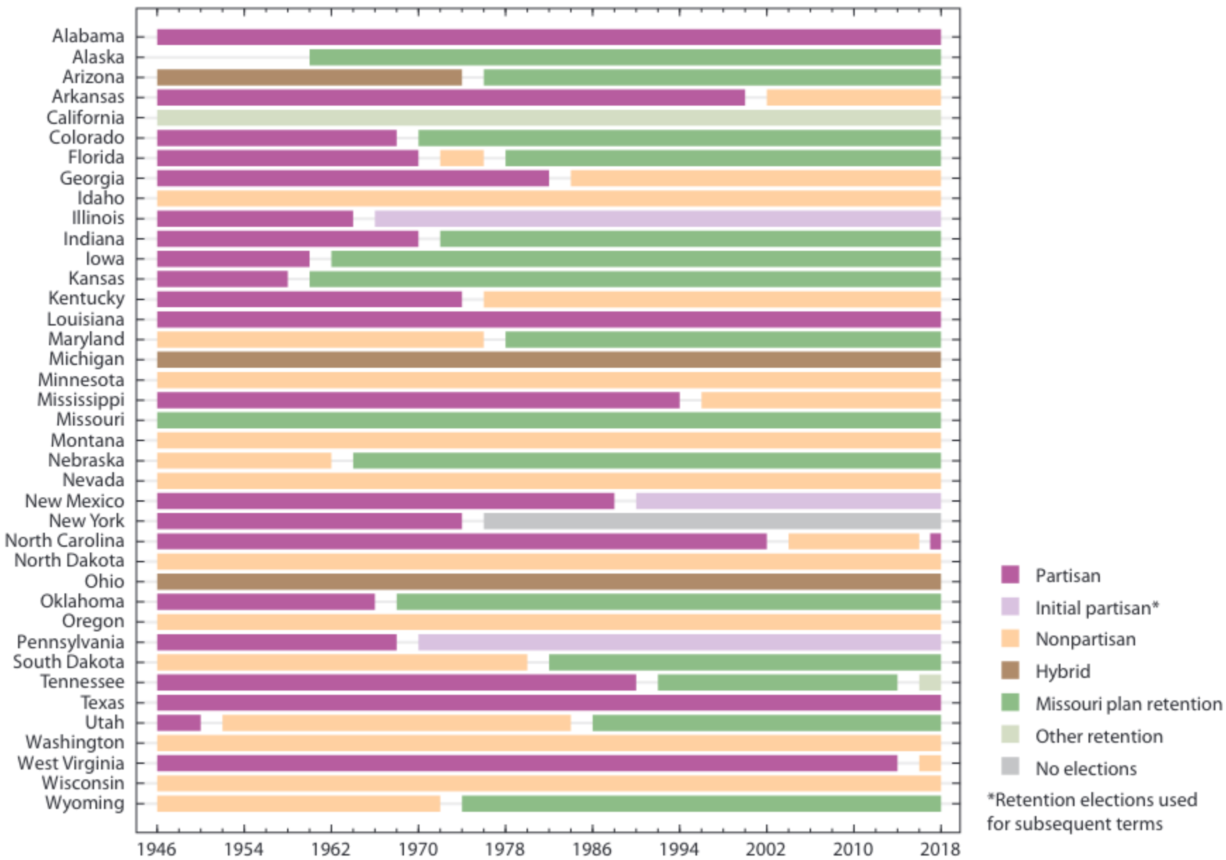
Corresponding author: Mason Reece, Ph.D. Student, MIT. mpreece@mit.edu

Introduction

Judges in the United States are normatively expected to be objective, unbiased interpreters of the law. Indeed, the Code of Conduct for United States federal judges states in Canon 5 that "A Judge Should Refrain From Political Activity." Yet, research since the 1940s has consistently challenged this expectation and shown that judges at all levels are ideological and that their ideology substantively affects their decisions (Martin and Quinn 2002; Bailey 2007; Epstein et al. 2007; Clark and Lauderdale 2010; Lauderdale and Clark 2014; Epstein, Landes, and Posner 2012; Bonica and Sen 2017). Although much research has focused on how a judge's ideology affects their decisions, comparatively little has focused on what shapes their ideology. One notable exception is Ash, Chen, and Naidu (2022), who show that judges who attended a specific economics course reflected their training in the choices and language they made in subsequent decisions. I focus instead on a separate institutional influence – the selection process for judges. At the federal level judges are appointed via a nomination and confirmation process that must go through the US Senate. However, the way that state level judges arrive to the bench varies widely. Figure 1, reproduced from Hasen (2019), shows variation between states and over time in how states elect their Supreme Court justices. Not only is there considerable variation across states, but these methods are constantly in flux even within states. This variation raises the obvious question – how does the selection process influence the court?

