Measuring the Impact of Elections on Judge Behavior Using Machine Learning and Economics Tools

by

Caroline Chin

B.S., Computer Science and Engineering, Massachusetts Institute of Technology (2016)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author . . . .
Department of Electrical Engineering and Computer Science
August 16, 2021

Certified by . . . .
David Autor
Ford Professor of Economics
Thesis Supervisor

Accepted by . . . .
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Measuring the Impact of Elections on Judge Behavior Using Machine Learning and Economics Tools

by

Caroline Chin

Submitted to the Department of Electrical Engineering and Computer Science
on August 16, 2021, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Judges play a critical role in maintaining a fair and independent criminal justice system. Using a combination of empirical tools from Computer Science and Economics, this paper examines the effects of judicial elections on decisions by magistrate court judges in Pennsylvania. I find that judges who are running in contested primary races dismiss fewer cases in the months leading up to their election. This effect is driven mostly by changes in the treatment of misdemeanor cases. Judges running in competitive primary races dismiss 16.2% fewer misdemeanor cases in the three months preceding their election date. This effect is consistent across estimates derived from linear regression methods as well as machine learning methods including lasso, decision tree and random forest models. In the context of prior research, these findings suggest that electoral pressure induces harsher treatment by judges across all stages of the judicial system.

Thesis Supervisor: David Autor
Title: Ford Professor of Economics

Acknowledgments

I would like to thank David Autor for his guidance and support throughout this project. I would also like to thank Hannah Shaffer, Natalia Emanuel and Emma Harrington for generously sharing the data used for this research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Judges play a critical role in maintaining a fair and independent criminal justice system. Thirty nine states use some form of election process to select judicial officials (Brennan Center for Justice 2021). However, the practice of electing judges remains a controversial issue. On the one hand, elections allow local residents to hold judicial officials accountable for inequitable or poor decision-making. On the other hand, electoral pressure can incentivize judges to pander to the political opinions of their voters at the expense of delivering impartial treatment to defendants. In this paper, I use a combination of empirical tools from Computer Science and Economics to examine the effects of judicial elections on decisions by magistrate court judges in Pennsylvania.

This paper builds off of an existing body of research investigating electoral patterns in judicial decision-making. Gordon and Huber (2007), Berdejó and Yuchtman (2013), Abrams et al. (2019), and Dippel and Poyker (2019) demonstrate that cases that occur closer to an election are associated with longer sentences. Park (2017) and Harvey and Yntiso (2020) find that minority defendants disproportionately experience the more punitive effects from changes in criminal justice outcomes driven by judicial electoral cycles. Lim, Snyder Jr, and Strömberg (2015) compare behavior of judges under partisan election, non-partisan election and appointment selection systems and find that sentences for violent offenses increase among cases assigned to judges who are in non-partisan competitive elections. Overall, the literature suggests that judges who are selected through elections issue harsher decisions in the months leading up to their election.

Most of the existing literature is focused on judges who occupy higher courts, largely ignoring the behavior of lower court judges who are responsible for pretrial proceedings. Magistrate court judges in Pennsylvania set bail and decide whether cases are dismissed or sent to the Court of Common Pleas where trials are held. Recent evidence has established a causal link between pretrial outcomes and subsequent changes in sentencing lengths and conviction rates (Dobbie, Goldin, and Yang 2018; Heaton, Mayson, and Stevenson 2017). Thus, they play an important role in the evolution of a case throughout the court system. In this paper, I focus specifically on investigating how electoral cycles affect case dismissal rates at the magistrate court level. Dismissals can lead to the elimination of further pretrial detention or possible conviction[1]. In addition, dismissing a case curtails the duration of a criminal proceeding, which can otherwise span many months. I find that magistrate court judges dismiss fewer charges when they are up for election and that this pattern is primarily driven by judges who are running in contested races. The effect begins in the months leading up to the election's filing deadline and peaks between the filing deadline and the primary election date. Finally, cases that involve misdemeanor offenses are the most impacted by changes in judicial behavior. In the context of results from existing literature, this suggests that electoral pressure induces harsher treatment across all levels of the judicial system.

To examine the stability of the estimated treatment effect under more flexible functional form assumptions, I draw on literature at the intersection of Computer Science and Economics that employs machine learning procedures to estimate causal parameters. Machine learning algorithms are typically optimized for creating precise predictions. While the final prediction model may contain information about the contribution of different covariates in the final outcome, attempting to recover average treatment effects directly from a machine learning prediction model leads to unsta-

[1]The prosecutor is able to refile charges if they disagree with the dismissal, but will have to restart the process from the beginning

ble and biased estimators (Mullainathan and Spiess 2017; Athey and Imbens 2019). Another hurdle in applying machine learning models to causal inference questions is the lack of standard errors on coefficients (Mullainathan and Spiess 2017). Recent papers by Chernozhukov et al. (2017) and Chernozhukov et al. (2018) introduce a set of methods that overcome these concerns using a combination of orthogonalization and data-splitting. I use these methods to supplement my baseline estimates. Finally, I predict dismissal risk using data from cases assigned to judges in non-election years to compare differences in the distribution of cases that are dismissed by judges under electoral pressure.

The remainder of this paper proceeds as follows: Section 2.1 explains the election process and the responsibilities of magistrate court judges. Section 2.2 describes the data used in this analysis. Chapter 3.1 presents the main results using a fixed effects model. Chapter 3.2 supplements these findings using machine learning tools. Chapter 3.3 describes initial exploration of changes in the suitability of dismissed cases across election cycles.

Chapter 2

Setting and Data

2.1 Magisterial District Courts

The Pennsylvania court system consists of five levels: Supreme court, Superior Court, Commonwealth Court, Court of Common Pleas and Minor Courts. The Minor Courts include 512 magisterial district courts, and the Philadelphia Municipal Court, which acts as the minor court in Philadelphia county. The minor courts are the entry point of the Pennsylvania judiciary. Judges in minor courts are responsible for setting bail, holding preliminary arraignments and preliminary hearings, and deciding whether criminal cases proceed to the higher courts.

2.1.1 Election Process

Judicial elections: Magistrate judges are elected into six-year terms through competitive partisan elections. These elections are held every two years. Judges and potential challengers file their candidacy and party affiliation in March for the primary election in May. Winners of each primary compete in the general election in November. Candidates can be affiliated with both the Democratic and Republican parties and run in both primaries. If they win both primaries, they are uncontested in general elections.

At the end of their term, magistrate court judges must be re-elected in partisan elections to keep their positions. All other judges in Pennsylvania, including those in Philadelphia's municipal court, are re-elected in non-competitive retention elections where constituents express approval (or disapproval) for a judge's continued tenure in a simple "Yes" or "No" question on the ballot. Judges who step down during their term are replaced temporarily by a governor appointed judge. However, the appointed judge must compete and win in the subsequent election to stay in their seat (The Unified Judicial System of Pennsylvania 2021a).

Eligibility: Magistrate judges are required to live in the district in which they serve. Unlike judges in other courts, they do not need to have formal legal training. Those who are not admitted to the Bar in Pennsylvania can complete a training program administered by the Minor Judiciary Education Board to be eligible for the position (The Unified Judicial System of Pennsylvania 2021b).

2.1.2 Criminal Procedure in Magisterial District Courts

There are three primary stages of criminal proceedings that occur in the magistrate court level:

- Arrest: A defendant is arrested and charged with a criminal offense.
- Preliminary Arraignment: The defendant is read their charges and is assigned bail. The preliminary arraignment should occur within 72 hours of arrest.
- Preliminary Hearing: The magistrate judge determines whether a charge is serious enough to be transferred to the Court of Common Pleas

This paper focus on events in the preliminary hearing, which is the last stage of the case process that is managed by magisterial judges.

At a preliminary hearing the prosecutor presents their argument to justify that "an offense has been committed and the defendant has committed it" (234 Pa. Code § 543 2021). Prosecutors may also withdraw charges at this stage. The magistrate judge evaluates whether there is sufficient evidence to link the defendant to each charge.

