

Judicial Favoritism of Politicians: Evidence from Small Claims Courts*

Andre Assumpcao[†]

Julio Trecenti[‡]

September 19th, 2019

Abstract

REWRITE While there are many studies investigating judicial favoritism on the basis of race, ethnicity, or political connectedness, there is scant evidence on judicial collusion with other branches of government in development settings. This paper address this gap by looking at pro-politician bias in small claims cases in the state of São Paulo, Brazil. Using an empirical strategy developed by Abrams et al. (2012), a regression discontinuity design, and machine learning techniques, we document a positive pro-politician bias and point to the amount claimed in court, judge tenure, judge pay, and politician vote share as the main factors driving judicial bias. To our knowledge, this is the first study providing evidence of power collusion in developing countries.

Keywords: judicial politics; judicial bias; judicial decision-making; political economy of development; law and politics.

JEL classification: D73; K42; P48; H83.

*We are grateful to the insightful feedback from Brigitte Seim, Scott Desposato, George Avelino, Ciro Biderman, Ashu Handa, Tricia Sullivan, Doug MacKay, and seminar participants at CEPESP-FGV, the Paraná State Bar Association, and the 2019 Canadian Law and Economics Association. We acknowledge financial support from the Institute for Humane Studies (IHS) and the UNC-Chapel Hill Department of Public Policy. All remaining errors are our own.

[†]Ph.D. Candidate, Department of Public Policy, The University of North Carolina at Chapel Hill. Contact details: aassumpcao@unc.edu.

[‡]Executive Secretary of the Brazilian Association of Jurimetrics and Ph.D. Candidate, Department of Mathematics and Statistics, The University of São Paulo. Contact details: jtrecenti@abj.org.br.

1 Introduction

REWRITE Suppose a case involving a politician is brought before a fair judge. The judge will hear both sides and make her decision uniquely based on case merits. Further, assume lawyers are equally skilled and unlikely to influence the judge’s ruling. Under these simplifying conditions, we could reasonably expect a fair trial, one in which political power plays no role in any court decision. But is this a realistic depiction of judicial systems? Besides individual bias, there are institutional constraints that prevent a fair trial. In political regimes with separation of powers, the executive and the legislative have some control over the judiciary, and thus judges could adjust their decisions to account for this oversight mechanism. In developing countries, where institutional constraints are stronger, the deviations in justice might even be more pronounced. Surprisingly, however, there are not many studies measuring judicial favoritism in cases involving politicians in development settings. This paper fills in the gap by investigating judicial independence when politicians appear before judges in the Brazilian judiciary system.

There is extensive evidence supporting other discrimination effects. Shayo and Zussman (2011) document a positive in-group bias, or preferential treatment, of 17 to 20 percentage points (p.p.) when the judge and litigants have the same ethnicity. Abrams et al. (2012) find that African American defendants are 18 p.p. more likely to be incarcerated than white defendants. Lu et al. (2015) show that politically connected firms are more likely to have favorable rulings in property rights cases in China. Rachlinski and Wistrich (2017) report several other discrimination situations, and though there are multiple studies on the political ideology of judges, there are very few studies looking into the legal treatment of individual politicians.

In recent years, there has been growing interest in the relationship between judicial decision-making and politics in developing countries. Using data on employment claims filed in Venezuela between 2006 and 2017, Sanchez-Martinez (2017) looks at whether defendant employers are more likely to see a favorable outcome when they are affiliated with the United Socialist Party of Venezuela (PSUV), in power since 2007. The author finds that employers who share the same party affiliation as judges are 20 percent more successful at trial. In Lambais and Sigstad (2018), the authors identify a 50 percentage point advantage in the win rate at trial for elected versus non-elected candidates when both are defendants in corruption cases filed after election day in Brazil. This project supplements the recent literature by investigating whether favoritism persists in cases where judges and politicians have less at stake (e.g., small claims cases), testing a theory of power collusion across branches of government.

Besides this theoretical component, this paper makes a series of data analysis contributions to the literature in law and politics. In order to measure the effect of political bias, we scrape and code small claims decisions from the São Paulo State Court System in Brazil for all mayor and city council candidates since 2008. We then apply the methodology in Abrams et al. (2012) to evaluate whether politicians have the upper hand in their cases. To further investigate the direction and magnitude of bias, we employ a regression discontinuity (RD) strategy comparing court outcomes across politicians who have barely won and lost a municipal election in the state. Finally, we

implement various machine learning regression algorithms to predict court outcomes and identify the most critical factors driving bias. To our knowledge, this is the first paper proposing such comprehensive analysis of judicial favoritism of politicians in developing countries.

We find that state court judges rule relatively and significantly more in favor of politicians in small claims cases than not. Applying the methodology in Abrams et al. (2012), we identify less variation in how they rule cases involving politicians when compared to simulated rulings. In other words, this result suggests that court outcomes for politicians are less vulnerable to judge type than for everyone else; the former consistently see more predictable court outcomes. In the regression discontinuity analysis, we find no difference in rulings for mayor candidates who barely lose or win a municipal election. We identify a positive correlation between vote share and favorable court outcomes, but as we narrow in on smaller vote margins, which suggest more competitive elections, the effect of holding office becomes statistically insignificant at the five percent level. Though this might seem surprising, we believe it is a direct consequence of the design of local governments in Brazil. Since municipal governments are responsible for implementing high salience policy, such as health and primary education, mayor candidates enjoy high visibility and might see favorable outcomes even if they do not take up office. In the court outcome prediction exercise, we lastly find that judge tenure and politician vote share are among the most important (positive) drivers of judicial bias, consistent with indicative evidence from previous studies and with the collusion argument developed here.

These results supplement the empirical legal studies literature in multiple ways. First, we document judicial favoritism where there should be none. Small claims cases have limited political consequences, so judges should have lower incentives to benefit politicians in court. Having documented bias in low salience cases, however, we can only conclude that the overall judicial favoritism is widespread. As legal cases increase in importance, judges and politicians would respond accordingly, and we would expect more pro-politician bias. Along these lines, Lambais and Sigstad (2018) report a 50 percentage point increase in the probability of acquittal in corruption cases filed against politicians in Brazil, which could be interpreted as the upper bound of favoritism in high-profile cases. By presenting evidence of critical factors driving judicial decisions, we also provide a benchmark for future analysis of pro-politician bias. Judge tenure, the amount claimed in court, judge pay, politicians as defendants, and politician vote share are the primary drivers of bias and should be the subject of further scientific investigation. These factors are also first-order issues to be brought to the attention of judges as sources of bias, and should significantly help state court systems adjust their rulings going forward.

The remainder of this paper is as follows. Section 2 discusses the theoretical framework behind judicial favoritism and previous evidence; section 3 presents the institutional design of the court system in Brazil. Section 4 summarizes the dataset used for analysis; section 5 discusses the analytical strategy and results, ROBUSTNESS CHECKS. We conclude in section 6.

