# The Hidden Effects of Judge Ideology in Federal District Courts*

Ryan Hübert† and Ryan Copus‡

September 10, 2019

In contrast to robust findings of partisan decision-making by federal appellate judges, researchers have frequently found that judges' partisanship does not substantially shape outcomes in district courts. Why? We use a newly constructed dataset of all civil rights cases in seven courts between 1995 and 2016 to demonstrate that partisan effects are largely *hidden*. First, we find modest partisan effects in our entire dataset, but they are largely buried in litigant-driven outcomes, such as settlements. We argue that prior research may not have detected these effects because of the common, but incorrect, practice of dropping settled cases. Second, we find large partisan effects in some subsets of cases, but they are hidden by a large number of potentially "frivolous" cases with small partisan effects. By figuring out why partisanship appears to matter so little in district courts, we paradoxically demonstrate that it actually matters *a lot*.

**Keywords:** judicial politics, trial courts, causal inference, heterogeneous effects

Word count: 11,853

*Replication materials to be made available online.*

*This is one of several joint papers by the authors on judicial decision-making in the federal courts; the ordering of names reflects a principle of rotation. We thank Joshua Boston, Alex Coppock, Mark Hurwitz, John Kastellec, A.K. Shauku, Jeffrey Yates and participants at the 2019 Conference on Institutions and Lawmaking at Emory University, the 2019 Conference on Data Science and Law at ETH Zürich, the 2019 meeting of the Southern Political Science Association and the 2018 and 2019 meetings of the American Political Science Association for helpful comments.

†Ryan Hübert is Assistant Professor of Political Science at the University of California, Davis, One Shields Avenue, Davis CA 95616. Email: `rhubert@ucdavis.edu`. *Corresponding author*.

‡Ryan Copus is a Climenko Fellow at Harvard Law School, 1525 Massachusetts Avenue, Cambridge MA 02138. Email: `rcopus@law.harvard.edu`.

A puzzling feature of research on federal district judges is its inability to consistently or reliably detect differences between the decisions made by Democratic and Republican appointees (for a review, see Boyd 2017).[1] This stands in contrast to well-known differences among judges at the appellate level, and it has led to growing view that political ideology is a relatively unimportant factor in district court decision-making. And yet, political actors apparently care about who gets appointed to the district courts. Presidents vet nominees to gauge their judicial philosophies, senators question nominees on politically salient issues, and ideologically motivated advocacy groups like the Federalist Society and the Alliance for Justice provide public and private input on the nominees' records.

This paper revisits the claim that political ideology is a relatively unimportant factor in federal district court decision-making. Our core contention is that the effect of district judge partisanship is *hidden* in ways that cause researchers to understate the impact of political ideology in these courts. This hiddenness is partially due to research design choices and partially due to distinct features of district courts (namely, procedural complexity and the volume of cases, see Kim et al. 2009).

First, when focusing specifically on judicial decisions, some prior research has not accounted for the possibility of litigant adaptation (Priest and Klein 1984). In anticipation of ideological decision-making, litigants may alter their behavior in response to the judge they are assigned. This too is an "effect" of the judge's ideology. To the extent that traditional analyses systematically ignore litigants' decisions (by, for example, dropping settled cases), these kinds of effects will be hidden. Doing this also poses a serious methodological problem since it invalidates the assumption that judges are as-if randomly assigned to the cases in a dataset. For example, if litigants differentially select out of litigation on the basis of judges' political ideologies, then cases ending with a judgment from a Democrat will be quite different than cases ending with a judgment from a Republican. By ignoring this litigant behavior, researchers break randomization and introduce

1. We will often refer to judges appointed by Democratic presidents as "Democrats" and judges appointed by Republican presidents as "Republicans."

post-treatment bias that can hide partisan effects.

Second, even if political ideology plays little role in *all* cases, there may be substantively important ideological effects hidden in specific subsets of cases. This is not an entirely new point. For example, researchers have frequently based their analyses of judicial decision-making on the subset of cases that are resolved with a published judicial opinion, partly on the grounds that they represent the "non-frivolous" cases where ideology can impact outcomes. But, again, selecting on the dependent variable in this manner will undermine the assumption that judges are as-if randomly assigned. We provide a novel machine-learning method for finding the subsets of cases where we expect the largest effects, e.g., the "non-frivolous" cases. Importantly, our technique is designed to find these subsets without breaking randomization.

In the sections below, we marshal a new dataset of all civil rights cases in the seven district courts in Washington, Oregon and California to show that the above concerns are not merely theoretical. Our findings suggest that the effect of ideology in district courts has, at least partially, been hidden. In our full dataset, the largest effect of judge party is on an outcome that pertains to litigant adaptation: whether a case settles. Our results are moderate, but substantively important: across these seven courts, plaintiffs in approximately 184 civil rights cases per year received some sort of settlement from defendants *solely because they were assigned to a Democrat* and plaintiffs in approximately 71 cases per year voluntarily withdrew their civil rights claim (receiving nothing) *solely because they were assigned to a Republican*. We further demonstrate that if we were to drop settled cases from our dataset (as other researchers have done), we would recover effects that are biased significantly downward.

Nonetheless, the results from our full dataset are broadly compatible with prior findings demonstrating a small effect of judge party on case outcomes. But in light of their heavy caseloads, we next explore the possibility that these small effects hide much larger effects in subsets of cases. When we use our novel machine learning technique to focus on subsets of cases where we predict the largest effects, we indeed see very large effects, especially for outcomes driven by litigant adaptation. In the 5% cases where we predict the largest partisan effects (roughly 4,200 cases), assignment to a Republican appointee increases the probability of voluntary dismissal by 10.8 percentage points and decreases the probability of settlement by 21.1 percentage points.

Taken together, these findings provide ample evidence that there are hidden effects of ideology in district courts. We also adjudicate an important debate about the *reason* that ideology appears to matter less than expected in district courts. Some have claimed that it's the judges: district court judges are just less ideological than other kinds of judges (e.g., Zorn and Bowie 2010), either because of the appointment process or because they are constrained by appellate judges. Others have claimed that it's the cases: district courts just hear a large number of frivolous cases where ideology isn't particularly salient (e.g., Epstein, Landes, and Posner 2013). Existing research has had limited success distinguishing between these two explanations. While we are not the first to make this critique (see Cameron and Kornhauser 2017), we are the first to provide a causally identified way to resolve it. Indeed, the large effects we find in subsets of cases is convincing evidence that it's the cases and not the judges or institutions.

By establishing the reason that ideology plays a limited role in the district courts, we paradoxically demonstrate that ideology actually plays a significant role in district court decision-making. Indeed, our analysis demonstrates that partisan effects have typically been so low in prior research due to an artifact of estimation. Even though judge partisanship appears to be a very important factor in the way that many cases get resolved in the district courts, this is not obvious when estimating average treatment effects on pooled datasets since those courts *also* resolve a large number of "frivolous" cases. Our analysis suggests that average treatment effects on pooled datasets may not the most appropriate way to test whether there are partisan effects in district courts.

# 1 Judge Ideology in District Judge Decision-Making

In political science, the conventional wisdom has long been that judges in the U.S., as a whole, are influenced by their personal political ideologies when they make decisions. Researchers have amassed a great deal of empirical evidence that politics matters in judging, but much of it has been concentrated in federal appellate courts (for a review, see chapter 2 of Epstein, Landes, and Posner 2013).

Existing empirical research focused specifically on district courts presents mixed results, but the view among many researchers is that political ideology plays a relatively small role at the district court level (Ashenfelter, Eisenberg, and Schwab 1995; Howard 2002; Perino 2006; Keele et al. 2009; Zorn and Bowie 2010; Epstein, Landes, and Posner 2013; but see Boyd and Hoffman 2010; Boyd and Spriggs II 2009; Carp and Rowland 1996; Schanzenbach and Tiller 2007; Fischman and Schanzenbach 2011). For example, Epstein, Landes, and Posner (2013) presents statistical analyses of district court decision-making, concluding that "ideology plays only a small role at the district court level, even though district judges have considerable discretionary authority" (p. 253).

There are two main explanations offered for why ideology does not appear to matter as much in the district courts. First, the relative lack of ideological effects in the district courts could be explained by *the cases* that district courts decide. Under this theory, "ideology doesn't matter" in the district courts because district courts are tasked with deciding easy, routine cases that leave no room for ideological disagreement. Circuit courts, in contrast, decide the more controversial set of cases that are appealed. This is the argument made in Epstein, Landes, and Posner (2013). For example, the authors point out that a "substantial fraction of cases filed in the district courts (many by persons who do not have legal representation) have no possible merit and so really are just noise in the data" (p. 207).

Second, the relative absence of ideological effects could be explained not by cases, but by differences in the *institutions or judges*. Under these theories, "ideology doesn't matter" because, for example, their decisions do not create new law (and thus offer a lower ideological payoff), they comply with the circuit court's preferences due to fear of reversal, they are less ideologically inclined as a group, or they are simply too busy to worry about advancing their ideology. This is the argument made in Zorn and Bowie (2010), which concludes that the article's "findings robustly support the widely held perception that judges' policy preferences influence their decision-making to a greater extent at higher levels of the judicial hierarchy than at lower ones" (p. 1212).

At a theoretical level, adjudicating between these mechanisms is crucial. Confirming the judge- and institution-based explanation would indeed suggest that ideology really doesn't matter in district courts. By this account, judges are unable to exercise ideological influence no matter which case is before them. Confirming the case-based explanation, however, paradoxically suggests that ideology *does* matter. For example, suppose that 90% of cases are "frivolous" claims where Democratic and Republican appointees prefer the same decision. In the remaining 10% of cases, suppose Democrats and Republicans disagree (on average) about how cases should come out. In particular, suppose that in this 10% of cases, Republicans are 40% more likely to rule for defendants than Democrats. It would strain credulity to argue that this is not a very large and meaningful effect. However, if one were to estimate the effect of Republicans in the entire set of cases, it would seem much less impressive: $0.9 \times 0 + 0.1 \times 0.4 = 0.04$.