If so, they transfer the case from the local magistrate district court to the Court of Common Pleas where the formal arraignment and trial process takes place. If not, the judge can dismiss the some or all charges in the case. Misdemeanor charges can also be dismissed by the magistrate judge if they show that "public interest will not be adversely affected" and both the prosecutors and victims agree to the dismissal (234 Pa. Code § 546 2021).

There are a number of factors that make magisterial judges particularly susceptible to behavioral changes from election pressure. First, magistrate judges in Pennsylvania have discretion in dismissing charges. Unlike judicial decision making in other settings, such as sentencing, there are no mandatory minimum or maximum guidelines that constrain judge behavior. Second, magisterial districts are small. In many districts, the margins between winning and losing an election can be extremely narrow compared to those of state-wide elections. In 2015, there were 16 races where the difference between the incumbent magistrate judge and their nearest competitor was less than 100 votes. With these margins, small changes in the rates of favorable or unfavorable decisions could realistically change election outcomes. Third, magisterial judges are required to live in the district where they serve. Since districts are small, they are more easily able to assess the political leanings of their constituents. This means judges can respond more accurately to the preferences of their electorate.

2.2 Pennsylvania Court Records

I use administrative data from the Administrative Office of Pennsylvania Courts from 2008 – 2015. The data contains individual level records of criminal cases as they evolve throughout the court system. This includes details related to defendant demographic characteristics (eg. race, age, sex), and case characteristics (eg. list of charges, charge disposition, judge assigned).

2.2.1 Case data

Dismissed Charges: Every case record is associated with a set of offenses corresponding to each charge filed against the defendant. The offense data includes the specific statute linked to the offense and the final disposition of the charge. I define a dismissed case as a case where all charges are dismissed.

Offense Severity Classification: Each offense is associated with a classification that describes the severity of the crime. The least serious crime classification is a summary offense, which includes loitering or underage drinking. Misdemeanors are more serious than summary offenses and are classified into three levels: The least serious classification is a 3rd degree misdemeanor offense (eg. marijuana possession), followed by a 2nd degree offense (eg. shoplifting), and finally a 1st degree offense (eg. assault). Felony offenses are categorized in a similar fashion.

Offense Type Categorization: Offenses are manually classified into categories based on the statute associated with the charge: assault, theft, drugs, fraud or forgery, disorderly conduct, motor vehicle, robbery, sexual abuse, theft, weapons, property, and other. Over 95% of charges fall under one of the manually labeled offense categories.

Case-Level Data: I collapse my data to the case level since judges likely evaluate each charges in the context of other charge within a case. For each case, I retain the most serious charge by crime classification and binary indicator for each offense category denoting the presence of such a charge in the case. In records where the crime classification is missing or if the offense is simply classified as "Misdemeanor" or "Felony", I impute offense severity using the crime classification of all other charges associated with the same statute. Approximately 10% of offense severity classifications are imputed.

2.2.2 Elections Data

Election records are obtained from 2011, 2013 and 2015 magisterial elections data from Ballotpedia, which contains the names of candidates in each district race, a flag indicating whether the candidate was an incumbent during the election, and the final election outcome. Overall there are 394 elections with incumbent judges. Of these, 27% of races have contested primary elections and 5% have contested general elections.

The level of electoral pressure fluctuates throughout the election cycle, but there are two main events that are likely to induce the largest change in judge behavior. The first is prior to the filing date in March. In anticipation of increased public scrutiny of their judicial record, judges may adjust their behavior ahead of the filing deadline if they expect to compete against a primary election challenger. The second is before a primary election. Electoral pressure in this setting should be largest for sitting judges who in a contested primary election. The vast majority of judges win their primary elections, so I focus on proximity to the primary election as my main source of variation in election pressure rather than proximity to a general election.

2.2.3 Sample Selection

Since elections occur every two years, I limit my sample to cases where the offense disposition occurs after May 2009 (2 years before the 2011 primary election), and the three months after the 2015 primary election. Philadelphia holds Yes/No retention elections for their minor court judges, rather than partisan elections. I exclude all cases assigned to judges in Philadelphia since they do not face the same type of electoral pressures as magisterial judges in other counties. I also exclude cases where key case information is missing, such as age of the defendant or offense description. The final dataset contains 983,116 rows.

2.2.4 Descriptive Statistics

Table 2.1 presents descriptive statistics on the defendant demographics and case characteristics by election status of the assigned judge. The observable distribution of defendant characteristics appear to be quite similar across all samples. Black defendants make up about 25% of all defendants. About 75% of all defendants are men. The middle and bottom panels contain case characteristics. Since each case contains multiple charges, the values in each column do not sum to 1. About 80% of all cases contain at least one misdemeanor charge. Many of these cases also include lower level summary charges. Felony charges are the least common offense classification and make up slightly more than a quarter of all charges. The bottom panel shows the distribution of common case categories. Approximately one quarter to one fifth of all cases contain charges related to drugs, theft, assault or motor vehicle violations. These are typically low-level offenses, which is consistent with the high rates of misdemeanor and summary charges.

From Table 2.1 we conclude that cases assigned to judges are similar regardless of their incumbency status. Figures 2-1 and 2-2 show case characteristics assigned to judges by incumbency status as the election approaches. The x-axis is the number of quarters before a primary election, where "-1" denotes the 3 months immediately preceding a primary election and "+1" denotes the 3 months immediately following a primary election. Cases are generally stable across judges by incumbency status, suggesting that potential changes in decision making between judge groups is not attributed to changes in underlying cases.

Table 2.1: Descriptive Statistics

	Not In Election	Uncontested Incumbent	Contested Incumbent
Defendant Demographics			
Black	0.243	0.225	0.267
Male	0.743	0.742	0.753
Age	33.264	33.202	33.242
Charge Severity			
Includes Felony	0.271	0.260	0.268
Includes Misdemeanor	0.803	0.801	0.806
Includes Summary	0.443	0.457	0.460
Charge Category			
Drug	0.263	0.254	0.262
Motor Vehicle	0.199	0.208	0.209
Assault	0.210	0.208	0.216
Theft	0.248	0.253	0.235
Disorderly Conduct	0.152	0.145	0.150
Observations	791,339	138,657	53,120