2 Motivating Judicial Favoritism of Politicians

Suppose there are three representative public agents, one for each branch of government: the executive, the legislative, and the judiciary. Though these agents are independent, they interact with one another over time. The executive agent serves on one or two four-year mandates (pending reelection at the end of the first term). They control the majority of government budgets, and have the discretion to set wages and resources allocated to the other branches. The legislative agent also serves for a four-year period,¹ but their mandate is renewable as many times as they are reelected; they have no term limit. They are responsible for passing law and determining budget levels but not its composition. In other words, they approve the amount of money available for other branches of government but do not have a say on how to spend the money. The judiciary agent serves on life-long mandates and yields power in restrictive but steady ways. They have limited control over resources as they only oversee budgets in the courts at which they serve, but resolve disputes between the other two branches of government and other economic agents (individuals, companies, etc). In this simplified model, the judiciary pleases or upsets the executive and legislative by settling their disputes.

We are interested in the behavior of the judiciary with respect to other branches of government. Power collusion could exist if the judiciary were using sentences to please the executive and legislative as a means of buying out those who make the calls on judiciary resources. Under this hypothesis, the representative judiciary agent derives utility in each period t according to equation (1), which describes the benefit f as a function of k observable characteristics $\sum_{x=1}^k x_k$, such as their time in post, their wages, their working conditions, and unobservable characteristics ε , such as reputation and their happiness in serving justice; costs c are a function of m_t working conditions, executive and legislative utilities u_e^t and u_l^t ; δ_t are exogenous, stochastic shocks that impact judicial work. These per-period utilities are computed in perpetuity in accordance with a judge's mandate:

$$u_j^t = \frac{1}{r} \times \sum_{t=1}^t \left[f\left(\sum_{x=1}^k x_k, \varepsilon_t\right) - c\left(\sum_{m=1}^k m_t, u_e^t(p), u_l^t(p), \delta_t\right) \right] \quad (1)$$

Since the executive and the legislative agents have primary responsibility for government budget, we can expect the judiciary to strategically maximize future net benefit by pleasing other agents. If the executive and the legislative agents are less likely to lose court cases, their utility increases. This incentive, however, is not uniformly distributed across all types of judicial cases. Because electoral, political, and even legal costs are higher in high salience cases, such as corruption, politicians are more likely to expect and respond to favoritism in such cases than in low salience cases. This makes small claims cases, which are one example of relatively less important court cases, an ideal object for judicial independence research. Any sign of favoritism in these cases is likely to carry over to high profile cases, and indicate a widespread deviation of justice. In addition, small claims cases are more homogeneous, and suffer less from unobservable factors which could drive court outcomes.

¹Senators serve for eight years but this does not change the theoretical predictions in this section.

In the majority of cases, litigants do not need an attorney to file a claim, defendants pay a small monetary compensation if they lose in court, and the duration of cases is shorter than in other courts in the state system. In other words, we are testing whether $\partial u_j^t / \partial u_{e,t}^t = 0$ while holding ε_t and δ_t relatively constant.

2.1 Previous Evidence on Judicial Bias

A key principle of liberal democracies is the separation of powers. It creates mechanisms of checks and balances aiming at preserving individual liberties and avoiding abuse of government power against its citizens (Persson et al., 1997; La Porta et al., 2004). An important issue in the literature is whether courts are independent, and whether an independent judiciary improves social, political, and economical outcomes in line with the principle of separation of powers. Fortunately, most of the macro-level evidence points to that direction: there is documentation of the positive impact of judicial independence on growth (Voigt et al., 2015); enforcement of contracts and private property protection (Acemoglu and Johnson, 2005); bureaucratic control (McCubbins and Rodriguez, 2009; Hanssen, 2000); the alternation of power (Ramseyer, 1994); and preservation of policies (Hanssen, 2004). From a comparative perspective, Helmke and Rosenbluth (2009) show that democracy is a necessary but insufficient condition for judicial independence and the rule of law.

In this study, we supplement the literature on judicial independence and separation of powers in two ways. First, we look at disaggregated, individual-level data on judicial bias regarding the political status of litigants. There are many studies focusing on ethnicity, race, gender, or religion bias (Abrams et al., 2012; Shayo and Zussman, 2011; Arnold et al., 2018, and many others), but only a handful of these papers focus on the political status of litigants. Helmke (2002) documents strategic defection of judges in how they rule government cases in Argentina when governments are weak; Lambais and Sigstad (2018) examine bias in corruption cases against politicians in Brazil; Sanchez-Martinez (2017) looks at favoritism of government agency defendants in Venezuela when the agency head is a member of the ruling socialist party; Poblete-Cazenave (2019) investigates whether office-holding impacts court outcomes for candidates to state legislatures in India. Departing from these studies, we also observe politicians as plaintiffs, such that we can break down the favoritism effect even further. If there is any selective litigation, we can reasonably expect that favoritism would play out differently conditional on the status of politicians as plaintiff or defendant. Secondly, we look at low salience cases, in which no favoritism is expected – as described in the previous section. Most of the literature focuses on a narrow set of criminal cases, but there is no reason to believe that political favoritism would be restricted to these cases, and this study also addresses this question.

3 Court Systems in Brazil

Brazil’s judiciary system is divided into general and limited jurisdiction courts. Federal and State Courts form the general system and Electoral, Military, and Labor Courts form the limited juris-

diction system. There are up to three instances of judicial review in either system, and the court of last resort is the Federal Supreme Court (STF). It takes up cases under its jurisdiction as set out by the Brazilian Constitution, cases in which there are conflicting norms or jurisprudence issued by lower courts, and cases where there is a direct violation of constitutional norms. To limit the sources of heterogeneity, this paper focuses on cases heard in state court systems. In particular, we focus on the state of São Paulo, the most economically and politically important state in the country.

There are 319 judicial districts in the state, and each district has one or more courthouses. These courthouses host at least one judge with either a broad mandate, meaning that they can rule on any issue within the state court system’s attribution, or a narrow mandate, which means they only oversee certain types of cases within the system, e.g., commercial or family law. Within the state system, there are specialized small claims courts called *Juízados Especiais Cíveis* (Special Civil Tribunals, in free translation, and SCTs henceforth). SCTs replaced the primary small claims courts across Brazil upon the passage of the most recent Brazilian Constitution in 1988.² Their goal is to simplify and increase access to justice across states by removing many procedural requirements present in other litigation instruments. SCTs are the primary judicial body for low complexity cases, defined as cases in which claims do not exceed 40 times the minimum wage³, and involve issues such as lease breach, consumer rights, debt executions, torts, and others. There is no need for hiring an attorney if claims are under 20 times the minimum wage. SCTs are only open to individual or small company plaintiffs.

