

Precedent and Bargaining on the U.S. Supreme Court: An Empirical Evaluation

Isaac Cape*
Department of Politics
Princeton University
isaacmcape@gmail.com

Jonathan P. Kastellec
Department of Politics
Princeton University
jkastell@princeton.edu

April 13, 2026

Abstract

Which justices control the content of Supreme Court opinions? Theoretical and empirical debates persist about whether judicial policy falls at the location of the median justice, the median of the majority coalition, or the opinion author. We propose a new empirical strategy to disentangle these competing theories using the Court's negative treatments of its own precedents. We measure ideological distance from a precedent-setting court to a precedent-treating court under each theory (e.g. the distance from the median justice in *Roe* to the median justice in *Dobbs*), then evaluate which distance best predicts negative treatment of existing precedent. The results show that median-based theories substantially outperform author-based theories in predicting negative treatments; furthermore, the bulk of evidence places the locus of judicial power at the median of the majority coalition rather than the Court median. These findings have important implications for understanding the ideological trajectory of the current conservative super-majority.

*The initial research for this paper was originally conducted by Cape (with Kastellec as his adviser) for his 2025 Princeton University senior thesis, "[The Evolution of Precedent and Partisanship on the Roberts Court.](#)" We thank James Spriggs for sharing his data with us. We also thank Ethan Bueno de Mesquita, Tom Clark, Eitan Sapiro-Gheiler, and Anthony Taboni for helpful comments and suggestions.

1 Introduction

Who controls the content of Supreme Court opinions? This question lies at the heart of longstanding debates in judicial politics about bargaining, influence, and the institutional locus of power on the Supreme Court in particular and collegial courts more generally. As with any multimember body, answering this question requires both the careful delineation of competing theories of bargaining on the Court and empirical strategies capable of adjudicating among them. Under a straightforward median-voter logic, judicial outcomes reflect the preferences of the Court’s median justice. Alternative accounts emphasize the median of the majority coalition, reflecting the institutional reality that justices vote not only on case dispositions but also on the rules and rationales that give opinions precedential force. Finally, other theories highlight the role of opinion authors, arguing that authors exercise disproportionate control over doctrinal content through their agenda-setting powers. Yet despite many innovative tests (reviewed in Section 2), considerable uncertainty remains about which family of theories best describes substantive decisionmaking on the Supreme Court.

In this paper, we offer an empirical contribution to this debate by leveraging the Court’s treatment of its own precedents as a window into judicial bargaining. Specifically, we use the Court’s *negative treatments* of its own precedents—such as limiting, distinguishing, or overruling prior decisions—as a proxy for policy disagreement. For example, when the post-1937 New Deal Court weakened or overruled *Lochner*-era precedents, it signaled a shift toward greater acceptance of government intervention in the economy. Similarly, the Roberts Court’s negative treatment of a number of liberal precedents in recent years—most notably the overruling of *Roe v. Wade* in *Dobbs v. Jackson Women’s Health Organization*—reflects a conservative shift away from the doctrines of earlier Courts.¹

The basic logic of our tests is straightforward. We begin by assuming that an opinion establishing a precedent—for example, *Roe*—can be characterized as having been written at the ideological location implied by each bargaining model—we call these “origin” opinions.

¹*Roe v. Wade*, 410 U.S. 113 (1973); *Dobbs v. Jackson Women’s Health Organization*, 597 U.S. 215 (2022).

We then measure the ideological “distance” between that location and the estimated location of the Court at the time the precedent is later treated—for example, in *Dobbs*—which we call “progeny” opinions. Crucially, for each origin–progeny pair, the distance measure is allowed to vary across the relevant points of comparisons between origin and progeny cases—e.g. the distance from the median justice in *Roe* to the median justice in *Dobbs*. We then empirically evaluate the question of which of these distances best explains variation in the likelihood that a future Court will negatively cite an existing precedent.

To implement this strategy, we assemble and extend existing data on Shepard’s treatments of Supreme Court majority opinions from 1953 through 2023. We operationalize ideological distance using dynamic ideal point estimates and evaluate competing theories using two complementary data structures: a term-level panel that tracks whether a precedent is negatively treated in a given term, and a citation-level dataset that examines the probability of a negative citation, conditional on any citation. We then use regression-based analyses and out-of-sample validation to assess both the substantive predictive effects of ideological distance and the comparative predictive performance of the competing models.

Across specifications and datasets, our results point to a consistent conclusion. Measures based on Court medians—especially the median of the majority coalition—substantially outperform author-based measures in predicting negative treatment of precedent; indeed, the latter perform quite poorly, both substantively and in terms of predictive accuracy. Overall, distance from the median of the majority coalition provides the strongest and most consistent explanation for when the Court chooses to move away from its prior decisions.

These results contribute to our understanding of bargaining and power on the Supreme Court in general. But they also have important implications for understanding the current Supreme Court in particular. The emergence of a 6-3 conservative majority following the 2020 appointment of Amy Coney Barrett has raised the question of whether power (on average) resides more in the “centrist” bloc of Barrett, Chief Justice Roberts, and Justice Kavanaugh, or whether the more conservative views of Justices Thomas, Alito, and Gorsuch

are likely to prevail (see e.g. Howe 2021, Ward 2022, Re 2025). Our results—which build on earlier research finding support for median-of-the-majority theories (Clark and Lauderdale 2010, Carrubba et al. 2012)—suggest that power is more likely to reside toward the conservative wing of the Court rather than at the middle, which has important implications for the Court’s future ideological decisionmaking and trajectory.

2 Bargaining on the Supreme Court: Theory and Evidence

Dating back at least to Murphy (1964), scholars of judicial politics have long sought to pin down the relationship between bargaining among the justices and how that corresponds to the Court’s outputs. However, the last two decades have witnessed the crystallization of specific theories about the nature of bargaining. Because this paper makes an empirical contribution rather than advancing the theoretical literature, our summary of the three core models is fairly brief (see Clark (2012), Lax and Rader (2015), and Cameron and Kornhauser (2017) for thorough reviews).²

2.1 Median justice theory

The most straightforward theory of bargaining simply imports the logic of the median voter theorem to the Supreme Court. Under the theory—formalized in Hammond, Bonneau and Sheehan (2005) and Jacobi (2009)—if we make the standard assumptions of a unidimensional policy space with single-peaked preferences, then we get the prediction that the preferences of the “median justice” should be decisive in determining both case dispositions and opinions on the Court.

Median justice theory has the twin theoretical virtues of simplicity and elegance. It also generally accords with popular qualitative understandings of the importance of the “swing justice” on the Court. For example, following the retirement of Justice Sandra Day O’Connor

²The formal theories discussed below have built upon the canonical informal rational choice theories of bargaining presented in Epstein and Knight (1997) and Maltzman, Spriggs and Wahlbeck (2000).

in 2005, it was clear that in many cases, Justice Kennedy was the median justice. In the Court’s 2006 term, Kennedy was in the minority in *two* cases out of the 81 merits decisions the Court decided that term. And, in the 24 cases that term decided by either a 5-to-4 or 5-to-3 vote, he was in the majority in all 24 (Harrow 2007).

As these statistics show, it certainly makes sense that the median justice should be pivotal when it comes to *dispositions*—that is, which side wins or loses the case. But once we move beyond that outcome, the predictions of the median justice theory quickly fall short against the reality of bargaining on the Court. In particular, because the median justice alone is decisive, this means the theory predicts that *none* of the following matter at all:³

- Whether the vote on the case disposition is liberal or conservative does not affect the content of the majority opinion;
- The composition of the dispositional majority likewise has no effect on opinion content—in other words, a unanimous vote, a 7–2 vote, and a 5–4 vote all yield the same opinion;
- Opinion authorship does not matter—the content of the majority opinion is the same whether it is written by Justice Alito or Justice Kagan;
- Case selection—in particular, the substantive characteristics of the case—does not affect the content of the majority opinion;
- Finally, changes in the composition of the Court matter only insofar as they “move the median,” which is a relatively rare occurrence (Cottrell, Shipan and Anderson 2019, Cameron and Kastellec 2016, Krehbiel 2007).

Because of these limitations, scholars have turned to theories that seek to generate more realistic predictions, especially with respect to opinion content and locations.

2.2 Author-based theories

One of the starkest limitations of median justice theory, as noted above, is that it offers no role for the influence of opinion authors. This prediction runs directly against the widespread belief among legal scholars and observers that, at least in some cases, Supreme Court opinions reflect the preferences (and tastes) of the opinion author, above and beyond what would

³This list is drawn from Cameron and Kastellec (2023, 322-3).

be predicted by the joint preferences of the author and the other justices in the majority coalition. For example, reflecting upon the legacy of Chief Justice Earl Warren, Justice Abe Fortas wrote:

“When the Chief Justice is in the majority, he assigns the writing of the opinion of the Court in particular cases. This is a function of great significance. If the Chief Justice assigns the writing of the opinion of the Court to [Justice A], a statement of profound consequence may emerge. If he assigns it to [Justice B], the opinion of the Court may be of limited consequence” (Fortas 1975, 405).

Given this perception, scholars have sought to incorporate a role for opinion authors.

Following Clark (2012) and Lax and Rader (2015), it is useful to distinguish between “author monopoly” and “author influence” theories. Under the notion of author monopoly, an opinion author is able to achieve exactly his or her preferred opinion location regardless of the preferences of the other justices—even those in the majority coalition. This is obviously a strong prediction, and to the best of our knowledge, there is no existing theory where such a equilibrium exists. The closest thing to a true author monopoly theory is the take-it-or-leave-it model of bargaining (i.e., a Romer-Rosenthal model (Romer and Rosenthal 1978)) presented in Hammond, Bonneau and Sheehan (2005); depending on the location of the status quo and the median justice, in some cases the opinion author can achieve exactly their ideal point. But even that setup assumes too much power for authors, as it prohibits other justices from writing a “counter opinion” that the median justice might prefer. At the same time, the predictions from an author monopoly theory are very precise: opinion locations should systematically reflect author ideal points, even when those differ from the median.

Because of the very strong assumptions that would have to underlie an author monopoly approach, frontier models that incorporate a role for authors have taken the author influence approach. In an early such effort, Schwartz (1992) makes the strong assumption that, in any case, the Court can choose among only two exogenous policy choices, along with an endogenous level of precedent that an opinion contains; under these conditions, authors

retain influence over the final opinion location.

Relaxing many of these strong assumptions, the formal theory in Lax and Cameron (2007) provides authors with a first-mover advantage, under which they can invest costly effort to produce high quality opinions, thereby preventing counter-opinions that would attract the support of the median justice. In equilibrium, the final opinion tends to fall ideologically somewhere between the ideal point of the author and the median justice.

Finally, Parameswaran, Cameron and Kornhauser (2021) develop a model in which author influence arises from formal proposal power within a structured bargaining game. In their framework, once a dispositional majority forms, the opinion author is a recognized proposer who selects a doctrinal rule, subject to acceptance by pivotal members of the majority coalition. Because coalition members can reject proposals that fall outside their acceptance sets, the author cannot impose his or her ideal point unilaterally; instead, equilibrium opinion content reflects the most favorable policy that satisfies the constraints imposed by the coalition. As a result, opinion outcomes need not coincide with the preferences of either the Court median or the coalition median, but instead reflect the interaction between author preferences and majority bargaining constraints. In particular, the author’s influence is a function of the justices’ discount rates and how they affect the cost of delay in the larger bargaining game.

All in all, while the mechanics differ across these models, in each authors retain some agenda-setting advantage that privileges them over the other justices, thereby generating some non-median policy outcomes.

2.3 Median-of-the-majority coalition theories

The final family of theories—median-of-majority coalition theory—builds upon the institutional reality that the justices vote on two things in every merits case: which side should win or lose (the disposition), and the rule or policy location that the final opinion endorses. Importantly, the rule announced by the Court must be consistent with the disposition—i.e., the rule announced cannot imply that the opposite disposition should be reached in

the instant case—a property Cameron and Kornhauser (2017) label as “dispositional consistency.” Finally, unlike in legislative voting, where legislators are constrained to vote “yea” or “nay,” judges and justices can either write or join concurrences (as well as dissents), which complicates equilibrium outcomes in the production of final outcomes.

The first formal theory to emphasize the rule-disposition distinction and how it may produce non-median outcomes was Carrubba et al. (2012).⁴ In their model, if one assumes there are cases where the median justice will not strategically vote against their preferred dispositional outcome, that choice means they cannot credibly threaten to defect to the minority dispositional coalition; this reduces the median justice’s bargaining power, thereby moving the opinion location *away* from the overall median and *towards* the median of the majority coalition. Parameswaran, Cameron and Kornhauser (2021) build upon Carrubba et al. (2012) by allowing for opinion authors and formalizing the bargaining protocol more precisely. In their model, both case locations and the degree to which the justices care about the disposition affect the extent to which the final opinion should be located closer to the overall median or the center of the majority coalition. In addition, the degree of bargaining intensity determines in equilibrium whether the opinion is located closer to the center of the majority coalition or the author. Higher discount rates—i.e., higher “bargaining intensity”—privilege the median of the majority coalition, while lower discount rates favor the author. These models therefore imply that equilibrium doctrine may systematically deviate from the overall Court median, even in the absence of strong assumptions about author influence.

2.4 Existing empirical tests

Disentangling the three families of theories is tricky, as it requires moving beyond the usual outcome measure of dispositional votes—i.e. the typical dependent variable in a study that uses the Supreme Court Database (Spaeth et al. 2024)—and instead measuring either opinion locations directly or relying on indirect tests. Despite this difficulty, a number of

⁴While it was not the main focus of Schwartz (1992), the equilibrium in that model also points to a median-of-the-majority result; see pp. 240-41.

scholars in recent years have undertaken this endeavor, which we summarize in this subsection (see Table 1 for an overview of the designs and key findings in this literature).

While the research designs differ in both approach and general conclusions, a striking pattern is that they all tend to reject median justice theory as a strong predictor of opinion locations. Overall, the evidence consistently suggests that the unusual structure of voting on the Court—i.e., the combination of dispositions, setting rules through policy, and the availability of concurrences—helps to drive outcomes.

The overall evidence, however, is less conclusive as to whether author-based theory or median-of-the-majority coalition theories better predict opinion locations on the Court. Clark and Lauderdale (2010), Carrubba et al. (2012), and Cameron and Kastellec (2023, 328-32) find stronger support for median-majority theory compared to either median justice theory or author-based theories, while Cameron, Park and Beim (2009), Lax and Rader (2015), and Lauderdale and Clark (2016) find more support for author-based theories.

While these studies are all valuable, a notable feature of most of these empirical tests is they are either indirect or limited in important ways. On the indirect side, Carrubba et al. (2012) study concurrence behavior to infer opinion location, while Cameron, Park and Beim (2009) and Lauderdale and Clark (2016) examine joins of majority opinions to do the same. On the limited side, Lax and Rader (2015) use voting fluidity—i.e. changes between the justices’ conference votes and the final votes—to adjudicate between the three theories. While clever, voting fluidity is both relatively rare and is not easily explained by existing theoretical models.⁵

The one exception to these limitations is the empirical tests employed in Clark and Lauderdale (2010), who use a model of citations to develop estimates of opinion locations and justices’ ideal points on the same scale. This allows them to directly test which theory best predicts opinion location. Their data, however, comprise fewer than 600 combined search-and-seizure and freedom-of-religion cases, leaving uncertain how well their results

⁵Specifically, no existing formal theory predicts that bargaining should go past the “first round,” so in some sense any fluidity is off-the-equilibrium path.