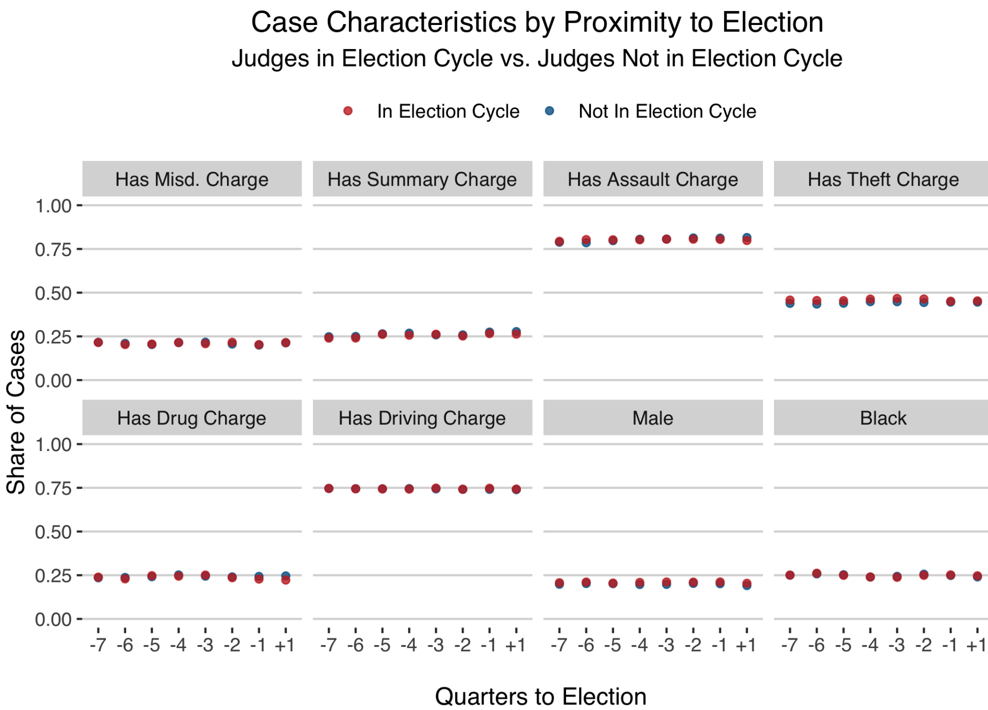

Figure 2-1: Cases Assigned to Judges by Incumbency Status

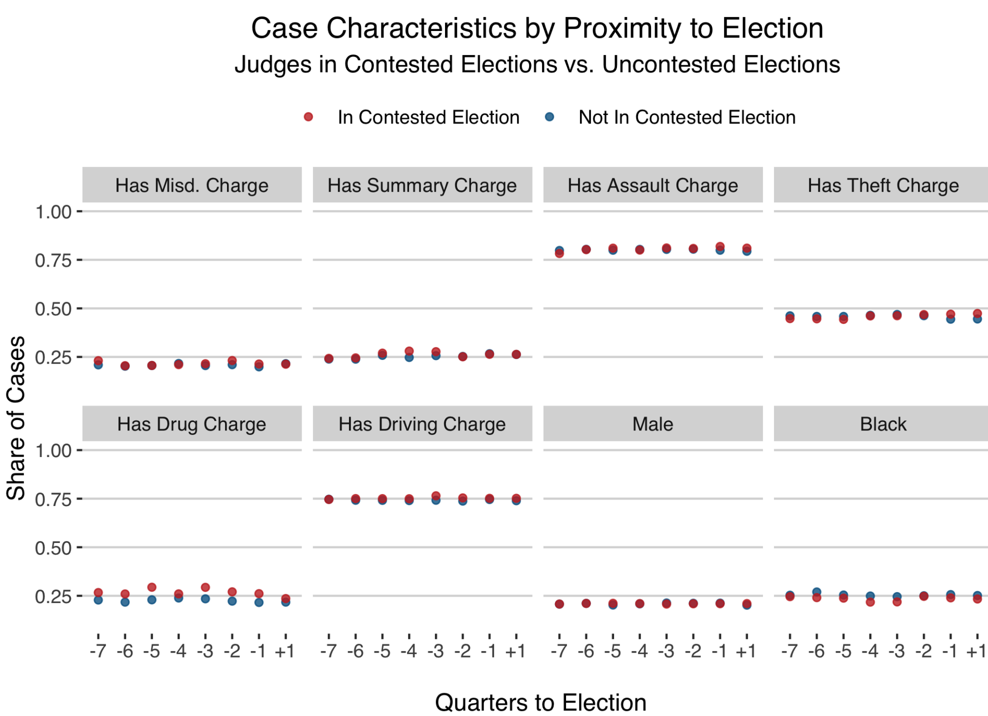

Figure 2-2: Cases Assigned to Judges by Contested v. Uncontested Election Status

Chapter 3

Methods and Results

3.1 Linear Model

This section describes the core results of the paper. I start by specifying a linear model to identify the effects of electoral pressure on case dismissal rates. For each case i presented to judge j at time t within q quarters of the judge's election, I estimate:

$$y_{ijt} = \sum_{q \neq -6} \beta_q I(q_{jt}) + \alpha X_i + \gamma_j + \gamma_t + \epsilon_{ij} \tag{3.1}$$

where y_{ijt} is case dismissals, the outcome of interest. $I(q_{jt})$ is the proximity to an election in three-month intervals where the judge j is an incumbent. Indicators – $I(q_{jt})$ – span from the 6th quarter before to the first quarter after the election. X_i is a vector of controls, including defendant age, sex, race, offense category and offense severity. γ_j contains judge fixed effects to account for underlying baseline variation in judge leniency; γ_t contains election quarter-year fixed effects to account for time trends and seasonality.[1]

The coefficients of interest (β_q) report the average change between time q and the reference period among incumbent judges relative to that same change over time among judges not up for election. Given that electoral pressure is likely to be strongest

[1]Since all elections occur on the same date across the state, I construct quarters t such that there are eight quarters between each election.

in the months leading up to an election, we expect to see a treatment effect in $q = 0$ and $q = -1$. Coefficients for $q < -1$, on the other hand, allow for visual inspection of the parallel trends assumption in the pre-period.

In a variant of the baseline regression, I interact the proximity fixed effects with an indicator for whether the judge is engaged in a contested primary race to isolate the effect of having an electoral competitor.

In order to identify the effect of election pressures, we assume that the change in types of cases presented to magistrate judges are unaffected by their proximity to an election. In specifications comparing contested and uncontested elections, I assume that criminal activity does not change in response to a competitive judicial election and that police officers are not changing their arresting behavior. While I cannot directly test this assumption, evidence from Figures 2-1 and 2-2 indicate that observable case characteristics are stable across treatment groups.

3.1.1 Results

Results for changes in dismissed cases are presented in Tables 3.1 and 3.2. Table 3.1 defines treatment variable as all judge who are up for re-election, including those who are running in uncontested elections. Column (1) shows changes in dismissal rates for all cases, while column (2) shows results for cases where the most serious charge is a misdemeanor. Each quarter represents the difference in likelihood of dismissing a case when it is assigned to a judge who is up for re-election relative to a judge who is not. If an election effect exists, we would expect to see the largest change in row 2, the coefficients representing the three months immediately preceding a primary election. We see modest reductions in case dismissal rates in the quarter before the primary election across cases overall. The effect appears to be more pronounced in cases without felony charges. Relative to a baseline dismissal rate of 6.9 percent, a misdemeanor case is 5.9% (0.41 percentage points) less likely to be dismissed when it is disposed in the quarter before a primary election when the judge is up for re-election.

Table 3.2 combines the election responses of judges in both contested and uncontested elections. For judges in uncontested races, there is no detectable effect in likelihood of dismissing a case. Intuitively, judges are most likely to change behaviour when their actions are scrutinized closely by their constituents, and when those constituents have leverage over the judge's career. Both factors are magnified in a contested election: the challenger is incentivized to highlight unfavorable or controversial decisions in the incumbent's track record, and voters have an alternative candidate to support on the ballot. Judges in uncontested elections face little to no threat of involuntarily losing their seat: if a challenger has not submitted their filing requirements by the deadline in March, they will have to run as a write-in candidate, virtually guaranteeing their election loss.

To explore the possibility that competitive elections induce larger responses, I show the changes in judge behavior for those in contested elections compared with those in uncontested elections. Results are reported in Table 3.2, where the values in row 2 are the main coefficients of interest. Each quarter represents the difference in likelihood of dismissing a case when it is assigned to a judge who in a contested election relative to a judge who is running in an uncontested election. This table only reports estimates for the interaction term between the proximity to an election and being in a competitive race. The full set of quarter-level estimates are reported in A.1. Column (1) shows the effect on cases of all offense severity classifications. As with Table 3.1, the estimated change is small and negative. These differences are not detectable even at the 0.01 level. However, columns (2) and (3) suggest that relative to those in uncontested elections, judges who face a primary challenger are less likely to dismiss a case when the most serious charge is a misdemeanor. Estimates from column (2) imply that judges in contested elections reduce case dismissal rates by 14% (1.1 percentage points relative to a baseline of 7.9).