An example helps illustrate a typical SCT case. Suppose your mobile phone service provider has been overbilling you for international phone calls that were never made. You, unfortunately, could not resolve this issue with the company’s customer service and now want to take legal action and receive financial compensation for the wrongful charges. You walk up to an SCT office, speak to a courthouse clerk, and file your claim along with any supporting documentation. The clerk then provides a court date for a conciliation hearing. At the hearing, you and the phone company will try to reach an agreement; if that fails, the judge sets trial for either later that same day or in the following days. At trial, the judge issues a sentence which can be appealed within ten days; on appeal, a three-judge panel then issues the final ruling. This entire process might take less than three months, representing a substantial improvement when compared to cases in the regular judicial process at state courts, which take up to 38 months on average to conclude (CNJ, 2018).

SCT’s structure dramatically reduces the number of dimensions driving judicial decisions. According to the São Paulo State Court website, there are less than 15 types of cases that can be brought before SCTs. It is then easier for judges and lawyers, when hired by the litigants, to specialize and reduce any skill discrepancy that could substantially drive a case outcome. The sentence is also standard across cases: the losing side will pay the claim amount to the winning side, which

²More evidence of this in Lichand and Soares (2014).

³There are no state minimum wages in Brazil, so this is the federal, nationwide minimum wage of R\$ 1,040 per month beginning in 2020. This makes the maximum claim amount R\$ 41,600, or ~\$9,900 in current dollars using the 2019 end-of-year exchange rate.

is capped at 40x the minimum wage. The standard and relatively low salience punishment removes additional sources of heterogeneity from high-profile cases, such as corruption cases in Lambais and Sigstad (2018) or violent crimes in Lim et al. (2015). Small claims cases do not attract as much media attention as corruption or violent crime, for instance. In fact, the use of small claims courts is an approach first introduced by Shayo and Zussman (2011) that takes advantage of the relative homogeneity of small claims cases in Israel to isolate the effect of ethnicity and religion on court outcomes. Lastly, judges have no control over which SCT cases they take. In single-judge court-houses, all cases are presented before the same judge; in multiple-judge courthouses, the cases are randomly distributed to judges serving on each SCT. These distribution rules reduce the sources of external variability of outcomes and prevent that cases are differently assigned to systematically more lenient (or harsher) judges in the state system.⁴

4 Data

We construct a case-level dataset (5,224 observations) with information from two sources. First, we collect SCT case and judge level information from the São Paulo State Court (TJ-SP). The court publishes all judicial decisions on their website, and the information available is the case duration, type (breach of contract, debt execution, and others above), the court where it was filed, ruling judge, the amount claimed, litigants and their lawyers (if hired), and sentences. Second, we use the information on candidates running for municipal office in the state of São Paulo in 2008, 2012, and 2016 from the Brazilian Electoral Court (TSE). TSE has jurisdiction over the entire electoral process in Brazil, from registering candidates, ruling over breach of electoral law, and overseeing the voting process on election day, to counting votes and authorizing that elected politicians take up office. It collects individual-level data on politicians and publishes everything online. We use TSE electoral results, candidate information, and electoral district data for every elected candidate in the state of São Paulo in the municipal elections of 2008, 2012, and 2016. Table 1 contains the descriptive statistics of this sample of candidates in the state.

Though the targeted case duration is three months, we can see that cases will last on average 361 days to conclude (12 months), and the average amount claimed by plaintiffs is R\$11,750 (~US\$2,790). In 64.5 percent of the cases, judges rule in favor of politicians. Sixty-one percent of judges are male and have held their position for over ten years. Their income is R\$35,152 per month, or US\$8,370, on average. The average age of candidates involved in SCT cases in the state is 44 years old, 90.1 percent are male, and 24.3 percent have previous political experience, measured as an indicator variable for candidates who have been reelected or have declared their occupation to the Electoral Court as politician of any kind (city councilor, mayor, governor, member of Congress, senator, president). Forty percent of the politicians in our sample were holding office at the time these cases were heard by the state court. On average, they spent R\$18,212 (US\$4,336) on the

⁴Yet, for robustness purposes, I replicate the process in Abrams et al. (2012) producing random distributions of court outcomes to serve as a check on the quality of the case assignment system implemented by the state of São Paulo.

campaign trail. We have also collected categorical variables for educational attainment and marital status for all elected officials, but they are omitted from the table. The most frequent educational level and marital status are a four-year college degree or equivalent (41.2 percent) and married (70.3 percent), respectively.

5 Empirical Analysis

We analyze judicial favoritism in five steps. We start off by confirm random assignment of cases. Next, we test whether the distribution of rulings in favor of politicians resembles a random distribution of court rulings. Any deviation would indicate judges do vary how they try cases having politicians as litigants. Third, we estimate the causal effect of holding office using a regression discontinuity design. This result is supplemental to judicial favoritism because it allows us to check heterogeneous effects for those who run for office and those who run and take up office. The four step is checking pro-politician bias results against a sample of similar small claims cases, decided by the same judges around the same time, but whose litigants are not politicians. Our goal is to provide a counterfactual estimate of judicial outcomes in the absence of litigant politicians. Finally, we use four machine learning algorithms to predict court outcomes and recover the most critical factors driving judges’ decisions in SCT cases.

5.1 Are Cases Assigned at Random?

To reasonably compare court outcomes across judges, our first concern is establishing that judges cannot select into cases they try. The design of the Brazilian state court system in general and SCTs in particular is the first guarantee in support of random assignment of cases. First, plaintiffs file small claims lawsuits where the wrongful act took place or where they (or the defendant) live. Judges do not choose whether to hear the case. Plaintiffs walk up to the court, speak to clerks, and are instructed on how to proceed. State judges come in only when the case has been opened by clerks and cannot choose which of the open cases to hear. Second, in districts where there is more than one judge trying cases, the allocation of lawsuits by judge is random; the distribution of cases across judges happens immediately after the case has been included in the caseload management system. Together, these two legal features indicate limited, if any, control of case allocation across judges.

Nevertheless, we provide an empirical support to random case assignment. We follow the strategy in Abrams et al. (2012), who analyze racial bias in felony cases sentencing across judges in Cook County, IL. They suggest that random assignment could be tested regressing a case characteristic (e.g., politician’s age) on multiple control variables, as below:

$$\text{Age}_{ijc} = \alpha + \sum_{k=1}^k \beta x_{kijc} + \sum_{n=1}^n \Gamma D_n + \sum_{c=1}^c \lambda_c + \varepsilon_{ijc} \quad (2)$$

Where Age_{ijc} is the politician age, x_k are k control variables, D_n is a matrix of n judge fixed-

effects, λ_c is a matrix of court fixed-effects, and subscripts i, j, c are indexing case i , judge j , and court c . Under random assignment, the F -test on the joint distribution of judge fixed-effects should fail to reject the null (i.e., fixed-effects have the same effect on politician's age). However, such direct F -test regression would likely lead to overrejection of the null as described by Abrams et al. (2012). Since the number of judges per SCT court is relatively small (an average of 1.74 judges per court and 10 cases per judge), this F -test would not meet the asymptotic properties of the F -statistic and would suffer from finite-sample bias (the judge variability within courts is small), over-rejecting the null effect of judge fixed-effects.