It is very challenging to find these effects. As a result, we argue that the role of ideology in district courts is hidden. Unfortunately, many existing studies make only limited progress at finding these effects. First, research designs often do not exploit the random assignment of judges to cases, generating biased estimates. The statistical bias introduced by these research designs could explain why estimated effects are all over the place in studies of district courts. Second, there has been no causally identified way to systematically find the subsets of cases where ideology matters the most. Again, this is important since the presence of large effects in subsets of cases suggests that ideology matters. In the remainder of this section, we discuss these two problems.

## 1.1 Causally Estimating the Effect of District Judge Ideology

In federal courts, judges are randomly assigned to cases. This provides a unique, and rare, opportunity to estimate causally identified effects in an institutional setting. While some research has exploited this opportunity (e.g., Boyd, Epstein, and Martin 2010; Boyd 2013; Kastellec 2013), much of the research on district courts has not. We identify two ways in which past research on district courts has neglected to exploit randomization in district courts.

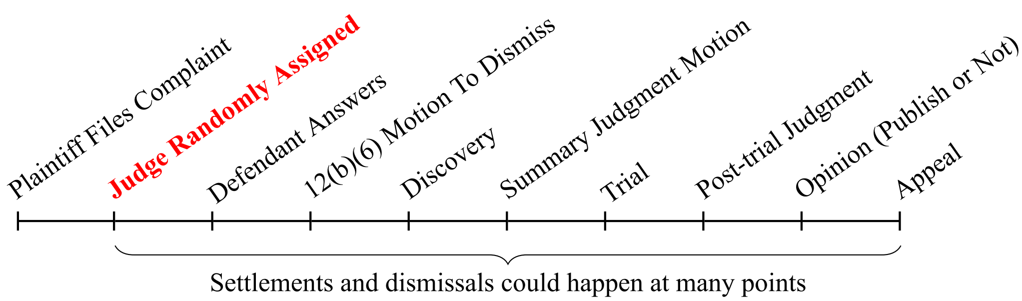
**Bundle of sticks problem** Political ideology is just one feature of a judge's identity. As a result, it is one of the sticks that makes up the "bundle of sticks" defining judges' identities (Sen and Wasow 2016). Then, to isolate the specific causal effect of ideology, one must have reasonable confidence that judges' ideologies are at least conditionally independent of judges' other characteristics. In observational settings, this is not only impossible, it is unreasonable. The specific components of a judge's identity are co-determined long before those judges are randomly assigned to hear cases. Inferences about the effect of judges on case outcomes are causally valid; claims that these effects are driven by specific features of a judge's identity may not be.

The best we can do without introducing the potential for bias is to exploit the fact that judges, *as whole bundles of sticks*, are randomly assigned to cases. One approach is to define the treatment to be whether a Republican appointee was assigned to a case (instead of a Democratic appointee). This is substantively interesting in its own right. How Republican appointees—whether due to their ideologies, ages, genders, races, judicial philosophies etc.—affect litigation outcomes as compared to Democratic appointees informs the stakes of presidential elections for the functioning of the courts. We take this approach and accordingly refer to *partisan* or *Republican* effects rather than *ideological* effects.

**Sampling errors and post-treatment bias** Even with a well defined causal estimand, it is still difficult of estimate effects using case-level data from the district courts. There are two specific issues: too many outcomes and too many cases. Each issue poses specific problems.

District court cases can end in many more ways than appellate cases can. Using the Federal Judicial Center's Integrated Database, which we describe in Section 2, we identify 35 distinct case outcomes in federal district courts, but only eight in federal appeals courts. For researchers keen on finding partisan effects, the complexity of federal litigation raises reasonable questions about what exactly to study. A common approach to simplify this complexity is to gather data on and analyze a particular "decision point" in litigation, such as motions for summary judgment, cases with opinions or cases that were appealed. But such research design choices can undermine the assumption that judges are randomly assigned to cases, yielding biased estimates.

**Figure 1:** *Litigation in federal district courts is complex and has multiple stages. A judge's partisanship can affect outcomes at any decision point after her initial assignment to the case.*




To see why, consider Figure 1, which presents a stylized depiction of the litigation process in federal courts. Random assignment of a judge is one of the very first things that occurs after a plaintiff files a lawsuit. In order to exploit this source of randomization, a researcher must perform analysis on a dataset that contains *all* cases that were randomly assigned to a judge. This is not common in studies of district courts. Perino (2006) and Keele et al. (2009) study a subset of cases that have written opinions. Randazzo (2008) and Zorn and Bowie (2010) study a subset of cases that were eventually appealed. Epstein, Landes, and Posner (2013) studies both a subset of cases with motions to dismiss, and a subset of cases that were appealed. In each of these studies, the analysis was performed on datasets that dropped all settled cases, many or most dismissed cases, and a sizeable number of judgments (e.g., which did not have an opinion or were not appealed). But actions taken by both judges and litigants earlier in litigation mean that these samples of cases do not reflect the initial randomized case assignment.

Some courts scholars have warned about the dangers of drawing inferences from non-representative samples of cases. Clermont and Eisenberg (2002) points out how researchers can make faulty inferences from selected samples of cases: settlement patterns "mean[] that the win rate [in court judgments] reveals something about the set of adjudged cases, a universe dominated by close cases—but reveals little about the underlying, variegated mass of disputes and cases" (p. 138). In a study of veil-piercing suits, Boyd and Hoffman (2010) points out: "This Article asserts that current veil piercing scholarship is founded on sand [...] sampling errors have produced an incoherent picture of veil piercing doctrine; until now, we have been predicting the iceberg by its odd, biased tip" (p. 855). Hoffman, Izenman, and Lidicker (2007) laments that scholars who focus solely on cases that end with opinions "proceed to claim that the dataset of opinions is good enough for statistical inference. After all, scholars should focus on "difficult" cases, not "easy" ones" (p. 687-8).

The specific kind of "sampling" problem that concerns us (dropping cases that have specific outcomes) has implications for casual inference. An equivalent way to express the problem is that a researcher is conditioning on a post-treatment variable. There is a large (and growing) literature warning researchers about the potential of post-treatment bias (see for example, Rosenbaum 1984; Acharya, Blackwell, and Sen 2016; Montgomery, Nyhan, and Torres 2018; Aronow, Baron, and Pinson 2019; Coppock 2019; Knox, Lowe, and Mummolo 2019).

We propose a relatively easy fix for this problem: conduct analyses on the universe of cases that were assigned a judge (or a random sample of this universe). In Section 3.1, we provide evidence that, conditional on courthouse and year, judges are as-if randomly assigned to cases. In Section 3.3 we demonstrate that dropping specific kinds of cases (like those settled or not appealed) will bias partisan effects downward toward zero. Of course, the cost of our approach is figuring out how to study the myriad outcomes that could occur in each case. We opt to study the four most frequent outcomes separately, and leave to future work a more unified analysis. This cost is well worth paying: to our knowledge, we provide the first causally identified effect of judge partisanship in civil litigation in federal district courts.

**Adjudicating between causal mechanisms** To adjudicate between the two mechanisms we describe in the beginning of this section (i.e., cases and judges), we need to do more than estimate unbiased effects. We need to identify the subsets of cases where we would expect partisanship to matter most, and see if there actually are large partisan effects. Past studies of district court decision-making have had limited success. In a review of Epstein, Landes, and Posner (2013), Cameron and Kornhauser (2017) articulate the two mechanisms that explain small partisan effects in district court decision-making (which they term the "DII result"), but emphasize the difficulty adjudicating between these mechanisms: "empirical findings like the DII result do not speak for themselves" (p. 546).

Using advances in machine learning, we are able to adjudcate mechanisms without violating our assumption that judges are randomly assigned to cases. The idea is simple: we subset cases using pre-treatment variables, which are independent of whether a Republican or Democrat is assigned to a case. However, the implementation is not simple. We construct statistical models of how cases would end if assigned to a Democrat and if assigned to a Republican. These models are based solely on pre-treatment variables. Then, we are able to characterize, for each case, the extent to which we predict (again, based on pretreatment variables only) Democrats and Republicans to "disagree" on the appropriate outcome of the case. When we look at cases where we predict a high degree of disagreement, we indeed find it in the actual case outcomes.

This technique allows us to provide convincing and causally identified empirical evidence for the case-based (and against the judge- and institution-based) explanation. In effect, our novel data and novel analytical approach enables us to find the hidden partisan effects. In the next section, we describe our dataset. In the following sections, we present and interpret our main results.

# 2 Data

Our analysis is based on an original dataset of the universe of civil rights cases filed in the U.S. District Courts in Washington, Oregon and California between 1995 and 2016.[2] To construct the dataset, we collected every docket sheet stored in the courts' online record system and used automated text processing methods to extract key pieces of information. Because we rely on docket sheets collected directly from the courts, our dataset does not have missing data problems that are known to introduce bias in datasets derived from other data sources, such as Westlaw and Lexis-Nexis (see Boyd and Hoffman 2010; Burbank 2004; Hoffman, Izenman, and Lidicker 2007; Kim et al. 2009).