The treatment group in columns (1) and (2) include both the judges who continue to be challenged in general elections as well as those who won both primary elections. Individuals who remain in competitive races after the primary may inflate our estimate in the post-primary election period as they continue to face election pressure while most judges do not. Alternatively, judges who lost their primary election may have been handicapped by their *lack* of response to voter pressure. This explanation

Table 3.1: Proximity to Elections on Dismissed Cases

	(1) All Charges	(2) Misdemeanor Charges
1 quarter after election	-0.0007	0.0010
	(0.0022)	(0.0024)
1 quarter before election	-0.0034^{+}	-0.0041^{+}
	(0.0018)	(0.0024)
2 quarters before election	0.0006	0.0017
	(0.0023)	(0.0026)
3 quarters before election	-0.0010	-0.0010
	(0.0026)	(0.0029)
4 quarters before election	-0.0033^{+}	-0.0019
	(0.0019)	(0.0028)
5 quarters before election	-0.0075*	-0.0052
	(0.0028)	(0.0032)
6 quarters before election	0.0002	0.0026
	(0.0024)	(0.0032)
7 quarters before election	-0.0023	-0.0029
	(0.0019)	(0.0022)
Race:Black	0.0046	0.0020
	(0.0028)	(0.0032)
Sex:Male	-0.0051***	-0.0046***
	(0.0010)	(0.0011)
Age	0.0003***	0.0003***
	(0.0001)	(0.0001)
Num. Charges	-0.0002***	-0.0009***
	(0.0000)	(0.0002)
N	983,116	691,110
r2	0.11	0.13

[a] Change in likelihood of dismissing a criminal case in the quarters preceding and following a primary election between judges who are in an election cycle compared with those who are not. Column (1) contains all offenses. Column (2) contains cases where the most serious offense is a misdemeanor. All specifications contain judge and month-year fixed effects as well as controls for race, sex, age, the severity and offense category of charges in the case. Standard errors are clustered at the county level.

would predict a suppression of estimated treatment effect in the pre-primary election period. Column (3) excludes judge who are in competitive general elections from the sample. The effect in the months before a primary election is larger in magnitude and more significant when I exclude incumbents who are in contested primary elections but not contested general elections as illustrated in column (3), suggesting that incumbents who do not respond to electoral pressure fare worse in electoral outcomes.

Estimates from column (3) show that being in a competitive primary election reduces the likelihood of dismissing a misdemeanor case by 16.4% (or 1.29 percentage points). Though both are significant at the 0.05 level, the estimated effects of contested elections vs. uncontested elections reported in Table 3.2 are more precise and larger in magnitude than the estimated effect between judges in contested elections and judges who are not up for elections. This supports the hypothesis that differences in judge behavior is driven by competition in elections rather than any auxiliary consequence of being in an election cycle.

This analysis relies on the assumption that there are no pre-trends between contested judges, uncontested judges and judges who are not in their election year. I plot the coefficients for the specification in (3) from Table 3.2 in Figure 3-1. The difference between judges who are in contested vs. uncontested elections are indistinguishable from zero prior to the six months preceding a primary election. Those who are in contested elections reduce their rates of charge dismissals starting from six months before the primary elections. The effect is largest in the quarter immediately preceding a primary race, and subsides after the the judge's primary victory.

Table 3.2: Proximity to Contested Elections on Dismissed Cases

	(1) All Offenses	(2) Misdemeanors	(3) Misdemeanors Uncontested General
1 quarter after election × contested	-0.0024	-0.0070	-0.0067
	(0.0048)	(0.0062)	(0.0067)
1 quarter before election × contested	-0.0070	-0.0110*	-0.0129*
	(0.0042)	(0.0054)	(0.0058)
2 quarters before election × contested	-0.0033	-0.0099+	-0.0109+
	(0.0049)	(0.0057)	(0.0062)
3 quarters before election × contested	0.0013	0.0027	0.0037
	(0.0057)	(0.0078)	(0.0089)
4 quarters before election × contested	0.0004	-0.0020	-0.0044
	(0.0049)	(0.0068)	(0.0072)
5 quarters before election × contested	0.0025	-0.0004	0.0035
	(0.0057)	(0.0070)	(0.0072)
6 quarters before election × contested	0.0003	0.0017	0.0043
	(0.0069)	(0.0102)	(0.0106)
N	983,116	691,110	676,489
r2	0.11	0.13	0.13

[a] Change in likelihood of dismissing a criminal case in the quarters preceding and following a primary election for judges who in contested re-election races relative to judges who are in uncontested re-election races. Column (1) contains all cases. Column (2) contains cases where the most serious offense is a misdemeanor offense. Column (3) contains cases where the most serious offense is a misdemeanor and the assigned judge has no general election challenger. All specifications contain judge and month-year fixed effects as well as controls for race, sex, age, the severity and offense category of charges in the case. Standard errors are clustered at the county level.

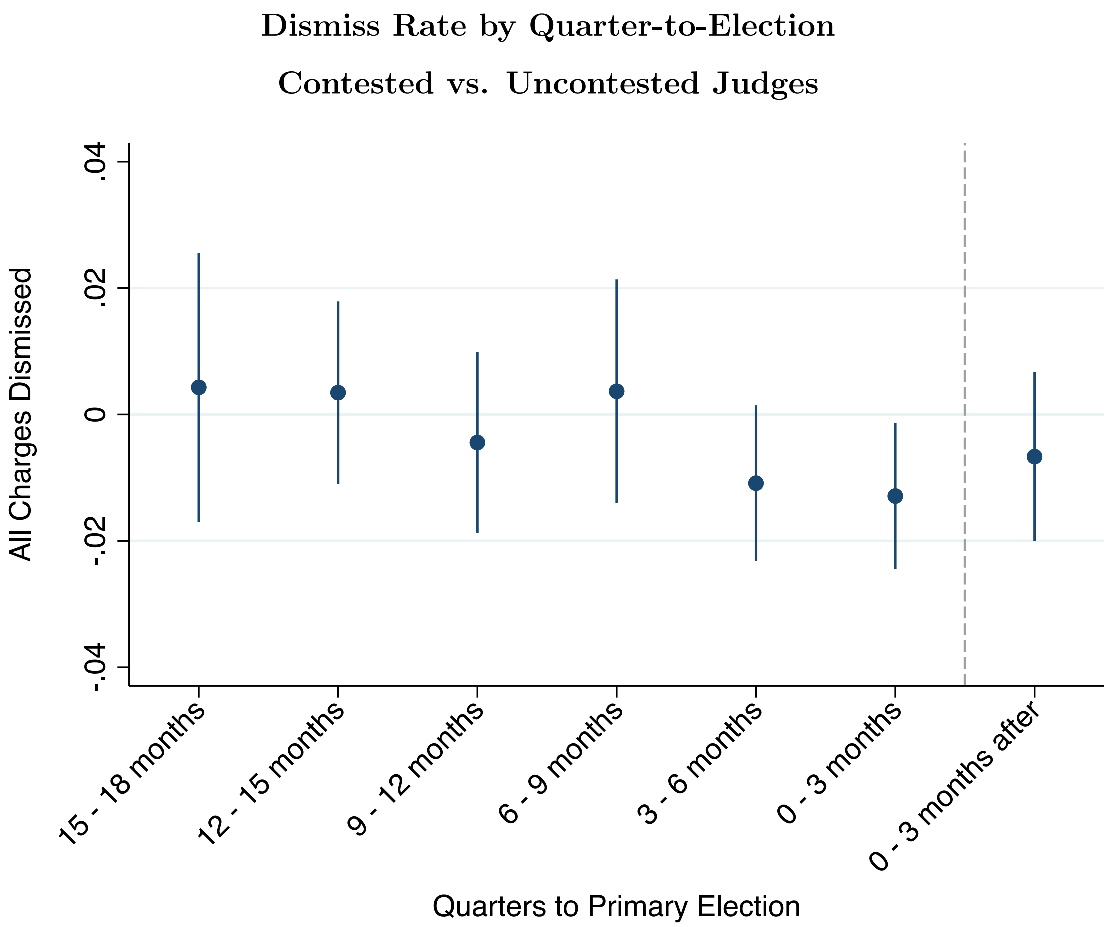

Figure 3-1: Change in case dismissals where the most serious charge is a misdemeanor.