The solution to the overrejection problem is the construction of simulated datasets where the assignment of cases is indeed random and the subsequent comparison of statistical moments in the empirical distribution versus the simulated moments. By comparing the empirical realization of politician age per judge against its simulated distribution, we construct an unbiased test of favoritism. This exercise is as follows. First, the researcher should group the actual (empirical) sample into the randomization units, which are the many SCT courts in this study. Within these units, the researcher creates simulated observations, for each judge, from random draws (with replacement) of each of the variables in the data. Suppose there are 20 cases heard by four judges (five cases each) in a given SCT in the state of São Paulo. Each case has a set of observed characteristics, e.g., plaintiff gender, age, claim type, claim amount, whether the politician was the plaintiff or defendant in that court case, and so on. The researcher then creates 20 simulated cases, five per judge, keeping the same proportion as in the original data, where each case characteristic is randomly drawn from the sample of 20 observed cases. Once this process is replicated for all randomization units (SCTs), a simulated dataset of the same size as the empirical dataset has been created. This process is then repeated so that there are 1,000 simulated datasets.

Armed with these datasets, we compare the statistical moments in the empirical and simulated datasets without fearing overrejection. Similarly to Abrams et al. (2012), we compute the means of politician ages for all cases tried by any judge and compute the 25-75 interquartile range (IQR) for the entire distribution of judges. We report the results in figure 1. The empirical IQR for age is 7.75. This value summarizes the difference in politician age for the middle 50 percent of cases across all judges. If this empirical moment is not statistically different from a random distribution of politician age across the 1,000 datasets, then there is evidence in support of the random allocation of cases – which is what we expected from the design of SCTs in São Paulo. The p -value for an IQR of 7.75 is .072, and thus there is support for random assignment at the 5 percent level. Though one could claim that .072 is very close to the .050 threshold, we should point out that there is a small number of judges in our sample, with each judge overseeing approximately 10 cases (514 judges for 5,224 SCT cases). Therefore, we are still under the threat of overrejection of the null when compared to Abrams et al. (2012), who have an average of 489 cases per judge. In fact, if we are not able to reject random assignment with this sample, we are not likely to reject it in larger samples of small claims cases in the state.

5.2 Heterogeneous Convictions Across Judges

To measure judicial favoritism, we carry out a similar process. We produce the same simulated datasets and IQRs, but instead of looking at the distribution of a case covariate, we examine the distribution of the outcome of interest, i.e., whether the case has been ruled in favor of a politician. We present the results in figure 2, where we plot the simulated IQR range across all judges in gray. These judges vary in the harshness with which they rule cases involving politicians. The IQR can thus be interpreted as the difference in politician win probability across the harsh-lenient judge spectrum. In the simulated data, this harshness-lenient distance is 50.3 percentage points. It means that if a politician’s case were shifted away from a harsh (lenient) judge, their win probability should have increased (decreased) by 50.3 percentage points. We expected this simulated result: the 25 percent harshest judges should try in favor of politicians 25 percent of the times; the 25 percent most lenient judges should try in favor of politicians 25 percent of the times. The 50 percent in the middle of the harshness-lenient spectrum is the simulated IQR. The distribution of win rates matches the quantiles of the spectrum.

The empirical IQR, however, is 38.5 percentage points. This distance between the more strict and more lenient judges is statistically different ($p = .0001$), and smaller, than the simulated IQR. There is less variation in court outcomes in the empirical data when compared to the simulated data, which means that the distribution of court outcomes for litigant politicians is not random. While we document more homogeneous outcomes for politicians, we do not know the direction of the court decisions. In other words, it means we still need to answer whether these systematic, homogeneous decisions are in favor or against politicians. To answer this question, we examine the politician win rate across all 514 judges. The results are presented in figure 3, where we run the same simulation exercise and compare the empirical distribution of pro-politician decisions against a simulated distribution of pro-politician decisions. We use the same process as before to reflect judges’ heterogeneous harshness-lenient preferences, in which we draw outcomes for each judge from the same randomization units (e.g., the SCT court). In figure 3, we plot a empirical win rate of 60.8 percent of all cases against the distribution of simulated win rates (mean = 64.5; standard deviation = .006). Politicians see a 3.7 percentage points. smaller probability of winning a small claims court case compared to their simulated win probability. Therefore, we find evidence of systematic punishment of politicians in SCT court.

Contrary to our expectation, state judges do not favor politicians in anticipation of future benefits. Indeed, they are harder than expected on politicians. A potential explanation for this result is that these unfavorable decisions are power moves. Judges punish politicians in these low salience cases as warning signs of future hostility in case politicians pass on their judicial benefits. The judiciary is still using decisions as means of extracting benefits, but rather than using a carrot (favorable decisions), they are using a stick (unfavorable decisions) to keep politician behavior in check. There are no ways of testing this mechanism in the absence of more data on judicial decisions and the involvement of politicians in court cases. Future studies should investigate this mechanism in other court systems and types of cases.

5.3 Does Holding Office Matter?

In addition to the negative effect documented above, an interesting question is whether judges respond differently to politicians who are holding office at the time of trial. Since office-holding politicians yield more power compared to politicians seeking office because the former are effectively making policy decisions, a plausible hypothesis is that judges would target elected politicians more often. The ideal experiment to answer this question would be to randomly assign office across politicians involved in court cases and compare their outcomes. We would be able to compare how non-politicians fare against politicians and pin down the exact bias, in favor or against office-holders. For ethical reasons, however, this is an unrealistic experiment, and the regression discontinuity design we adopt in this paper is its best approximation yet.

We start by subsetting our sample to mayor candidates. They face each other off in majoritarian elections, and the candidate who reaches 50 percent plus one of the votes is elected to office. This design allows us to compute the vote share for each mayor candidate and compare to the 50 percent threshold: politicians elected to office with narrow margins approximate random sorting to office when compared to politicians who barely lost their election (Lee, 2008). Vote margin is the running variable for a standard regression discontinuity design where the treatment condition is holding elected office at the time of trial. Formally, we estimate the following equation:

$$y = \alpha + \gamma_1(x - c) + \gamma_2(x - c)^2 + \rho_1(x \geq c) + \rho_2(x - c)(x \geq c) + \varepsilon \quad (3)$$

Where y is whether the elected politician has received a favorable SCT ruling; $(x - c)$ is the vote share centered at 50 percent, and the squared term captures decreasing returns to scale in vote share; $(x \geq c)$ is the treatment indicator variable, i.e., whether the candidate has been elected; the interaction $(x - c)(x \geq c)$ summarizes differential trends on both sides of the vote share threshold. Using the bandwidth selection model from Calonico et al. (2015), ρ_1 represents the causal effect of holding office on SCT outcome. Figure 4 plots the results of equation 3 using the optimal bandwidth of 8.4 percentage points as calculated following the Calonico et al. (2015) model as well. Though there seems to exist a positive correlation between holding office and favorable SCT rulings, the causal effect is not significant at $p = .05$. Mayor candidates who barely lost or won an election seem to perform equally well in court. We force a discontinuity at vote margin zero to make the causal effect explicit, but the potential causal impact is just masking the non-linear relationship between x (vote margin) and y (court outcome). The dashed line depicts the polynomial regression on the entire sample and serves as evidence of this non-linearity.