Study	Type of Evidence	Author Model	Court Median	MMC
Cameron, Park & Beim (2009)	Stewart → O'Connor case study; join behavior	Supported (most favorable to author influence)	Not supported	Mixed/limited
Clark & Lauderdale (2010)	Citation-based doctrine scaling; opinion location estimation	Rejected	Weaker relative to MMC	Strongly supported (best predictor of doctrine)
Carrubba et al. (2012)	Concurrence behavior	Limited role	Limited role	Supported
Lax & Rader (2015)	Conference-to-final vote switching (“fluidity”)	Strongly supported (authorship affects switching)	Incomplete (median-only models inconsistent with switching)	Compatible but not directly tested
Lauderdale & Clark (2016)	Case-specific ideal points; assignment and joining analysis	Strongly supported (distance from author predicts joining)	Weakly supported	Weakly supported
Cameron and Kastellec (2023, Ch. 12)	Bloc-composition logit models of case dispositions	Not modeled	Loses significance once blocs included	Strongly supported (bloc size predicts outcomes)

Table 1: Summary of Existing Empirical Tests of Author, Court-Median, and Majority-Coalition Median (MMC) Theories of Supreme Court Bargaining.

generalize to the broader population of the Court’s output.

3 Using Precedent and Legal Change for Theory Adjudication

While in many ways our empirical approach builds upon these earlier efforts, our key point of departure is that we shift attention from where doctrine—as set forth in opinions—is located at the moment it is produced to how later Courts treat that doctrine over time. To

do so, we build on the extensive literature on precedent and legal change (see, e.g., Hansford and Spriggs 2006, Spriggs and Hansford 2001, Black and Spriggs 2013). Although the degree to which the justices are constrained by horizontal *stare decisis* remains debated (see, e.g., Spaeth and Segal 1999, Richards and Kritzer 2002, Lax and Rader 2010), it is well established that policy and precedent go hand-in-hand on the Supreme Court. Lower courts are formally bound to respect Supreme Court precedents under the principle of vertical *stare decisis*, and there is consistent evidence that they generally do so (see, e.g., Songer and Sheehan 1990, Westerland et al. 2010). Thus, when the Court negatively treats its own precedents (which we define more precisely shortly), this has policy implications not just for the instant case but for the broader legal issue going forward.

Accordingly, the importance of precedence suggests a point of leverage for adjudicating among bargaining theories. If precedents are written at different implied locations under competing models then the pattern of subsequent negative treatment should vary systematically with ideological distance between the origin precedent and later Courts (Hansford and Spriggs 2006, Spriggs and Hansford 2001, Black and Spriggs 2013). Our general approach is to use this insight by examining the Supreme Court’s negative treatment of its own precedents to evaluate competing models of bargaining and influence.

We will explain our data and statistical tests adjudicating between the three theories of bargaining shortly. But the logic of the tests is straightforward. Essentially, we ask the following: let’s assume an origin opinion was best characterized as written at the ideological location implied by each model in a given case. Then, measure the “distance” between those locations and the estimated location of the progeny court, using either the estimated ideal point of the median justice or the estimated mean of the conservative bloc (we define this more precisely below)

Our central question is simple: which of these distances best explains variation in the likelihood that a progeny Court will negatively cite an origin precedent?

For an illustration of this approach, consider Figure 1. In the 1990 case of *Austin v.*

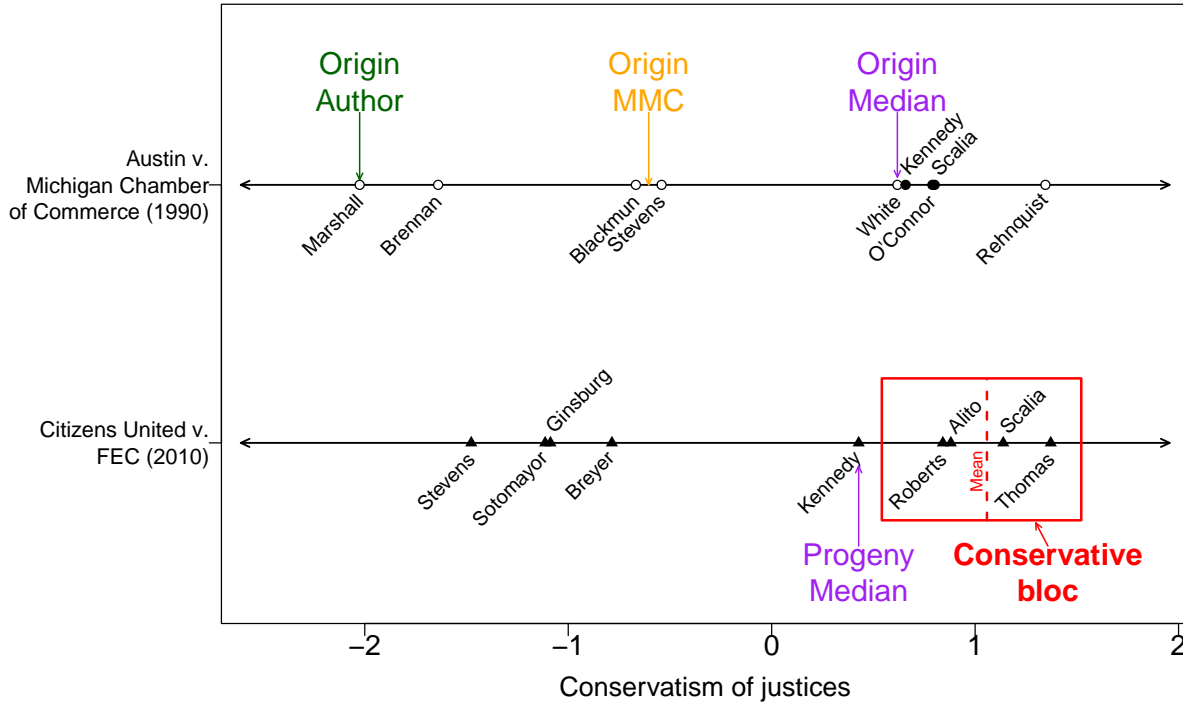


Figure 1: An illustration of origin-progeny distances across two cases. The top panel depicts the estimated ideal points of the justices who decided *Austin v. Michigan Chamber of Commerce* in 1990, which upheld restrictions on corporate expenditures; the plot also indicates the location of the opinion author, the median of the majority coalition (MMC), and the Court median. The bottom panel depicts the justices' preferences as of the 2010 case of *Citizens United v. FEC*; the red box depicts the justices in the conservative bloc, while the dashed vertical line depicts the mean of those justices' ideal points.

Michigan Chamber of Commerce (494 U.S. 652), the Supreme Court upheld a Michigan law that banned corporations from using money from their treasuries toward independent expenditures to support or oppose candidates in state elections. This decision stood as an important precedent in the Court's doctrine toward campaign finance regulation, as it held that limits on corporate expenditures were broadly constitutional. The case was decided 6-3, with Justice Thurgood Marshall writing the majority opinion, which was joined by Chief Justice Rehnquist and Justices Brennan, White, Blackmun, and Stevens, with Justices Scalia, Kennedy, and O'Connor in dissent. The top panel in Figure 1 depicts the estimated ideal point of each justice (defined shortly), along with the location of the author (Marshall), the median of the majority coalition (the midway point between Blackmun and Stevens), and the median (Justice White).

As the Court became more conservative over the next 20 years, it steadily chipped away at the holding in *Austin*. Finally, in 2010, the Court formally overruled *Austin* in *Citizens United v. Federal Election Commission* (558 US 310), holding that the First Amendment protects corporate speech and thus it is not constitutional to restrict independent expenditures by corporations.

The bottom panel in Figure 1 depicts the 2010 Court’s preferences. With the caveat (discussed more below) that there may be some issue-specific measurement error at play (most notably, campaign finance was an area where Justice Kennedy was quite conservative), we can see that the center of the Court had not moved much at all in the intervening decades, as measured by the Court median. However, the additions of Thomas, Roberts, and Alito after 1990 made the conservative bloc much more conservative and homogenous. But, more importantly for our tests, the variation in the estimated locations of the author, median of the majority coalition, and the Court median, give us useful “identification” for testing the bargaining theories. Imagine, for example, Justice Marshall had been able to write the opinion in *Austin* exactly at his ideal point. Then, the distance between that location of the progeny median at the time of *Citizens United* would have been substantial. On the other hand, if the median voter theorem prevailed in *Austin*, the distance between the policy in that case and the Progeny Court would have been much smaller. Finally, suppose the median of the majority coalition had prevailed. The *Austin* policy would still be much more liberal than the 2010 Court, even though the distance would be not be as great as under an author monopoly model.

4 Data and Measures

In this section we explain the data and measures we use to conduct our analyses below.

4.1 *Shepard’s Citations of Supreme Court Opinions*

To evaluate the theories of bargaining, we follow a long tradition in judicial politics and use *Shepard’s Citation Services’* coding of Supreme Court opinions to examine legal change.

As explained nicely by Benjamin, Kim and Quinn (2024, 3), “*Shepard’s* is a widely used commercial legal research service that employs attorneys to examine and code the content of every citation within each state and federal court opinion. *Shepard’s* classifies citations that meaningfully engage with an earlier opinion (i.e., that discuss and substantively respond to an opinion rather than just mentioning it) as neutral, negative, or positive. The legal literature often refers to such citations and accompanying discussions that are coded by *Shepard’s* as *Shepard’s treatments*.”

To be sure, using *Shepard’s* to study legal change is something of a blunt instrument, as it collapses what can be a nuanced judgment—how much does a new judicial precedent deviate from existing precedent—into a set of discrete categories (described shortly). But the advantage is that Spriggs and Hansford (2000) systematically demonstrated the reliability of *Shepard’s* citation. And while LexisNexis (which owns and implements *Shepard’s*) has changed some of its practices since 2000 (mainly in terms of how it automates parts of its data collection and digitizes the results), the core logic and practices of how it codes the treatment of precedent remains unchanged.⁶ Perhaps not surprisingly then, *Shepard’s* has been used extensively by political scientists and legal scholars to study legal change across multiple courts.(see Spriggs and Hansford 2001, Fowler et al. 2007, Cross et al. 2010, Westerland et al. 2010, Black and Spriggs 2013, Hansford, Spriggs and Stenger 2013, Benjamin and Vanberg 2016, Hinkle 2017, Benjamin, Quinn and Kim 2023, Benjamin, Kim and Quinn 2024, *inter alia*)

For our analysis, we build upon the pioneering analysis of Hansford and Spriggs in their 2006 book, *The Politics of Precedent on the Supreme Court*. Hansford and Spriggs collected *Shepard’s* citations to the majority opinions in every orally argued Supreme Court case between the 1946 and 1999 terms. Black and Spriggs (2013) subsequently (using additional data collected by Fowler et al. (2007)) updated this data through the 2004 term; Black and Spriggs (2013, see p. 334, fn. 3) also added some per curiam opinions that did not follow oral

⁶See LexisNexis (n.d.a) and LexisNexis (n.d.b) for a history of *Shepard’s* and a broad overview of the service, respectively.

arguments, if they contained any legal reasoning. We then updated this dataset through the end of the 2023 term. We choose to drop cases decided before the 1953 term; our measures of justice ideal points (described shortly) only begin in the early 1950s, and 1953 is a natural starting point, as it marks the onset of the Warren Court.

Following Hansford and Spriggs (2006) and Benjamin, Kim and Quinn (2024), we focus on citations that can be defined as “positive” or “negative,” and exclude “neutral” citations (which are relatively rare).⁷ We follow standard practice and code a citation as “positive” if *Shepard’s* assigns a “Following label” to it; according to LexisNexis, this means the “citing opinion relies on the [cited] case ... as controlling or persuasive authority” (LexisNexis 2025, 33). Negative precedents, by contrast, are any that fall into the following categories, ordered from the least severe to the most severe form of negative treatment: “Distinguished,” “Criticized,” “Limited,” “Questioned,” and “Overruled.” (We give precise definitions of each shortly in Table 2). For simplicity, when an origin case receives one or more positive or negative *Shepard’s* treatments within the same progeny case, we collapse these to a single origin-progeny treatment dyad. Accordingly, the data comprise unique origin-progeny dyads defined over positive and negative treatments only (though, to be clear, multiple origin precedents can be cited in the same progeny case and a progeny case can treat multiple origin precedents).

We then merged these cases with the Supreme Court Database (Spaeth et al. 2024). The purpose of this merge is severalfold. First, we can examine which Supreme Court decisions are ever cited, as well as how and how often they are cited. Second, we gather information on the composition of both of the original precedent setting court and future courts that do or could potentially cite on a case. Finally, we gather relevant information on the precedent-setting case, such as issue area and the number of concurrences (we describe these in detail below).

⁷One concern is that *Shepard’s* employs relatively restrictive, rule-based criteria in assigning positive or negative treatments versus neutral citations; some citations coded as “neutral” may nevertheless reflect more subtle approval or criticism (LexisNexis 2025). We acknowledge this limitation, but it is unclear how any resulting measurement error would systematically favor one bargaining theory over another.

Coding	Definition	Citations	Percent (%)
Followed	The progeny opinion relies on the origin case as controlling or persuasive authority.	5,254	48.3%
Distinguished	The progeny case differs from the origin case, either involving dissimilar facts or requiring a different application of the law.	4,908	45.2%
Criticized	The progeny opinion disagrees with the reasoning or result of the origin case, although the progeny court may not have the authority to materially affect its precedential value.	283	2.6%
Limited	The progeny opinion restricts the application of the origin case, finding its reasoning applies only in specific, limited circumstances.	42	0.4%
Questioned	The progeny opinion questions the continuing validity or precedential value of the origin case because of intervening circumstances, including judicial or legislative overruling. The origin case itself may be questionable due to a later action that affects its precedential value.	190	1.7%
Overruled	The progeny case expressly or effectively overrules or disapproves all of the origin case.	190	1.7%

Table 2: The distribution of citations, by positive and negative categories. The table depicts the distribution of the type of citations, among progeny cases cited by the Supreme Court, 1953 to 2023. The category definitions come from LexisNexis (2025); note, however, that the “origin” and “progeny” terminology is ours and not theirs.

All told, the Supreme Court handed down about 8,400 decisions between the 1953 and 2023 terms. Of these decisions, 4,155 received at least one positive or negative citation; these 4,155 were cited a total of 10,867 times. As mentioned above, we denote a precedent-setting case as an “origin case” and a citing case as a “progeny case” (see e.g. Segal and Spaeth 1996). Precedents, of course, are dynamic, and any given case can be both an origin case (if it is cited in the future by the Court) and a progeny case (if it cites a past Supreme Court opinion). But analyzing opinions in this way provides for a clear structure for testing theories of bargaining.

Table 2 depicts the distribution of citations, by positive and negative categories. For each category, we provide a definition, as taken from LexisNexis (2025). The right two columns depict the number of citations that fall into each category, along with the percentage of each category out of all total citations. The table makes clear that the vast majority of citations

Decision Type	0 <i>Shepard's</i> Treatments	At least 1 <i>Shepard's</i> Treatment
Signed opinion of the Court	2,828 (43.1%)	3,736 (56.9%)
Judgment of the Court	794 (81.6%)	179 (18.4%)
Per curiam opinion	65 (98.5%)	1 (1.5%)
Decree	93 (100.0%)	0
Summary disposition (cert/stay/etc.)	442 (81.9%)	98 (18.1%)
Miscellaneous order	87 (38.2%)	141 (61.8%)

Table 3: *Shepard's Treatments by Supreme Court Decision Type*. The table depicts the percentage of origin cases that are ever cited, by type of origin decision. (Percentages are row percentages).

are either of the “Followed” or “Distinguished” variety. At the other extreme, the Court overruled 190 decisions in this period. Because the vast majority of negative citations fall into the distinguished category, we collapse all the negative categories (i.e. everything other than “Followed”) into a single negative treatment indicator, and then use that variable as our key outcome of interest throughout.