We would expect to see the largest effect of election pressure when exploring the changes in likelihood that *all* charges are dismissed. The scenario where a judge transfers all charges and when a judge dismisses all but the least serious charge are both categorized the same way under the "all dismissed" classification. This measure does not provide insight into differences in behavior for judges who dismiss only a subset of charges. To explore this possibility further, I calculate two additional outcome measures. First, I find the share of all charges within a case that are dismissed. Second, I record whether or not the most serious charge in a case is dismissed. The first measure helps us detect finer changes in judge behavior by distinguishing between settings where most charges are dismissed and those where only a small share of charges are dismissed. The second measure differentiates between instances where the judge dismisses a low level charge, which may only prevented a minor fine, and the most serious charge in a case, which would likely bear the most severe punishment.

Results for share of charges dismissed and likelihood that the worst charge is dismissed are presented in Table 3.3 and Figures 3-2 and 3-3. This table only reports estimates for the interaction term between the proximity to an election and being in a competitive race. The full set of quarter-level estimates for judges in non-competitive races are reported in A.2. The difference in dismissal behavior between judges in contested elections and whose who are not remains constant until the six months before the primary election, when dismissal rates begin to fall. Between the filing date and the primary election, the likelihood dismissing the most serious charge in the cases decreases by share of charges dismissed within a case decreases by 12.9% (1.33 percentage points from a baseline of 10.3 percent) and the share of charges dismissed within a case decreases by 15.3% (1.52 percentage points from a baseline of 9.9 percent). Taken together, the evidence suggests that judges are less lenient in the months preceding an election when they are running in a competitive election.

Table 3.3: Proximity to Contested Elections on Dismissed Charges

	Worst Charge Dismissed		Percent of Charges Dismissed	
	(1)	(2)	(3)	(4)
	All Charges	Misd.	All Charges	Misd.
1 quarter after election × contested	-0.0049	-0.0124^{+}	-0.0042	-0.0099
	(0.0059)	(0.0065)	(0.0048)	(0.0062)
1 quarter before election × contested	-0.0082	-0.0133^{+}	-0.0089^{+}	-0.0152*
	(0.0051)	(0.0076)	(0.0049)	(0.0073)
2 quarters before election × contested	-0.0013	-0.0046	-0.0016	-0.0065
	(0.0065)	(0.0100)	(0.0059)	(0.0078)
3 quarters before election × contested	0.0005	0.0036	0.0001	0.0038
	(0.0074)	(0.0112)	(0.0064)	(0.0095)
4 quarters before election × contested	0.0038	0.0010	0.0033	0.0007
	(0.0068)	(0.0091)	(0.0061)	(0.0076)
5 quarters before election × contested	0.0055	0.0083	0.0051	0.0072
	(0.0093)	(0.0107)	(0.0078)	(0.0089)
6 quarters before election × contested	0.0013	0.0080	0.0013	0.0078
	(0.0076)	(0.0123)	(0.0071)	(0.0111)
N	983,116	676,489	983,116	676,489
r2	0.13	0.15	0.13	0.15

[a] Change in likelihood of dismissing a criminal case in the quarters preceding and following a primary election for judges who in contested re-election races relative to judges who are in uncontested re-election races. Columns (1) and (3) contain all cases. Columns (2) and (4) contain cases where the most serious offense is a misdemeanor offense. All specifications contain judge and month-year fixed effects as well as controls for race, sex, age, the severity and offense category of charges in the case. Standard errors are clustered at the county level.

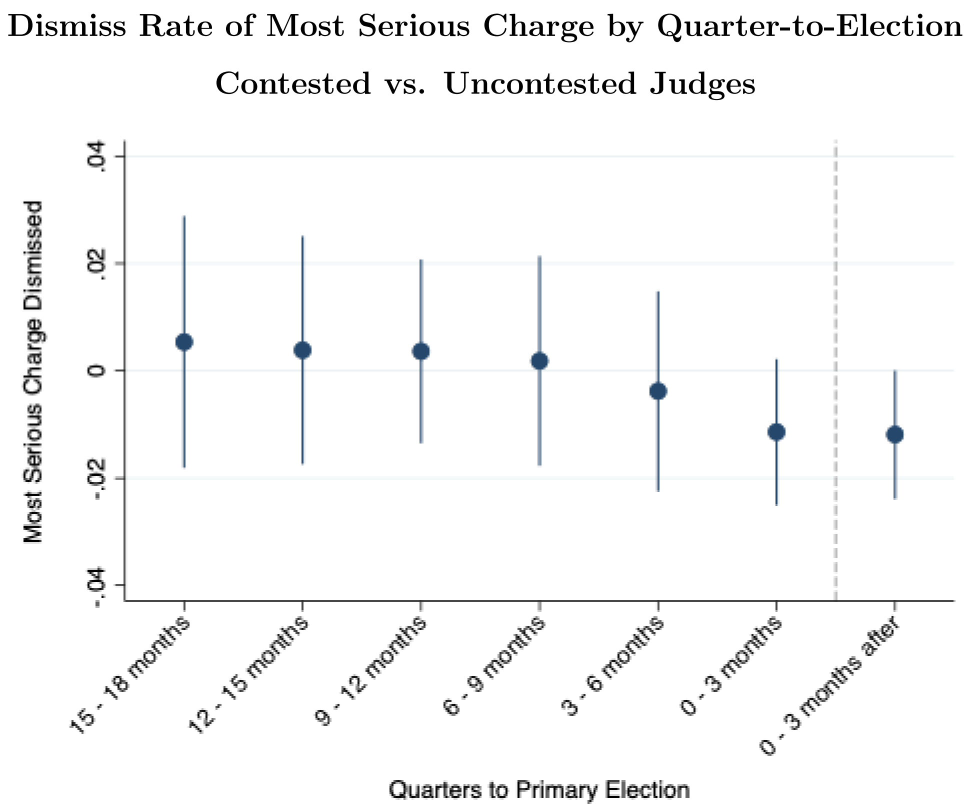

Figure 3-2: Change in Dismissals of the Most Serious Charge (Misdemeanor Cases Only)

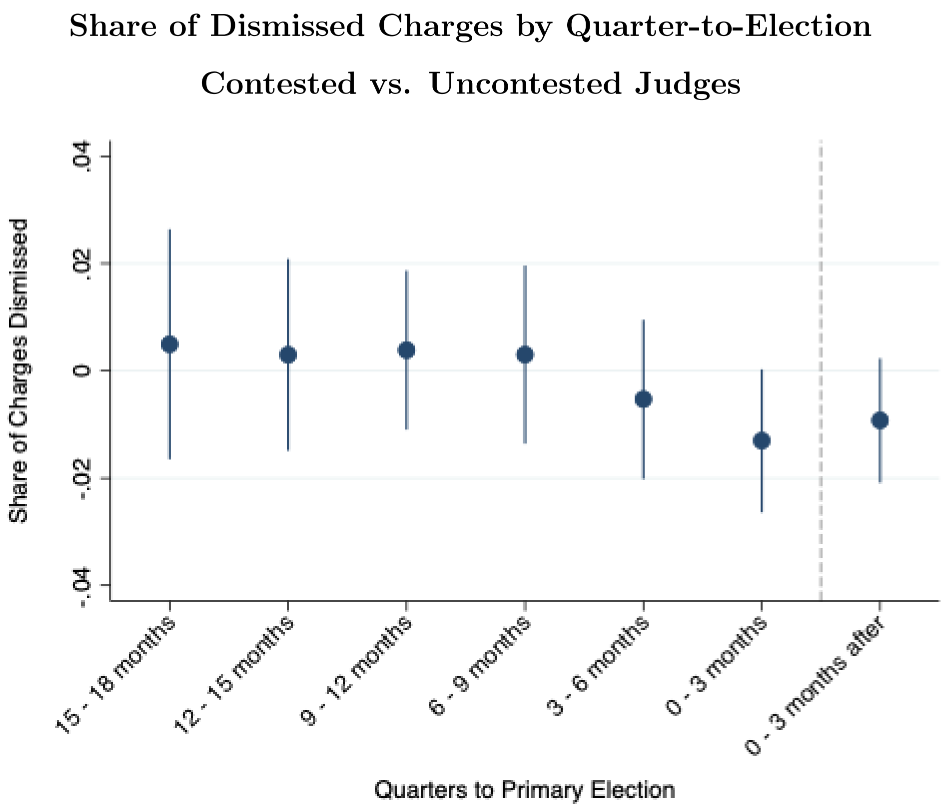

Figure 3-3: Change in the Share Dismissed Charges (Misdemeanor Cases Only)

3.2 Double/Debiased Machine Learning

The fixed effects linear regression model from section 3.1 imposes a linear relationship between each covariate and charge dismissal probability. However, it is possible that this form of linearity does not reflect decision-making processes in reality. Complexities in the underlying structure of how electoral pressure affects judicial decision are difficult to predict apriori. Testing many combinations of covariates in different formats can uncover spurious relationships in the data due to data mining. In this section, I use "Double Machine Learning" to identify the effect of election proximity in case dismissal rates.

