

Judicial Favoritism of Politicians: Evidence from Small Claims Courts*

Andre Assumpcao[†]

Julio Trecenti[‡]

September 19th, 2019

Abstract

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

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 when the judge and litigants have the same ethnicity. Abrams et al. (2012) find that African American defendants are 18 percentage points 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 (2018) 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 presents the institutional environment of Brazil’s State Court Systems. Section 3 summarizes the dataset used for analysis; the theoretical framework is presented in section 4; section 5 discusses the analytical strategy, and section 6 predicts court outcomes for our sample of politicians. We conclude in section 7.

2 Institutional Background

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 jurisdiction 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 1988.¹ 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 wage², 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

¹More evidence of this in Lichand and Soares (2014).

²There are no state minimum wages in Brazil, so this is the federal, nationwide minimum wage at R\$ 954.00 in 2018. This is equivalent to \$10,500 in current dollars using the 2019 exchange rate average

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 alter a case outcome. The sentence is also standard across cases: the losing side will pay the claim amount to the winning side, which 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 courthouses, 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.³

3 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\$3,000). 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\$9,000, 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

³Yet, 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.

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 elected to office at the time these cases were heard by the state court. On average, they spent R\$18,212 (US\$4,500) on the 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.

4 Theory

Suppose there are three representative agents in Brazil, 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 serves on a four-year mandate,⁴ which 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

⁴Senators serve for eight years but this does not change the theoretical predictions in this section.

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. 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.

5 Empirical Strategy

We analyze judicial favoritism in three steps. First, we test whether the distribution of politician-favorable outcomes resembles a random distribution of court outcomes. Any deviation would indicate judges do vary how they try cases having politicians as litigants. Next, we estimate the direction and magnitude of the effect using a regression discontinuity design. Using RD allows us to simultaneously test both the significance and magnitude of the causal effect of holding office on court outcomes. 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. This is guaranteed by the design of the state court system in general and SCTs in particular. 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 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. Still, we provide an empirical test of 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., 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 plaintiff 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 are equal). 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, this F -test would not meet its asymptotic properties 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 provides an unbiased test of favoritism. The process is as follows. First, the researcher should group the actual (empirical) sample into the randomization units, which are the many SCTs in this case. Within these units, the researcher creates simulated observations, for each judge, from random draws (with replacement) of each of the variables in the empirical sample unit. Suppose there are 20 observed 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, whether the plaintiff was a politician at the time the case was heard, claim amount, 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 actual 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 7.73. 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 7.73 is .065 across the distribution of simulated IQRs, and thus supports random assignment at the 5 percent level. Though one could claim that .065 is very close to the .05 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 make random draws of the outcome of interest, i.e., whether the case has been ruled in favor of the politician. We present the results in figure 2, where we plot the simulated interquartile range across all politicians in gray. The average distance between harsher judges (25 percentile of favorable outcomes) and more lenient judges (75 percentile of favorable outcomes) is .502. In other words, a simulated, random distribution of pro-politician outcomes should vary by about 50 percentage points for the middle 50 percent of state judges. The empirical IQR, however, is .342, which means that the distance between the harsher and more lenient judges is statistically different ($p = .0001$), and smaller, in the observed data instead of the random data. There is less variation in court outcomes in the empirical data when compared to the simulated data, which means that the distribution of pro-politician outcomes is not random. Politicians see more consistent court outcomes than everyone else when litigating in SCTs in São Paulo. Though this result signals judicial favoritism, we are not yet sure of the direction of such bias. The variability of court outcomes is small but could be pro or against politicians. To test this, we use a regression discontinuity strategy in the following subsection.

5.3 Does Holding Office Matter for Judges?

The ideal experiment to answer the question on the causal effect of being a politician on court outcomes would be to randomly assign politicians as litigants in an SCT case. We would be able to compare how non-politicians fare against politicians and pin down the exact judicial favoritism. For ethical reasons, however, this is an unrealistic experiment – the approach adopted in this paper is its best approximation yet.

We start by focusing on the subsample of 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 of each mayor candidate and compare to the 50 percent mark: politicians elected to office with narrow margins approximate random sorting to office when compared to politicians who barely lost their election (Lee, 2008). The running variable for this regression discontinuity design is vote margin, and the treatment condition is being elected. 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 election on SCT outcome.