Also of interest is the type of Supreme Court decisions that are more likely to be cited at all. Table 3 depicts the distribution of whether a case is ever cited by the type of opinion the Court issues in an origin case, as coded in the Supreme Court Database. Not surprisingly, signed opinions of the Court (i.e. “full” merits opinions) are the origin cases most likely to be cited in progeny cases—though fully 43% of such opinions never get cited. All other types of opinions are cited by progeny courts very rarely—except miscellaneous orders, though there are very few of those to begin with. (Because per curiam opinions and decrees exhibit essentially no variation—nearly all are never cited, with only a single per curiam opinion ever receiving a citation—we drop them from the analysis from this point forward.)⁸

4.2 Measuring preferences and opinion locations

To measure the preferences of the justices, we use the ideal point estimates developed by Bailey (2007) and updated by Bailey (2020), which run from 1950 to 2019.⁹ The Bailey scores

⁸Recall we are excluding neutral citations; if we did not, there would surely be more treatments of per curiam opinions.

⁹To carry these scores forward to the 2023 term, we do the following. For justices who were on the bench in 2019, we carry forward their 2019 scores through 2023. For Justices Barrett and Jackson, who were appointed in 2020 and 2022, respectively, we impute their ideal points using contemporaneous scores of ideologically similar justices; for Barrett, we take the mean of Kavanaugh and Gorsuch; for Jackson, we

have the virtue of being based on both inter-institutional and inter-temporal bridging, which more cleanly models changes in the Court’s agenda over time. This approach allows for more valid dynamic estimates of justices’ ideal points and, in turn, more meaningful comparisons of ideal points and majority-opinion coalitions across time.¹⁰ The primary limitation of the Bailey scores is that they are unidimensional. While this may introduce measurement error in some issue areas, it is unclear *ex ante* why such error would systematically privilege any of the bargaining models we evaluate.

To measure opinion locations, we proceed as follows. For each origin decision, we first calculate the estimated ideal point of the median of the Court by taking the median ideal point among all justices who participated in the decision. Next, we identify the justices who formed the majority coalition in each origin case and calculate the median ideal point of those justices (the “median of the majority”).¹¹ (In most unanimous cases, then, the location of the overall median and the median of the majority will be equivalent.) Finally, for cases with a signed majority opinion, we record the estimated ideal point of the majority opinion’s author.

The main test we will implement below examines how far distant a progeny court is from an origin court, and how well that distance predicts negative citations. There is a complication, however, in that we cannot use any information from the progeny case itself, since the outcome of the case is endogenous to decision of whether and how to treat an existing precedent. Most importantly, we cannot use the median of the majority coalition in the progeny case at all, since obviously that can’t be known until after the case is decided (the same is true with the author of a progeny opinion). Accordingly, we use the following two-fold strategy. First, we follow Hansford and Spriggs (2006) and code the estimated ideal

impute Justice Kagan’s score.

¹⁰By contrast, Martin–Quinn scores rely solely on overlap among justices to identify agenda changes and their effects on dynamic ideal point estimates. Bailey (2007) shows that Martin–Quinn scores produce facially questionable estimates of the location of the median justice over time, whereas those produced by the Bailey scores are facially valid.

¹¹For this calculation, we exclude any justices who specially concurred in the judgment (i.e., cases in which the `vote` variable in the Supreme Court Database equals 4), since special concurrences indicate agreement with the disposition but not with the rule or rationale of the majority opinion.

point of the median justice of the Court at the time a progeny case was decided, since that is exogenous. Second, we take advantage of the innovation of Cameron and Kastellec (2023, ch. 12) and divide the justices into liberal, moderate, and conservative “blocs” for each year, based on the Bailey scores.¹² The theoretical basis for this measure—see Parameswaran, Cameron and Kornhauser (2021)—is that any majority opinion author needs to write an opinion that will receive at least four additional joins to become fully precedential. As the majority bloc becomes greater in size, members of that bloc can accomplish this by offering their most preferred position or something near it (Cameron and Kastellec 2024). Accordingly, we can think of variation in bloc sizes as providing an *ex ante* proxy for what a dominant majority might prefer, without relying on the *ex post* measure of the majority of the median coalition.

It is, of course, well known that the Supreme Court has become much more conservative over time, beginning with the onset of the Burger Court in 1969, continuing with the Rehnquist and Roberts Court, before culminating in the 6-3 conservative super-majority that emerged following Amy Coney Barrett’s replacement of Ruth Bader Ginsburg in 2020 (Bailey 2013, Cameron and Kastellec 2024, Re 2025). Cameron and Kastellec (2023, 324) show that the size and conservatism of the conservative bloc, as measured using the Bailey scores, has increased since the 1970s. In some regression models below we will use the distance between the origin opinion estimates and the estimated location of the conservative bloc in the progeny case to predict the likelihood of negative treatment by the progeny Court. Because we would expect this measure to be most predictive during the period where conservative justices controlled the Court, we focus on progeny cases decided 1969 and beyond for these models.

¹²Specifically, following Cameron and Kastellec (2023, 323), we use the full distribution of ideal points for every justice from 1950 to 2023; in other words, “justice-years” are the unit of analysis. Then we divide this distribution into thirds: justices who, in a given year, fall in the bottom third are coded as being the liberal bloc, justices in the middle third are coded as being in the moderate bloc, and justices in the top third are coded as being in the conservative bloc. Because the Bailey scores are dynamic, a justice may appear in a different bloc in different years if their score happens to cross one of the thresholds over time.

4.3 Measuring distance between origin and progeny cases

How best should we measure ideological distance between origin opinions and progeny courts? Most spatial models of politics assume *symmetric loss*: outcomes equidistant from a decision maker are treated as equally costly in either direction. Operationally, researchers implement this assumption by taking the absolute value of the difference between two ideological measures. For example, Hansford and Spriggs (2006) use the absolute distance between the origin median of the majority coalition and the progeny Court’s overall median.

This is certainly a plausible assumption about how the Court evaluates precedents, but it is not the only possibility. To see this, return to Figure 1. Suppose the opinion in *Austin* was written exactly at Justice Marshall’s ideal point—substantially to the left of the Court’s center in 2010. Under the symmetric-loss assumption captured by absolute distance, the *Citizens United* Court would evaluate an opinion written at the *reflection point* around the progeny Court’s median—that is, far to the ideological right—the same as it would evaluate Justice Marshall’s original opinion.¹³ This may be plausible, but we can readily imagine a world in which a Court that is strongly skewed in one ideological direction evaluates precedents directionally as well as with respect to proximity—focusing especially on altering precedents that lie far from it on the opposite side of the ideological spectrum. Given that most of our data covers periods in which the Court was either trending conservative or firmly conservative, such directional behavior would imply greater attention to liberal precedents than to conservative ones, *ceteris paribus*.

Because this is a theory-testing rather than a theory-building paper, we adopt an agnostic approach. In all analyses that follow, we measure ideological distance between the origin Court and progeny Court using both absolute distance and raw (i.e., signed) distance. Figure 2 depicts the distribution of the 12 origin-progeny distance combinations, using both raw distance (top row) and absolute distance (bottom row); the vertical lines are drawn at

¹³More specifically, Justice Marshall’s Bailey score in 1990 was approximately -2 , and the Court median in 2010 (Justice Kennedy) was about 0.4 . The symmetric point around Kennedy is therefore $2 \times 0.4 - (-2) \approx 2.8$. This value exceeds the maximum Bailey score for any justice, but the substantive point remains.

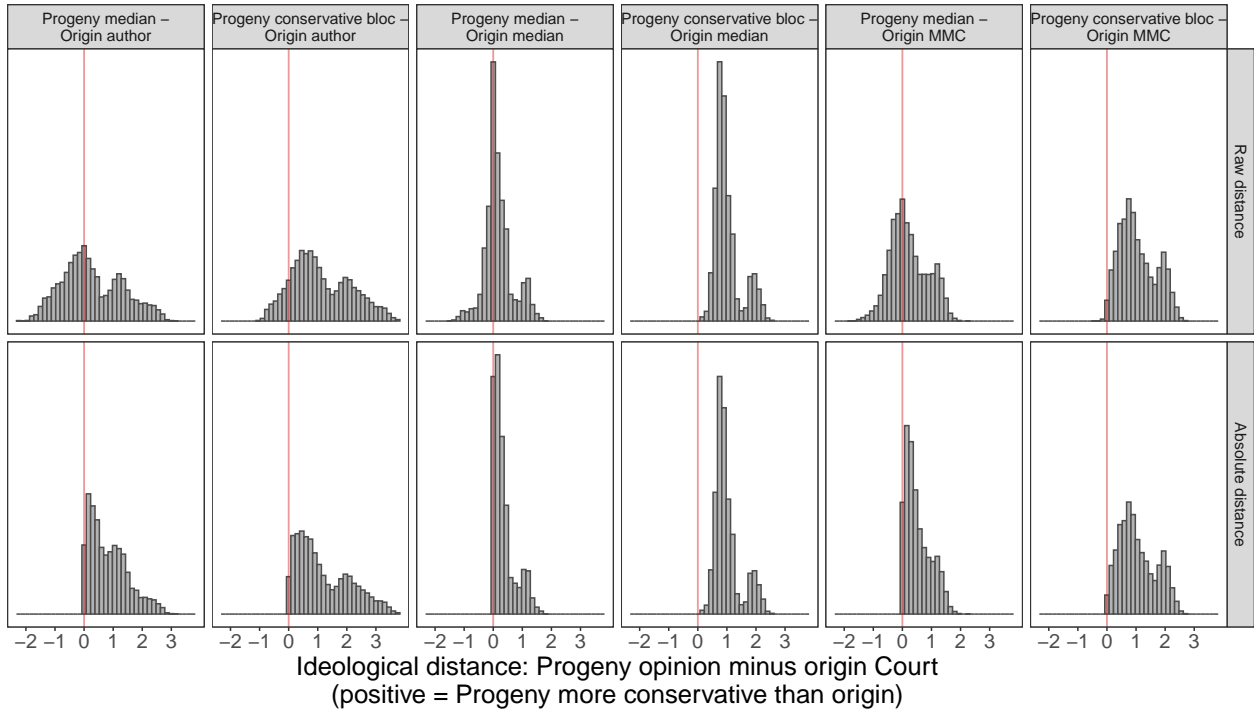


Figure 2: The distribution of distances among origin and progeny pairs, using raw distance (top row) and absolute distance (bottom row). The vertical red lines are drawn at zero.

zero. Two patterns stand out. First, looking at the raw distances, most of the density is positive, meaning that the progeny court is more conservative than the origin court. This is particularly true using the mean of the conservative bloc as the progeny measure—the conservative bloc is rarely more liberal than the origin author and almost never more liberal than either the origin Court median or median of the majority. This is not surprising, given the rightward drift of the Court since 1969, and the increased conservatism of Republican appointees in the 21st century. Second, there is much more variation in the author-based measures, since the variation in medians is constrained by the composition of the court. These patterns will help inform our interpretation of the regression analyses below.

4.4 Negative citations over time

How have the Court’s citation patterns changed across the time period of our study? Figure 3 depicts the Court’s negative citations over time. Panel A depicts the number of

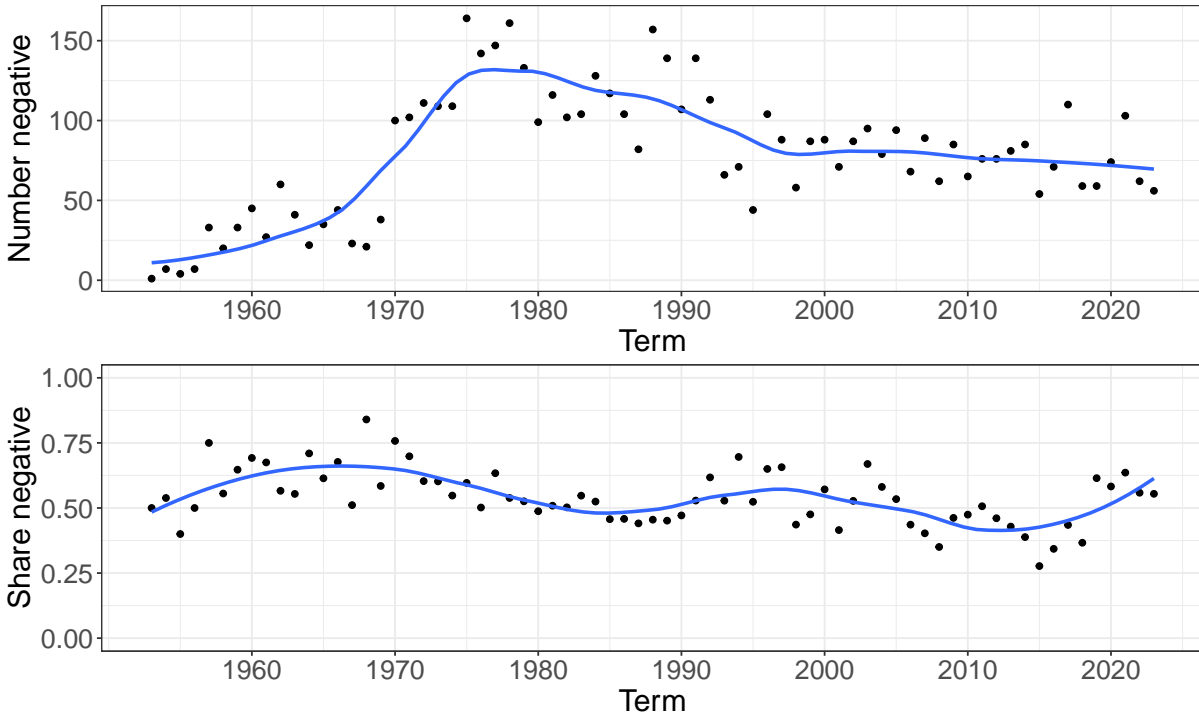


Figure 3: The Supreme Court’s negative treatment of precedent, 1953-2023. The top panel depicts the number of negative treatments per term, while the bottom panel depicts the percentage of negative treatments among all treatments (i.e. $\text{negative}/(\text{negative} + \text{positive})$). The blue lines are all loess lines (with a span of .4).

negative citations per term, between 1953 and 2023; the blue line is a loess line. We can see that the number of negative citations increases substantially until the middle of the 1970s; some of that rise is likely mechanical, however, given that our origin data starts in 1953; following the mid-1970s, the number of negative citations trends downwards, though it has been largely flat over time.

However, as is well known, the number of *cases* the Supreme Court hears each term has dropped significantly in this period, meaning it will have fewer opportunities to negatively treat precedent. Panel B accounts for this by depicting the share of negative citations among all treatments; we can see that the rate has been fairly level over time, though it seemingly ticked upwards after the Trump appointees joined the Court.

5 Analyzing Negative Citations

We now turn to our core empirical analyses. Modeling citations is somewhat tricky, given the structure of the data. Accordingly, we pursue two complementary empirical strategies, each with their own strengths and weaknesses.

First, following Hansford and Spriggs (2006, ch. 4), we create an unbalanced panel data set where the unit of analysis is the origin case/term. Specifically, for each origin precedent, we trace whether and how it was cited in any given term following the introduction of the precedent, through 2023 (the 2022 term). (We call this the “term panel” dataset.) For example, returning to *Austin v. Michigan Chamber of Commerce*, that origin case was cited in four separate terms by progeny courts (and in each instance was treated negatively). Those four terms are therefore coded as containing a negative treatment (i.e. coded “1”), while all other post-Austin terms are coded as zero.¹⁴ The estimand here is thus the probability that an origin case receives a negative treatment in any given term.