Machine learning algorithms typically employ non-parametric models to answer questions about prediction. While the final prediction model may contain information about the contribution of different covariates in the final outcome, these "coefficients" do not have the consistent and unbiased properties that are critical estimating parameters for causal inference. The two main sources of bias introduced stem from regularization and over-fitting. "Double Machine Learning" is a method introduced in Chernozhukov et al. 2018 and Chernozhukov et al. 2017 that uses a machine learning framework to estimate causal effects while addressing both regularization and over-fitting bias. I first briefly describe the Double Machine Learning procedure and the concerns it addresses, then present results estimated using the method.

Regularization Bias: Regularization is a process of increasing the generalizability of a model by reducing its complexity. One of the risks of creating non-parametric models with high dimensional data is over-fitting the model to the noise in the existing data set, leading to poor predictive power. Without penalizing complexity in these settings, we will likely see large fluctuations in estimators depending on the dataset used to train the model, leading to low overall predictive power. Regularization generates simpler models that are highly effective for prediction, but also simultaneously decreases the variance to the estimator and introduces bias.

Chernozhukov et al. 2018 proposes a method to overcome regularization bias through orthogonalization. Consider a linear model: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u$. The coefficient β_1 can either be estimated using a multivariate linear regression, or by regressing Y on X_2, regressing X_1 on X_2 and regressing the residuals from the former regression on the residuals from the latter regression. Double Machine Learning uses a similar procedure that leverages the predictive capabilities from machine learning algorithms. Consider an analogous set up, where we want to find the relationship between independent variable D, outcome variable Y with covariates Z:

$$\begin{aligned} Y &= D\theta_0 + g_0(X) + U, \quad && E[U|X,D] = 0 \\ D &= m_0(X) + V, && E[V|X] = 0 \end{aligned}$$

The parameter of interest is θ_0, which is the treatment effect of D. Function m_0 describes the relationship between X and D; Function g_0 describes the relationship between X and Y given D. Note that this is similar to the OLS setup when m_0 and g_0 are linear. The Double Machine Learning procedure recovers θ_0 by: (1) using machine learning to predict Y using X, (2) using machine learning to predict D from X, and finally regressing the residuals from (1) on the residuals from (2). The procedure to generalize this example to eliminate the linear dependency between Y and D is described in more detail in Chernozhukov et al. 2018.

Over-fitting: Double Machine Learning manages biases that arise from over-fitting using cross-fitting. The sample is split randomly into two partitions. Predictions for Y using X and D using X are obtained from one partition. Then θ_0 is estimated using the second sample. The process is repeated using the second sample to generate predictions for Y and D, and the first sample to estimate θ_0. The average of θ_0 values in both stages is used as the final estimator. Generalizing this procedure to K-fold version is described in detail in Chernozhukov et al. 2018.

3.2.1 Results

Using the Double Machine Learning procedure, I estimate the impact of assignment to a judge in a contested primary election in the six months prior to the election date on the likelihood of dismissing a case. The point estimates for Double Machine Learning results using lasso regression, decision trees and random forest algorithms are presented in Figure 3-4.

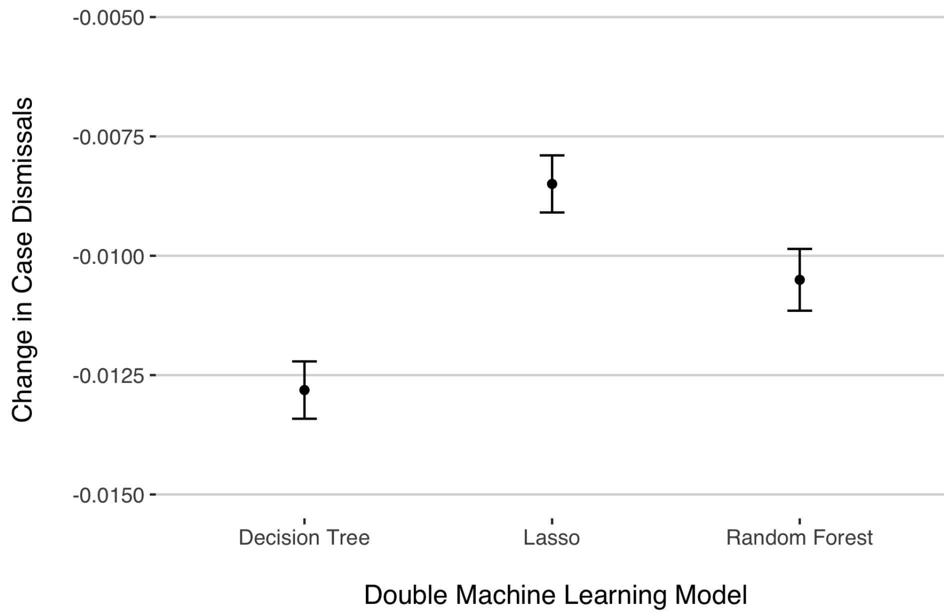

Figure 3-4: Estimated Effect of Election Pressure on Case Dismissal Rates using Double Machine Learning

The point estimates from each Double Machine Learning specification are all very similar to the treatment effect of the linear fixed effects model (-0.0129), but are much more precisely estimated. One possible explanation is that the fixed effects model is picking up noise from the sample when partialling out the effect of covariates on the outcome and treatment variables. Regularization in the predictive machine learning models may reduce overfitting on excess randomness, thereby decreasing the standard error on the final treatment estimates.

3.3 Estimating the Marginal Dismissed Case Using Machine Learning Predictions

In this section, I explore one method to identify the marginal dismissed case by leveraging the exogenous decrease in case dismissals prior to an election period. Since the estimated effect of elections is small and this exercise requires strong assumptions about the nature of judge decision making, I treat this portion of the paper as a supplementary exploratory analysis.

Overview: I start by assuming that the collective decision making of magistrate judges in non-election years reflects the true "suitability" that a case should be dismissed. Under this assumption, I create a measure of dismissal suitability by predicting the probability that a case will be dismissed.

The suitability of the marginally dismissed case can be estimated by comparing changes in the average suitability scores of dismissed cases, and changes in the number of cases dismissed. The change in average suitability score from a reduction in case dismissals is the difference between the suitability of the marginal case and the average case scaled by the number of dismissed cases. For small changes, this approximates the implicit threshold that judges employ when deciding which cases to dismiss. This procedure assumes that case dismissal decisions operate around a cutoff of the "true" suitability score. It also makes the strong assumption that judges dismiss cases monotonically – in other words, if all cases were ranked by the true suitability of dismissal, judges will dismiss cases in descending order of the ranked cases. If this assumption is violated, the change in suitability no longer reflects the characteristics of the case at the margin.

Since I predict suitability scores, I should be able to observe the relationship between suitability scores and case dismissals directly in the predicted data. I pursue this strategy to compare implied threshold estimates with the observed distribution of predicted scores.