Previous scholars have approached the responsiveness of judges to institutional features from a variety of angles. Broadly, Choi, Gulati, and Posner (2010) show that how judges are selected influences the judges produced. For example, one focus has been on whether judges who are appointed are more independent than judges who are elected. Several studies conclude that they are (Cann 2007; Shepherd 2009; La Porta et al. 2004), but one prominent study finds that they may not be (Choi, Gulati, and Posner 2010). With regard to jurisdictions with retention elections, at least one study, Canes-Wrone, Clark, and Park (2012), finds that they lessen judicial independence. Research has also linked

Figure 1. Judicial Selection Method Over Time, by State



Note: Figure is reprinted from Hasen (2019), which uses data from Kritzer (2015). On the x-axis, time in years is represented. On the y-axis, each state in the United States is represented. The bars indicate the type of judicial selection method used in each state at the time indicated. There are gaps between some bars because the timelines for judicial selection are not continuous, so when a procedure changes there is a gap before the new procedure is adopted.

variation in formal judicial selection mechanisms with differences in quality, but these findings are mixed. Some studies document that elected judges are more productive than appointed judges (Choi, Gulati, and Posner 2010), while others find no relationship (Landes and Posner 1980). In a long-term analysis, Berkowitz and Clay (2006) show that states settled initially by civil-law countries are more likely to have partisan elections, which result in lower-quality judges. Bonica and Sen (2017) estimate judge's ideology in relation to the ideology of the lawyer pool and the ideology of politicians, conditional on the selection mechanism. They find that in appointment and partisan systems, there is little to no relationship between lawyer and judicial ideology, but that there is a significant and substantive relationship with politician's ideology. In merit and nonpartisan elections, they find that these relationships are reversed.

There are two pieces of research directly related to the relationship between judicial ideology and selection methods. In the online appendix of Windett, Harden, and Hall (2015), the authors estimate an OLS regression between a judge's estimated ideology at time $t-1$ and at time t. They condition the relationship by the public mood at $t-1$ and an interaction between two variables – whether it is the year before an election and whether it is a year before the judge is forced into mandatory retirement. The only significant result is that judges in the year before their mandatory retirement are more conservative in year t than in year $t-1$. I think there are two ways this research can be improved. First, Windett, Harden, and Hall (2015) do not use fine-grained temporal measures of judicial ideology. Research on other elites (like members of Congress) shows that representatives are temporally aware and will ramp up their electoral behavior close to the time of an election, not necessarily just in the full year beforehand. Especially since elections are typically held in November, using the decisions from the previous year obscures important variation that might occur closer to the election. The second improvement is related to the measurement of judicial ideology. Judges who are appointed or elected will often need to communicate in more subtle ways to other elites their positions on topics. Although judges do make rulings, they are somewhat limited by the exact nature of the cases that appear before them. I theorize that instead, judges can credibly signal their positions on a larger range of topics through the content of their opinions they write about those decisions. Automated text analysis of judicial opinions is a promising field (Oldfather, Bockhorst, and Dimmer 2012) but has been limited by the lack of large-scale opinion data. Hausladen, Schubert, and Ash (2020) and Lauderdale and Clark (2016) are two examples of research that has used the text of judicial opinions. Hausladen, Schubert, and Ash (2020) only focuses on creating a classification algorithm for whether the case is conservative/liberal, while Lauderdale and Clark (2016) is focused on predicting the internal decision-making of judges who serve *en banc*. My research would contribute to the nascent textual analysis of judicial opinions literature and would add important nuance to our understanding of the relationship between judicial ideology and institutional features of judicial selection. In the next section, I clearly set out my expectations for what results I expect to produce with this research.

Theoretical Expectations

There is a long history of research showing that other types of policymakers are responsive to institutional features. Mayhew (1974) and Ansolabehere, Snyder, and Stewart III (2001) both demonstrate how Members of Congress are responsive to electoral considerations in competitive elections and the timing of elections. Voters and interest groups themselves are also responsive to the timing of elections (Anzia 2013). Similarly, presidents are responsive to public opinion when reelection is imminent or when the topic is particularly relevant (Canes-Wrone and Shotts 2004). Even when policymakers themselves are not elected, they can be responsive to public mood because they are selected by elected officials (Stimson, Mackuen, and Erikson 1995). This research prompts several hypotheses about how judges would respond to electoral dynamics.

First, I hypothesize that judges who are elected will be responsive to the timing of their elections, but only when those elections are competitive. By responsive, I mean that they will moderate their opinions to more closely match the relevant groups at hand. There are two important notes to clarify this hypothesis. I do not expect that the average voter will be highly attuned to judicial elections nor especially to the opinions that judges decide. However, I do expect that interest groups are highly responsive to judicial decision-making and will modify their support of the candidate depending on the decision and the wording of the specific decision. In particular, I argue that not only is the exact decision relevant to businesses, but the text of the opinion that a judge writes communicates important nuance about the decision and can both signal to interested groups how a judge would rule on specific adjudications of this law for a group and how the judge might rule on future cases as well. These interested groups pay close attention in low-turnout elections and are often the ones mobilizing core constituencies of voters who care about certain issues (Campbell 2012).