However, because we rely on data collected directly from the courts, we are limited to the courts where we received permission to access to PACER. We then use PACER's Nature of Suit codes to identify all civil rights cases, which yielded 94,768 unique docket sheets. We focus on civil rights cases for two reasons.[3] First, after prisoner petitions, civil rights cases make up the largest share of our dataset. More broadly, civil rights is an important area of federal law. Second, limiting our analysis to civil rights cases provides an opportunity for clearer interpretations. We then dropped cases that (1) were denials of *in forma pauperis* status, (2) had no presiding district judge, and (3) were filed in a division of a court that had fewer than 100 cases heard by either Democratic or Republican judges. All in all, we dropped 9,579 cases, yielding a dataset of 85,189. Most of the cases we dropped had no district judge listed. Docket sheets contain a wealth of information, which is divided into three broad categories: case metadata, litigant information and docket entries.

The case metadata is easily collected from each docket sheet. For example, every docket sheet clearly lists the case's jurisdiction, *i.e.*, diversity, federal or U.S. government litigant. We supplement our docket sheet data with data in the Federal Judicial Center's Integrated Database (IDB),

2. For the U.S. District of Oregon, our dataset includes all cases filed between 1997 and 2014. For exact dates that our data covers for each court, see Table A.1 in the appendix.

3. Subsetting to civil rights cases does not break randomization since a case's suit category is determined before a judge is assigned.

available at `https://www.fjc.gov/research/idb`. This database contains additional metadata about each case, which overlaps somewhat with the data we collected directly from the docket sheets. However, there were a handful of variables that we could not capture from the docket sheets but which are available in the IDB. Most importantly, we use the IDB's measure of outcomes for each case.[4] Our case outcome variables were derived from the IDB's `DISP` variable (how the case ended) and its `JUDGMENT` variable (which party won). We merged these together so that, for example, jury verdicts favoring the defendant are considered a different outcome than jury verdicts favoring the plaintiff.

In Table 1, we present sample means for the outcomes in our dataset. Perhaps the most surprising feature of adjudication that is evidenced by our dataset is the relatively small number of civil rights cases that end in a judgment on the merits. For example, nearly half (approximately 46%) of cases are terminated with a transfer, remand or dismissal, while another third (approximately 34%) end with a private settlement between the litigants. Around 13% of cases end with a judgment on a pretrial motion, around 2% of cases end with a trial and the remaining 5% of cases end with another kind of judgment.

The IDB has important limitations, and our docket-based dataset represents an improvement in two ways. First, the Federal Judicial Center does not report judge-identifying information in its publicly released version of the IDB. One of our major contributions is to include this information. Each docket sheet contains a field indicating which judge a case is assigned to. We determined that this field is "dynamic" in the sense that it gets updated every time a case is reassigned to a new judge. Because we wish to preserve the random assignment (and we worry that reassignments could be judge-driven), we base our analysis on the *first* judge assigned to a case. We extract this information by scanning the docket entries of each docket sheet and collecting name of the first judge who was mentioned. If none was mentioned, we default to the judge named in the previously

4. While we do not go into detail here, the IDB was helpful for us in two additional ways. First, it allowed us to verify that our docket sheet database is not missing cases. Second, it allowed us to more cleanly determine when the district court's decision was made, and thus the case's termination date.

**Table 1:** *Case Outcomes in Sample*

| Outcome | Proportion |
|---|---|
| Settlements | 0.3388 |
| Voluntary Dismissals | 0.1830 |
| Involuntary Dismissals | 0.1560 |
| Judgments for Defendant on Pretrial Motion | 0.1103 |
| Remands to a State or an Agency | 0.0498 |
| Transfers | 0.0203 |
| Other Miscellaneous Judgments | 0.0211 |
| Default Judgments for Plaintiff | 0.0025 |
| Dismissals for Failure to Prosecute | 0.0328 |
| Judgments for Plaintiff on Pretrial Motion | 0.0064 |
| Miscellaneous Judgments for Plaintiff | 0.0042 |
| Miscellaneous Judgments for Defendant | 0.0145 |
| Dismissals for Lack of Jurisdiction | 0.0138 |
| Consent Judgments for Plaintiff | 0.0039 |
| Other Judgments on Pretrial Motion | 0.0081 |
| Judgments for Both on Pretrial Motion | 0.0039 |
| Judgments for Defendant after Jury Trial | 0.0127 |
| Consent Judgments for Both | 0.0018 |
| Miscellaneous Judgments for Both | 0.0026 |
| Judgments for Plaintiff after Jury Trial | 0.0054 |
| Judgments for Defendant after Bench Trial | 0.0020 |
| Other | 0.0059 |

mentioned "Assigned to" field. For approximately 7.8% of the cases in our dataset, the party of the presiding judge determined by our method is different than the party of the judge listed as the assigned judge.

Second, the IDB does not contain detailed data on the docket entries or on the parties to the case. However, each docket sheet lists the litigants and litigants' attorneys. We code several variables from this information. First, we count the number of litigants and attorneys on the plaintiff and defendant sides. Then, we generate two variables that allow us to proxy for asymmetries among the litigants. We code whether the litigants and the attorneys on the plaintiff and defense sides are "repeat players" in the seven courts in our dataset.[5] We also code whether the attorneys on

5. We construct this variable before subsetting to civil rights cases, so it identifies repeat players across all civil cases in our data's time-frame.

the plaintiff and defendant sides come from top revenue law firms (using 2018 data on law firm revenue).

**Table 2:** *Case Level Summary Statistics for Litigants*

| Variable | Sample Mean |
|---|---|
| *Litigant and Attorney Counts* | |
| Defendants | 4.12 |
| Plaintiffs | 1.51 |
| Defendant Attorneys | 1.87 |
| Plaintiff Attorneys | 1.64 |
| *Litigant and Attorney Quality* | |
| Pro Se Defendants | 0.03 |
| Pro Se Plaintiffs | 0.25 |
| Top Revenue Law Firms Representing Defendants | 0.01 |
| Top Revenue Law Firms Representing Plaintiffs | 0.00 |
| Repeat Players: Defendants | 1.22 |
| Repeat Players: Plaintiffs | 0.28 |
| Repeat Players: Defendant Attorneys | 1.66 |
| Repeat Players: Plaintiff Attorneys | 1.06 |
| *Litigant Types* | |
| Defendants: Businesses | 0.67 |
| Defendants: Governments | 0.89 |
| Defendants: Individuals | 2.51 |
| Defendants: Other | 0.07 |
| Plaintiffs: Businesses | 0.04 |
| Plaintiffs: Governments | 0.01 |
| Plaintiffs: Individuals | 1.45 |
| Plaintiffs: Other | 0.01 |

Note: Repeat Players variables present the average number of parties/attorneys on a case who appeared at least 20 times in our dataset.

In Table 2, we present basic summary statistics about the litigants in our dataset. First, note that there are many more defendants per case than there are plaintiffs. Moreover, there are indications that defendants in our sample are better resourced and more experienced. There are more pro se plaintiffs per case than pro se defendants, and defendants have on average more attorneys per case than plaintiffs. Defendants and their attorneys are repeat players in the legal system more often than plaintiffs. For example, the average number of defendants per case that are repeat players[6] is 1.66. This statistic is 1.06 for plaintiffs. Finally, the defendants' attorneys are more likely to be from top revenue law firms (using 2018 revenue figures).

6. We define "repeat players" as parties/attorneys who appear in our full dataset at more than 20 times.

In the bottom section of Table 2, we present statistics about the types of litigants in our data set. We used machine learning classification to code these variables. We began with a random sample of 3,632 parties from our data set, which were hand coded into several categories by a research assistant. These categories correspond to whether a party is a government, business, individual or other. We then trained a classifier on this hand-coded sample using the text of the party names to predict their category. We then used these classification models to categorize each party in our entire dataset.

# 3 Analysis of Full Dataset

We now move to our main empirical analysis. In this section, we do three things. First, we present evidence that our original dataset of district court cases does not break randomization. We have qualitative knowledge that suggests there are violations of as-if random assignment in the pooled dataset since random assignment occurs within court divisions and has a temporal aspect. To demonstrate our full dataset preserves as-if random assignment conditional on these known violations, we ask whether *other* pre-treatment variables are additionally predictive of treatment beyond a benchmark model predicting treatment with court division and year. We find they are not additionally predictive, which is evidence that judge assignment within court divisions and years is as-if random.

Second, we perform our main analysis of the effect of judge party on the full dataset. Because there are many possible outcomes in district courts, we focus our attention on the four most prevalent, which together account for 79% of cases: settlements (34% of cases), voluntary dismissals (18% of cases), involuntary dismissals (16% of cases)[7] and judgments for defendants on pretrial motions (11% of cases). Settlements are largely self explanatory: the plaintiff(s) agree with the defendant(s) to end the case in exchange for some benefit to the plaintiff (e.g., money or an in-

7. The IDB classifies these as dismissals for other reasons. However, we opt to label them involuntary dismissals to distinguish them from voluntary dismissals.

junction). Voluntary dismissals involve the plaintiff withdrawing their case, either with or without consent of the defendant, and generally mean that the plaintiff can refile at a later date (and perhaps in a different court). Involuntary dismissals and judgments granting pretrial motions end a plaintiff's case through judicial action.We perform each analysis separately on the entire dataset and reveal modest effects of judge party on case outcomes, with the largest effects for settlements.

Third we empirically demonstrate how the common practice of subsetting to specific kinds of "interesting" outcomes (e.g., cases that were not settled, or cased that were appealed) can generate biased estimates of partisan effects.