Figure 5 presents a more detailed picture of the relationship between politicians and judges. We re-estimate equation 3 using different bandwidth sizes (as suggested by Lee and Lemieux, 2010, and others), which are reported on the x -axis. The Calonico et al. (2015) optimal bandwidth is again 8.4 percentage points (in blue). The y -axis displays the point estimates for ρ_1 in each regression equation. The results confirm a positive and significant relationship (at $p = .05$) between vote share and favorable court outcomes across the wider bandwidths (40-15 percentage point margins).

However, since the significance is present in larger bandwidths, and candidates in these wider ranges away from the zero cutoff are not likely comparable against each other, we cannot conclude this effect is causal. As we narrow in on smaller margins, we lose statistical power, and the election coefficient becomes insignificant.⁵ We can conclude that there is a positive correlation between vote share and pro-politician bias, but we cannot know the exact magnitude of this effect.

In sum, there is evidence of a positive and significant relationship between vote margin and SCT outcome, but we cannot recover the unbiased, causal effect of holding office on court outcomes.⁶ A straightforward explanation to the null effect is the visibility of mayors in Brazil. Since municipal governments are in charge of health and education policies, mayor candidates are well-known local figures, and much more important than counterparts in U.S. local elections. Candidates run for office multiple times and could be perceived as politically important even when they are not in office. This visibility would explain why office might not matter for court outcomes. Finally, though the positive correlation here might seem at odds with the negative bias in section 5.2, we note that the sample used for the causal test in this section is only composed of mayor candidates. The original sample contains people running for both mayor and city council seats. It seems possible that judges adjust their decision-making conditional on these power differences across politicians, and that city council candidates are punished more harshly. Unfortunately, we cannot run the same simulations from the previous section to capture the win probabilities for mayors due to the small number of these candidates involved in SCT cases ($n = 489$). Since city councilors are elected following proportional rule, neither can we use vote share to identify the causal effect of holding office for the city council sample.

5.4 Comparing Politician vs. Non-politician Rulings

An important concern with our results is the extent to which our sample is representative of the universe of decisions handed out by SCT judges in the State of São Paulo. If there is any selection into the SCT courts included in this study, then we might not be catching the true case win probability for politicians. For instance, state judges serving at these courts might be naturally harsher, or have a lesser opinion of politicians, and thus the outcomes we observe reflect a single draw from the distribution of court outcomes across the state. The results here have not external validity.

To address these concerns, we reproduce the results in section 5.2 using another sample of small claims cases, decided by the same judges around the same time, but whose litigants are not politicians. Except for the absence of litigant politicians, these cases share all other characteristics with the primary sample. We construct this sample using the unique identification number assigned

⁵Despite the larger causal effect at the smallest margin (1 percentage point), the small sample size prevents us from making stronger inference claims about the relationship between holding office and seeing a favorable result in court.

⁶We also find null causal effects using difference-in-differences as an alternative identification strategy. We use an indicator for trial decision issued after previous election (time variable) and another for when politicians were elected to office (treatment variable). Their interaction, i.e., the causal effect, is null regardless of the model we estimate (including or excluding covariates and fixed-effects). These models and their results are available upon request.

to all court cases, in all states, as required by law in Brazil. This case ID is a sequence of 20 digits structured as follows: the case filing order (7 digits), a control sequence (2 digits), the filing year (4 digits), the court system identifier (3 digits), and a judicial district identifier (4 digits). To construct the counterfactual sample, we marginally change the case number to recover lawsuits filed immediately before and after every politician case, but which were still decided by the same judge at the same time. In other words, we add and subtract one to each case number, keeping all other digits the same, and use these new case IDs to find lawsuits and their information at the São Paulo state court website. Table 2 displays a sample case ID from the politicians sample: case 3002615 was filed in 2013 at court system 8, state 26, district 0510. In this case, the IDs for constructing the counterfactual sample are ~~3002614~~-95.2013.8.26.0510 and ~~3002616~~-65.2013.8.26.0510.⁷ Since our primary sample has 5,224 observations, we recover 10,448 potential lawsuits to serve as counterfactual to the main analysis. After excluding invalid IDs, lawsuits for which there is no information available online, and non-SCT cases tried at the same judicial district as SCT cases, our final counterfactual sample contains 3,233 cases.

Table 3 displays descriptive statistics of the counterfactual sample (panel A) and the tests for mean differences across original and counterfactual sample (panel B). To compare across groups, we keep the same variables except for the court decision measure. Since there are no politicians in the alternative sample, we cannot compare pro-politician rulings. The percentage of cases ruled in favor of claimants or defendants can be calculated and compared across samples. We chose the former. Panel B shows there are significant differences (at $p = .05$) for case claim amount and the share of claimant win, but no differences in judge characteristics across samples. These case-level variables, however, ignore shared variation across outcomes due to judges trying multiple cases at a judicial district. We thus suggest measuring outcomes at the judge-level, computing average win rates for all judges in the sample. The results of this exercise are presented in table 4. When claimant win rates are computed across judges, we observe no significant difference between the politician and the non-politician samples ($p = .24$). The only remaining difference is the average amount claimed in court. Politician cases are associated with a higher claim amount than in the alternative cases. We believe this difference comes from the relationship between claim amount and defendant type (individual or corporation). Since companies tend to settle higher claims before trial, we observe lower claim amounts for corporate defendants. The correlation between claim amount and corporate defendant is $-.247$. The politician sample has fewer corporate defendants (49.3 percent), thus higher claims, compared to the non-politician sample, where corporate defendants are 66.2 percent of the total.

We believe this difference comes from the lower share of corporate defendants, who tend to settle higher claims before trial, in the primary sample. We identify this negative correlation between claim amount and corporate defendants ($\rho = -.247$), supporting the higher claims in the politician sample (49.3 percent of corporate defendants) compared to the non-politician sample (66.2 percent

⁷The control sequence changes but its calculation is public, so we can, and effectively do, reverse-engineer these two digits for all cases.

of corporate defendants).

5.5 Predicting Drivers of Court Outcomes

In the previous sections, we provided evidence showing that the distribution of SCT cases is as good as random. In addition, we identified a positive correlation between vote share and pro-politician rulings, supporting the hypothesis of judicial favoritism. In this section, we investigate whether politician characteristics are more relevant than case characteristics for predicting court outcomes.

To this end, we test four machine learning models and compare how well they predict court outcomes. In every model, the court outcome is the dependent variable, and the characteristics of the politician and case are the predictors. We apply an 80-20 percent split across train and test datasets, and use the latter dataset to construct performance measures of each model. Table 5 reports accuracy and Cohen’s Kappa scores (Landis and Koch, 1977). The four models are (1) Logistic regression with Lasso regularization (Tibshirani, 1996); (2) Random Forest (Breiman, 2001); (3) Gradient Boosting (Friedman, 2001); and (4) Deep Neural Networks using dense layers (Goodfellow et al., 2016). The best performing model is random forest, with 77% out of sample accuracy. Thus, we choose such a model for recovering the most critical factors driving judicial decisions.