Second, I hypothesize that there are important differences in how judges respond depending on the election mechanism that gets them on the bench. When judges are selected via partisan elections, I expect the strongest response to relevant electoral groups. When judges are selected via nonpartisan elections, I still expect a response but not as strong of an effect. The difference is because I expect that parties are the most intense "interest group" in judicial elections, and they are interested parties who will closely evaluate all of a judge's opinions and decide whether to support them or not in future elections – and then communicate this information to their base of voters. In nonpartisan elections, some party apparatus may be present but it will be limited, and other interest groups will play a relatively larger role in the evaluation of judges. In systems where judges are appointed by the governor or legislature, I would expect that judges are still responsive to elections, but would respond more closely to the party organization and individual members (either the governor or the pivotal members of the legislature) of the nominating and appointing bodies.

One concern with the theory is that judges do not have perfect control over what cases appear on their docket. However, especially at the state court of last resort, judges often have the option to pick and choose the cases they adjudicate. In addition, judges can negotiate among themselves who writes the opinion of the case, an additional way for judges to individually communicate their beliefs to relevant outside groups.

Estimating Judicial Partisanship

A significant amount of research has already been dedicated to the methodological challenge of estimating judicial ideology (for an overview, see Bonica and Sen 2021). Original methods for estimating judicial ideology mostly relied on manual classification of decisions into either a liberal or conservative direction, then proceeding with qualitative or simple quantitative evaluation. This approach is unfeasible for state courts of last resort, since thousands of decisions are published each year. Therefore, I only focus on unsupervised methods to learn a latent dimension underlying the decisions. I choose to highlight two unsupervised approaches to estimate state supreme court ideology here, although there are many others, before proposing a novel unsupervised machine learning method for estimating ideology using judicial opinions.

First, the prevailing unsupervised method around the turn of the century for estimating judicial ideology relied solely on the votes of justices. This method, pioneered by Martin and Quinn (2002) and extended in Lauderdale and Clark (2012), uses the distribution of votes on cases, which were Yes/No votes, to estimate a binary item-response theory model for each justice on the Supreme Court. The model was later extended to include a dynamic component. Using novel data, Windett, Harden, and Hall (2015) use this method to model ideology for state supreme court justices.

Second, some authors have expanded beyond the simple decisions and have instead focused on the network of positive citations between judicial opinions (Clark and Lauderdale 2010; Schmid, Chen, and Desmarais 2022). This research relies on the assumption that cases which positively cite other cases must be related on some latent dimension. The resulting network structure, which increases the number of observations available for each case (since each case cites many other cases), provides more precise estimates of decisions in judicial cases.

Third, and perhaps the most widely used metric, are the DIME scores created by Bonica (2014) and applied to state supreme court justices by Bonica and Woodruff (2015). Indeed, Windett, Harden, and Hall (2015) project their estimated scores onto a common-space defined by the DIME scores, which are themselves projected into the space defined by the famous DW-NOMINATE scores for members of Congress (Poole and Rosenthal 1985). DIME scores take advantage of campaign finance records to estimate the relative ideological projection of justices. A positive of DIME scores is that they are used to simultaneously scale a variety of political actors, and help make comparison between different policymakers possible. However, there are several problems with the DIME scores that make them challenging to use specifically for state supreme court justices. First, the underlying database that DIME relies on for campaign finance information for state supreme court justices begins in 1990.[1] Second, some judges do not make nor receive campaign contributions, especially in non-election states. Bonica and Woodruff (2015) get around this issue by creating scores with a stepwise procedure. First, "if a justice ran for election, [they] assign an ideal point based on her CFscore as a candidate. If a justice has not run for judicial office, [they] look to whether she campaigned for a different elected office during her political career." (477). Next, they search the database to see if a judge has contributed to campaigns themselves. Third, if the justice has been appointed, they assign "a score based on the CFscores of the appointing governor or legislative body" (478). Only 31% of judges are assigned ideal points based on their candidacy, which is

1. See https://www.followthemoney.org for more details

the most reliable method. Bonica and Woodruff (2015) suggest in the conclusion of their paper that a better method for estimation should collect ”data on state judicial decisions in recent decades for all 52 state Supreme Courts” (494).

I take the step of collecting judicial opinions, and in the next section I explain my method for unsupervised estimation of ideology using judicial decisions. I do not see my approach as replacing the other methods discussed, since modeling the text of the opinions may omit important decision, citation, or campaign dynamics in the measurement of judicial behavior. Instead, I consider my method a complement to these other approaches, and future work should seek to carefully combine together their strengths.