The principal advantage of this approach is that it avoids any selection bias that may arise from the Court having discretion over which cases to cite and when to cite them. The disadvantage is that it creates a rather large dataset with a rare events structure—there are about 339,000 case–term observations, with about 5,300 negative citations across all case-term pairs—potentially making it harder to tease out any substantive differences across the different theories. In addition, it inherently ignores any information when the Court treats the same origin case multiple times in the same term. For example, *Michigan* was negatively treated twice in the 2006 term, but those two progeny cases are collapsed into a single indicator.

To account for these weaknesses, we create and analyze a second dataset where the citation is the unit of analysis, and the outcome of interest is, whether conditional on a citation, is the citation negative, versus positive. We call this the “case-level citation” dataset.

¹⁴Hansford and Spriggs (2006) separately analyze the incidence of positive citations, but the theoretical basis for connecting positive citations directly to ideological distance between origin and progeny courts seems weaker to us.

Returning to *Austin v. Michigan Chamber of Commerce v Michigan*, it was cited a total of five times by progeny courts; thus there are five progeny observations for that case in the data. All told, the case-level citation dataset has about 11,000 observations, of which 51% feature a negative treatment. We think the probability of negative citation, conditional on any citation, is a more intuitive estimand. Of course, the downside of this dataset is that it conditions on citations, and thus is subject to selection bias, so it is important to keep in mind what this estimand is when we work with this dataset.

5.1 Descriptive relationships: distance and negative citations

Before turning to our regression analyses, we begin by showing the descriptive relationship between ideological distance and negative citations. Using the case-level citation data, Figure 4 depicts the relationship between the two using the same 12 distance/opinion location pairs as seen in Figure 2. The x-axis depicts distance, while the y-axis depicts negative citations at 1 and positive citations at zero (using the “rugs”). The blue lines are loess lines, so the lines can be interpreted as the probability of a negative citation, conditional on a citation occurring.

Several interesting patterns emerge. Beginning with the absolute distance measures on the bottom row, perhaps not surprisingly we see generally monotonic relationships between increased ideological distance across origin precedents and progeny courts and the likelihood of a negative citation. However, of note is that the relationship for the two median measures appears to be stronger than the author measures. Turning to the raw distances, we do in fact see non-monotonic relationships for some—but not all—of the distance measures. In particular, the panels using progeny median minus origin median and progeny median minus origin MMC show that the Court is in fact likely to negatively treat origin precedents that are much more conservative than the progeny Court (even if such cases are overall much rarer in our data). Conversely, for all the measures using either origin author or progeny conservative bloc, we do not find this symmetric relationship (again, for the conservative bloc measures, there are almost no cases where it is more liberal than any origin precedent).

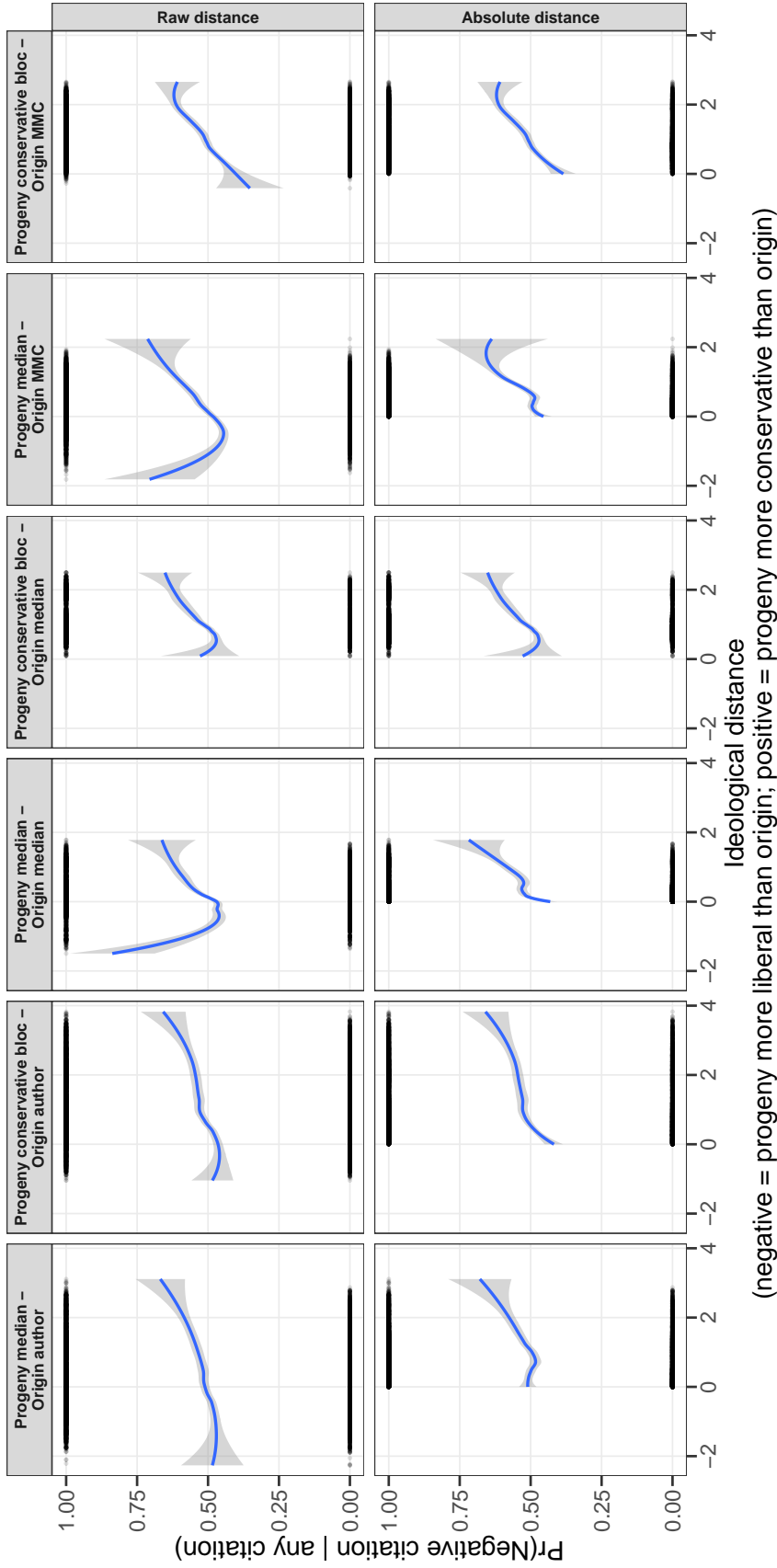


Figure 4: Ideological distance versus negative citations. Each panel plots the relationship between ideological distance and the probability that a citation is coded as negative, conditional on an origin case being cited. The “rugs” depict the distribution of origin–progeny citations. The solid lines are loess smoothers ($span = 0.75$), estimated separately for each distance definition. The top row depicts the signed (raw) distance measures, while the bottom row depicts absolute-value measures.

5.2 Regression-based analyses

Our empirical strategy is as follows. First, we estimate two sets of parallel regression models using the case-level citation dataset and the term-panel dataset. For each dataset, then we estimate two regression models for each of the 12 distance/opinion location pairs; the first model is simply a bivariate logit model where the only predictor is the relevant distance measure, while the second model adds a set of control variables (described shortly). We mainly use these results to assess the statistical significance of each distance coefficient across the various models.

Second, in order to understand the substantive significance of these coefficients, we calculate and depict the average predicted effect of distance for each of the 12 distance/opinion location pairs. Because this probability is more intuitive using the case-level citation dataset, we focus on those results in the main text and present the parallel results for the term-panel dataset in the Appendix.

Finally, the core theoretical question we seek to answer empirically is what model of bargaining on the Court best explains the Court's treatment of precedents? To answer this question, we compare the model fit of each of the 12 regression models from each dataset in question. This task is not straightforward, given that a) the outcome variable is binary; b) the sample sizes are often uneven across model and c) the number of predictors varies across the absolute value-based and signed-distance based models. (The last two factors mean we can't use straightforward comparisons of information criteria statistics (e.g. AIC).) Accordingly, we use a cross-validation procedure to conduct this analysis and estimate the out-of-sample predictive performance of each model. This allows us to compare models on a common predictive metric while avoiding reliance on in-sample fit statistics or significance tests.

5.2.1 Regression results

We begin with our core regression results. For the models with controls, our general strategy is to only control for variables that were defined at the time of the origin case. We

do this in order to isolate the predictive effect of changing ideological distance over time, rather than bundling that with other characteristics of progeny cases. Even though we are not pursuing a strict causal inference design, as an analogy, controlling for any thing that occurs after the origin case would potentially introduce “post-treatment” bias. For example, one of the key predictors in Hansford and Spriggs (2006) is the “vitality” of a precedent—that is, how often has an origin precedent been treated positively and/or negatively at the moment the Court is deciding a new progeny case. While the question of whether and how vitality predicts future treatments is descriptively of interest, essentially vitality can be thought of an “intermediate variable,” since how the Court treats an origin decision in the past will surely inform how it treats it in the future. Accordingly, we focus only on origin-based predictors as control variables, with the idea of accounting for various factor that may lead the Court to be more likely to negatively cite any case, *ceteris paribus*.

Specifically, we use the following control variables:

- Whether the disposition of the origin case was liberal (coded 1) or conservative (coded 0). While dispositions are of course correlated with the ideological location of the majority opinion, future Courts may focus on ideologically divergent dispositions (i.e. a conservative court targeting liberal dispositions).
- Whether an origin case is anything other than a signed opinion of the court; we call these “non-full decisions.” Recall from Table 3 that such decisions are much less likely to be cited by the Court.
- We include indicators capturing whether the authority of the origin case can best be described as constitutional, statutory, administrative, or other (using the Supreme Court Database’s standard coding). Constitutional precedents are generally more prone to negative treatments.
- We code the issue area of the origin case, again based on the Supreme Court Database, which codes each case into 14 broad categories: Criminal Procedure; Civil Rights; First Amendment; Due Process; Privacy; Attorneys / Government Officials’ Fees or Compensation; Unions; Economic Activity; Judicial Power; Federalism; Interstate Relations; Federal Taxation; Miscellaneous; Private Law.
- Finally, we code for the age of the origin precedent at the time a progeny case is decided, as well as age-squared to account for any possible non-linearities. (While age is technically post-treatment, it is obviously mechanically determined by time.)

	Auth-Med		Auth-CBloc		Term panel data – absolute distance		MMC-Med		MMC-CBloc			
					OMed-Med	OMed-CBloc						
Intercept	-3.876*** (0.032)	-3.142*** (0.091)	-3.899*** (0.036)	-3.191*** (0.095)	-4.009*** (0.029)	-3.393*** (0.093)	-3.954*** (0.058)	-3.566*** (0.103)	-4.167*** (0.031)	-3.449*** (0.094)	-4.124*** (0.047)	-3.475*** (0.100)
Distance	-0.073* (0.028)	-0.021 (0.030)	-0.053* (0.023)	0.028 (0.023)	-0.305*** (0.057)	0.180** (0.056)	-0.189*** (0.054)	0.251*** (0.048)	0.059 (0.043)	0.259*** (0.045)	-0.030 (0.039)	0.175*** (0.040)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	261805	259811	250810	248856	331704	328790	316824	314009	331511	328597	316641	313826
	Auth-Med		Auth-CBloc		Term panel data – signed distance		MMC-Med		MMC-CBloc			
					OMed-Med	OMed-CBloc						
Intercept	-3.895*** (0.026)	-3.124*** (0.089)	-3.909*** (0.036)	-3.205*** (0.095)	-4.059*** (0.023)	-3.340*** (0.093)	-3.496*** (0.142)	-3.686*** (0.159)	-4.137*** (0.025)	-3.349*** (0.094)	-3.754*** (0.077)	-3.338*** (0.112)
Distance	-0.039 (0.027)	0.061* (0.027)	-0.030 (0.057)	0.096 (0.057)	-0.121** (0.043)	0.225*** (0.043)	-1.007*** (0.246)	0.470* (0.234)	-0.119** (0.037)	0.117** (0.037)	-0.814*** (0.145)	-0.149 (0.140)
Distance ²	-0.024 (0.016)	-0.030 (0.017)	-0.008 (0.019)	-0.026 (0.019)	-0.142** (0.044)	0.007 (0.045)	0.307*** (0.093)	-0.082 (0.089)	0.076* (0.032)	0.097** (0.033)	0.312*** (0.057)	0.129* (0.054)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	261805	259811	250810	248856	331704	328790	316824	314009	331511	328597	316641	313826
	Auth-Med		Auth-CBloc		Case-level citation data – absolute distance		MMC-Med		MMC-CBloc			
					OMed-Med	OMed-CBloc						
Intercept	-0.054 (0.040)	-0.482*** (0.098)	-0.196*** (0.042)	-0.678*** (0.100)	-0.114*** (0.033)	-0.517*** (0.092)	-0.507*** (0.060)	-0.944*** (0.103)	-0.186*** (0.038)	-0.578*** (0.095)	-0.444*** (0.047)	-0.851*** (0.098)
Distance	0.144*** (0.035)	0.111** (0.036)	0.198*** (0.027)	0.160*** (0.028)	0.516*** (0.060)	0.406*** (0.063)	0.541*** (0.054)	0.422*** (0.056)	0.481*** (0.053)	0.361*** (0.056)	0.462*** (0.039)	0.397*** (0.046)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	10374	10374	9730	9730	10866	10866	10188	10188	10866	10866	10188	10188
	Auth-Med		Auth-CBloc		Case-level citation data – signed distance		MMC-Med		MMC-CBloc			
					OMed-Med	OMed-CBloc						
Intercept	0.007 (0.031)	-0.409*** (0.097)	-0.142*** (0.040)	-0.623*** (0.100)	-0.036 (0.028)	-0.448*** (0.092)	-0.629*** (0.160)	-1.065*** (0.183)	-0.069* (0.030)	-0.429*** (0.095)	-0.441*** (0.074)	-0.859*** (0.113)
Distance	0.125*** (0.029)	0.083*** (0.030)	0.141* (0.063)	0.085 (0.064)	0.219*** (0.055)	0.155** (0.057)	0.766** (0.283)	0.647* (0.289)	0.272*** (0.044)	0.201*** (0.049)	0.461** (0.149)	0.423*** (0.161)
Distance ²	0.021 (0.019)	0.027 (0.019)	0.012 (0.022)	0.018 (0.022)	0.251*** (0.058)	0.202*** (0.060)	-0.087 (0.108)	-0.087 (0.110)	0.168*** (0.047)	0.151** (0.047)	-0.001 (0.062)	-0.012 (0.065)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	10374	10374	9730	9730	10866	10866	10188	10188	10866	10866	10188	10188

Table 4: Regression models of negative treatments. The table presents four sets of regressions, corresponding to each combination of data structure (term panel and case-level citation data) and distance specification (absolute and signed). For each set, we report two models for each combination of origin location (listed first in the column header) and progeny location. The first model in each pair is a bivariate specification; the second includes the full set of controls, which are omitted here for brevity but reported in the Appendix. Signed-distance models include both distance and its square, while absolute-distance models do not. Statistical significance is denoted by * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. All models include standard errors clustered by origin case. Models using the conservative bloc as the progeny location only use progeny cases beginning with the 1969 term.