## 3.1 Causal Identification

A nice feature of studying adjudication in U.S. federal courts is that cases are (as-if) randomly assigned to judges. This provides an opportunity to estimate the causal effect of assigning different judges to cases. In our context, in order to exploit the random assignment of judges to cases, we must be careful to account for potential violations of this randomization. In district courts, random assignment occurs within each district court's divisions[8] after a case is filed. For example, General Order No. 16-05 in the Central District of California says that "when a case with a civil case number [...] is assigned to a district judge, it will be randomly assigned from a *division-specific* General Civil Assignment Deck" (emphasis added).[9] This means that we can only assume judges are randomly assigned to cases *within court divisions* and *within specific intervals of time*. To account for this, we control for division-year fixed effects in all of our analyses.

To evaluate our causal identification strategy, we provide evidence that suggests that, conditional on division-year fixed effects, judges are as-if randomly assigned to cases in our dataset. To assess balance on the pre-treatment variables between the treatment and control groups, it is

8. Each district court can have multiple divisions. For example, the Central District of California is divided into three divisions (Western, Eastern and Southern) spread across four courthouses (two in Los Angeles, one in Riverside and one in Santa Ana).

9. Available at `https://www.cacd.uscourts.gov/sites/default/files/general-orders/16-05.pdf`.

not generally sufficient to examine balance on the marginal distributions of each pre-treatment variable (Chen and Small 2016). Indeed, the entire purpose of demonstrating statistical balance between treatment and control groups is to provide evidence of statistical balance on the *joint distribution* of the pre-treatment variables. Rosenbaum and Rubin (1985) propose using the propensity score, which is the probability of a subject being assigned to a particular treatment given a set of covariates, to check covariate balance on the joint distribution. We use a variation of this approach. Because the effectiveness on such an approach depends on choosing a good model, we turn to machine learning.

We use a random forest algorithm to generate out-of-sample predicted probabilities[10] that each case will be assigned to a Republican, *i.e.*, propensity scores.[11] We do this predictive exercise twice. First, we generate propensity scores using *only* division-year fixed effects, thus correcting for known violations of randomization. This provides us with a useful benchmark since we already expect to have covariate imbalance before controlling for division-year fixed effects. Second, we generate propensity scores using division-year fixed effects as well as all other pre-treatment variables available to us. If this second model provides substantially more predictive power for treatment assignment than the benchmark model, then we have evidence that our causal identification assumptions do not hold. This is due to the fact that the additional pre-treatment variables are helping us predict whether a Democrat or Republican is assigned to cases, which should be impossible since pre-treatment variables by definition measure features of cases that are fixed *before* they are assigned to judges.

10. Specifically, we use ten fold cross-validation and retrieve the predicted probabilities from the held-out set in each fold.

11. In many contexts, a "propensity score" is typically assumed to be a predicted probability generated from a logistic regression. We use an optimized non-parametric approach to generate predicted probabilities, so our predictions will generally be more accurate than logistic regression. In spite of this (and with some abuse of language), we still refer to our predicted probabilities as "propensity scores" to emphasize that they are conceptually identical to those generated from logistic regression even though they are more accurate.

**Figure 2:** *We plot an ROC curve (left) and an eQQ plot (right), which present evidence that assignment of judges to cases is as-if random conditional on court-division and year fixed effects.*

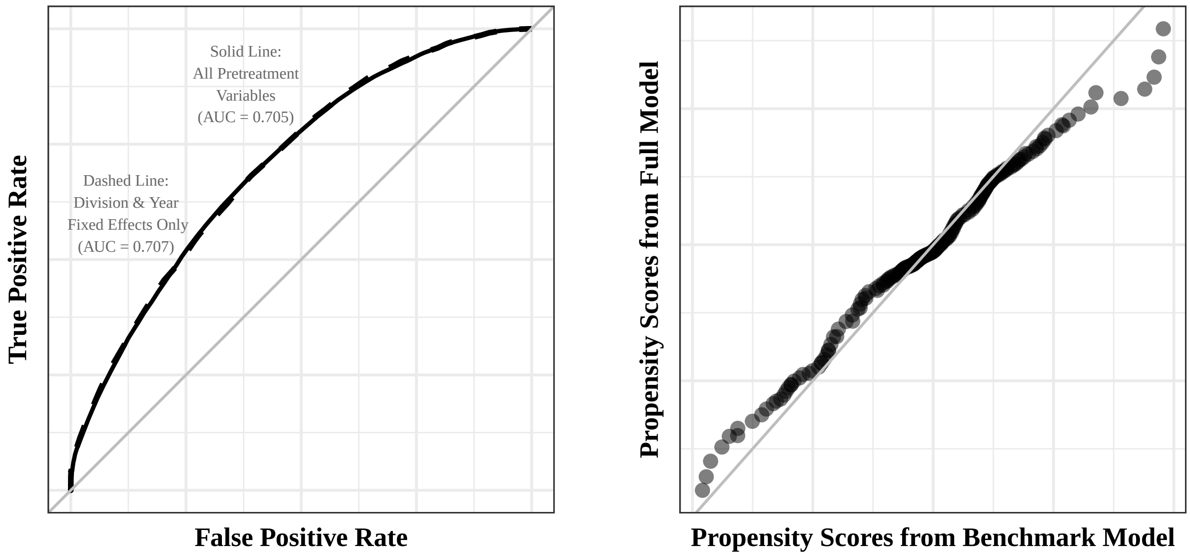



Figure 2 presents our results. In the left panel, the dashed ROC curve—as well as its associated area under the curve (AUC) metric—provide an indication of the predictiveness of our benchmark division-year fixed-effect models. In machine learning contexts, the goal is an ROC curve that is highly protruded from the 45 degree line, since this indicates a model is highly predictive. In this specific context, since we are testing for covariate balance, the goal is the opposite: *i.e.*, that pre-treatment variables are *not* predictive of whether a case is assigned to a Republican or a Democrat. As we expect, the division-year fixed effects are predictive of which judge is assigned since we know that randomization occurs within court divisions and over time. The solid ROC curve (and associated AUC) demonstrates the predictiveness of the model that includes all pre-treatment variables including the division-year fixed effects. The model that includes all the pre-treatment variables is no more predictive of whether cases are assigned to Republicans or Democrats than the benchmark model. This provides evidence that, conditional on court division and year, judges are as-if randomly assigned to cases.

In the right panel, we present this information in a different way, using an empirical quantile-

quantile (eQQ) plot. This plot compares the distribution of propensity scores for the benchmark model with division-year fixed effects against the distribution of propensity scores for the full model that uses all pre-treatment variables.[12] If there are no substantial imbalances after controlling for the known violations of randomization (*i.e.*, after including division-year fixed effects) then the points should be aligned along the 45 degree line. This would indicate that the distribution of propensity scores does not shift substantially between the benchmark model and the model with all pre-treatment variables. The eQQ plot demonstrates only minor differences between these two distributions, providing further evidence that judges are randomly assigned to cases.

## 3.2 Estimating the Influence of Partisanship

Our goal in this paper is to empirically evaluate whether judges' partisanship matters for outcomes in our dataset of civil rights cases. In this section, we mimic traditional approaches and estimate average treatment effects in the entire dataset. More specifically, for each of the four outcomes we study (e.g., settlement, voluntary dismissal, involuntary dismissal, and judgment for the defendant on pretrial motion), we estimate the average effect of assigning Republican judges to cases (instead of Democratic judges). While this is typically interpreted as the average effect of judge "ideology" on case outcomes, for reasons we discuss in Section 1, we interpret and describe it as the ***average Republican effect***.

In light of our discussion in the preceding section, we implement our causal identification strategy by estimating a linear regression of the outcome on the treatment plus division-year fixed effects. Formally, for each outcome $k$, we estimate the following model:

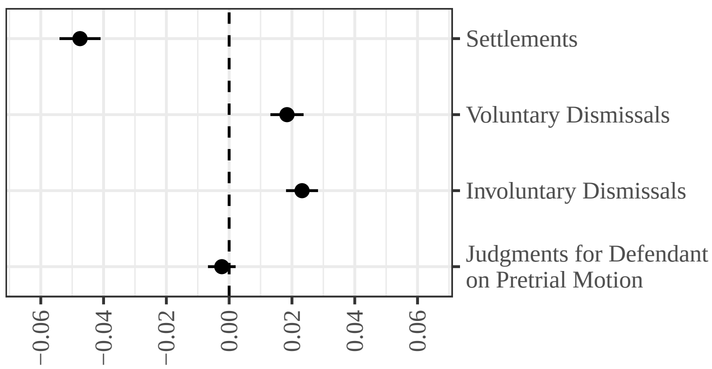
$$Y_i^k = \beta_0^k + \beta_1^k Republican_i + \beta_2^k Division_i \times Year_i + \varepsilon_i^k \tag{1}$$

where $i$ indexes cases. Estimates of $\beta_1^k$ give the (causally identified) average Republican effect on

12. To generate this plot (and make it more intelligible), we bin each percentile of each distribution by taking the mean of the propensity scores in that percentile and plotting those means.

outcome $k$. Because we study several outcomes, we perform our analyses outcome-by-outcome. For example, when looking at settlements, the outcome variable takes a value of 1 if the case settled and 0 if it did not settle. As a result, for each outcome, $Y_i = 0$ can be interpreted as "all other possible outcomes."[13] We report the average Republican effect on each outcome in Figure 3. In all plots, including Figure 3, we include 95% confidence intervals derived from robust standard errors of model (1). All effects we report on dot plots are also reported as regression tables in the appendix.

**Figure 3:** *Average Republican effects for each outcome in the entire dataset of civil rights cases.*




In our full dataset, the outcome with the largest average Republican effect is settlement. On average, being assigned to a Republican (instead of a Democrat) causes settlements to occur 4.8 percentage points less often, while they cause voluntarily[14] or involuntary dismissals of plaintiff cases at increased rates of 1.8 and 2.3 percentage points, respectively. While we explore possible mechanisms in more detail in the final section of the paper, there are a number of plausible explanations that are consistent with the general belief that Republicans disfavor civil rights claims. It may be the case that defendants, once assigned to a Republican, positively update their assessment of prevailing in litigation, thus making defendants less likely to provide benefits to the plaintiff

13. Given the size of our dataset, multiple comparison adjusted standard errors are not substantially different than those reported here. As a result, we do not include them.