Figure 6 shows the variable importance plot based on the Mean Decrease in Gini score. The Gini coefficient of a decision tree measures the contribution of each variable to the homogeneity of the tree’s nodes and leaves. Variables that result in nodes with higher purity (more homogeneous) have a higher value in the Gini coefficient. The Mean Decrease in Gini is just the average of Gini coefficients for all decision trees used to build the model, and higher values also indicate higher variable importance. The most critical factors are judge tenure, followed by the amount claimed in court, judge pay, and candidate vote share, which align well with the original hypothesis in section 2.⁸ We posited that politicians and judges develop a mutually beneficial relationship; i.e., they collude with one another when their utility of cooperation is higher than their utility of competition. For instance, ruling in favor of a politician in an SCT case might create a positive attitude towards increasing resources available to the judicial branch. Knowing that judge tenure and pay are signals of experience, politician vote share is a signal of ability, and both politicians and judges repeat the litigation interaction multiple times during their careers, the random forest classification is supportive of the theoretical model discussed in section 2.

Interpreting non-linear models such as random forests is not easy. To evaluate the direction of the effect of these explanatory variables on the predicted outcome, we have to analyze partial dependence plots, or PDPs in short (Friedman, 2001). PDPs recover the marginal effect of an explanatory variable x_s on the outcome of the model by integrating x_s over the range of values x_c

⁸We ignore the indicator variable for when politician is the defendant because this makes up for no more than ten percent of all cases, so it is not relevant for all other cases. All other variables, including candidate age, are have a much smaller contribution to predictions.

(equation 4):

$$\hat{f}_{x_s}(x_s) = \mathbb{E}_{x_c}[\hat{f}_{x_s}(x_s, x_c)] = \int \hat{f}_{x_s}(x_s, x_c) d\mathbb{P}(x_c) \quad (4)$$

Where \hat{f} is the adjusted prediction function, and x_c is the set of all other explanatory variables used to build \hat{f} . In the case of a linear model, for example, \hat{f}_{x_s} is always a linear function. Figure 7 displays the PDPs of the four most important variables detected in the random forest model. The y -axis displays how the probabilities of politicians winning their SCT cases change with marginal changes on each of the four variables. The shaded area under the x -axis shows the concentration of observations in the test data at each value, and larger blocks indicate more observations at a particular level of the explanatory variable.

Since these are non-linear marginal effects, and they are a function of the values of the other covariates, there is no straightforward interpretation of effects. The best we can do is observe trends and compare to linear models or previous results in the literature. Judge tenure and election share, for instance, are generally increasing over the range of their values, indicating a positive contribution to win probability. Lambais and Sigstad (2018) have the most comparable result, where the probability of winning in corruption cases is larger for politicians who are tried by more experienced judges. The positive contribution of election share here matches the positive relationship reported in the regression discontinuity analysis. Together, these factors strongly support judge tenure and election as the main drivers of pro-politician bias. On the other hand, both the marginal effect of the total amount claimed in court and judge pay display a downward trend on pro-politician rulings. This effect is not surprising, however, and summarizes the decreasing returns to scale from the marginal *real* on judges' utilities.

6 Conclusion

This paper investigates whether there is differential treatment of politicians in small claims courts in the state of São Paulo, Brazil. To our knowledge, it is the first paper to produce clear evidence on judicial bias in favor of politicians even in relatively low importance settings. Previous studies in the literature measure similar effects, such as how the political ideology of judges and the salience of cases might bias judicial decision-making. This project supplements these previous initiatives by investigating a more nuanced yet undetected judicial favoritism. Having documented bias in small claims courts, we certainly anticipate more deviation in high profile cases, such as corruption or electoral crimes.

Besides the evidence of judicial favoritism of politicians, this project also makes predictions of court outcomes and pinpoints the most critical factors driving judicial bias, i.e., judge tenure, amount claimed in case, judge pay, and politician vote share. Though we only focus on a small set of cases, those filed in São Paulo's special civil tribunals (SCTs), the predictions should serve as a benchmark for the deviations in other judiciary systems. The majority of studies in the political economy literature posit that an independent judiciary is crucial for checking the power of the

executive and the legislative, and for supporting economic development (Baland et al., 2010, and many others). Thus, this study provides valuable insights for policymakers in developing countries sharing similar institutional designs.

Future projects should adopt new strategies to identify the causal effect of holding office on court outcomes. We find a positive correlation for mayor candidates, but we are unable to partial out the unobservable bias, such as a politician’s ability, or the experience they have with state court systems. Thus, we only provide the magnitude of the correlation between holding office and pro-politician rulings in SCT courts. Projects identifying the causal effect or widening the scope to also include the effect for city council candidates would be great contributions to this literature.

References

- Abrams, D. S., Bertrand, M., and Mullainathan, S. (2012). Do Judges Vary in Their Treatment of Race? *The Journal of Legal Studies*, 41(2):347–383.
- Acemoglu, D. and Johnson, S. (2005). Unbundling Institutions. *Journal of Political Economy*, 113(5):949–995.
- Arnold, D., Dobbie, W., and Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, 133(4):1885–1932.
- Baland, J.-M., Moene, K. O., and Robinson, J. A. (2010). Chapter 69 - Governance and Development. In Rodrik, D. and Rosenzweig, M., editors, *Handbook of Development Economics*, volume 5, pages 4597–4656. Elsevier.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2015). Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association*, 110:1753–1769.
- CNJ (2018). Justiça Em Números. Technical report.
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):1189–1232.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT press.
- Hanssen, F. A. (2000). Independent Courts and Administrative Agencies: An Empirical Analysis of the States. *Journal of Law, Economics, & Organization*, 16(2):534–571.
- Hanssen, F. A. (2004). Is There a Politically Optimal Level of Judicial Independence? *American Economic Review*, 94(3):712–729.
- Helmke, G. (2002). The Logic of Strategic Defection: Court-Executive Relations in Argentina under Dictatorship and Democracy. *American Political Science Review*, 96(2):291–303.
- Helmke, G. and Rosenbluth, F. (2009). Regimes and the Rule of Law: Judicial Independence in Comparative Perspective. *Annual Review of Political Science*, 12(1):345–366.
- La Porta, R., López-de Silanes, F., Pop-Eleches, C., and Shleifer, A. (2004). Judicial Checks and Balances. *Journal of Political Economy*, 112(2):445–470.
- Lambais, G. and Sigstad, H. (2018). *Judicial Subversion: Evidence from Brazil*. Mimeo, Harvard University.
- Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.
- Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, 142:675–697.
- Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48:281–355.
- Lichand, G. and Soares, R. R. (2014). Access to Justice and Entrepreneurship: Evidence from Brazil’s Special Civil Tribunals. *The Journal of Law and Economics*, 57(2):459–499.
- Lim, C. S. H., Snyder, J. M., and Strömberg, D. (2015). The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems. *American Economic*