Table 4 presents four sets of regressions, corresponding to each combination of data structure (term panel and case-level citation data) and distance specification (absolute and signed). For each set, we report two models for each combination of origin location (listed first in the column header) and progeny location. The first model in each pair is a bivariate specification; the second includes the full set of controls, which are omitted here for brevity but reported in the Appendix (Section A.1). Signed-distance models include both distance and its squared value, while absolute-distance models do not. To account for the non-independence among progeny cases that treat the same origin precedent, all models include standard errors clustered by origin case. Finally, models using the conservative bloc as the progeny location only use progeny cases beginning with the 1969 term—that is the year that the Warren Court ended and the Court began its long and steady shift to the right.

The models are organized such that the models using the origin author measure appear first, followed by the origin Court Median (“Omed”), followed by the origin median of the majority of coalition. While the results vary somewhat across the datasets and distance metrics, in general we find that the coefficients on distance for the two median measures are both more likely to be statistically significant (in the expected, positive direction); in addition, the coefficients are generally larger in magnitude for the median measures than the author measures. All told, these results provide preliminary evidence against a strict author-influence theory of bargaining on the Supreme Court.

5.2.2 Average predicted effects of distance

Because logit coefficients are difficult to interpret substantively (especially for the term-panel data), we convert the models to predicted probabilities to ease interpretation. Using 500 simulations of each of the 24 models with controls reported in Table 4, we calculate the “average predictive effect” (alternatively, averaged predictive comparisons) of ideological distance on the probability of negative citation. The advantage of this approach, compared to using predicted probabilities evaluated at pre-specified values of the other predictors, is that it incorporates the empirical distribution of the data directly into the predictions.

Specifically, at discrete levels of ideological distance, we calculate the probability of negative citation for every observation, using each observation’s observed values on all other covariates, and then average the resulting predicted probabilities across observations. Because the procedure is simulation-based, it also allows for the construction of uncertainty intervals.¹⁵

Figure 5 presents these average predicted effects. (Essentially, it can be thought of a model-based analogue to the descriptive plots in Figure 4.) Results from the case-level citation data appear on top, while results from the term-panel dataset appear on bottom. Note the y-axis differs across the two; the probability of negative treatment is much lower in the term-panel data because, as noted above, most origin cases are cited quite infrequently across the span of the data. In each panel, the curve predicts the average probability of citation across distance, with the shaded regions depicting 95% confidence intervals. Each curve runs from the smallest observed distance (which is, of course, zero in the absolute distance measures) to the largest observed distance for a given dataset/distance metric pair.

Beginning with the case-level citation data, we can see the curves all generally indicate a positive correlation between increased distance between origin and progeny courts and the probability of negative citation, conditional on any citation. However, we can see that the relationship for the two median measures is much stronger compared to the origin author measure. For example, using the absolute distance measure for the progeny median – opinion author model, the probability of a negative citation increases from about 0.49 [95% CI 0.48, 0.51] to 0.58 [0.54, 0.61]. That is certainly substantively significant, but the corresponding change in average predicted probabilities for the Progeny median – origin median model is 0.48 [0.46, 0.5] to 0.65 [0.62, 0.68], a much larger predicted effect. Similarly, for the Progeny median – origin MMC model, the corresponding shift in probabilities is from 0.47 [0.46, 0.49] to 0.66 [0.62, 0.70].

¹⁵See Gelman and Pardoe (2007) and Hanmer and Kalkan (2013) for general discussions of average predicted effects and simulation-based inference in nonlinear models. To estimate uncertainty, for each model, we draw repeated samples of the coefficient vector from a multivariate normal distribution centered at the estimated coefficients, with covariance given by the cluster-robust variance–covariance matrix. From these repeated samples, we generate 95% confidence intervals by taking the .025 and .975 quantiles of the estimated average predicted probabilities.

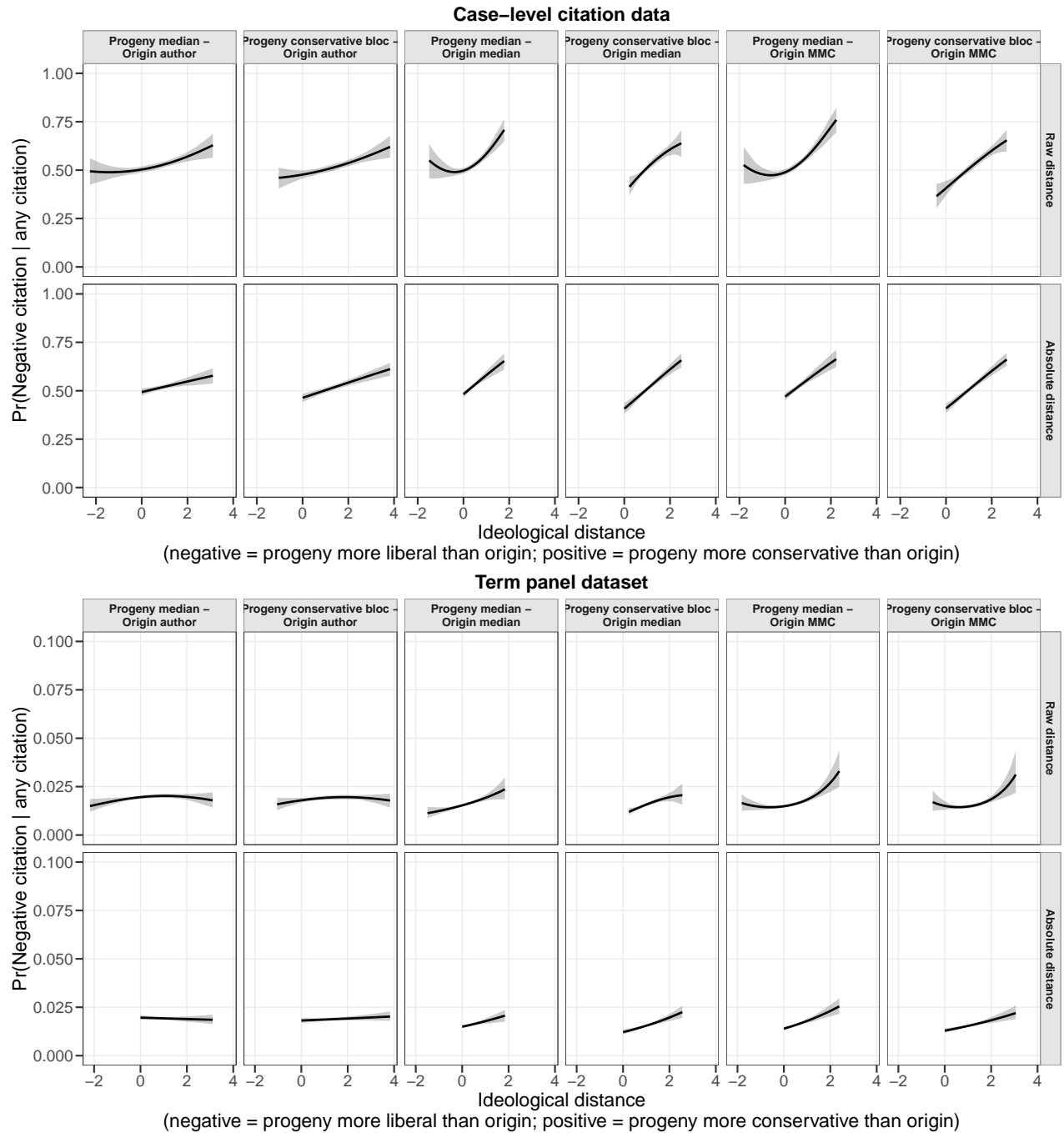


Figure 5: Average predicted effects of ideological distance on negative citation. Each panel plots the average predicted probability of negative treatment across the empirical range of ideological distance, holding all other covariates at their observed values. Solid lines denote mean predicted probabilities, while shaded regions depict 95% confidence intervals based on simulation from the estimated coefficient distribution. The top panels use the citation-level data, while the bottom panels use the term-panel dataset. Note that the y-axis differs across datasets due to the lower baseline probability of negative treatment in the term-panel data.

Turning to the term-panel data, the difference across models is even starker. We see that for the author models, the average predicted effect of increased distance on the probability

of a negative citation is essentially zero. By contrast, we still see a positive correlation with the two median measures. Specifically, in the Progeny median – origin median model (again using absolute distance), the predicted probability rises from 0.015 [0.014, 0.016] to 0.02 [0.018, 0.024] at maximum distance. For the Progeny median – origin MMC model, the corresponding shift in probabilities is even larger, moving from 0.014 [0.013, 0.015] at zero distance to 0.025 [0.022, 0.03] at maximum distance. Although these changes are modest in absolute terms, they are still meaningful given the low base rate in citations. Thus, both datasets point to the median models as being more predictive of negative citations than the author.

Finally, of interest in Figure 5 is that the non-monotonicity we saw in the descriptive plots in Figure 4 do not appear to be statistically identifiable, once we control for other factors. In particular, in the case-level citation data, in the progeny median – opinion author and progeny median – opinion MMC models with signed distance, the curves do appear to shift upwards when the progeny court is more liberal than the origin court (i.e. when distance is negative). However, the confidence intervals reveal that a null hypothesis of monotonicity cannot be rejected, as they each extend to well below the global minimum of the predicted probability lines. Of course, as discussed earlier, this null result could be due to a lack of statistical power, since in our data most progeny courts are more conservative than the origin median measures. But it is also consistent with an increasingly conservative court becoming more skeptical of liberal precedents, and negatively treating them accordingly, *ceteris paribus*.

5.2.3 Model performance

The last step in our analysis is to use model performance metrics to examine which bargaining model best explains the treatment of precedent by the Supreme Court. Due to the nature of our data, standard approaches for comparing theories are not available. First, one approach would be to use a nested model with interactions to see whether one of the key distance predictors can be statistically differentiated from the others. However, because the

distance measures are all highly correlated and conceptually closely related, this approach would suffer from severe multicollinearity and yield unstable estimates, and therefore is not appropriate. Second, a straightforward model-based comparisons using information criteria (e.g. AIC) also is not possible, since those statistics only are valid when applied to models that are estimated on the same sample, and our samples differ across models.

Accordingly, we take a cross-validation approach for comparing the out-of-sample performance of each of the 24 regression models in Table 4 with controls, an approach that is specifically tailored to our data (see e.g. Hastie, Tibshirani and Friedman 2009, Cranmer and Desmarais 2017). The basic idea behind cross-validation is that one splits the data into a number of subsets. Then, the data from one subset is held out (denoted the “test data”), and the model is run on the remaining subsets of the data (denoted the “training data”). Next, the estimated model is used to generate predictions in the test data, and then these predictions are compared to the actual data in the test (i.e. out-of-sample) observations. The better these predictions, the better the model performance. This process is then repeated across subsets, and the resulting performance metrics are averaged across these repetitions.

Specifically, for the citation-level models, we implement clustered 10-fold cross-validation, where each fold comprises approximately 10% of the data. In each iteration of the cross-validation process, one fold is held out to form a test set, and the model is estimated on the remaining data (i.e. the other 90% of observations). Importantly, folds are not formed by simple random sampling. Instead, we cluster by progeny case, such that all citations originating from the same progeny case are assigned to the same fold. This design ensures that observations from a given opinion never appear in both the training and test sets, which would otherwise lead to overly optimistic performance estimates due to within-opinion dependence (Roberts et al. 2017).

For the term-panel models, our procedure differs as to reflect the time-series cross-sectional design of that data. Specifically, we use what’s known as an “evaluation on a rolling forecasting origin” (Hyndman and Athanasopoulos 2021, ch. 5.10). In this proce-

dure, the models are trained on all terms up to term t , and then evaluated on observations from the subsequent term ($t + 1$). The training window is then expanded sequentially one term at a time, up until the penultimate term. This approach ensures that predictions are made using only information available prior to the progeny term being predicted, such that no future information enters model estimation (also known as “look-ahead bias” (Yae 2024)). As with the citation-level models, performance metrics are aggregated across validation iterations. Because test sets differ in size across iterations, we weight performance metrics by the number of out-of-sample observations in each test set. This weighting is equivalent to pooling all out-of-sample predictions across iterations and computing performance metrics on the pooled set, ensuring that each held-out observation contributes equally to the overall evaluation. In the term-panel data, later terms typically contain more observations because the number of origin precedents increases over time; consequently, iterations using later terms as test sets contribute more observations to the pooled evaluation.

We evaluate model performance using three complementary metrics. First, log loss measures the quality of probabilistic predictions by penalizing confident but incorrect predictions; lower values indicate better model fit. Second, Brier scores capture the mean squared error of predicted probabilities relative to the observed outcomes; again lower values indicating better performance. Finally, area under the ROC curve (AUC) measures each model’s ability to discriminate between negative and non-negative treatment across all possible classification thresholds; higher values indicate better discrimination.

Because no single metric fully captures predictive performance, we emphasize patterns that are consistent across three measures. In addition, because the scores themselves are not intuitive, we focus mainly on relative rankings of model performance rather than on metric values themselves.¹⁶ Table 5 presents a summary of the rankings across the 24 models with controls. For the term-panel and citation-level data, we separately sort each of the 12 models by their average rank across the three metrics, from best performing to worst performing.

¹⁶See Section A.4.3 in the Appendix for the full results.

The 2nd and 3rd columns indicate which origin-progeny measure the model uses, along with which distance metric (signed vs. absolute), while the last three columns show the rankings of each of the three metrics (with some small deviations, the rankings are very consistent).

The results in Table 5 do not point to a single dominant specification. But collectively they do point to some firm takeaways.¹⁷ First, not surprisingly, the origin author models tend to perform relatively poorly. In the term-panel data, they comprise the four worst performing models, and in the citation-level data, they appear no higher than 5 in any of the rankings (and comprise the two worst performing models.) Taking together with the results above, the evidence does not support a strong theory of author influence in Supreme Court bargaining.

Differentiating among the two models is less straightforward. However, quite interestingly, among the models that do *not* include the author origin measure, the models that include the mean of the conservative bloc as the progeny measure uniformly performs better than the models that use progeny median. To the extent that the conservative bloc measure is a useful proxy for where the median of the majority lies in many cases, this suggests that the Court’s decision to negatively treat existing precedent is more consistent with power existing within the median of the majority rather than the median of the Court overall. With respect to *origin* medians, here the results differ across the two datasets. In the term-panel data, the “Progeny conservative bloc–Origin median” model performs better than the “Progeny conservative bloc–Origin MMC model,” but the reverse is true in the citation data. But both perform better than a median-to-median model, again suggesting that the bargaining and voting institutions on the Court point away from a strong median voter theorem explanation.

¹⁷Note the choice of signed versus absolute distance generally does not matter, as the rankings do not differ meaningfully across the two.

