

A Silent Corrupting Force?

Criminal Sentencing and the Threat of Recall

Sidak Yntiso*

Sanford C. Gordon†

Abstract

39 U.S. states authorize recall elections, but the incentives they create are understudied. We examine how changes in the perceived threat of recall alter behavior of one set of officials: judges. In 2016, outrage over the sentence imposed on a Stanford athlete following his sexual assault conviction sparked a drive to recall the presiding judge. Using case disposition data from six California counties and matched arrest records for a subset of defendants, we examine whether critical events in the recall campaign were accompanied by corresponding changes in other judges' sentences. We find a large, discontinuous increase in severity associated with the campaign's announcement, but not the recall itself – suggesting the announcement shifted judges' beliefs about their political environment. The increase may have indirectly burdened minority defendants disproportionately. Our findings are the first to document incentive effects of recall, and suggest that targeted political campaigns may have far-reaching, unintended consequences.

Word Count: 11,121

*Department of Politics, New York University, New York, New York, USA. Email: sidak.yntiso@nyu.edu

†Department of Politics, New York University, New York, New York, USA. Email: sanford.gordon@nyu.edu

1 Introduction

Electoral accountability requires that voters have information to evaluate incumbent performance, but voters may be only selectively attentive, and the availability of information may itself be contingent on incumbent choices. Either of these considerations may distort the incentives of elected officials, and so assessing the extent to which electoral institutions mitigate or exacerbate such distortions is a critical task for empirical political science.

One electoral institution that enjoys widespread use in the United States is the recall election, in which voters may remove an incumbent from office before the expiration of his or her ordinary term. 39 states authorize recall elections for at least some offices. In the 19 states that allow for gubernatorial recall, five governors have or are currently facing recall threats as of writing.¹ And of the 45 state-level recall elections in U.S. history, nearly half have occurred in the past ten years.²

Despite the availability and increasing prominence of the recall option, there exists, to our knowledge, no study systematically assessing how the threat of recall affects incumbent behavior. This lacuna may stem in part from a host of methodological challenges. Most obviously, recall campaigns are not randomly assigned, and so comparing officials who do and do not experience them is likely to suffer from substantial omitted variables bias. Relatedly, if public officials rationally anticipate the consequences of a recall threat, they may take pains to avoid it. And finally, the behavior of officials in jurisdictions with and without the recall institution may differ in innumerable ways besides the availability of that specific institution, and the durability of the institution makes within-jurisdiction comparisons infeasible.

¹State constitutions and/or statutes permit gubernatorial recall elections in AK, AZ, CA, CO, GA, ID, IL, KS, LA, MI, MN, MO, NV, NJ, ND, OR, RI, WA and WI. Recall elections against governors in CA, CO, OR, NJ and NV have recently been attempted. Former WI governor Scott Walker faced an unsuccessful recall election in 2012.

²Source: <https://www.ncsl.org/research/elections-and-campaigns/recall-of-state-officials.aspx>.

To circumvent these difficulties, we examine the effect of a potential shock to the salience of the recall threat brought by a widely publicized and eventually successful recall campaign. In June 2016, Santa Clara Superior Court Judge Aaron Persky achieved notoriety for imposing an unusually lenient sentence on Brock Turner, an affluent, white Stanford student athlete convicted of two counts of sexual assault and one count of attempted rape. Two years later, 61.5% of voters in Santa Clara County, CA took the highly unusual step of voting to recall Persky.

Several weeks before the 2018 vote, Persky delivered a speech that included the following warning:

We promise as judges to rule on the facts and on the law, not on public opinion... When public opinion influences a juror's decision or a judge's decision, it corrupts the rule of law. This recall, if successful, will make it harder for judges to keep that promise ... The judicial recall, if successful, will be a silent force, a silent corrupting force. A force that will enter the minds of judges as they contemplate difficult decisions.

A host of elected officials, political activists, and legal academics echoed Persky's warning about the incentive effects of the recall effort. These observers argued that it might push judges to become more punitive in their sentencing decisions, even while condemning the leniency of the specific sentence that instigated the campaign.³ Several of these observers argued that the burden of any change in electoral incentives would be borne disproportionately by minority defendants rather than white ones – a cruelly ironic prediction given that the behavior the recall aimed to sanction was leniency toward a white defendant.