- Journal: Applied Economics*, 7(4):103–135.
- Lu, H., Pan, H., and Zhang, C. (2015). Political Connectedness and Court Outcomes: Evidence from Chinese Corporate Lawsuits. *The Journal of Law and Economics*, 58(4):829–861.
- McCubbins, M. D. and Rodriguez, D. B. (2009). The Judiciary and the Role of Law.
- Persson, T., Roland, G., and Tabellini, G. (1997). Separation of Powers and Political Accountability. *The Quarterly Journal of Economics*, 112(4):1163–1202.
- Poblete-Cazenave, R. (2019). *Crime and Punishment: Do Politicians in Power Receive Special Treatment in Courts? Evidence from India*. Mimeo, University-College London.
- Rachlinski, J. J. and Wistrich, A. J. (2017). Judging the Judiciary by the Numbers: Empirical Research on Judges. *Annual Review of Law and Social Science*, 13(1):203–229.
- Ramseyer, J. M. (1994). The Puzzling (In)Dependence of Courts: A Comparative Approach. *The Journal of Legal Studies*, 23(2):721–747.
- Sanchez-Martinez, C. (2017). *Dismantling Institutions: Court Politicization and Discrimination in Public Employment Lawsuits*. PhD Dissertation, Stanford University.
- Shayo, M. and Zussman, A. (2011). Judicial Ingroup Bias in the Shadow of Terrorism. *The Quarterly Journal of Economics*, 126(3):1447–1484.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Voigt, S., Gutmann, J., and Feld, L. P. (2015). Economic Growth and Judicial Independence, a Dozen Years On: Cross-country Evidence Using an Updated Set of Indicators. *European Journal of Political Economy*, 38:197–211.

Tables and Figures

Table 1: Descriptive Statistics

	<i>n</i>	Mean	St. Dev.	Min	Max
<i>Case Level</i>					
Case Duration (in days)	5,224	361	433	1	5,416
Amount Claimed (in R\$)	5,224	11,750	10,633	35	40,000
Pro-Politician Ruling	5,224	.645	.478	0	1
<i>Judge Level</i>					
Male	514	.611	.488	0	1
Tenure (in days)	514	4,385	2,892	13	12,987
Wage (in R\$)	514	35,152	10,797	13,156	145,616
<i>Candidate Level</i>					
Age	2,943	44	10.3	18	78
Male	2,943	.901	.298	0	1
Political Experience	2,943	.243	.430	0	1
Elected to Office	2,943	.402	.500	0	1
Campaign Expenditures (in R\$)	2,943	18,212	77,890	11	1,770,315