Average rank	Progeny-origin measure	Distance type	Rank: log loss	Rank: brier	Rank: AUC
<i>Term-panel data</i>					
1.7	Progeny conservative bloc – Origin median	Absolute	1	1	3
2.3	Progeny conservative bloc – Origin MMC	Signed	3	3	1
2.7	Progeny conservative bloc – Origin median	Signed	2	2	4
3.3	Progeny conservative bloc – Origin MMC	Absolute	4	4	2
5.0	Progeny median – Origin MMC	Absolute	5	5	5
6.3	Progeny median – Origin median	Absolute	7	6	6
6.7	Progeny median – Origin MMC	Signed	6	7	7
8.0	Progeny median – Origin median	Signed	8	8	8
9.0	Progeny conservative bloc – Origin author	Absolute	9	9	9
10.0	Progeny conservative bloc – Origin author	Signed	10	10	10
11.0	Progeny median – Origin author	Absolute	11	11	11
12.0	Progeny median – Origin author	Signed	12	12	12
<i>Citation-level data</i>					
1.0	Progeny conservative bloc – Origin MMC	Absolute	1	1	1
2.0	Progeny conservative bloc – Origin MMC	Signed	2	2	2
3.0	Progeny conservative bloc – Origin median	Absolute	3	3	3
4.0	Progeny conservative bloc – Origin median	Signed	4	4	4
5.7	Progeny median – Origin MMC	Signed	5	5	7
5.7	Progeny conservative bloc – Origin author	Absolute	6	6	5
6.7	Progeny conservative bloc – Origin author	Signed	7	7	6
8.3	Progeny median – Origin median	Absolute	9	8	8
9.0	Progeny median – Origin MMC	Absolute	8	9	10
9.7	Progeny median – Origin median	Signed	10	10	9
11.0	Progeny median – Origin author	Signed	11	11	11
12.0	Progeny median – Origin author	Absolute	12	12	12

Table 5: Predictive performance of ideological distance models across datasets

6 Conclusion

Who controls the content of Supreme Court opinions? For many year, scholars have debated whether doctrine is best understood as reflecting the preferences of the median justice, the median of the majority coalition, or the opinion author. Although a rich theoretical and empirical literature has developed around these competing accounts, adjudicating among them has proven difficult, in part because their observable implications often overlap when measured using case outcomes alone.

This paper advances that debate by shifting attention from contemporaneous opinion locations to the subsequent treatment of precedent. By examining how later Courts negatively treat earlier decisions, we exploit the temporal dimension of Supreme Court decision-making to generate new leverage for distinguishing among bargaining models. If doctrine is written at different implied ideological locations under competing theories, then ideological shifts in Court composition should differentially affect the durability of those precedents. Using both term-panel and citation-level data on negative treatments of Supreme Court majority opinions between 1953 and 2023, we show that measures based on Court medians—especially the median of the majority coalition—consistently outperform author-based measures in predicting negative treatment. These results are robust across alternative distance specifications and are reinforced by out-of-sample cross-validation comparisons.

Our findings contribute to the literature in three ways. First, they provide new empirical support for median-of-the-majority coalition theories of Supreme Court bargaining. While author-based models allow for meaningful agenda-setting advantages, the evidence suggests that doctrinal outcomes are more tightly constrained by coalition dynamics than by individual author preferences. Second, by leveraging precedent durability rather than contemporaneous voting behavior, we demonstrate a novel strategy for theory adjudication that may prove useful in other institutional settings. Finally, our use of cross-validation places competing models on a common predictive footing, helping to move the debate beyond in-sample significance tests toward comparative model performance.

These results contribute to our understanding of bargaining and power on the Supreme Court in general. But they also have important implications for understanding the current Supreme Court in particular. The emergence of a 6–3 conservative majority following the 2020 appointment of Justice Amy Coney Barrett has raised the question of whether influence, on average, resides in a relatively centrist conservative bloc—often associated with Chief Justice Roberts and Justices Kavanaugh and Barrett—or whether the Court’s more conservative members, including Justices Thomas, Alito, and Gorsuch, are more likely to shape doctrinal outcomes. Our results—building on earlier research finding support for median-of-the-majority theories (Clark and Lauderdale 2010, Carrubba et al. 2012)—suggest that influence is more likely to be concentrated toward the conservative wing of the Court rather than at its ideological center.

More broadly, our findings suggest that the institutional structure of opinion production—specifically, the requirement that precedential opinions command majority support—continues to shape the locus of judicial power. While the median justice may be pivotal for case dispositions, the content of doctrine appears more closely aligned with the center of the governing coalition. As the Court evolves in composition and ideology, understanding the size, cohesion, and internal bargaining dynamics of that coalition will remain central to predicting the durability of precedent and the direction of legal change.

References

- Bailey, Michael A. 2007. “Comparable Preference Estimates Across Time and Institutions for the Court, Congress, and Presidency.” *American Journal of Political Science* 51(3):433–448.
- Bailey, Michael A. 2013. “Is Today’s Court the Most Conservative in Sixty Years? Challenges and Opportunities in Measuring Judicial Preferences.” *The Journal of Politics* 75(3):821–834.
URL: <https://www.journals.uchicago.edu/doi/abs/10.1017/S0022381613000443>
- Bailey, Michael A. 2020. “Bridge Ideal Points.” <https://michaelbailey.georgetown.domains/bridge-ideal-points-2020/>. Accessed: 2025-12-15. Supreme Court, Congress, and Presidential ideal points extending estimates through 1950–2020.
- Benjamin, Stuart Minor, ByungKoo Kim and Kevin M. Quinn. 2024. “Partisan Panel Com-

- position and Reliance on Earlier Opinions in the Circuit Courts.” *Journal of Law and Empirical Analysis* 1(1):1–13.
URL: <https://journals.sagepub.com/doi/10.1177/2755323X241240384>
- Benjamin, Stuart Minor and Georg Vanberg. 2016. “Judicial Retirements and the Staying Power of U.S. Supreme Court Decisions.” *Journal of Empirical Legal Studies* 13(1):5–26.
URL: <https://onlinelibrary.wiley.com/doi/10.1111/jels.12095>
- Benjamin, Stuart Minor, Kevin M. Quinn and ByungKoo Kim. 2023. “Twenty-First Century Split: Partisan, Racial, and Gender Differences in Circuit Judges Following Earlier Opinions.” *Brigham Young University Law Review* 49(2):367–464. Available at: Digital Commons, Brigham Young University Law Review.
URL: <https://digitalcommons.law.byu.edu/lawreview/vol49/iss2/7/>
- Black, Ryan C. and II Spriggs, James F. 2013. “The Citation and Depreciation of U.S. Supreme Court Precedent.” *Journal of Empirical Legal Studies* 10(2):325–358.
URL: <https://onlinelibrary.wiley.com/doi/10.1111/jels.12012>
- Cameron, Charles M., Jee-Kwang Park and Deborah Beim. 2009. “Shaping Supreme Court Policy through Appointments: The Impact of a New Justice.” *Minnesota Law Review* 93:1820–1870.
- Cameron, Charles M and Jonathan P. Kastellec. 2016. “Are Supreme Court Nominations a Move-the-Median Game?” *American Political Science Review* 110(4):778–97.
- Cameron, Charles M. and Jonathan P. Kastellec. 2023. *Making the Supreme Court: The Politics of Appointments, 1930–2020*. United Kingdom: Oxford University Press.
URL: <https://global.oup.com/academic/product/making-the-supreme-court-9780197680537>
- Cameron, Charles M. and Jonathan P. Kastellec. 2024. “The Supreme Court: How Did We Get Here? And What Comes Next?” *Annals of the American Academy of Political and Social Science* 713(1):35–56. First published online June 13, 2025.
URL: <https://journals.sagepub.com/doi/10.1177/00027162251324879>
- Cameron, Charles M and Lewis A Kornhauser. 2017. *Theorizing the US Supreme Court*.
- Carrubba, Cliff, Barry Friedman, Andrew D Martin and Georg Vanberg. 2012. “Who Controls the Content of Supreme Court Opinions?” *American Journal of Political Science* 56(2):400–412.
- Clark, Tom. 2012. Bargaining and Opinion Writing in the US Supreme Court. In *New Directions in Judicial Politics*, ed. Kevin McGuire. Routledge pp. 186–204.
- Clark, Tom S. and Benjamin Lauderdale. 2010. “Locating Supreme Court Opinions in Doctrine Space.” *American Journal of Political Science* .

- Cottrell, David, Charles R. Shipan and Richard J. Anderson. 2019. "The Power to Appoint: Presidential Nominations and Change on the Supreme Court." *Journal of Politics* 81(3):1057–68.
- Cranmer, Skyler J. and Bruce A. Desmarais. 2017. "What Can We Learn from Predictive Modeling?" *Political Analysis* 25(4):437–449.
- Cross, Frank B., II Spriggs, James F., Timothy R. Johnson and Paul J. Wahlbeck. 2010. "Citations in the U.S. Supreme Court: An Empirical Study of Their Use and Significance." *Illinois Law Review* (2):489–576.
- Epstein, Lee and Jack Knight. 1997. *The Choices Justices Make*. Sage.
- Fortas, Abe. 1975. "Dedication: Chief Justice Warren: The Enigma of Leadership." *Yale Law Journal* 84(3):405–412.
- Fowler, James H., Timothy R. Johnson, II Spriggs, James F., Sangick Jeon and Paul Wahlbeck. 2007. "Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court." *Political Analysis* 15(3):324–346.
- Gelman, Andrew and Iain Pardoe. 2007. "Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components." *Sociological Methodology* 37(1):23–51.
URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9531.2007.00181.x>
- Hammond, Thomas H., Chris W. Bonneau and Reginald S. Sheehan. 2005. *Strategic Behavior And Policy Choice On The U.S. Supreme Court*. Palo Alto, CA: Stanford University Press.
- Hanmer, Michael J. and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1):263–277.
- Hansford, Thomas G. and II Spriggs, James F. 2006. *The Politics of Precedent on the U.S. Supreme Court*. Princeton, NJ: Princeton University Press.
- Hansford, Thomas G., II Spriggs, James F. and Anthony A. Stenger. 2013. "The Information Dynamics of Vertical Stare Decisis." *The Journal of Politics* 75(4):894–906.
URL: <https://doi.org/10.1017/S0022381613000686>
- Harrow, Jason. 2007. "Justice Kennedy's remarkable OT06.". Accessed February 10, 2026.
URL: <https://www.scotusblog.com/2007/06/justice-kennedys-remarkable-ot06/>
- Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *The Elements of Statistical Learning*. 2 ed. New York: Springer.
- Hinkle, Rachael K. 2017. "Panel Effects and Opinion Crafting in the U.S. Courts of Appeals." *Journal of Law and Courts* 5(2):313–336.
URL: <https://doi.org/10.1086/692731>

- Howe, Amy. 2021. “In Barrett’s first term, conservative majority is dominant but divided.” Accessed January 30, 2026.
URL: <https://www.scotusblog.com/2021/07/in-barretts-first-term-conservative-majority-is-dominant-but-divided/>
- Hyndman, Rob J. and George Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3 ed. Melbourne, Australia: OTexts.
URL: <https://otexts.com/fpp3/>
- Jacobi, Tonja. 2009. “Competing Models of Judicial Coalition Formation and Case Outcome Determination.” *Journal of Legal Analysis* 1(2):411–458.
- Krehbiel, Keith. 2007. “Supreme Court Appointments as a Move-the-Median Game.” *American Journal of Political Science* 51(2):231–40.
- Lauderdale, Benjamin E. and Tom S. Clark. 2016. “Estimating Vote-Specific Preferences from Roll-Call Data: A Supervised Statistical Model for Latent Ideologies.” *The Journal of Politics* 78(4):1153–1169.
- Lax, Jeffrey R. and Charles M. Cameron. 2007. “Bargaining and Opinion Assignment on the U.S. Supreme Court.” *Journal of Law, Economics, and Organization* 23(2):276–302.
- Lax, Jeffrey R and Kelly T Rader. 2010. “Legal Constraints on Supreme Court Decision Making: Do Jurisprudential Regimes Exist?” *The Journal of Politics* 72(02):273–284.
- Lax, Jeffrey R and Kelly T Rader. 2015. “Bargaining Power in the Supreme Court: Evidence from Opinion Assignment and Vote Switching.” *Journal of Politics* 77(3):648–63.
- LexisNexis. 2025. “Shepard’s Editorial Phrases–Alphabetical List.” <https://www.lexisnexis.com/pdf/lexis-advance/Shepards-Editorial-Phrases-Alphabetical-List.pdf>. Accessed: 2026-2-22.
URL: <https://www.lexisnexis.com/pdf/lexis-advance/Shepards-Editorial-Phrases-Alphabetical-List.pdf>
- LexisNexis. n.d.a. “Shepard’s Citation Guide, Part 1: History.” <https://www.lexisnexis.com/community/insights/legal/b/product-features/posts/shepards-citation-guide-part-1-history>. Accessed: 2025-02-11.
- LexisNexis. n.d.b. “Shepard’s Citations Explained by the Experts.” <https://www.lexisnexis.com/community/insights/legal/b/product-features/posts/shepards-citations-explained-by-the-experts>. Accessed: 2025-02-11.
- Maltzman, Forrest, II Spriggs, James F. and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. Cambridge: Cambridge University Press.
- Murphy, Walter. 1964. *Elements of Judicial Strategy*. Chicago: University of Chicago Press.
- Parameswaran, Giri, Charles M Cameron and Lewis A Kornhauser. 2021. “Bargaining and Strategic Voting on Appellate Courts.” *American Political Science Review* .

- Re, Richard M. 2025. “To a Conservative Warren Court.” *Harvard Law Review* 139(1):1–. Discusses ideological trends in the Supreme Court and characterizes the current Court’s conservative orientation, including the conservative supermajority after Justice Barrett’s appointment.
URL: <https://harvardlawreview.org/print/vol-139/to-a-conservative-warren-court/>
- Richards, Mark J. and Herbert M. Kritzer. 2002. “Jurisprudential Regimes in Supreme Court Decision Making.” *American Political Science Review* 96(2):305–20.
- Roberts, David R, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller et al. 2017. “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.” *Ecography* 40(8):913–929.
- Romer, Thomas and Howard Rosenthal. 1978. “Political Resource Allocation, Controlled Agendas, and the Status Quo.” *Public choice* 33(4):27–43.
- Schwartz, Edward P. 1992. “Policy, Precedent, and Power: A Positive Theory of Supreme Court Decision-Making.” *Journal of Law, Economics, & Organization* 8(2):219–252.
- Segal, Jeffrey A. and Harold J. Spaeth. 1996. “The Influence of Stare Decisis on the Votes of United States Supreme Court Justices.” *American Journal of Political Science* 40(4):971–1003.
- Songer, Donald R. and Reginald S. Sheehan. 1990. “Supreme Court Impact on Compliance and Outcomes: Miranda and New York Times in the United States Courts of Appeals.” *Western Political Quarterly* 43:297–316.
- Spaeth, Harold J. and Jeffrey A. Segal. 1999. *Majority Rule or Minority Will: Adherence to Precedent on the U.S. Supreme Court*. Cambridge University Press.
- Spaeth, Harold J., Lee Epstein, Michael J. Nelson, Andrew D. Martin et al. 2024. “Supreme Court Database, Version 2024 Release 1.” <https://scdb.psu.edu>. Accessed: 2025-02-11.
- Spriggs, James F., II and Thomas G. Hansford. 2000. “Measuring Legal Change: The Reliability and Validity of Shepard’s Citations.” *Political Research Quarterly* 53(2):327–341.
- Spriggs, James F., II and Thomas G. Hansford. 2001. “Explaining the Overruling of U.S. Supreme Court Precedent.” *Journal of Politics* 63(4):1091–1111.
- Ward, Alex. 2022. “Thomas, Kavanaugh, Roberts: Conservatives dominate Supreme Court but lack clear leader.”. Accessed January 30, 2026.
URL: <https://thehill.com/regulation/court-battles/3703025-thomas-kavanaugh-roberts-conservatives-dominate-supreme-court-but-lack-clear-leader/>
- Westerland, Christopher, Jeffrey A. Segal, Lee Epstein, Charles M. Cameron and Sharon A. Comparato. 2010. “Strategic Defiance and Compliance in the U.S. Courts of Appeals.” *American Journal of Political Science* 54(4):891–905. Available via JSTOR.
URL: <http://www.jstor.org/stable/20788776>

Yae, James. 2024. "Unintended look-ahead bias in out-of-sample forecasting." *Applied Economics Letters* 31(10):953–957.