Methodology

Judicial opinions have been used before for supervised learning problems. Ash, Chen, and Naidu (2022) show that when judges attend an influential economics course, their opinions reflect that training. Their method differs than one I would use since they focus on comparison between a ground truth (the content of the course) and classifications of judicial decisions themselves. Similarly, Hausladen, Schubert, and Ash (2020) classify US Supreme Court judicial decisions into either the ”liberal” or ”conservative” direction, based on expert classifications of a small corpus of texts. Vatsal, Meyers, and Ortega (2023) develop a more complex classification model for US Supreme Court opinions that is able to produce more fine-grained classifications with high accuracy. Although there are no examples of using judicial opinions for unsupervised learning, other types of text have been used for this purpose. Herrmann and Döring (2023) scale over 2000 political parties to the same latent dimension based on the text tags assigned to each party on Wikipedia. Lauderdale and Herzog (2016) use an unsupervised estimator to identify multiple latent dimensions of the Irish Dáil and US Senate based on the content of floor speeches. Finally, Fagni and Cresci (2022) use social media posts and an unsupervised deep learning pipeline to estimate the political party of Italian Twitter users. Fagni and Cresci (2022) derive the centroids of each political party based on official party tweets, then classify individual users by classifying their tweets into each party's cluster using the model trained on the party tweets.

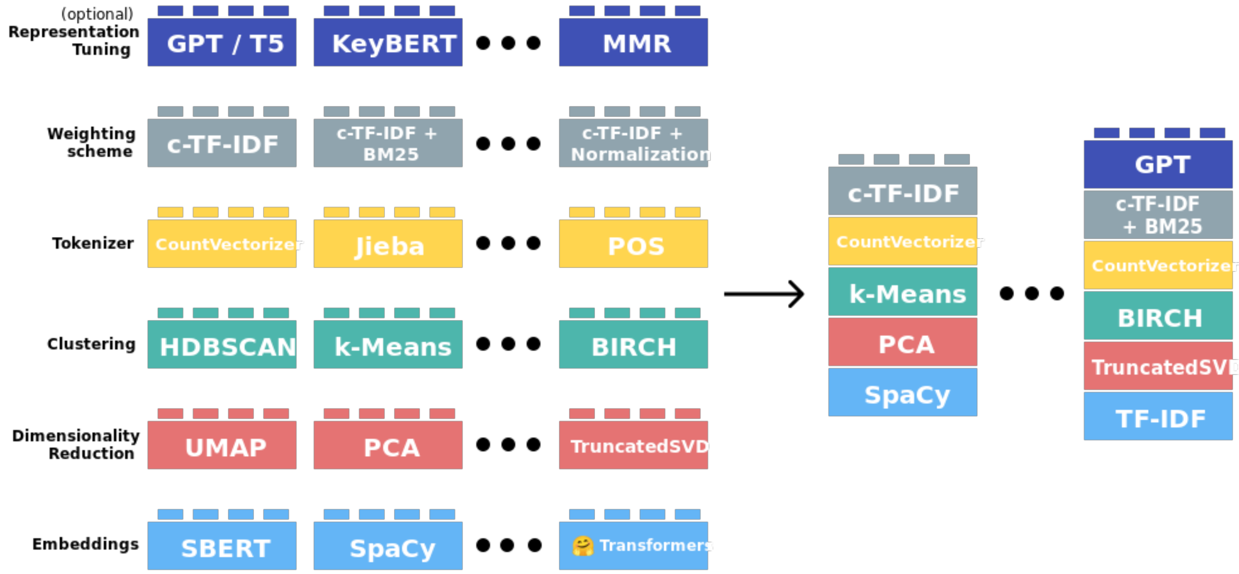
Supervised and unsupervised methods have different advantages and disadvantages, and I choose to combine them together into a semi-supervised model. Supervised methods, such as classification, have the advantage of being easily interpretable and are clearly defined by the researcher. However, they rely either on onerous expert classification (which is not error-proof) or force specific distinctions between data that are based in assumptions and may not play out in the data itself (such as classification between only liberal and conservative). Unsupervised methods can be run on data without a cumbersome pre-analysis process and rely on the data to determine the uncovered structure. However, they are often challenging to understand and tune to outcomes of interest for a researcher. The semi-supervised method I propose allows me, the researcher, to provide "suggestions" that nudge an unsupervised model towards outcomes of interest that I am focused on, while still giving it the flexibility to discover unknown structures in the data that I could not have foreseen.

The semi-supervised method I rely on is called BERTopic (Grootendorst 2022). BERTopic builds on a long history of text analysis[2] and is most closely related to the field of topic modeling. BERTopic is a framework for analysis that uses a series of building blocks that I can manipulate depending on my needs. I reproduce an example of the modularity of BERTopic in Figure 2, which demonstrates how the framework can be used with a variety of algorithms for each of the steps. Next, I provide more detail on what each step of the process does and which algorithm I have chosen to use.