Figure 1: Interquartile Range of Candidate Age by Judge

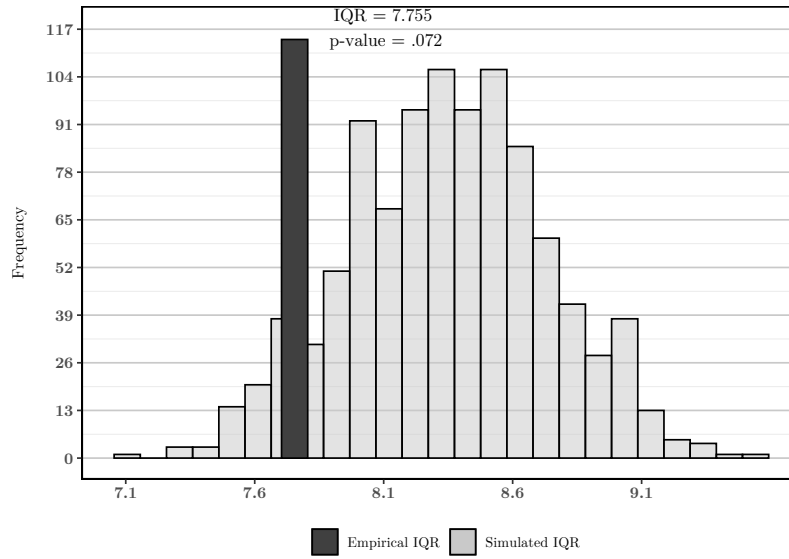


Figure 2: Interquartile Range of Favorable Ruling by Judge

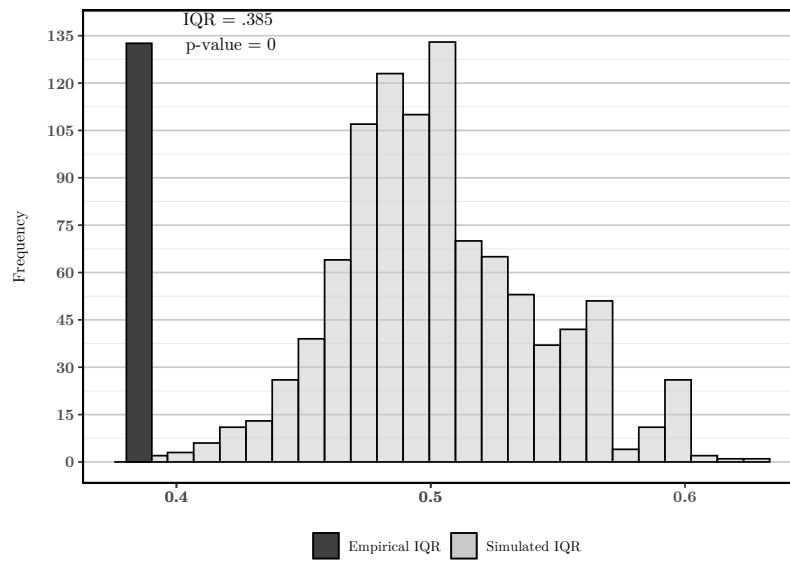


Figure 3: Comparison of Politician Win Probabilities Per Judge

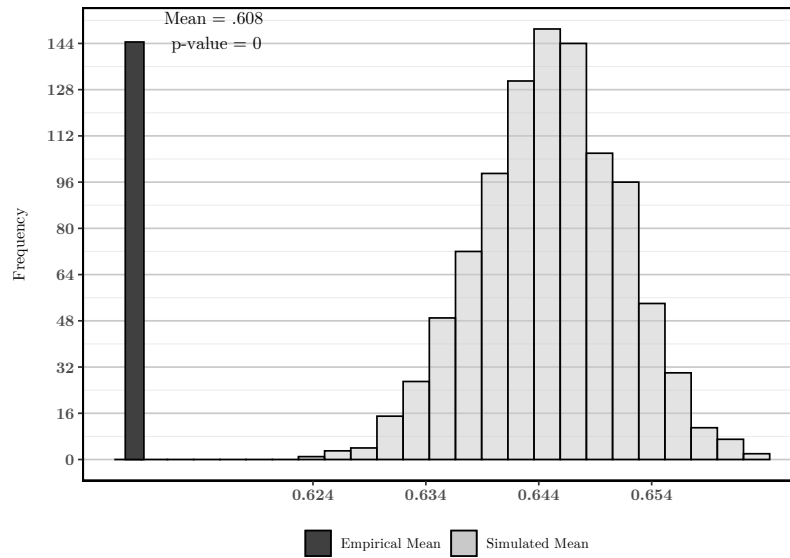


Figure 4: The Causal Effect of Election on Court Outcomes

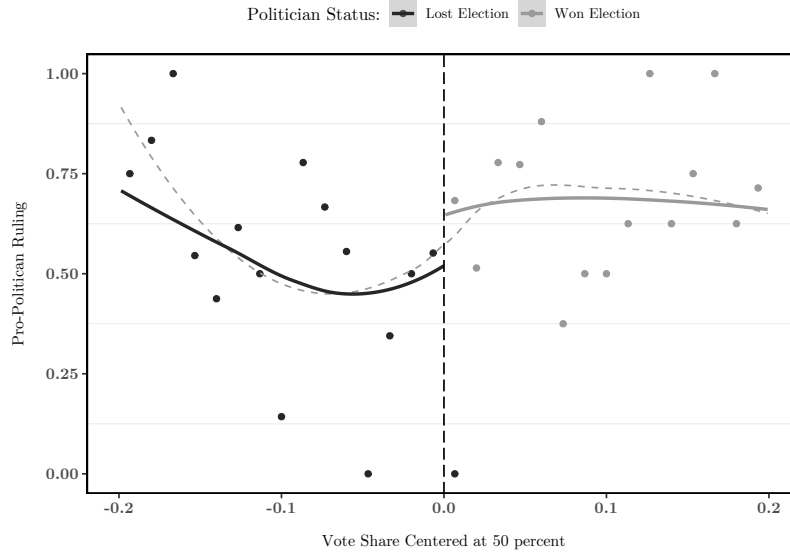


Figure 5: Election Point Estimates (and their 95% CIs) on Court Outcomes

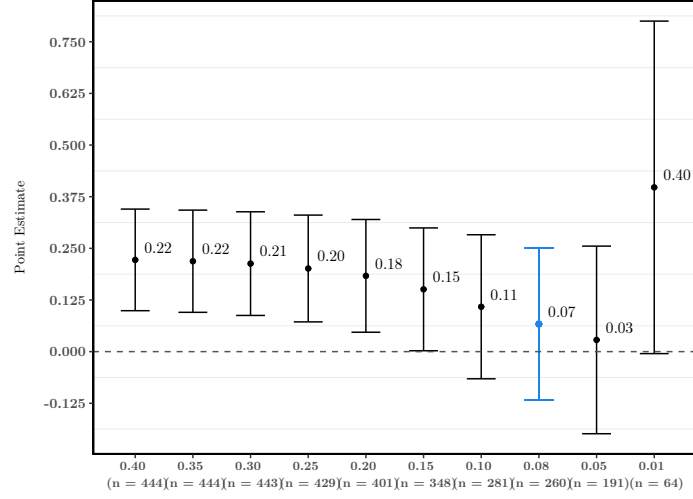


Table 2: Individual Identifier for Lawsuits in Brazil

3002615–80.2013.8.26.0510	{	3002615: case filing order
		80: control sequence
		2013: filing year
		8: court system ID
		26: state ID
		0510: district ID

Table 3: Descriptive Statistics Across Politician and Non-Politician Sample

<i>Panel A: Non-Politician Sample</i>					
	<i>n</i>	Mean	St. Dev.	Min	Max
<i>Case Level</i>					
Case Duration (in days)	3,233	353	380	1	3,336
Amount Claimed (in R\$)	3,233	9,565	9,631	10	40,000
Claimant Win Rate	3,233	.749	.434	0	1
<i>Judge Level</i>					
Male	389	.622	.485	0	1
Tenure (in days)	389	4,298	2,824	64	12,987
Wage (in R\$)	389	35,238	11,334	13,156	145,616
<i>Panel B: Mean Difference Across Samples</i>					
	Mean Politician Sample (<i>n</i> = 5,224)	Mean Non-politician Sample (<i>n</i> = 3,233)	Mean Difference	<i>t</i> -statistic	<i>p</i> -value
<i>Case Level</i>					
Case Duration (in days)	360	353	7.516	.837	.402
Amount Claimed (in R\$)	11,749	9,565	2,184	9.737	.000
Claimant Win Rate	.729	.749	-.019	-1.979	.048
<i>Judge Level</i>					
Male	.611	.622	-.011	-.343	.732
Tenure (in days)	4,384	4,297	87.055	.454	.650
Wage (in R\$)	35,151	35,237	-86.074	-.115	.908

Table 4: Mean Difference of Outcomes at the Judge-Level

	Mean Politician Sample (<i>n</i> = 514)	Mean Non-politician Sample (<i>n</i> = 389)	Mean Difference	<i>t</i> -statistic	<i>p</i> -value
Claimant Win Rate	.734	.757	-.023	-1.184	.237
Amount Claimed (in R\$)	12,500	10,308	2,192	4.401	.000
Claimant Politician Win Rate	.721	.757	-.036	-1.787	.074
Defendant Politician Win Rate	.220	.243	-.023	-.954	.340

Table 5: Performance Measures For All Models

<i>Model</i>	Accuracy	Kappa
1. Random Forest	77.31%	19.43%
2. Gradient Boost	75.00%	15.86%
3. Deep Neural Networks	73.65%	13.78%
4. Lasso	73.46%	13.49%

Figure 6: Random Forest Variable Importance

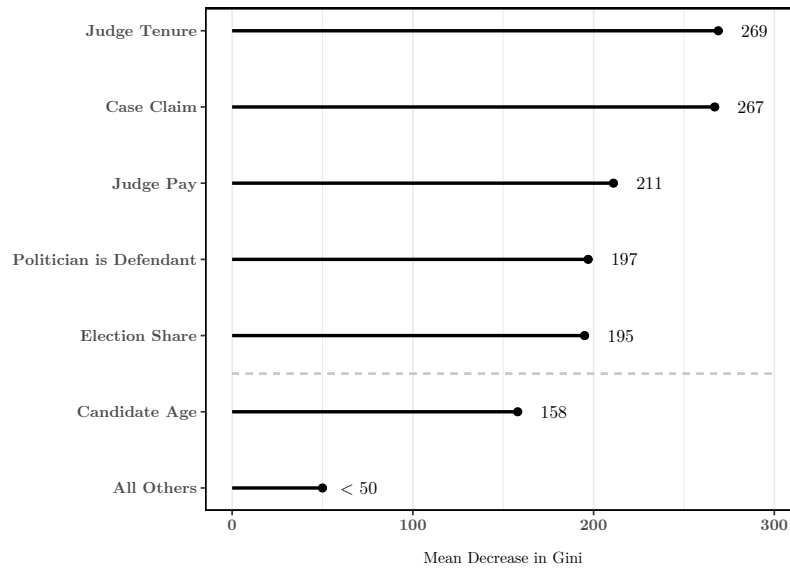


Figure 7: PDPs for the four most critical predictors