A Appendix

This appendix provides additional results, as well as information on the data and methods used in the paper.

A.1 Full Regression Tables

Tables A-1, A-2, A-3, and A-4 present the full regression results from the models in Table 4 with controls; each table corresponds to each of the four data-distance metric pairs in Table 4.

	Auth-Med	Auth-CBloc	OMed-Med	OMed-CBloc	MMC-Med	MMC-CBloc
Intercept	-3.142*** (0.091)	-3.191*** (0.095)	-3.393*** (0.093)	-3.566*** (0.103)	-3.449*** (0.094)	-3.475*** (0.100)
Distance	-0.021 (0.030)	0.028 (0.023)	0.180** (0.056)	0.251*** (0.048)	0.259*** (0.045)	0.175*** (0.040)
Age (term)	-0.063*** (0.003)	-0.064*** (0.004)	-0.065*** (0.003)	-0.068*** (0.004)	-0.065*** (0.003)	-0.067*** (0.004)
Age ²	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Issue area: Federalism	-0.000 (0.095)	-0.003 (0.096)	-0.019 (0.094)	-0.003 (0.095)	-0.006 (0.093)	-0.001 (0.094)
Issue area: Federal Taxation	-1.009*** (0.148)	-0.921*** (0.160)	-1.047*** (0.148)	-0.962*** (0.159)	-1.012*** (0.148)	-0.936*** (0.159)
Issue area: Miscellaneous	0.535 (0.499)	0.548 (0.500)	0.679 (0.632)	0.708 (0.631)	0.618 (0.642)	0.656 (0.654)
Issue area: Private Action	-8.478*** (0.785)	-8.455*** (0.779)	-7.839*** (0.651)	-7.802*** (0.651)	-7.935*** (0.687)	-7.892*** (0.672)
Issue area: Civil Rights	0.114 (0.063)	0.138* (0.064)	0.103 (0.061)	0.126* (0.062)	0.106 (0.061)	0.131* (0.062)
Issue area: First Amendment	0.276*** (0.069)	0.285*** (0.071)	0.280*** (0.068)	0.285*** (0.070)	0.275*** (0.068)	0.294*** (0.070)
Issue area: Due Process	-0.334** (0.108)	-0.309** (0.111)	-0.346** (0.106)	-0.307** (0.109)	-0.332** (0.106)	-0.305** (0.109)
Issue area: Privacy	0.319 (0.178)	0.343 (0.179)	0.316 (0.176)	0.343 (0.175)	0.317 (0.175)	0.323 (0.176)
Issue area: Attorneys	0.323* (0.153)	0.338* (0.157)	0.389** (0.150)	0.402** (0.155)	0.389** (0.150)	0.400** (0.154)
Issue area: Unions	-0.177 (0.131)	-0.160 (0.143)	-0.208 (0.130)	-0.191 (0.140)	-0.193 (0.129)	-0.175 (0.140)
Issue area: Economic Activity	-0.325*** (0.072)	-0.302*** (0.075)	-0.318*** (0.070)	-0.294*** (0.073)	-0.305*** (0.070)	-0.291*** (0.073)
Issue area: Judicial Power	-0.094 (0.086)	-0.060 (0.088)	-0.152 (0.082)	-0.122 (0.085)	-0.138 (0.082)	-0.119 (0.085)
Authority: Administrative	-0.337* (0.163)	-0.297 (0.175)	-0.123 (0.161)	-0.097 (0.173)	-0.119 (0.161)	-0.102 (0.173)
Authority: Constitutional	0.494*** (0.078)	0.507*** (0.082)	0.743*** (0.080)	0.750*** (0.083)	0.741*** (0.079)	0.746*** (0.083)
Authority: Statutory	-0.169* (0.076)	-0.181* (0.080)	0.064 (0.079)	0.057 (0.082)	0.069 (0.079)	0.049 (0.082)
Origin disposition (liberal)	0.267*** (0.043)	0.271*** (0.045)	0.236*** (0.042)	0.248*** (0.044)	0.196*** (0.042)	0.178*** (0.049)
Non-full opinion	-0.166 (0.089)	-0.162 (0.091)	-1.138*** (0.081)	-1.122*** (0.084)	-1.131*** (0.080)	-1.104*** (0.083)
N	259811	248856	328790	314009	328597	313826

Table A-1: Full regression models for models with controls presented in Table 4, using the term panel data, with absolute distance. Statistical significance is denoted by * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. All models include standard errors clustered by origin case. Models using the conservative bloc as the progeny location only use progeny cases beginning with the 1969 term.

	Auth-Med	Auth-CBloc	OMed-Med	OMed-CBloc	MMC-Med	MMC-CBloc
Intercept	-3.124*** (0.089)	-3.205*** (0.095)	-3.340*** (0.093)	-3.686*** (0.159)	-3.349*** (0.094)	-3.338*** (0.112)
Distance	0.061* (0.027)	0.096 (0.057)	0.225*** (0.043)	0.470* (0.234)	0.117** (0.037)	-0.149 (0.140)
Distance ²	-0.030 (0.017)	-0.026 (0.019)	0.007 (0.045)	-0.082 (0.089)	0.097** (0.033)	0.129* (0.054)
Age (term)	-0.063*** (0.003)	-0.064*** (0.004)	-0.067*** (0.003)	-0.068*** (0.004)	-0.066*** (0.003)	-0.066*** (0.004)
Age ²	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Issue area: Federalism	-0.008 (0.096)	-0.002 (0.096)	-0.011 (0.094)	-0.002 (0.095)	-0.001 (0.093)	0.005 (0.093)
Issue area: Federal Taxation	-1.009*** (0.148)	-0.923*** (0.160)	-1.034*** (0.148)	-0.963*** (0.159)	-1.001*** (0.148)	-0.923*** (0.159)
Issue area: Miscellaneous	0.533 (0.506)	0.537 (0.503)	0.668 (0.637)	0.712 (0.631)	0.607 (0.659)	0.631 (0.648)
Issue area: Private Action	-8.483*** (0.778)	-8.465*** (0.777)	-7.897*** (0.650)	-7.795*** (0.652)	-7.982*** (0.686)	-7.882*** (0.671)
Issue area: Civil Rights	0.108 (0.063)	0.139* (0.064)	0.095 (0.061)	0.125* (0.062)	0.103 (0.061)	0.133* (0.062)
Issue area: First Amendment	0.275*** (0.069)	0.286*** (0.071)	0.279*** (0.067)	0.284*** (0.070)	0.283*** (0.068)	0.289*** (0.070)
Issue area: Due Process	-0.337** (0.109)	-0.306** (0.111)	-0.342** (0.106)	-0.310** (0.109)	-0.328** (0.107)	-0.300** (0.109)
Issue area: Privacy	0.323 (0.177)	0.344 (0.179)	0.314 (0.175)	0.340 (0.176)	0.309 (0.175)	0.333 (0.176)
Issue area: Attorneys	0.319* (0.153)	0.337* (0.157)	0.384* (0.151)	0.400* (0.155)	0.391** (0.149)	0.399** (0.154)
Issue area: Unions	-0.188 (0.132)	-0.160 (0.143)	-0.208 (0.130)	-0.192 (0.140)	-0.190 (0.129)	-0.166 (0.140)
Issue area: Economic Activity	-0.332*** (0.072)	-0.299*** (0.076)	-0.313*** (0.070)	-0.295*** (0.073)	-0.307*** (0.070)	-0.284*** (0.073)
Issue area: Judicial Power	-0.098 (0.086)	-0.057 (0.088)	-0.154 (0.082)	-0.122 (0.085)	-0.142 (0.082)	-0.109 (0.085)
Authority: Administrative	-0.338* (0.163)	-0.299 (0.175)	-0.124 (0.161)	-0.097 (0.173)	-0.125 (0.161)	-0.094 (0.173)
Authority: Constitutional	0.498*** (0.078)	0.509*** (0.082)	0.737*** (0.080)	0.749*** (0.083)	0.740*** (0.079)	0.745*** (0.083)
Authority: Statutory	-0.168* (0.076)	-0.180* (0.080)	0.065 (0.079)	0.057 (0.082)	0.066 (0.079)	0.054 (0.082)
Origin disposition (liberal)	0.242*** (0.044)	0.267*** (0.045)	0.230*** (0.042)	0.249*** (0.044)	0.153*** (0.046)	0.195*** (0.050)
Non-full opinion	-0.168 (0.089)	-0.167 (0.091)	-1.146*** (0.081)	-1.122*** (0.084)	-1.135*** (0.081)	-1.106*** (0.083)
N	259811	248856	328790	314009	328597	313826

Table A-2: Full regression models for models with controls presented in Table 4, using the term panel data, with signed distance. Statistical significance is denoted by * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. All models include standard errors clustered by origin case. Models using the conservative bloc as the progeny location only use progeny cases beginning with the 1969 term.

	Auth-Med	Auth-CBloc	OMed-Med	OMed-CBloc	MMC-Med	MMC-CBloc
Intercept	-0.482*** (0.098)	-0.678*** (0.100)	-0.517*** (0.092)	-0.944*** (0.103)	-0.578*** (0.095)	-0.851*** (0.098)
Distance	0.111** (0.036)	0.160*** (0.028)	0.406*** (0.063)	0.422*** (0.056)	0.361*** (0.056)	0.397*** (0.046)
Age (years since origin)	0.012* (0.005)	0.016** (0.005)	0.008 (0.005)	0.014** (0.005)	0.010* (0.005)	0.015** (0.005)
Age ²	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Issue area: Federalism	0.107 (0.113)	0.114 (0.114)	0.120 (0.111)	0.151 (0.114)	0.115 (0.110)	0.163 (0.114)
Issue area: Federal Taxation	0.068 (0.174)	0.305 (0.185)	0.046 (0.179)	0.241 (0.198)	0.107 (0.177)	0.351 (0.195)
Issue area: Miscellaneous	0.278 (0.645)	0.364 (0.643)	0.222 (0.649)	0.291 (0.682)	0.298 (0.632)	0.459 (0.689)
Issue area: Civil Rights	0.039 (0.081)	0.101 (0.083)	0.047 (0.078)	0.097 (0.081)	0.048 (0.078)	0.118 (0.082)
Issue area: First Amendment	0.351*** (0.089)	0.372*** (0.090)	0.348*** (0.087)	0.354*** (0.091)	0.355*** (0.085)	0.405*** (0.086)
Issue area: Due Process	-0.150 (0.111)	-0.111 (0.117)	-0.100 (0.111)	-0.055 (0.114)	-0.087 (0.111)	0.001 (0.118)
Issue area: Privacy	0.020 (0.174)	0.103 (0.172)	0.037 (0.172)	0.095 (0.167)	0.038 (0.168)	0.119 (0.161)
Issue area: Attorneys	0.230 (0.212)	0.301 (0.215)	0.245 (0.204)	0.319 (0.208)	0.246 (0.204)	0.358 (0.206)
Issue area: Unions	0.096 (0.128)	0.128 (0.146)	0.096 (0.130)	0.102 (0.144)	0.123 (0.130)	0.137 (0.144)
Issue area: Economic Activity	-0.035 (0.083)	-0.013 (0.085)	0.001 (0.080)	0.017 (0.082)	0.019 (0.081)	0.041 (0.083)
Issue area: Judicial Power	0.028 (0.100)	0.059 (0.102)	0.047 (0.094)	0.077 (0.094)	0.074 (0.094)	0.105 (0.094)
Authority: Administrative	0.021 (0.157)	0.026 (0.169)	0.001 (0.150)	0.024 (0.156)	0.006 (0.156)	-0.026 (0.170)
Authority: Constitutional	0.225* (0.088)	0.277** (0.091)	0.229** (0.084)	0.277** (0.087)	0.229** (0.084)	0.274** (0.088)
Authority: Statutory	0.009 (0.086)	-0.006 (0.089)	0.034 (0.083)	0.044 (0.086)	0.038 (0.084)	0.024 (0.087)
Origin disposition (liberal)	0.275*** (0.052)	0.242*** (0.054)	0.288*** (0.050)	0.301*** (0.052)	0.247*** (0.051)	0.103 (0.060)
Non-full opinion	0.010 (0.119)	0.040 (0.118)	0.035 (0.084)	0.051 (0.085)	0.037 (0.084)	0.056 (0.088)
N	10374	9730	10866	10188	10866	10188

Table A-3: Full regression models for models with controls presented in Table 4, using the case-level citation data, with absolute distance. Statistical significance is denoted by * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. All models include standard errors clustered by origin case. Models using the conservative bloc as the progeny location only use progeny cases beginning with the 1969 term.

	Auth-Med	Auth-CBloc	OMed-Med	OMed-CBloc	MMC-Med	MMC-CBloc
Intercept	-0.409*** (0.097)	-0.623*** (0.100)	-0.448*** (0.092)	-1.065*** (0.183)	-0.429*** (0.095)	-0.859*** (0.113)
Distance	0.083** (0.030)	0.085 (0.064)	0.155** (0.057)	0.647* (0.289)	0.201*** (0.049)	0.423** (0.161)
Distance ²	0.027 (0.019)	0.018 (0.022)	0.202*** (0.060)	-0.087 (0.110)	0.151** (0.047)	-0.012 (0.065)
Age (years since origin)	0.011* (0.005)	0.016** (0.005)	0.008 (0.005)	0.014** (0.005)	0.009 (0.005)	0.015** (0.005)
Age ²	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Issue area: Federalism	0.109 (0.113)	0.111 (0.114)	0.131 (0.112)	0.154 (0.114)	0.137 (0.110)	0.164 (0.114)
Issue area: Federal Taxation	0.085 (0.169)	0.302 (0.185)	0.045 (0.178)	0.242 (0.197)	0.135 (0.175)	0.350 (0.194)
Issue area: Miscellaneous	0.284 (0.650)	0.345 (0.639)	0.232 (0.654)	0.290 (0.685)	0.351 (0.655)	0.460 (0.690)
Issue area: Civil Rights	0.036 (0.081)	0.103 (0.083)	0.039 (0.078)	0.093 (0.081)	0.048 (0.079)	0.118 (0.082)
Issue area: First Amendment	0.352*** (0.089)	0.370*** (0.091)	0.341*** (0.088)	0.352*** (0.091)	0.370*** (0.085)	0.406*** (0.086)
Issue area: Due Process	-0.149 (0.113)	-0.105 (0.117)	-0.098 (0.110)	-0.060 (0.114)	-0.059 (0.112)	-0.001 (0.118)
Issue area: Privacy	0.028 (0.172)	0.093 (0.172)	0.026 (0.169)	0.093 (0.166)	0.036 (0.163)	0.118 (0.162)
Issue area: Attorneys	0.218 (0.211)	0.291 (0.215)	0.243 (0.202)	0.318 (0.208)	0.262 (0.201)	0.357 (0.206)
Issue area: Unions	0.091 (0.132)	0.131 (0.146)	0.090 (0.131)	0.100 (0.144)	0.121 (0.131)	0.136 (0.145)
Issue area: Economic Activity	-0.046 (0.084)	-0.015 (0.085)	0.000 (0.081)	0.015 (0.082)	0.020 (0.081)	0.040 (0.083)
Issue area: Judicial Power	0.029 (0.100)	0.059 (0.102)	0.050 (0.093)	0.076 (0.094)	0.076 (0.093)	0.103 (0.094)
Authority: Administrative	0.024 (0.161)	0.034 (0.169)	0.005 (0.150)	0.025 (0.156)	-0.014 (0.159)	-0.027 (0.170)
Authority: Constitutional	0.227** (0.088)	0.276** (0.091)	0.222** (0.084)	0.277** (0.087)	0.220** (0.085)	0.274** (0.088)
Authority: Statutory	0.005 (0.086)	-0.005 (0.089)	0.036 (0.083)	0.044 (0.086)	0.029 (0.084)	0.023 (0.087)
Origin disposition (liberal)	0.234*** (0.053)	0.245*** (0.055)	0.286*** (0.050)	0.302*** (0.052)	0.153** (0.056)	0.103 (0.062)
Non-full opinion	-0.000 (0.120)	0.038 (0.118)	0.037 (0.085)	0.049 (0.085)	0.036 (0.086)	0.056 (0.088)
N	10374	9730	10866	10188	10866	10188

Table A-4: Full regression models for models with controls presented in Table 4, using the case-level citation data, with signed distance. Statistical significance is denoted by * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. All models include standard errors clustered by origin case. Models using the conservative bloc as the progeny location only use progeny cases beginning with the 1969 term.