14. Plaintiff's may voluntarily dismiss their own case through Federal Rule of Civil Procedure 41. Such dismissals generally do not preclude the plaintiff from refiling the case in another court. See FRCP 41(a)(1)(B)

through settlement and instead letting the judge dismiss the plaintiff's claim or the letting the plaintiff give up. Alternatively, Republicans may actually act differently than Democrats so as to be less likely to induce settlement. Republicans, might, for example, be less likely to hold settlement conferences or exert pressure on litigants during those conferences. But whatever the case, insofar as partisan effects exist, traditional research designs that exclude settlements and voluntary dismissals would fail to detect all of them.

## 3.3 Why You Can't Subset Cases Based on Outcomes

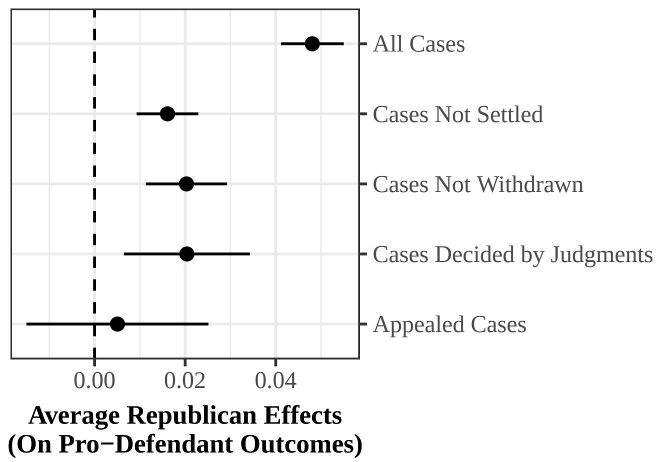
Given that our largest average Republican effects correspond to settlements, this raises serious questions about analyses of datasets that do not include settled cases. More broadly, we should worry that *any* dataset that subsets to specific kinds of outcomes (judgments, dismissals, appeals) will be introducing bias into results by breaking the random assignment of judges. We now demonstrate that this concern is warranted by dropping certain kinds of cases and re-running our analysis to reveal the extent of the bias. For this exercise, we focus on a new outcome: whether the resolution of the case favored the defendant.[15] We define a "pro-defendant" outcome to be any outcome that is either a dismissal or a judgment for the defendant. We use this coding since dismissals are effectively "wins" for the defendant.[16]

To conduct the analysis, we subset to several intuitively interesting sets of cases and re-estimate our average Republican effects on each of these subsets. These subsets are: (1) cases that were not settled, (2) cases that were not withdrawn either by settlement or by voluntary dismissal, (3) cases that were terminated with judgments issued by a judge, and (4) cases that were ultimately appealed. For a benchmark comparison, we also estimate the average Republican effect on our whole dataset (for which we have provided evidence that it satisfies our causal identification assumptions).

15. We focus on this outcome for a very simple reason: dropping cases based on outcomes eliminates those outcomes as potential dependent variables. We thus need to use a standardized, if imperfect outcome so that we may compare effects as we iteratively drop cases.

16. For example, 12(b)(6) motions to dismiss are frequently filed by defendants to terminate litigation.

**Figure 4:** *Focusing on theoretically or intuitively interesting subsets of cases based on their outcomes can bias results toward zero, hiding partisan effects.*




In Figure 4, we plot the results of this exercise. In the full dataset, there is a substantively meaningful average Republican effect of nearly 5%. However, after dropping specific kinds of cases, this effect is reduced by more than half. This provides one explanation for why some past research has failed to detect partisan effects: honing in on "interesting" kinds of outcomes may bias effects toward zero. Of particular interest is the average Republican effect on cases that were appealed. For example, Zorn and Bowie (2010) finds no effect of partisanship on a subset of district court cases that were appealed to the Supreme Court. Examining Figure 4, there is reason to worry that these effects could be driven entirely by the post-treatment bias introduced when subsetting to cases that were appealed.

# 4 It's the Cases, Not the Judges

We now turn to our analysis of the competing explanations for why partisan effects in district courts are relatively small. As explained in Section 1, there are two very different versions of "ideology doesn't matter" when studying district courts. Under one, district court judges do not make partisan decisions either because they do not have the proclivity, or because they are constrained by appellate courts. In other words, partisanship truly does not matter in district court decision-making. Under the other, district court judges do not, on average, make partisan decisions because the vast majority of cases they decide are so routine or frivolous that the cases do not allow for partisan decision-making: ideology matters, but not in most cases. In this section, we provide evidence in favor of the second hypothesis. District court judges show substantial partisan effects in some cases.

While our techniques are novel (and a bit technical), our insight is fairly simple. If partisanship matters less in district courts due to cases, then we should be able to detect substantial effect size heterogeneity in subsets of cases. Our goal, then is to aggressively *search* for the largest partisan effects in subsets of our dataset. If we cannot find substantial heterogeneity, then this suggests the judge-based story. If we can, it suggests the case-based story.

If our goal is to find subsets of cases with large effects, then the main challenge we face is to subset the data in a way that does not violate our causal identification assumptions. As we saw in Section 3.3, subsetting based on intuition can introduce bias into our estimates. In order to avoid this, we devise a technique to subset the data using *only* pre-treatment variables, and thus avoid "undoing" the random assignment of judges.

Since we are aggressively searching for large effects in subsets of cases in our dataset, we devise a statistic for each outcome $k$ that best captures our expectation about the likelihood that any given case will be resolved differently by Republicans than by Democrats. Formally, let $\hat{Y}_i^k(R, X_i)$ and $\hat{Y}_i^k(D, X_i)$ be predicted probabilities of an outcome $k$ on case $i$ when assigned to a Republican judge and Democratic judge, based on pre-treatment variables $X_i$. Then, for each case and outcome, we define a statistic $\hat{d}_i^k$:

$$\hat{d}_i^k = |\hat{Y}_i^k(R, X_i) - \hat{Y}_i^k(D, X_i)| \tag{2}$$

This statistic yields the predicted partisan disagreement for each case, which we will refer to as that case's *PPD score*.

Because each case will receive a PPD score, we can zoom in on subsets of cases in our dataset that have the highest PPD scores. Substantively, these are cases where we have the highest expectation (based on pre-treatment variables) of large partisan effects. We wish to emphasize two benefits of this approach. First, it targets exactly what we care about in the most efficient way possible. We do not know *ex ante* where to find the largest partisan effects, so our technique frees us from having to search over a large number of theory-driven subsets of cases to try to find partisan effects. Indeed, our technique shines a spotlight on precisely which cases feature these differences. Second, this subsetting technique does not introduce selection bias into our estimates. Since calculation of $\hat{d}_i^k$ *only* involves information from pre-treatment variables, it preserves causal identification.

In the remainder of this section, we describe how we generated the PPD scores, and then proceed to subset cases using the scores. We end by presenting our results.

**Predictions for PPD scores** The key components of the PPD scores are the predicted probabilities $\hat{Y}_i^k(R, X_i)$ and $\hat{Y}_i^k(D, X_i)$. To build intuition, let's focus on a specific outcome: settlements. We estimate predicted probabilities using machine learning.[17] For each case, we are able to generate a predicted probability that the case settles when a Republican judge is assigned to it and another predicted probability that the case settles when a Democratic judge is assigned to it. These predictions are based on a wide variety of pre-treatment variables, including variables relating to the case subject-matter, the case's litigants, and the contents of the initial docket entry.

We emphasize two important methodological components of our process. First, all of the predictions we generate are out-of-sample predictions (using 10-fold cross validation). If instead we generated our predictions using the same data on which we fit our model (i.e., in-sample), then we would run the risk of over-fitting and creating PPD scores that do not preserve random assignment of judges. Second, instead of using one specific algorithm, we use a stacked ensemble approach. In brief, we estimate models using several base algorithms (e.g., LASSO regression, random for-

17. Replication files will be made available online.

est, boosted CARTs, etc.) and then estimate a "meta" model that weights the results of each base algorithm in a way to maximize the accuracy of the predictions. While stacked ensembles require more computational resources (and time), they typically provide models that perform at least as well as the individual base models on which they are built.