First, the embeddings step uses a pre-trained model to convert a document to a

2. For an excellent overview of the historical trajectory of text analysis, including the build-up to the methods I show here, see Tsirmpas et al. (2024)

Figure 2. Modularity of BERTopic



Note: This figure showcases the modularity of BERTopic (Grootendorst 2022). On the left, text shows the generic name of each step in the process, whereas each building block on the right are samples of different algorithms or methods that can be used for each of the building blocks. Together, these building blocks come together in a unified software interface to estimate topics from documents.

numerical representation (Ash and Hansen 2023). Embeddings are useful since they project documents into a lower-dimensional space based on the context around words in their training set (which can be fine-tuned for specific use cases) and do a better job of maintaining important structures in documents than a typical bag-of-words approach. I use a model trained on legal documents to generate word embeddings, since I anticipate this type of model will keep the most important relationships in court opinions (Niklaus et al. 2023). The definition of a "document" is an important choice that is influenced by both topical knowledge of the area and limitations imposed by the embeddings model. I split each judicial opinion into paragraphs, which I believe capture entire sentiments and topics of a legal opinion, while still being within the limits of the embedding model. Although there are some models being developed that can handle long texts, even the latest models are not capable of processing an entire judicial opinion (Tsirmpas et al. 2024).

Second, the BERTopic framework takes the large matrix of embeddings and reduces the dimensionality of this resulting object since the third step, clustering, performs poorly using high-dimension data. I use the UMAP algorithm, which has good properties for maintaining both global and local structure in its resulting reduction (McInnes, Healy, and Melville 2018). I do not impose any limitations on UMAP, and let it reduce the dimensions to an appropriate level rather than requesting a specific dimensionality reduction. Any improvement is helpful for the clustering algorithm.

Third, after reducing the dimensionality, I cluster the objects into semantically similar documents. To do this, I use the HDBSCAN clustering algorithm, which has two convenient properties (Campello, Moulavi, and Sander 2013). One, it can find clusters of different shapes and structures, which is desirable since judicial opinions vary widely in terms of content and tone. Second, it does not force documents into clusters and will also identify a series of outliers, which is especially helpful since some content of judicial opinions will be nuisance words or unrelated to other documents due to their hyper-specificity (an example of this, perhaps, are names of the participants in a particular case). Once the clusters are estimated, each of the relevant ”topics” have been defined in an abstract sense.

Fourth, once the clusters have been established, the framework seeks to generate semantically useful terms, phrases, or descriptions that help a researcher understand what is contained in each cluster. This begins with a first step, which is based on the bag-of-words technique. This step takes every word in all of the documents in the cluster, then counts them up and generates a frequency matrix for each cluster. Then, the framework creates a TF-IDF matrix where each cluster is treated as a document, referred to as c-TF-IDF. In this step, I additionally follow the recommendations of the BERTopic maintainers and remove stopwords from the resulting TF-IDF matrix, since they may contaminate the resulting topic representations. Words that score more highly on the c-TF-IDF matrix for each cluster are then representative of ”topics” in each of the resulting matrix. Optionally, the BERTopic framework also allows users to pass the resulting matrix of words off to another model (for example, ChatGPT) to fine-tune the words into a cohesive description, but I do not take this step.

After the topic models are estimated, they can then be made dynamic by estimating the frequency of each cluster by year. I could force the model to cluster topics within year to enforce a stronger sense of dynamism, but long court timelines and the inability of judges to determine exactly what they hear each year leads me to apply the dynamic modeling in post-processing. An additional argument for imposing a dynamic model on the data could be that the semantic importance of certain words or phrases has changed over time. In general, this is an important concern, but there are two reasons I don't consider this an applicable issue to my data. First, the scope of my data only contains data from 1994-2024, which is not enough time for their to be substantial differences in semantics. Second, and more importantly, legal writing and styles of argumentation are particularly static and judges and lawyers regularly cite and use language from historical documents, preventing me from clearly placing each document in the "contemporary present" since much of the language comes from other time periods. Taken together, these reasons lead me to be relatively unconcerned about the choice to estimate topics over time in post-processing, but future work could consider this question in a more systematic way than I do here.

Data

I use opinions written by judges on state Supreme Courts from 1994-2024. I supplement this information with information on the judges, including their listed partisanship (if available). My main treatment variable is the type of procedure used to choose judges, which is measured at the state level, and relies on the data in Kritzer (2015). I gather the opinion data from *CourtListener*, an automated service that scrapes opinion texts from courts all over the United States (Free Law Project 2024). I follow other research on judicial opinions (Clark and Lauderdale 2010; Bonica and Woodruff 2015) and remove opinions that are unanimous or decisions that do not have opinions associated with them.

After this filtering process, my data contains 26,076 opinions written by 492 judges, from all 50 states. I perform minimal pre-processing of the opinions themselves, only removing citations and web links. This ensures that BERTopic can make use of the full context of each opinion.