A.2 Average predictive effects

This appendix provides details on the procedure used to compute the average predictive effects (APEs) of ideological distance depicted in Figure 5. The APEs summarize how the predicted probability of negative treatment changes as ideological distance varies across its observed support, while averaging over the empirical distribution of all other covariates.

A.3 Model and notation

For each dataset (citation-level or term-panel), let $i = 1, \dots, N$ index observations, let $Y_i \in \{0, 1\}$ indicate whether the observation is coded as a negative citation/treatment, let d_i denote the ideological distance variable of interest, and let \mathbf{z}_i denote the remaining covariates. The fitted model is a logistic regression of the form:

$$\Pr(Y_i = 1 \mid d_i, \mathbf{z}_i) \equiv p_i(d_i, \mathbf{z}_i; \boldsymbol{\beta}) = \Lambda(\eta_i) = \frac{1}{1 + \exp(-\eta_i)}, \quad (\text{A-1})$$

where the linear predictor η_i is

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}. \quad (\text{A-2})$$

\mathbf{x}_i denotes the row of the design matrix corresponding to observation i , including an intercept, the distance term, and all other covariates (and, where specified, a quadratic term in distance). In models with a quadratic distance term, the linear predictor includes both d_i and d_i^2 :

$$\eta_i = \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + \mathbf{z}_i^\top \boldsymbol{\gamma}. \quad (\text{A-3})$$

Next, we fix a distance value d and define the *average predicted probability* at d as

$$\bar{p}(d; \boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^N p_i(d, \mathbf{z}_i; \boldsymbol{\beta}), \quad (\text{A-4})$$

where $p_i(d, \mathbf{z}_i; \boldsymbol{\beta})$ denotes the model-implied probability when the distance variable is set to d for all observations, while each observation retains its observed covariate values \mathbf{z}_i . Equation (A-4) shows that the APE profile averages predictions over the empirical covariate

distribution rather than evaluating effects at pre-specified covariate values.

In practice, we evaluate (A-4) over a grid of distance values spanning the empirical support of the distance variable in the estimation sample. For signed (raw) distance measures, the grid spans the observed minimum and maximum:

$$d \in [\min_i d_i, \max_i d_i]. \quad (\text{A-5})$$

For absolute distance measures, which are bounded below by zero, the grid spans $[0, \max_i d_i]$.

To quantify uncertainty in $\bar{p}(d; \boldsymbol{\beta})$, we use simulation from an approximate sampling distribution for the coefficient vector. Let $\widehat{\boldsymbol{\beta}}$ denote the estimated coefficient vector and let $\widehat{\mathbf{V}}$ denote the cluster-robust variance-covariance matrix used for inference. We draw

$$\boldsymbol{\beta}^{(s)} \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{V}}), \quad s = 1, \dots, S, \quad (\text{A-6})$$

and for each draw compute

$$\bar{p}^{(s)}(d) \equiv \bar{p}(d; \boldsymbol{\beta}^{(s)}) = \frac{1}{N} \sum_{i=1}^N p_i(d, \mathbf{z}_i; \boldsymbol{\beta}^{(s)}). \quad (\text{A-7})$$

The reported APE curve at d is the mean across simulations

$$\widehat{\text{APE}}(d) = \frac{1}{S} \sum_{s=1}^S \bar{p}^{(s)}(d), \quad (\text{A-8})$$

and the corresponding 95% uncertainty interval is given by the 2.5th and 97.5th percentiles of $\{\bar{p}^{(s)}(d)\}_{s=1}^S$:

$$\left[\widehat{\text{APE}}_{0.025}(d), \widehat{\text{APE}}_{0.975}(d) \right] = \left[\text{Quantile}_{0.025}(\bar{p}^{(1)}(d), \dots, \bar{p}^{(S)}(d)), \text{Quantile}_{0.975}(\bar{p}^{(1)}(d), \dots, \bar{p}^{(S)}(d)) \right]. \quad (\text{A-9})$$

We use 500 simulations per model.

A.4 Cross-validation procedures

This section provides a mathematical description of the two cross-validation procedures used to evaluate out-of-sample predictive performance: (i) clustered K -fold cross-validation for citation-level models, and (ii) rolling-origin (forward-chaining) evaluation for term-panel models. In both settings, let $y_i \in \{0, 1\}$ denote the binary outcome (negative treatment), x_i the covariate vector, and let $\hat{p}_i \equiv \Pr(y_i = 1 \mid x_i; \hat{\theta})$ denote predicted probabilities from a fitted logit model with parameters $\hat{\theta}$.

Throughout, out-of-sample performance is evaluated using metrics computed on held-out test observations. When aggregating performance across iterations, we weight by test-set size so that each out-of-sample observation contributes equally to the overall reported performance.

For any given set of held-out test observations \mathcal{T} , predicted probabilities \hat{p}_i are evaluated using the following metrics:

Log loss (cross-entropy). Define the per-observation log loss

$$\ell_i(\hat{p}_i) = -\left(y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)\right). \quad (\text{A-10})$$

The test-set log loss is the average

$$\text{LogLoss}(\mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} \ell_i(\hat{p}_i). \quad (\text{A-11})$$

Brier score. The Brier score is the mean squared error of the probabilistic forecast:

$$\text{Brier}(\mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} (y_i - \hat{p}_i)^2. \quad (\text{A-12})$$

Area under the ROC curve (AUC). Let \hat{p}_i be the score used to rank observations. The AUC can be defined as the probability that a randomly drawn positive observation receives

a higher score than a randomly drawn negative observation:

$$\text{AUC}(\mathcal{T}) = \Pr(\widehat{p}_{i^+} > \widehat{p}_{i^-}), \quad (\text{A-13})$$

where i^+ indexes $y = 1$ observations and i^- indexes $y = 0$ observations within \mathcal{T} (with standard tie handling).

A.4.1 Citation-level models: clustered K -fold cross-validation

Let the citation-level dataset be indexed by $i = 1, \dots, N$, and let $g(i) \in \{1, \dots, G\}$ denote the progeny-case cluster associated with observation i . Clustered K -fold cross-validation partitions the set of clusters $\{1, \dots, G\}$ into K disjoint subsets (folds) $\mathcal{G}_1, \dots, \mathcal{G}_K$, so that each cluster belongs to exactly one fold.

Define the test set for fold k as

$$\mathcal{T}_k = \{i : g(i) \in \mathcal{G}_k\}, \quad (\text{A-14})$$

and the corresponding training set as its complement

$$\mathcal{R}_k = \{i : g(i) \notin \mathcal{G}_k\}. \quad (\text{A-15})$$

By construction, all observations from the same progeny case are assigned to the same fold, so no information from a single case appears in both training and test sets.

Model fitting and prediction. For each fold $k \in \{1, \dots, K\}$, we fit the model on \mathcal{R}_k to obtain parameter estimates $\widehat{\theta}^{(-k)}$, then compute out-of-sample predicted probabilities for all $i \in \mathcal{T}_k$:

$$\widehat{p}_i^{(-k)} = \Pr(y_i = 1 \mid x_i; \widehat{\theta}^{(-k)}), \quad i \in \mathcal{T}_k. \quad (\text{A-16})$$

We then compute a performance metric $M(\mathcal{T}_k)$ using only the held-out observations in \mathcal{T}_k .

Aggregation across folds. Let $n_k = |\mathcal{T}_k|$ denote the number of test observations in fold k . Overall performance is the test-set-size-weighted average:

$$\bar{M} = \sum_{k=1}^K \omega_k M(\mathcal{T}_k), \quad \text{where } \omega_k = \frac{n_k}{\sum_{j=1}^K n_j}. \quad (\text{A-17})$$

This weighting is equivalent to computing the metric on the pooled set of out-of-sample predictions when the metric is an average of per-observation contributions, and it ensures each held-out observation contributes equally even when fold sizes vary due to clustering.

A.4.2 Term-panel models: rolling-origin (forward-chaining) evaluation

Let the term-panel dataset be indexed by term. Let \mathcal{I}_t denote the set of observations in term t , for terms $t \in \{t_{\min}, \dots, t_{\max}\}$. With a one-term forecast horizon ($h = 1$), rolling-origin evaluation proceeds by selecting a sequence of cutoff terms t for which (i) training data exist and contain sufficient outcome variation to estimate the model, and (ii) the next term $t + 1$ exists as a test period.

For each admissible cutoff t , define the training set as all observations up to and including term t :

$$\mathcal{R}_t = \bigcup_{s \leq t} \mathcal{I}_s, \quad (\text{A-18})$$

and define the test set as the subsequent term:

$$\mathcal{T}_{t+1} = \mathcal{I}_{t+1}. \quad (\text{A-19})$$

This design respects temporal ordering: the test period always occurs strictly after the training period.

Model fitting and prediction. For each cutoff t , we fit the model on \mathcal{R}_t to obtain $\hat{\theta}^{(\leq t)}$, then predict outcomes in the next term:

$$\hat{p}_i^{(\leq t)} = \Pr(y_i = 1 \mid x_i; \hat{\theta}^{(\leq t)}), \quad i \in \mathcal{T}_{t+1}. \quad (\text{A-20})$$

We compute a performance metric $M(\mathcal{T}_{t+1})$ using only the held-out observations in term $t + 1$.

Aggregation across rolling iterations. Let $n_{t+1} = |\mathcal{T}_{t+1}|$ denote the number of test observations in term $t + 1$. Overall performance is aggregated as a test-set-size-weighted average across all evaluated test terms:

$$\bar{M} = \sum_{t \in \mathcal{S}} \omega_{t+1} M(\mathcal{T}_{t+1}), \quad \text{where } \omega_{t+1} = \frac{n_{t+1}}{\sum_{u \in \mathcal{S}} n_{u+1}}, \quad (\text{A-21})$$

and \mathcal{S} denotes the set of admissible cutoff terms. This weighting assigns greater influence to test terms with more observations and ensures that each out-of-sample observation contributes equally to the overall evaluation. By construction, all predictions are genuinely out-of-sample and avoid incorporating information from future terms into model estimation (i.e., avoiding look-ahead bias).

A.4.3 Full cross-validation results

Tables A-5 and A-6 present the full cross-validation results presented in Section 5.2.3.

Log loss		
Progeny-origin measure	Distance type	Value
Progeny conservative bloc – Origin median	Absolute	0.0748
Progeny conservative bloc – Origin median	Signed	0.0748
Progeny conservative bloc – Origin MMC	Signed	0.0748
Progeny conservative bloc – Origin MMC	Absolute	0.0749
Progeny median – Origin MMC	Absolute	0.0771
Progeny median – Origin MMC	Signed	0.0772
Progeny median – Origin median	Absolute	0.0774
Progeny median – Origin median	Signed	0.0776
Progeny conservative bloc – Origin author	Absolute	0.0882
Progeny conservative bloc – Origin author	Signed	0.0882
Progeny median – Origin author	Absolute	0.0918
Progeny median – Origin author	Signed	0.0995
Brier score		
Progeny-origin measure	Distance type	Value
Progeny conservative bloc – Origin median	Absolute	0.0151
Progeny conservative bloc – Origin median	Signed	0.0151
Progeny conservative bloc – Origin MMC	Signed	0.0151
Progeny conservative bloc – Origin MMC	Absolute	0.0151
Progeny median – Origin MMC	Absolute	0.0155
Progeny median – Origin median	Absolute	0.0155
Progeny median – Origin MMC	Signed	0.0155
Progeny median – Origin median	Signed	0.0156
Progeny conservative bloc – Origin author	Absolute	0.0181
Progeny conservative bloc – Origin author	Signed	0.0181
Progeny median – Origin author	Absolute	0.0187
Progeny median – Origin author	Signed	0.0189
AUC		
Progeny-origin measure	Distance type	Value
Progeny conservative bloc – Origin MMC	Signed	0.7153
Progeny conservative bloc – Origin MMC	Absolute	0.7151
Progeny conservative bloc – Origin median	Absolute	0.7148
Progeny conservative bloc – Origin median	Signed	0.7148
Progeny median – Origin MMC	Absolute	0.7147
Progeny median – Origin median	Absolute	0.7139
Progeny median – Origin MMC	Signed	0.7134
Progeny median – Origin median	Signed	0.7126
Progeny conservative bloc – Origin author	Absolute	0.6725
Progeny conservative bloc – Origin author	Signed	0.6725
Progeny median – Origin author	Absolute	0.6708
Progeny median – Origin author	Signed	0.6705

Table A-5: Cross-validated predictive performance by model, using the term-panel data.

Log loss		
Progeny-origin measure	Distance type	Value
Progeny conservative bloc – Origin MMC	Absolute	0.6800
Progeny conservative bloc – Origin MMC	Signed	0.6801
Progeny conservative bloc – Origin median	Absolute	0.6811
Progeny conservative bloc – Origin median	Signed	0.6812
Progeny median – Origin MMC	Signed	0.6833
Progeny conservative bloc – Origin author	Absolute	0.6835
Progeny conservative bloc – Origin author	Signed	0.6839
Progeny median – Origin MMC	Absolute	0.6842
Progeny median – Origin median	Absolute	0.6842
Progeny median – Origin median	Signed	0.6845
Progeny median – Origin author	Signed	0.6865
Progeny median – Origin author	Absolute	0.6873
Brier score		
Progeny-origin measure	Distance type	Value
Progeny conservative bloc – Origin MMC	Absolute	0.2435
Progeny conservative bloc – Origin MMC	Signed	0.2436
Progeny conservative bloc – Origin median	Absolute	0.2441
Progeny conservative bloc – Origin median	Signed	0.2441
Progeny median – Origin MMC	Signed	0.2452
Progeny conservative bloc – Origin author	Absolute	0.2452
Progeny conservative bloc – Origin author	Signed	0.2454
Progeny median – Origin median	Absolute	0.2455
Progeny median – Origin MMC	Absolute	0.2456
Progeny median – Origin median	Signed	0.2457
Progeny median – Origin author	Signed	0.2467
Progeny median – Origin author	Absolute	0.2471
AUC		
Progeny-origin measure	Distance type	Value
Progeny conservative bloc – Origin MMC	Absolute	0.5903
Progeny conservative bloc – Origin MMC	Signed	0.5900
Progeny conservative bloc – Origin median	Absolute	0.5882
Progeny conservative bloc – Origin median	Signed	0.5881
Progeny conservative bloc – Origin author	Absolute	0.5802
Progeny conservative bloc – Origin author	Signed	0.5787
Progeny median – Origin MMC	Signed	0.5758
Progeny median – Origin median	Absolute	0.5757
Progeny median – Origin median	Signed	0.5743
Progeny median – Origin MMC	Absolute	0.5740
Progeny median – Origin author	Signed	0.5655
Progeny median – Origin author	Absolute	0.5621

Table A-6: Cross-validated predictive performance by model, using the citation-level data.