**Figure 5:** *We demonstrate the performance of our predictive models using the "area under the curve" (AUC) metric.*

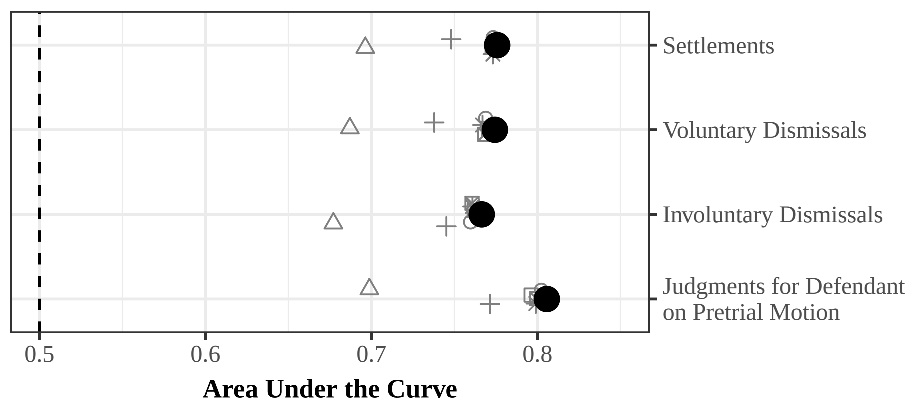
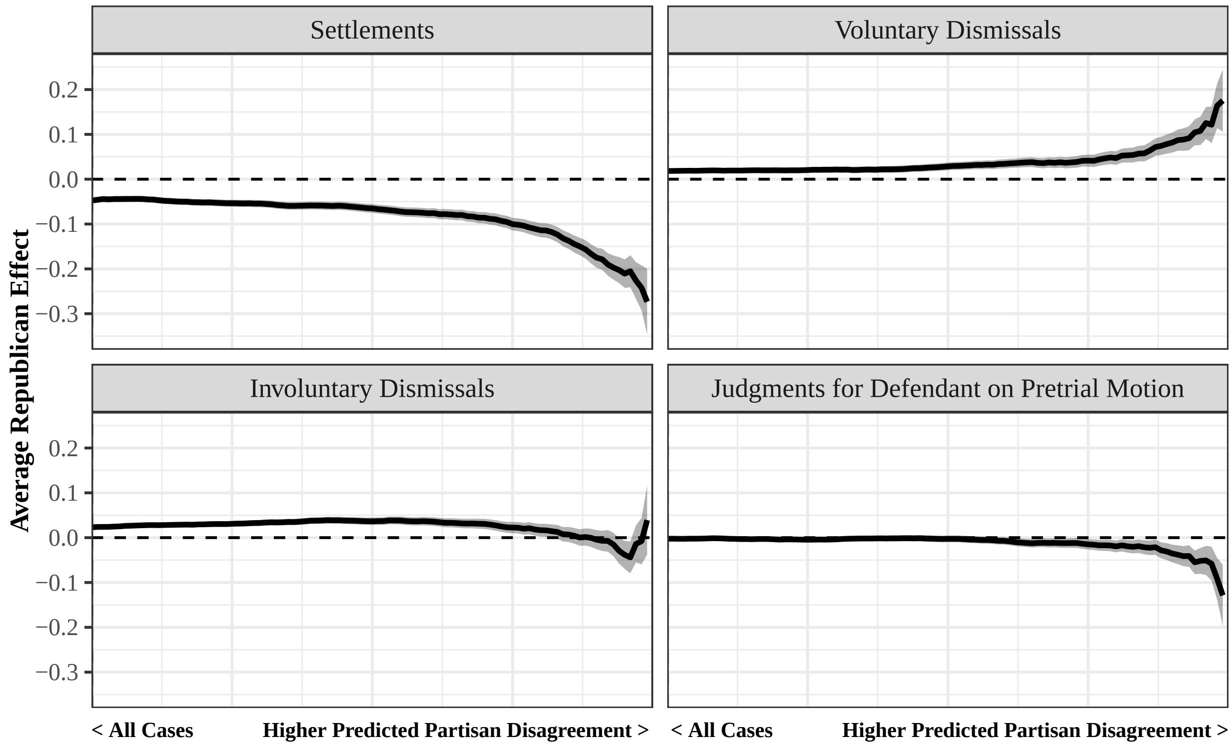



Figure 5 presents standard performance metrics. For each outcome, we plot the area under the curve (AUC) for our base algorithms (grey symbols), as well as the AUC for our ultimate stacked ensemble algorithm (solid circles). All of our stacked ensembles have cross-validated AUCs that are above 0.75 and are generally at least as high or higher than the best base algorithm.

**Results** For each outcome, we analyze sets of cases with successively larger PPD scores. We begin with all cases, estimating the average Republican effects for each outcome. We then drop the 1% of cases with the smallest PPD scores. With the remaining 99% of cases, we again calculate the average Republican effect for each outcome. We then drop the next 1% of cases with the smallest PPD scores and with the remaining 98% of cases again estimate the average Republican effect for each outcome. We continue this process, iteratively dropping an additional percentage of the cases with low PPD scores and estimating average Republican effects.

We plot the results of this exercise for all four outcomes in Figure 6. The most notable ef-

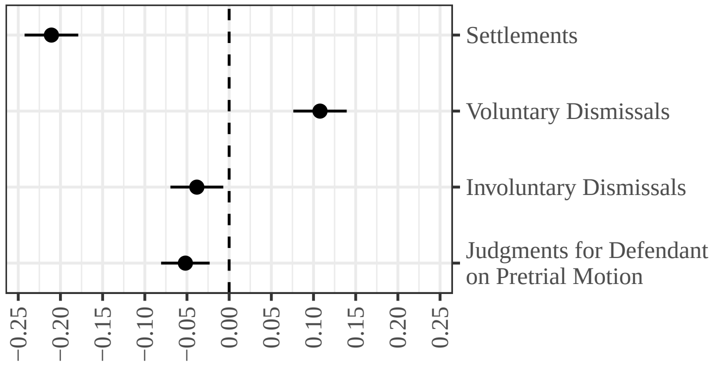
**Figure 6:** *Cases with the highest predicted partisan disagreement have very large average Republican effects.*




fects are on settlements and voluntary dismissals, which we show which in the top two panels. As one successively removes the cases with the lowest PPD scores (*e.g.*, move from left to right on the $x$-axis of each panel), larger differences between Republican and Democratic appointees emerge. Settlements become less and less likely with Republican judges relative to Democratic judges whereas voluntary dismissals become more common under Republican judges relative to Democratic judges. This both confirms the lesson that litigant adaptation is an important consideration when estimating partisan effects and shows that the small effects in the full dataset are masking much larger effects in subsets of cases. In other words, partisanship *does* matter a lot—in some cases.

Consider a specific subset of cases: the 5% of cases with the highest PPD scores (about 4,200 cases). In Figure 7, we plot the average Republican effects.

**Figure 7:** *Average Republican effects for each outcome in the 5% of civil rights cases that have the largest predicted partisan disagreement.*

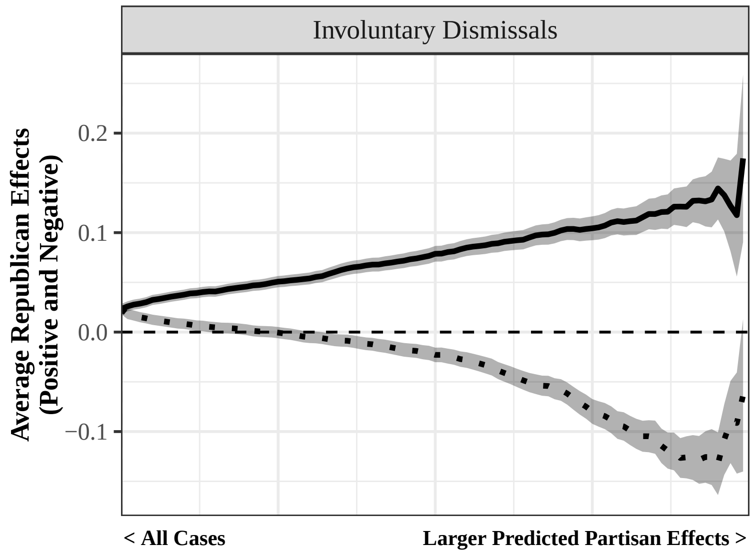



In this subset of cases, we estimate that assignment to a Republican judge reduces the probability of settlement by 21.1 percentage points and increases the probability of voluntary dismissal by 10.8 percentage points. Involuntary dismissals and judgments for the defendant, respectively, are 3.9 and 5.2 percentage points less likely when a case is assigned to a Republican.

While this technique reveals much larger partisan effects in subsets of cases, there may be *even larger* hidden partisan effects than we are able to find. For example, Copus and Hübert (2018) demonstrates that judges may make different decisions in a larger number of cases than an average treatment effect would reveal. We briefly consider the possibility this occurs in our analysis. If Republican assignment has positive effects in some cases, negative effects in others, and these effects are mostly "balanced" in the dataset, then the average Republican effect will look small even when partisan disagreement between judges is large.

To illustrate, we decompose the average Republican effects for the outcome with the lowest effect in Figure 7, involuntary dismissals. Recall that in our main analysis (as plotted in Figure 6), we look for cases with the largest partisan disagreement, $\hat{d}_i^k$. Notice that $\hat{d}_i^k$ does not distinguish the direction of disagreement. Now, we will look for the cases where we expect the largest *positive* effects and then look for the cases where we expect the largest *negative* effects. In Figure 8 we plot each of these for voluntary dismissals.

**Figure 8:** *We decompose the average Republican effects on involuntary dismissals into the largest positive effects (solid line above) and largest negative effects (dotted line below).*




This reveals that there are large positive effects canceling out large negative effects, which explains why the trend line in the bottom left panel of Figure 6 is relatively flat, and the estimate for involuntary dismissals in Figure 7 is close to zero. Figure 8 shows that there are indeed substantial effects—it is just that Republicans are more likely to involuntarily dismiss some cases and less likely to involuntarily dismiss others.

The above results collectively provide evidence that partisanship does indeed matter in district courts, at least if one looks at the right subsets of cases. But in order to piece together a more cohesive picture of how partisanship operates in the district courts, we need to do more than simply find the effects. We need to understand how the effects fit together: which effects trade off with other effects and what do cases with high PPD scores look like? In the next section, we explore these questions.

# 5 Exploring the Results

In the previous sections, we provide some of the first causally identified effects of judge partisanship on case outcomes in federal district courts. Our analysis challenges the view that judges' partisanship has little effect on outcomes. In this section, we begin to stitch together a more substantive story about the role of partisanship in district court decision-making. We emphasize that since the discussion in this section goes beyond our causal estimates, it is necessarily speculative. We do not intend to conduct a full investigation here, but instead hope to highlight a set of interesting issues for future research.