Results

Topic Estimation

Using the BERTopic framework, I estimate two models. First, I estimate an entirely unsupervised model using all judicial opinions. Second, I estimate a semi-supervised model, wherein for each judge who has a formal party affiliation, I assign them a ”topic” based on their party. This should at least ensure that judges who we know to be of the same party are nudged to also share topical similarity. Of the 492 judges in the data, only 49 have a given party affiliation, and together they only write about 4% of the documents in the data. The lack of data on judge’s given party affiliation is a sign that the semi-supervised approach will not add much to the topical representation, but it should at least help nudge the model towards a substantively useful model for my purposes.

In Table 1 I present the 10 most frequent topics for the opinion data using the fully unsupervised model, sorted by the frequency they appear in the data. The identified topics do not clearly fall along an ideological spectrum, and instead seem to focus more closely on substantive legal questions (especially rank 1, 4, 5, 7, 9, and 10). In addition, even though the modeling framework removes common stopwords that could have contaminated the model, they still appear to be present, as can be seen clearly in rank 2 and 8. These results are not suggestive of a common dimension underlying the opinion data.

I follow up this analysis by applying the semi-supervised model. In Table 2 I again present the 10 most frequent topics. A similar story to the fully unsupervised method

Table 1. Top-10 Topic Representations from the Unsupervised BERTopic Model

Rank	Top Words
1	defendant capped, questionnaire waiver, alford, nolo, no contest, colloquies, ngi, contendere, guilty plea, abeyance
2	opinion, court, thomas, point, of, the
3	yourself, youre, gonna, youve, youll, munck, arey, guys, youd, okay
4	post conviction relief, successive, post disposition, post commitment, bobos, pippitt, prior conviction, subclaim, post conviction post conviction, carridines
5	arbitrability, arbitrators, arbitrator, arbitrable, arbitral, arbitrate, nonsignatory, arbitrations, arbitra, nonsignatories
6	cross moved, cross motion, cross motions, mckennas, dismissal summary, gatsos, reasners, riptas, countermotion, unresisted
7	akerman, ufferman, brannock, kruidenier, tietmeyer, coral, shur, lauderdale, miami, gables
8	lawful, we
9	ethnicity, racial, ethnic, racially, retaliation, discriminate, profiling, gender, creed, ancestry
10	northpointe, siga, mrl, bta, pharmathene, bza, compacts, finag, upmc, evenflo

Note: This table includes the top-10 estimated topics from judicial opinion data using BERTopic, with the fully unsupervised parameters. The left column is the topic number, sorted by the frequency of the topic in the data. On the right are the top words associated with each model.

arises, where most topics are concerned with substantive legal questions, rather than an underlying dimension of ideology. Even with the nudging imposed by the semi-supervised method, the model still does not clearly arise on an ideological slant for any of the topics.

Table 2. Top-10 Topic Representations from the Semi-Supervised BERTopic Model

Rank	Top Words
1	questionnaire waiver, defendant capped, alford, ngi, no contest, colloquies, abeyance, bargains, colloquy, guilty plea
2	policyholder, claims made, insurers, policyholders, self insurance, insureds, insuring, exclusions, nutts, coverages
3	opinion, impression, ultimate, court, conclusions, of, the, this, for
4	arbitrability, arbitrators, arbitrator, arbitrate, arbitrable, arbitral, aaa, nonsignatory, arbitra, arbitrations
5	impaneled, reseating, deliberating, biased, inattentive, alternates, crislip, empaneled, excusing, sleeping
6	extended term, twenty five year, indeterminate, natural life, determinate, pronounced, commuted, mispronouncement, nonenhanced, upward
7	prenuptial, palimony, passably, surrogacy, trasters, defenses such, forbearance, understandings, lopsidedness, rewrite
8	ssdi, annuities, accidental disability, ttd, pension, disposable, pensions, pensioners, retirees, retiree
9	siga, bta, gtla, mrl, dnrc, pharmathene, bza, compacts, westinghouse, wcc
10	post appellate, awardable, first half, prevailing party, awards, attor, recalculation, actualized, fee shifting, they unlike

Note: This table includes the top-10 estimated topics from judicial opinion data using BERTopic, with the semi-supervised parameters. The left column is the topic number, sorted by the frequency of the topic in the data. On the right are the top words associated with each model.

Given the lack of ideological direction to any of the discovered topics with the BERTopic models, I see no reason to proceed with additional analysis in this vein until I can more clearly understand the underlying topics in these models. In some ways, my models highlighting substantive legal codes instead of a single ideological scale is similar to other research in the US on courts (Lauderdale and Clark 2014). Together, this work suggests that we should more carefully consider how to estimate ideology and responsiveness to electoral systems, beyond the typical assumption made in American politics about a single underlying dimension structuring much of our political process (Poole and Rosenthal 1985). In the next section, I propose several potential options that could help me more precisely identify latent dimensions and use them to understand the relationship with electoral systems.