Figure 3 plots equation 3. Though there is a positive correlation between vote share and receiving a favorable ruling in SCT cases, there is no causal relationship between election and favorable outcomes in court. Mayor candidates who barely lost or won an election seem to perform equally well in court. Although we force a discontinuity at vote margin zero to make the causal effect visible, we also plot the same polynomial regression on the entire sample (dashed line) to show that the discontinuity is just masking the non-linear relationship between x and y .

Figure 4 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 8.4 percentage points (in blue). The y -axis displays the point estimates for ρ_1 in each regression equation. The results confirm the positive and significant relationship between vote share and favorable court outcomes. However, since the significance is present in larger bandwidths, and candidates in these wider ranges from the zero cutoff are not comparable, 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. One straightforward hypothesis is that judicial favoritism comes from running for office instead of holding office, in particular when it comes to mayor candidates. Since municipal governments are in charge of health and education policies in Brazil, 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 local politicians have better odds in SCT cases.

6 Prediction 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

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.

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 2 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 5 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 4.⁶ 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 4.

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 (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 6 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

⁶We 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.

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.

7 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.
- 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.
- 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.
- 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.
- Sanchez-Martinez, C. (2018). *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.

Tables and Figures

Table 1: Descriptive Statistics

	N	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

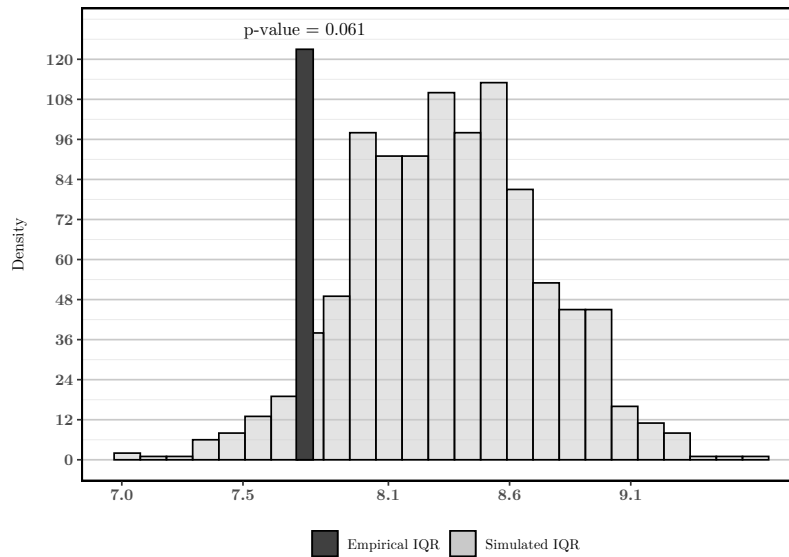


Figure 2: Interquartile Range of Favorable Ruling by Judge

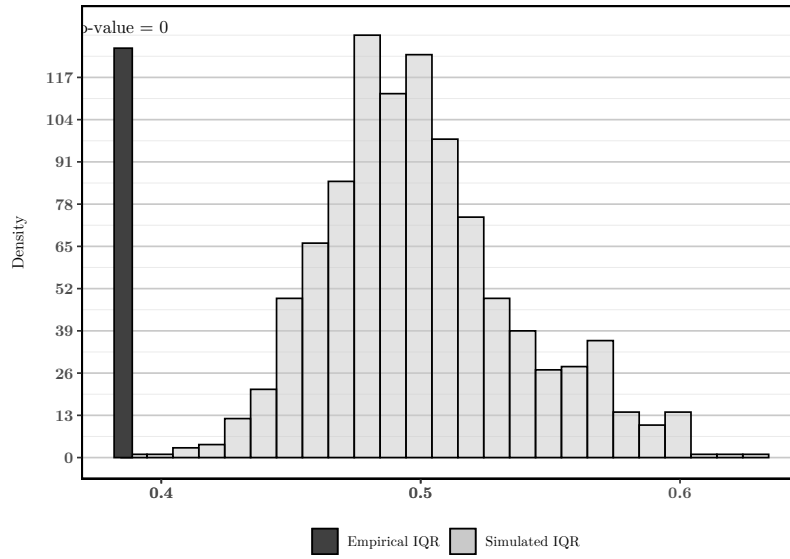


Figure 3: The Causal Effect of Election on Court Outcomes

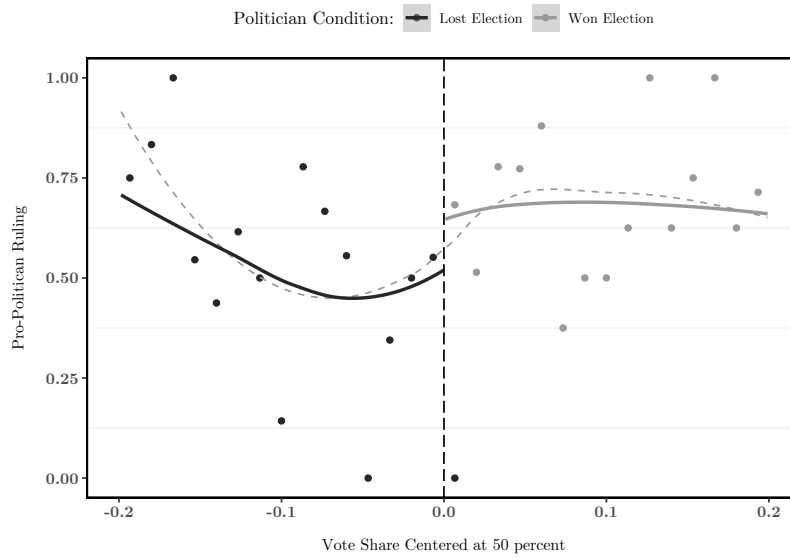


Figure 4: Election Point Estimates (and their 95% CIs)
on Court Outcomes

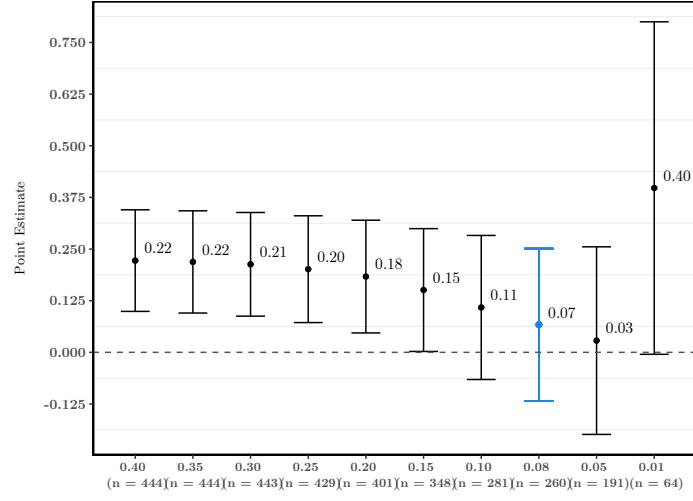


Table 2: Performance Measures For All Models

<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 5: Random Forest Variable Importance

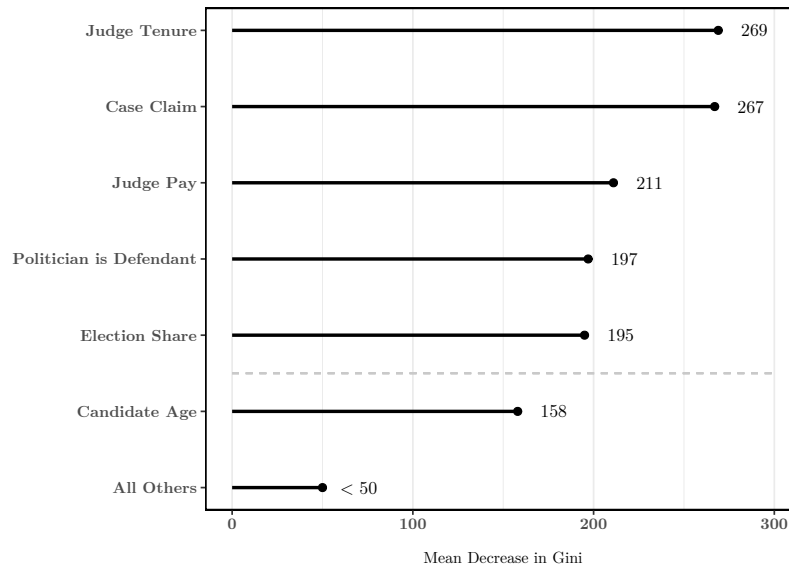


Figure 6: PDPs for the four most critical predictors