We address two key substantive questions that arise from our analysis. First, how do outcomes substitute for one another? Because there are many possible outcomes in district court litigation, it is not immediately clear how a partisan effect on one litigation outcome relates to effects on another outcome (if at all). For example, while we know that assignment of a Democrat causes some cases to settle when they would not have settled with a Republican, how *would* Republicans dispose of those cases? Second, in Section 4 we identified subsets of cases where partisan effects are quite large, but what about those cases makes them subject to larger partisan effects? In other words, what *substantively* distinguishes high PPD cases from low PPD cases?

**How do outcomes substitute for one another?** Recall that our PPD scores are built from a collection of case-level predicted probabilities for each outcome, which are based on (1) pre-treatment variables and (2) whether a Republican or a Democrat was assigned. So, for each case $i$, there are eight predictions: $\hat{Y}_i^k(R, X_i)$ and $\hat{Y}_i^k(D, X_i)$ for each of our four outcomes, indexed by $k$. With these in hand, we have effectively generated *predicted potential outcomes* for each case and each outcome. This enables us to calculate a predicted case-level treatment effect for each of our four outcomes. We emphasize that these are *predicted* effects, and should not be mistaken for *actual* unit-level treatment effect (which are impossible to calculate due to the fundamental problem of causal inference, see Holland 1986).

So that we can be clear about our next steps, we adopt the following notation. Let $\hat{\tau}_i^k$ be the predicted Republican effect on case $i$ for outcome $k$. Formally, it is defined by

$$\hat{\tau}_i^k = \hat{Y}_i^k(R, X_i) - \hat{Y}_i^k(D, X_i)$$

We use superscripts $k = \text{ST}$ for settlements, $k = \text{VD}$ for voluntary dismissals, $k = \text{ID}$ for involuntary dismissals and $k = \text{JD}$ for judgments on pretrial motions favoring the defendant. Consider a hypothetical example where $\hat{\tau}_i^{\text{ST}} = -0.2$ for a case $i$. This would mean that we predict settlement to be 20% less likely when Republican is assigned to that case instead of a Democrat.

Since each case has a predicted Republican effect for each of the four outcomes, we can look to see how the four predicted Republican effects correlate across the cases. We do this by regressing these predicted Republican effects on one another. For example, do our predicted Republican effects for settlements correlate with our predicted Republican effects for voluntary dismissals? If they correlate negatively, this means that on the cases where we predict voluntary dismissals to be *more* likely with Republicans (relative to Democrats), we also predict that settlements to be *less* likely with Republicans (relative to Democrats). Substantively, this would indicate that settlements and voluntary dismissals are "partisan substitutes." In other words, in the cases where we expect Republican assignment to lead to more voluntary dismissals, we expect Democratic assignment to lead to more settlements.

With respect to settlements and voluntary dismissals specifically, that is what we find. Consider Table 3, in which we present the regression coefficients for this exercise.

**Table 3:** *Regressing predicted unit-level treatment effects on one another provides some indication about how outcomes trade-off.*

| | $\hat{\tau}^{\mathrm{ST}}$ | $\hat{\tau}^{\mathrm{VD}}$ | $\hat{\tau}^{\mathrm{ID}}$ |
|---|---|---|---|
| $\hat{\tau}^{\mathrm{ID}}$ | $-0.167^{***}$ | | |
| $\hat{\tau}^{\mathrm{VD}}$ | $-0.380^{***}$ | | |
| $\hat{\tau}^{\mathrm{JD}}$ | $-0.117^{***}$ | | |
| $\hat{\tau}^{\mathrm{ID}}$ | | $-0.065^{***}$ | |
| $\hat{\tau}^{\mathrm{JD}}$ | | $0.008^{*}$ | |
| $\hat{\tau}^{\mathrm{JD}}$ | | | $-0.06^{***}$ |

$^{*}$ inidcates $p \leq 0.05$; $^{**}$ inidcates $p \leq 0.01$; $^{***}$ inidcates $p \leq 0.005$

Notice that the relationship between settlement and voluntary dismissal is indeed negative. Specifically, as our predicted Republican effect on voluntary dismissals goes up by 10% (e.g., go from predicting a 10% effect of Republican assignment to a 20% effect of Republican assignment), our predicted Republican effect on settlements goes down by 3.8%. Roughly speaking, this is suggestive evidence that in the cases voluntarily dismissed under a Republican, a good portion of them (we estimate around 40% of them) would have been settled under a Democrat.

After being assigned to a Republican, why might plaintiffs opt to voluntarily withdraw their case (via a voluntary dismissal) rather than try to come to a settlement with the defendant? Recall that we study civil rights cases where plaintiffs typically allege some form of civil rights violation, such as employment discrimination. Thus defendants are typically being accused of these civil rights violations. So, one explanation for the apparent substitution between settlements and voluntary dismissals is that plaintiffs might be aware that if they choose to proceed with a case once assigned to a Republican, their chances of losing have increased to the extent Republicans are skeptical of civil rights claims. It may not be worth the hassle or the expense to continue the case, especially because a voluntary withdrawal can provide the plaintiff with a chance to refile the case at a later point (and potentially get a different judge).

While this explanation is admittedly speculative, the other relationships depicted in Table 3 provide further clues that suggest Republican judges favor defendants (on average) and Democratic judges favor plaintiffs (on average). There is evidence of modest substitution from settlements to involuntary dismissals and judgments in favor of the defendant, both of which are judicial actions favoring defendants. This suggests that a Republican assignment reduces the prospect of settlements and increases the prospect of involuntary dismissals and judgments for the defendant.

There is little evidence in favor of a substitution between voluntary dismissals and pretrial judgments for the defendant, while there is evidence of moderate substitution between voluntary and involuntary dismissals. This suggests that plaintiffs are withdrawing mostly to avoid involuntary dismissals rather than pretrial judgments against them. One possible explanation is based on the fact that involuntary dismissals generally occur at a much earlier stage of the litigation than pretrial judgments. Sophisticated plaintiffs who can anticipate the likely actions of the judges may voluntarily withdraw shortly after being assigned to a Republican appointee. Less sophisticated plaintiffs may push on, forcing the judge to take the action that sophisticated plaintiffs may have foreseen long before.

**What do high PPD cases look like?** We also explore the nature of the cases where we find the largest partisan effects. The most straightforward interpretation is that partisanship has larger effects on more "difficult" cases—those where law does not provide a clear answer and provides judges with the room to fill gaps with their own partisan preferences. As an initial step to exploring this interpretation, we analyze the relationship between PPD scores and two intuitive proxies for whether a case has any merit: the presence of pro se litigants (which are almost all plaintiffs) and the number of attorneys on representing the defendant(s). If high PPD scores do indeed track with case case "difficulty," we should expect them to be negatively associated with the presence of pro se litigants. As Epstein, Landes, and Posner (2013) explain, a "substantial fraction of cases filed in the district courts (many by persons who do not have legal representation) have no possible merit and so really are just noise in the data" (p. 207). Similarly, we would also expect the number of defendant attorneys to be positively associated with the difficulty of a case, as the sophisticated parties who could afford to hire larger legal teams are unlikely to waste resources on those teams if the case did not involve contestable issues.

**Figure 9:** *Plotting predicted partisan disagreement against two important pre-treatment variables relating to case features provides some insight into how our casually identified procedure maps onto intuitions about what makes a case "difficult" or "easy."*

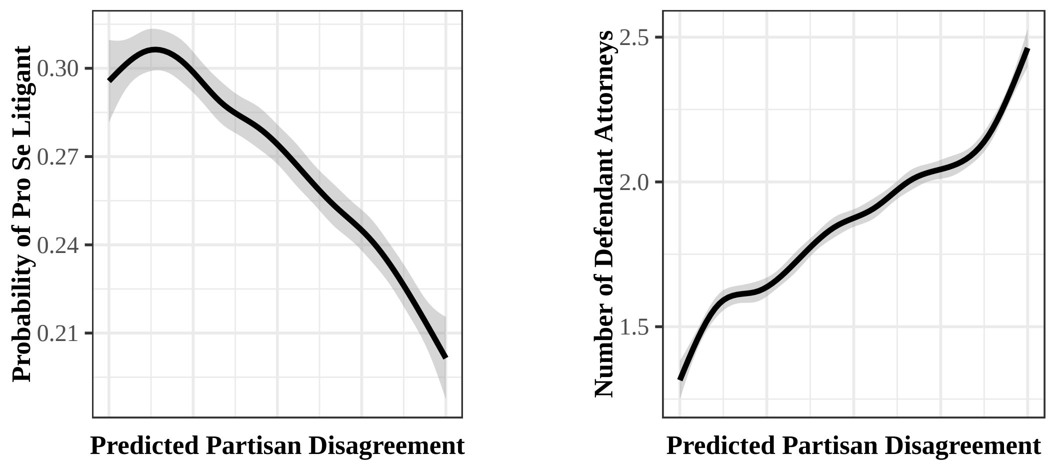



Figure 9 shows the relationship between these two "intuitive" indicators of whether a case is difficult and a weighted average of each case's PPD scores. We see evidence that PPD scores are associated in the expected directions with the presence of pro se litigants and number of defense attorneys. As we predict higher partisan effects (i.e., higher PPD scores), there are fewer pro se litigants, and more attorneys representing defendants.