Future Work

First, there are two simple steps that should increase the precision of these estimates. The opinion data provided by *CourtListener* is very clean, but does not always clearly identify the written section of the text that expresses a judge's views. After manually examining several opinions, there did not appear to be a clear way to automate the extraction of the relevant parts of an opinion without needing to follow up with substantial hand-coding. Future work would more carefully seek to extract the opinions. The second step is more concerning, which also relates to the cleanliness of the data. Some judicial opinions include both the majority opinion and dissenting opinions together in the same text, even when the text is coded by *CourtListener* as only being representative of the views of one of the authors. This type of data bleed is likely to seriously contaminate the results I've identified here, and may be why the substantive information stands out more, since it will be relatively consistent between the majority and minority opinions, while the opinions themselves will not be.

From a modeling perspective, I could take additional steps to encourage the model to more closely represent ideology. First, I could organize the model, similar to how a structural topic model operates, such that a single judge's opinions will be connected during the clustering algorithm, which would encourage the model to cluster opinions by the same author together. This might help cut across some of the substantive clustering, since a judge may communicate in a more subtly similar way across opinions, even when they are on disparate topics. Second, there is some research (Islam et al. 2016) that simultaneously infers ideal points (a form of dimension reduction) for justices on a multi-dimensional scale focused on substantive ideas at the same time as it places those judges somewhere on independent scales for each substantive issue. This approach would require a new, hand-coded, model for estimation, since the authors do not provide any replication code for estimating their quantities and I would prefer to rely on a BERTopic style framework, where possible.

Taken together, an improvement in the quality of the data and changes to the modeling framework should together improve the implementation and interpretation of these models. Assuming that I have success in the future with these modifications I would then proceed to relate those measures to my main outcomes of interest, the selection methods for each judge. If I end up needing a multi-dimensional model, perhaps over several substantive issues and ideology, then I will need to collect supplementary information on the relevant groups in each state for each of those topics (since judge's may react to specific groups regarding specific issues). Once I have collected this information, I would add to the topic modeling the dynamic component of how a judge's opinions vary over time. Then, those time series would be compared to the variation in group interests and proximity to a selection method. This proximity to selection combined with the relationship between the judge and relevant groups will become my main treatment outcome of interest.

References

Ansolabehere, Stephen, James M. Snyder, and Charles Stewart III. 2001. "The Effects of Party and Preferences on Congressional Roll-Call Voting." *Legislative Studies Quarterly* 26, no. 4 (November): 533.

Anzia, Sarah F. 2013. *Timing and Turnout: How Off-Cycle Elections Favor Organized Groups.* University of Chicago Press, December 3, 2013.

Ash, Elliott, Daniel L Chen, and Suresh Naidu. 2022. *Ideas Have Consequences The Impact of Law and Economics on American Justice.* Working Paper.

Ash, Elliott, and Stephen Hansen. 2023. "Text Algorithms in Economics." *Annual Review of Economics* 15, no. 1 (September 13, 2023): 659–688.

Bailey, Michael A. 2007. "Comparable Preference Estimates across Time and Institutions for the Court, Congress, and Presidency." *American Journal of Political Science* 51, no. 3 (July): 433–448.

Berkowitz, Daniel, and Karen Clay. 2006. "The Effect of Judicial Independence on Courts: Evidence from the American States." *The Journal of Legal Studies* 35, no. 2 (June): 399–440.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58 (2): 367–386.

Bonica, Adam, and Maya Sen. 2017. "The Politics of Selecting the Bench from the Bar: The Legal Profession and Partisan Incentives to Introduce Ideology into Judicial Selection." *The Journal of Law and Economics* 60, no. 4 (November): 559–595.

———. 2021. "Estimating Judicial Ideology." *Journal of Economic Perspectives* 35, no. 1 (February 1, 2021): 97–118.

Bonica, Adam, and Michael J. Woodruff. 2015. "A Common-Space Measure of State Supreme Court Ideology." *Journal of Law, Economics, and Organization* 31, no. 3 (August): 472–498.

Campbell, Andrea Louise. 2012. "Policy Makes Mass Politics." *Annual Review of Political Science* 15, no. 1 (June 15, 2012): 333–351.

Campello, Ricardo J. G. B., Davoud Moulavi, and Joerg Sander. 2013. "Density-Based Clustering Based on Hierarchical Density Estimates." In *Advances in Knowledge Discovery and Data Mining,* edited by Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, 160–172. Berlin, Heidelberg: Springer.

Canes-Wrone, Brandice, Tom S. Clark, and Jee-Kwang Park. 2012. "Judicial independence and retention elections." *The Journal of Law, Economics, & Organization* 28 (2): 211–234.

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.

Cann, Damon M. 2007. "Justice for sale? Campaign contributions and judicial decision-making." *State Politics & Policy Quarterly* 7 (3): 281–297.

Choi, Stephen J., G. Mitu Gulati, and Eric A. Posner. 2010. "Professionals or politicians: The uncertain empirical case for an elected rather than appointed judiciary." *The Journal of Law, Economics, & Organization* 26 (2): 290–336.