Predicting Case Dismissals: I create a suitability score using judge dismissal behavior during non-election years as a baseline. I train a gradient boosted trees model on judge dismissal behavior using 80% of the sample from cases disposed during non-election years. Since the number of dismissed cases is relatively low, I use the area under the precision-recall curve (AUPRC) as my optimization metric. This allows the algorithm to select a model based on the trade-off between high precision (the rate of correctly predicted positive classes of all positive predictions), and high recall (the rate of correctly positive predictions of all the actual positive values). This is particularly useful for classifying imbalanced data. A model trained using a more common metric like root mean square error may return a results with a low RMSE score simply by labeling every point as 0 in an unbalanced data set.

It is important to note that the suitability score internalizes any biases that judges exhibit in case dispositions since the scores are trained solely on case disposition data. There is a large body of literature studying methods to detect and rectify such biases, but investigating those methods is outside the scope of this project.

Using the trained model, I predict the suitability of dismissal on the test data. I convert the raw predicted probability into percentile ranks to increase spread of the final score. Figure 3-5 shows the distribution of predicted suitability for case dismissals in the test set. While there is substantial overlap between the two distributions, the scores for cases that are truly dismissed skews heavily to the right. This indicates that the suitability score contains some predictive power of actual case outcomes.

Figure 3-5: Predicted Probability of Case Dismissals

We proceed by finding the change in suitability of cases that are dismissed due to proximity to an election. I instrument for change in dismissals with an indicator for being assigned a judge in the six months before a competitive election and run a regression using controls for case characteristics and judge fixed effects. If the assumptions hold, a decrease in dismissal rates should mean that cases that continue to be dismissed are especially suitable. We would expect the average score to increase.

Unfortunately, the estimated coefficients are noisy and directionally negative (-0.0008115 with standard error of 0.0029940), suggesting that our monotonicity assumption does not hold, or that the predicted score is noisy. Indeed in the ideal setting, the score distribution in Figure 3-5 should show a threshold delineating cases that are dismissed and those that are not. Instead, we observe two overlapping distributions.

Instead, I plot the distribution of predicted suitability scores by case disposition and by election competition status in Figure 3-6. Although we establish that judges reduce the number of cases dismissed during their election cycle, the distribution of predicted scores for dismissed cases does not appear to be more concentrated at the upper tail as we would expect. This could imply that judges do not reduce dismissals in a linear fashion starting from the least suitable to the most suitable. Alternatively, the model may be more precise in predicting dismissed cases for judges during non-election years. Excess noise in predicted suitability scores can shift the distribution for contested judges away from the tail. Future work can improve on these prediction methods to more precisely identify group differences in marginal cases.

Predicted Dismissal Suitability Score by True Case Disposition and Election Status

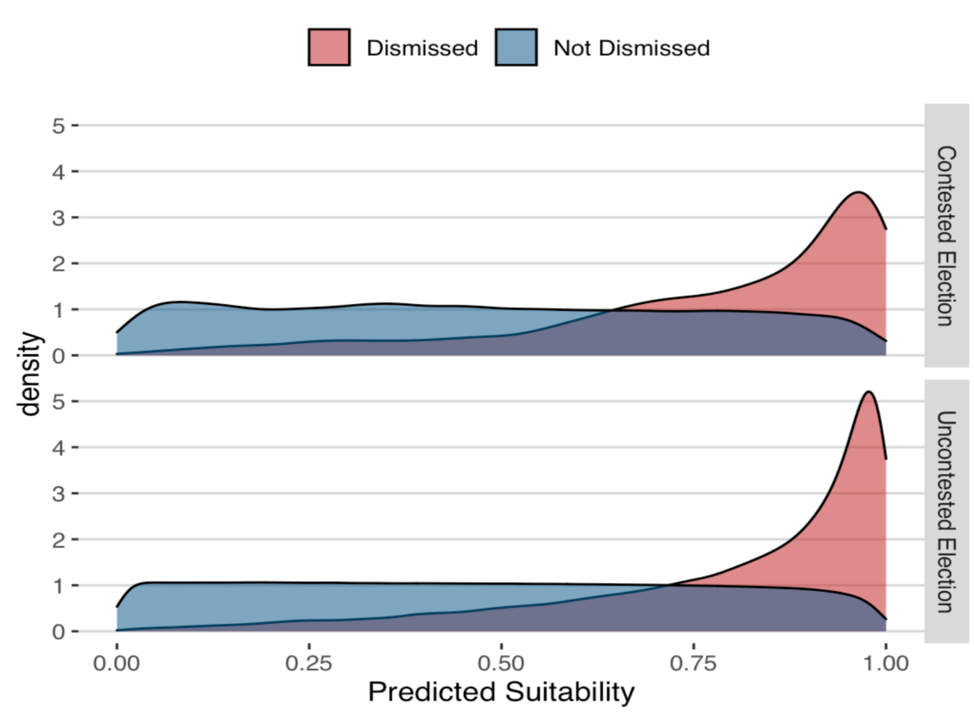

Figure 3-6: Predicted Probability of Case Dismissals by Election Status

Chapter 4

Conclusion

This paper presents evidence for electoral patterns in judicial decision making: Magistrate court judges who are running in contested elections in Pennsylvania dismiss fewer charges as their election approaches. Cases involving misdemeanors and other low-level offenses are particularly affected. These effects are estimated using a variety of methods in Computer Science and Economics with similar final results.

Prompt dismissals of appropriate criminal charges are important in preventing undue burden on innocent defendants. Future work can focus on exploring downstream consequences of the change in judge behavior. In particular, it would be valuable to investigate whether the decline in dismissals by magistrate judges lead to increased conviction rates, or merely a delay in case dismissals. This paper also does not evaluate whether the estimated reduction in dismissals is warranted. Further research determining a link (or lack thereof) between reduced dismissal rates and averted crime can inform assessments of the benfits and risks of judge elections.

Bibliography

234 Pa. Code § 543. Commonwealth of Pennsylvania. URL: https://www.pacodeandbulletin.gov/Display/pacode?file=/secure/pacode/data/234/chapter5/s543.html&d=reduce (visited on 08/10/2021).

234 Pa. Code § 546. Commonwealth of Pennsylvania. URL: https://www.pacodeandbulletin.gov/Display/pacode?file=/secure/pacode/data/234/chapter5/s546.html&d=reduce (visited on 08/10/2021).

Abrams, David et al. "Electoral sentencing cycles". In: *CEPR Discussion Paper No. DP14049* (2019).

Athey, Susan and Guido Imbens. *Machine Learning Methods Economists Should Know About*. 2019. arXiv: 1903.10075 [econ.EM].

Bach, P. et al. *DoubleML – An Object-Oriented Implementation of Double Machine Learning in R*. arXiv:https://arxiv.org/abs/2103.096032103.09603 [stat.ML]. 2021. arXiv: 2103.09603 [stat.ML].

Berdejó, Carlos and Noam Yuchtman. "Crime, punishment, and politics: an analysis of political cycles in criminal sentencing". In: *Review of Economics and Statistics* 95.3 (2013), pp. 741–756.

Brennan Center for Justice. URL: https://www.brennancenter.org/our-work/research-reports/judicial-selection-significant-figures (visited on 08/10/2021).

Chernozhukov, Victor et al. *Double/debiased machine learning for treatment and causal parameters*. Tech. rep. 2017.

Chernozhukov, Victor et al. *Double/debiased machine learning for treatment and structural parameters*. 2018.

Dippel, Christian and Michael Poyker. *How Common are Electoral Cycles in Criminal Sentencing?* Working Paper 25716. National Bureau of Economic Research, Mar. 2019. DOI: 10.3386/w25716. URL: http://www.nber.org/papers/w25716.

Dobbie, Will, Jacob Goldin, and Crystal S Yang. "The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges". In: *American Economic Review* 108.2 (2018), pp. 201–40.

Gordon, Sanford C and Gregory Huber. "The effect of electoral competitiveness on incumbent behavior". In: *Quarterly Journal of Political Science* 2.2 (2007), pp. 107–138.