However, the relationships are perhaps not as strong as one might have expected. For example, more than 20% of the cases with the highest PPD scores still have pro se litigants. The surprising rate of pro se litigants in cases with high PPD scores is subject, we think, to three possible explanations: (1) many cases with pro se litigants are not, in fact, merit-less cases, (2) judges make partisan decisions even in frivolous cases, or (3) our predictions of partisan effects in pro se cases do not actually represent real effects. While we cannot adjudicate between these explanations, it suggests that researchers interested in identifying and studying "non-frivolous" or "difficult" cases might consider using PPD scores instead of their intuitions about litigants. Such intuitions would not necessarily point researchers toward the cases where judges are most likely to come to different outcomes, a tell-tale feature of "difficult" cases.

Our efforts here are necessarily speculative, we think we have identified a fruitful area of research for understanding the operation of partisan effects in district courts. First, our discussion suggests that civil rights plaintiffs may anticipate negative treatment from Republicans, whereas civil rights defendants may anticipate negative treatment from Democrats. Our exploration of outcome substitution in our dataset is consistent with this. Second, our data-driven way to subset cases based on PPD scores maps onto scholars' intuitions about what makes for a "difficult" case. That said, the mapping is imperfect, which suggests to us that the PPD scores (which preserve random assignment of judges) offer a more fruitful way to find cases where the "law runs out."

# 6 Conclusion

In this paper, we analyze a new data set of all civil rights cases in seven district courts from 1995-2016 to reexamine the view that "ideology doesn't matter" in district court judging. We demonstrate that the effects of judge partisanship are substantial, but largely hidden. First, the largest effects of judge partisanship manifest in litigant behavior, e.g. whether a case settles or is voluntarily dismissed. Most prior research studies on district courts use empirical strategies, such as dropping settled cases from analyses, that purposefully ignore these effects. Moreover, we show that these research designs introduce a form of post-treatment bias that may push estimates toward zero.

Second, we implement a novel machine learning technique to find the subsets of cases where we predict the largest effects. Our technique allows us to see whether the many "frivolous" cases heard by district judges are swamping the cases where partisanship is very important. We indeed show there are very large partisan effects in subsets of cases. By doing so, we demonstrate it's those courts' case mixes (and not their judges or institutions) that generate small partisan effects. By revealing the large effects in these subsets, we paradoxically provide evidence that partisanship matters a lot.

Taken as a whole, we provide some of the first causally-identified evidence of substantial partisan effects in district courts. Because these effects are largely hidden, our analysis suggests that scholars seeking to study partisan effects in courts employ empirical strategies that account for heterogeneous effects. While we provide one way to do this, there are many promising avenues for future research. That said, one thing is clear: for thousands, if not millions, of litigants, whether they walk away from litigation getting anything depends on whether they are randomly assigned a Democratic or Republican judge.

# References


Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110 (3): 512–529.

Aronow, Peter M., Jonathon Baron, and Lauren Pinson. 2019. "A Note on Dropping Experimental Subjects who Fail a Manipulation Check." *Political Analysis.*

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab. 1995. "Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes." *The Journal of Legal Studies* 24 (2): 257–281.

Boyd, Christina L. 2013. "She'll Settle It?" *Journal of Law and Courts* 1 (2): 193–219.

———. 2017. "Gatekeeping and Filtering in Trial Courts." In *The Oxford Handbook of U.S. Judicial Behavior,* edited by Lee Epstein and Stefanie A. Lindquist, 129–148. Oxford University Press.

Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54 (2): 389–411.

Boyd, Christina L., and David A. Hoffman. 2010. "Disputing Limited Liability." *Northwestern University Law Review* 104 (3): 853–916.

Boyd, Christina L., and James F. Spriggs II. 2009. "An Examination of Strategic Anticipation of Appellate Court Preferences by Federal District Court Judges." *Washington University Journal of Law and Policy* 29:37–81.

Burbank, Stephen B. 2004. "Vanishing Trials and Summary Judgment in Federal Civil Cases: Drifting Toward Bethlehem or Gomorrah?" *Journal of Empirical Legal Studies* 1 (3): 591–626.

Cameron, Charles M., and Lewis A. Kornhauser. 2017. "Rational Choice Attitudinalism?" *European Journal of Law and Economics* 43 (3): 535–554.

Carp, Robert A., and C.K. Rowland. 1996. *Politics and Judgment in Federal District Courts.* Lawrence, KS: University Press of Kansas.

Chen, Hao, and Dylan S. Small. 2016. "New Multivariate Tests for Assessing Covariate Balance in Matched Observational Studies." `https://arxiv.org/pdf/1609.03686.pdf`.

Clermont, Kevin M., and Theodore Eisenberg. 2002. "Litigation Realities." *Cornell Law Review* 88 (1): 119–154.

Coppock, Alexander. 2019. "Avoiding Post-Treatment Bias in Audit Experiments." *Journal of Experimental Political Science* 6 (1): 1–4.

Copus, Ryan, and Ryan Hübert. 2018. "Detecting Inconsistency in Governance." doi:`10.2139/ssrn.2812914`.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice.* Cambridge, MA: Harvard University Press.

Fischman, Joshua B., and Max M. Schanzenbach. 2011. "Do Standards of Review Matter? The Case of Federal Criminal Sentencing." *Journal of Legal Studies* 40 (2): 405–437.

Hoffman, David A., Alan J. Izenman, and Jeffrey R. Lidicker. 2007. "Docketology, District Courts, and Doctrine." *Washington University Law Review* 85 (4): 681–751.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–960.

Howard, Robert M. 2002. "Litigation, Courts, and Bureaucratic Policy: Equity, Efficiency, and the Internal Revenue Service." *American Politics Research* 30 (6): 583–607.

Kastellec, Jonathan P. 2013. "Racial Diversity and Judicial Influence on Appellate Courts." *American Journal of Political Science* 57 (1): 167–183.

Keele, Denise M., Robert W. Malmsheimer, Donald W. Floyd, and Lianjun Zhang. 2009. "An Analysis of Ideological Effects in Published Versus Unpublished Judicial Opinions." *Journal of Empirical Legal Studies* 6 (1): 213–239.

Kim, Pauline T., Margo Schlanger, Christina L. Boyd, and Andrew D. Martin. 2009. "How Should We Study District Judge Decision-Making?" *Journal of Law and Policy* 29 (83): 83–112.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2019. "The Bias Is Built In: How Administrative Records Mask Racially Biased Policing." `https://ssrn.com/abstract=3336338`.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Post-treatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–775.

Perino, Michael A. 2006. "Law, Ideology, and Strategy in Judicial Decision Making: Evidence from Securities Fraud Actions." *Journal of Empirical Legal Studies* 3 (3): 497–524.

Priest, George L., and Benjamin Klein. 1984. "The Selection of Disputes for Litigation." *Journal of Legal Studies* 13:1–55.

Randazzo, Kirk A. 2008. "Strategic Anticipation and the Hierarchy of Justice in U.S. District Courts." *American Politics Research* 36 (5): 669–693.

Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a ConcomitantVariable ThatHas Been Affected by the Treatment." *Journal of the Royal Statistics Society. Series A (General)* 147 (5): 656–666.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. “Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score.” *The American Statistician* 39 (1): 33–38.

Schanzenbach, Max M., and Emerson H. Tiller. 2007. “Strategic Judging Under the U.S. Sentencing Guidelines: Positive Political Theory and Evidence.” *Journal of Law, Economics, and Organization* 23 (1): 24–56.

Sen, Maya, and Omar Wasow. 2016. “Race as a ’Bundle of Sticks’: Designs That Estimate Effects of Seemingly Immutable Characteristics.” *Annual Review of Political Science* 19:499–522.

Zorn, Christopher, and Jennifer Barnes Bowie. 2010. “Ideological Influences on Decision Making in the Federal Judicial Hierarchy: An Empirical Assessment.” *Journal of Politics* 72 (4): 1212–1221.

# A Tables

**Table A.1:** *Date Span of Dataset for Each Court*

| Court | Min. Filing Date | Max. Filing Date | Min. Termination Date | Max. Termination Date |
|---|---|---|---|---|
| CACD | 1994-12-27 | 2017-01-01 | 1995-01-03 | 2017-10-31 |
| CAND | 1994-12-27 | 2016-12-28 | 1995-01-04 | 2017-09-11 |
| CASD | 1994-12-27 | 2016-12-28 | 1995-03-09 | 2017-11-15 |
| CAED | 1994-12-27 | 2016-12-30 | 1995-01-03 | 2017-10-05 |
| WAWD | 1994-12-27 | 2016-12-21 | 1995-02-09 | 2017-11-28 |
| WAED | 1994-12-29 | 2016-12-30 | 1995-04-27 | 2017-11-21 |
| ORD | 1996-11-13 | 2015-02-10 | 1996-11-22 | 2017-12-15 |

**Table A.2:** *Average Republican effects in the full dataset.*

| | Settlements | Voluntary Dismissals | Involuntary Dismissals | Judgments for Defendant on Pretrial Motions |
|---|---|---|---|---|
| Republican Judge | -0.0475 | 0.01842 | 0.02321 | -0.00234 |
| | (0.00334) | (0.0027) | (0.0026) | (0.00225) |
| Division-Year Fixed Effects | ✓ | ✓ | ✓ | ✓ |
| $N$ | 85189 | 85189 | 85189 | 85189 |

**Table A.3:** *Average Republican effects in the top 5% of cases.*

| | Settlements | Voluntary Dismissals | Involuntary Dismissals | Judgments for Defendant on Pretrial Motions |
|---|---|---|---|---|
| Republican Judge | -0.21069 | 0.1077 | -0.03832 | -0.05189 |
| | (0.01623) | (0.01616) | (0.01598) | (0.01468) |
| Division-Year Fixed Effects | ✓ | ✓ | ✓ | ✓ |
| $N$ | 4260 | 4260 | 4260 | 4260 |