Clark, Tom S., and Benjamin Lauderdale. 2010. "Locating Supreme Court Opinions in Doctrine Space." *American Journal of Political Science* 54, no. 4 (October): 871–890.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2012. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice.* Harvard University Press, April 12, 2012.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland. 2007. "The judicial common space." *The Journal of Law, Economics, & Organization* 23 (2): 303–325.

Fagni, Tiziano, and Stefano Cresci. 2022. "Fine-Grained Prediction of Political Leaning on Social Media with Unsupervised Deep Learning." *Journal of Artificial Intelligence Research* 73 (February 18, 2022): 633–672. arXiv: 2202.12382[cs].

Free Law Project. 2024. *CourtListener,* March.

Grootendorst, Maarten. 2022. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint arXiv:2203.05794.*

Hasen, Richard L. 2019. "Polarization and the Judiciary." *Annual Review of Political Science* 22 (Volume 22, 2019 2019): 261–276.

Hausladen, Carina I., Marcel H. Schubert, and Elliott Ash. 2020. "Text classification of ideological direction in judicial opinions." *International Review of Law and Economics* 62:105903.

Herrmann, Michael, and Holger Döring. 2023. "Party Positions from Wikipedia Classifications of Party Ideology." *Political Analysis* 31, no. 1 (January): 22–41.

Islam, Mohammad, K. S. M. Hossain, Siddharth Krishnan, and Naren Ramakrishnan. 2016. "Inferring Multi-Dimensional Ideal Points for US Supreme Court Justices." *Proceedings of the AAAI Conference on Artificial Intelligence* 30, no. 1 (February 20, 2016).

Kritzer, Herbert M. 2015. *Justices on the Ballot: Continuity and Change in State Supreme Court Elections.* Cambridge University Press.

La Porta, Rafael, Florencio López-de-Silanes, Cristian Pop-Eleches, and Andrei Shleifer. 2004. "Judicial Checks and Balances." *Journal of Political Economy* 112, no. 2 (April): 445–470.

Landes, William M., and Richard A. Posner. 1980. "Legal Change, Judicial Behavior, and the Diversity Jurisdiction." *The Journal of Legal Studies* 9, no. 2 (March): 367–386.

Lauderdale, Benjamin E., and Tom S. Clark. 2012. "The Supreme Court's Many Median Justices." *American Political Science Review* 106, no. 4 (November): 847–866.

Lauderdale, Benjamin E., and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58, no. 3 (July): 754–771.

———. 2016. "Estimating Vote-Specific Preferences from Roll-Call Data Using Conditional Autoregressive Priors." *The Journal of Politics* 78, no. 4 (October): 1153–1169.

Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring political positions from legislative speech." *Political Analysis* 24 (3): 374–394.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political analysis* 10 (2): 134–153.

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

McInnes, L., J. Healy, and J. Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *ArXiv e-prints* (February). arXiv: 1802.03426 [stat.ML].

Niklaus, Joel, Veton Matoshi, Matthias Sturmer, Ilias Chalkidis, and Daniel E. Ho. 2023. "MultiLegalPile: A 689GB Multilingual Legal Corpus." *ArXiv* abs/2306.02069.

Oldfather, Chad M., Joseph P. Bockhorst, and Brian P. Dimmer. 2012. "Triangulating Judicial Responsiveness: Automated Content Analysis, Judicial Opinions, and the Methodology of Legal Scholarship." *Florida Law Review* 64 (5): 1189–1242.

Poole, Keith T., and Howard Rosenthal. 1985. "A spatial model for legislative roll call analysis." *American journal of political science,* 357–384.

Schmid, Christian S., Ted Hsuan Yun Chen, and Bruce A. Desmarais. 2022. "Generative Dynamics of Supreme Court Citations: Analysis with a New Statistical Network Model." *Political Analysis* 30, no. 4 (October): 515–534.

Shepherd, Joanna M. 2009. "The Influence of Retention Politics on Judges' Voting." *The Journal of Legal Studies* 38, no. 1 (January): 169–206.

Stimson, James A., Michael B. Mackuen, and Robert S. Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89, no. 3 (September): 543–565.

Tsirmpas, Dimitrios, Ioannis Gkionis, Georgios Th Papadopoulos, and Ioannis Mademlis. 2024. "Neural Natural Language Processing for Long Texts: A Survey on Classification and Summarization." *Engineering Applications of Artificial Intelligence* 133 (July): 108231. arXiv: 2305.16259[cs].

Vatsal, Shubham, Adam Meyers, and John E. Ortega. 2023. *Classification of US Supreme Court Cases using BERT-Based Techniques,* arXiv:2304.08649, July 24, 2023. arXiv: 2304.08649[cs].

Windett, Jason H., Jeffrey J. Harden, and Matthew E. K. Hall. 2015. "Estimating Dynamic Ideal Points for State Supreme Courts." *Political Analysis* 23 (3): 461–469.