Harvey, Anna and Sidak Yntiso. "Racial Disparities in Within-Judge Retention Effects". In: (2020).

Heaton, Paul, Sandra Mayson, and Megan Stevenson. "The downstream consequences of misdemeanor pretrial detention". In: *Stan. L. Rev.* 69 (2017), p. 711.

Lim, Claire SH, James M Snyder Jr, and David Strömberg. "The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems". In: *American Economic Journal: Applied Economics* 7.4 (2015), pp. 103–35.

Mullainathan, Sendhil and Jann Spiess. "Machine learning: an applied econometric approach". In: *Journal of Economic Perspectives* 31.2 (2017), pp. 87–106.

Park, Kyung H. "The impact of judicial elections in the sentencing of black crime". In: *Journal of Human Resources* 52.4 (2017), pp. 998–1031.

Pennsylvania judicial elections. Ballotpedia. URL: https://ballotpedia.org/Pennsylvania_judicial_elections (visited on 08/10/2021).

The Unified Judicial System of Pennsylvania. *How Judges are Elected.* URL: https://www.pacourts.us/learn/how-judges-are-elected (visited on 08/10/2021).

— *Judicial Qualifications.* URL: https://www.pacourts.us/courts/supreme-court/committees/supreme-court-boards/minor-judiciary-education-board/judicial-qualifications (visited on 08/10/2021).

Appendix A

Tables

A.1

Table A.1 reports the full interaction terms estimating the change in likelihood of dismissing a criminal case in the quarters preceding and following a primary election. Rows with labels containing an interaction term report estimates for the change in disposition behavior among judges who are in competitive primary elections relative to judges in non-competitive primary elections. Rows without the interaction term contain estimates for the change in disposition behavior among judges who are in non-competitive primary elections relative to judges who are not up for election at all. While there are some significant estimates among judges in uncontested elections in the 5-7 quarters before an election, the coefficients stabilize in the year before an election, suggesting that pre-trends are not a concern in this setting.

Column (1) contains all cases. Column (2) contains cases where the most serious offense is a misdemeanor offense. Column (3) contains cases where the most serious offense is a misdemeanor and the assigned judge has no general election challenger. All specifications contain judge and month-year fixed effects as well as controls for race, sex, age, the severity and offense category of charges in the case. Standard errors are clustered at the county level.

Table A.1: Proximity to Contested Elections on Dismissed Cases (Full Interaction Terms)

	(1) All Offenses	(2) Misdemeanors	(3) Misdemeanors Uncontested General
1 quarter after election × contested	-0.0024	-0.0070	-0.0067
	(0.0048)	(0.0062)	(0.0067)
1 quarter before election × contested	-0.0070	-0.0110*	-0.0129*
	(0.0042)	(0.0054)	(0.0058)
2 quarters before election × contested	-0.0033	-0.0099+	-0.0109+
	(0.0049)	(0.0057)	(0.0062)
3 quarters before election × contested	0.0013	0.0027	0.0037
	(0.0057)	(0.0078)	(0.0089)
4 quarters before election × contested	0.0004	-0.0020	-0.0044
	(0.0049)	(0.0068)	(0.0072)
5 quarters before election × contested	0.0025	-0.0004	0.0035
	(0.0057)	(0.0070)	(0.0072)
6 quarters before election × contested	0.0003	0.0017	0.0043
	(0.0069)	(0.0102)	(0.0106)
1 quarter after election	-0.0012	0.0012	0.0019
	(0.0022)	(0.0025)	(0.0026)
1 quarter before election	-0.0027	-0.0027	-0.0024
	(0.0022)	(0.0031)	(0.0031)
2 quarters before election	0.0004	0.0028	0.0026
	(0.0029)	(0.0034)	(0.0034)
3 quarters before election	-0.0026	-0.0034	-0.0034
	(0.0032)	(0.0039)	(0.0039)
4 quarters before election	-0.0046*	-0.0030	-0.0032
	(0.0021)	(0.0030)	(0.0030)
5 quarters before election	-0.0094**	-0.0068+	-0.0070+
	(0.0032)	(0.0038)	(0.0038)
6 quarters before election	-0.0011	0.0004	-0.0000
	(0.0031)	(0.0043)	(0.0042)
7 quarters before election	-0.0035+	-0.0047+	-0.0047+
	(0.0020)	(0.0027)	(0.0027)
N	983,116	691,110	676,489
r2	0.11	0.13	0.13

A.2

Table A.2 reports the full interaction terms estimating the change in share of charges dismissed and likelihood that the worst charge in a case is dismissed. Rows with labels containing an interaction term report estimates for the change in disposition behavior among judges who are in competitive primary elections relative to judges in non-competitive primary elections. Rows without the interaction term contain estimates for the change in disposition behavior among judges who are in non-competitive primary elections relative to judges who are not up for election at all. While there are some significant estimates among judges in uncontested elections in the 4th and 5th quarters before an election, the coefficients stabilize in the year before an election, suggesting that pre-trends are not a concern in this setting.

Column (1) contains all cases. Column (2) contains cases where the most serious offense is a misdemeanor offense. Column (3) contains cases where the most serious offense is a misdemeanor and the assigned judge has no general election challenger. All specifications contain judge and month-year fixed effects as well as controls for race, sex, age, the severity and offense category of charges in the case. Standard errors are clustered at the county level.

Table A.2: Proximity to Contested Elections on Dismissed Charges (Full Interaction Terms)

	Worst Charge Dismissed		Percent of Charges Dismissed	
	(1)	(2)	(3)	(4)
	All Charges	Misd.	All Charges	Misd.
1 quarter after election × contested	-0.0049	-0.0119^{+}	-0.0042	-0.0093
	(0.0059)	(0.0060)	(0.0048)	(0.0058)
1 quarter before election × contested	-0.0082	-0.0115^{+}	-0.0089^{+}	-0.0131^{+}
	(0.0051)	(0.0068)	(0.0049)	(0.0067)
2 quarters before election × contested	-0.0013	-0.0038	-0.0016	-0.0053
	(0.0065)	(0.0093)	(0.0059)	(0.0074)
3 quarters before election × contested	0.0005	0.0018	0.0001	0.0030
	(0.0074)	(0.0098)	(0.0064)	(0.0083)
4 quarters before election × contested	0.0038	0.0036	0.0033	0.0038
	(0.0068)	(0.0086)	(0.0061)	(0.0074)
5 quarters before election × contested	0.0055	0.0038	0.0051	0.0030
	(0.0093)	(0.0107)	(0.0078)	(0.0090)
6 quarters before election × contested	0.0013	0.0053	0.0013	0.0049
	(0.0076)	(0.0118)	(0.0071)	(0.0108)
1 quarter after election	0.0012	0.0037	0.0005	0.0028
	(0.0031)	(0.0033)	(0.0027)	(0.0030)
1 quarter before election	-0.0014	-0.0014	-0.0009	-0.0002
	(0.0026)	(0.0036)	(0.0025)	(0.0035)
2 quarters before election	0.0028	0.0046	0.0017	0.0035
	(0.0027)	(0.0038)	(0.0028)	(0.0035)
3 quarters before election	-0.0022	-0.0001	-0.0022	-0.0026
	(0.0036)	(0.0039)	(0.0034)	(0.0037)
4 quarters before election	-0.0058*	-0.0029	-0.0055*	-0.0047
	(0.0023)	(0.0039)	(0.0023)	(0.0037)
5 quarters before election	-0.0098**	-0.0084*	-0.0106**	-0.0085*
	(0.0031)	(0.0037)	(0.0032)	(0.0038)
6 quarters before election	-0.0010	-0.0006	-0.0017	-0.0018
	(0.0037)	(0.0048)	(0.0033)	(0.0046)
7 quarters before election	-0.0015	-0.0028	-0.0024	-0.0035
	(0.0026)	(0.0040)	(0.0021)	(0.0031)
N	983,116	691,110	983,116	691,110
r2	0.13	0.15	0.13	0.15