Using data on over 20,000 sentences handed down by over 158 Superior Court Judges

³In a similar spirit, we wish to strongly caution against interpreting any of the findings presented in this paper as speaking to the merits of the specific sentence that motivated Judge Persky's recall.

in six California counties from 2015 to 2018, we examine whether critical events in the recall campaign were accompanied by corresponding changes in other judges’ sentences. Specifically, using a regression discontinuity in time (RDit) approach (Hausman and Rapson, 2018), we examine the effects of two specific events: the initial announcement of the recall petition; and the recall election itself.

Our main results point to an instantaneous increase in average sentence length of over 30% in the immediate aftermath of the recall petition announcement. This result is robust to the inclusion of judge- and charge-level fixed effects, and a battery of placebo and specification tests. This effect is driven by increases in sentencing on non-sexual violent crimes. In contrast to the results for the petition announcement, we find no evidence that the recall election itself induced changes in sentencing. The conjunction of these findings suggests that the announcement of a well-organized, well-funded recall campaign against a Superior Court Judge signaled a new political reality for judges that was “priced in” by judges by the time the election took place. Next, we consider whether the effects of the observed shift were borne disproportionately by minority defendants. Drawing on recent research decomposing the sources of racial disparities in sentencing (e.g., Rehavi and Starr, 2014), we describe two channels through which disproportionate burdens might manifest themselves: a direct channel in which the post-announcement increase in sentencing was larger for minority defendants than for white ones; and an indirect channel in which, prior disparities in sentencing were amplified by the overall effect of the announcement. Contrary to expectations, we find no definitive evidence for the direct channel. However supplemental tests provide descriptive evidence that minority defendants in California tend to be charged with crimes bearing longer maximum confinement terms and, as a consequence, may have indirectly borne the brunt of any recall-induced increase in judicial punitiveness.

In the last part of the paper, we estimate the aggregate effect of the change in judicial precipitated by the petition announcement over a narrow (45 day) time frame. Our most

conservative estimates suggest that the petition announcement led to approximately 150 years of additional prison time in the six counties for which we have data.

In the most immediate sense, our findings corroborate the concern that the campaign to remove a sitting judge would affect the behavior of other judges, and amplify preexisting disparities in the criminal justice system. More generally, they contribute to our understanding of the role recall elections may play in contemporary political life. Arguments in favor of recalling an elected official invariably focus on a selection function: the recall gives voters the opportunity to remove a specific malfeasant public official. The findings presented here suggest broader incentive effects that may extend beyond the official in question, and that may operate counter to the objectives of the recall’s proponents.

2 Background

2.1 Institutional Setting

Recall elections in California and elsewhere. 39 states allow recall elections – those in which voters have the opportunity to remove a public official prior to the expiration of his or her term – in some form. Considerable variation exists, however, with respect to the particulars: whether state or local officials are eligible; the eligibility of appointed officials; signature requirements; permissible grounds for recall; and procedures for filling vacancies from successful recalls. Of the 39 states that permit recall, eight (AZ, CA, CO, MN, NV, ND, OR, and WI) specifically permit recall of judges.

California, the setting of the empirical analysis that follows, adopted recall elections by a constitutional amendment in 1911. Since the amendment went into effect in 1913, there have been 165 attempts to recall statewide officials, of which ten qualified for the ballot, and six were successful – the most well-known being the recall of Governor Gray Davis in 2003.⁴

⁴Source: Complete List of Recall Attempts, California Secretary of State. Available at

Far more ubiquitous in the state are recall efforts for local officials. Elected state legislators have been removed by voters in safe as well as very competitive districts.⁵

Since 1995, the earliest year for which we have systematic data, recall attempts for 333 local officials have qualified for the ballot (reflecting a fraction of the full set of recall attempts); of these 244 have been successful.⁶ Figure 1 displays qualified attempts and successes over time; as is evident from the figure, every year voters recall some local officials in California. The anti-Persky campaign, discussed in greater detail below, represents the only attempt to recall a superior court judge to qualify for the ballot during this time period.⁷

Judges and judicial discretion in California. California has the largest judicial system in the nation, with 1,743 authorized superior court judges sitting in 58 county courts. During 2016–2017, approximately 6 million cases were filed in these courts. Superior courts in California have jurisdiction over civil and criminal cases. These courts also hear appellate cases arising from certain civil cases worth under \$25,000 as well as some misdemeanor cases.⁸ Since 1998, superior courts are the only consolidated general jurisdiction trial courts. Superior court judges run in non-partisan competitive elections for six-year terms. In the event of a vacancy, judges are appointed by the Governor.

Judicial discretion over sentencing in California is constrained by a complex array of <https://www.sos.ca.gov/elections/recalls/complete-list-recall-attempts/>.

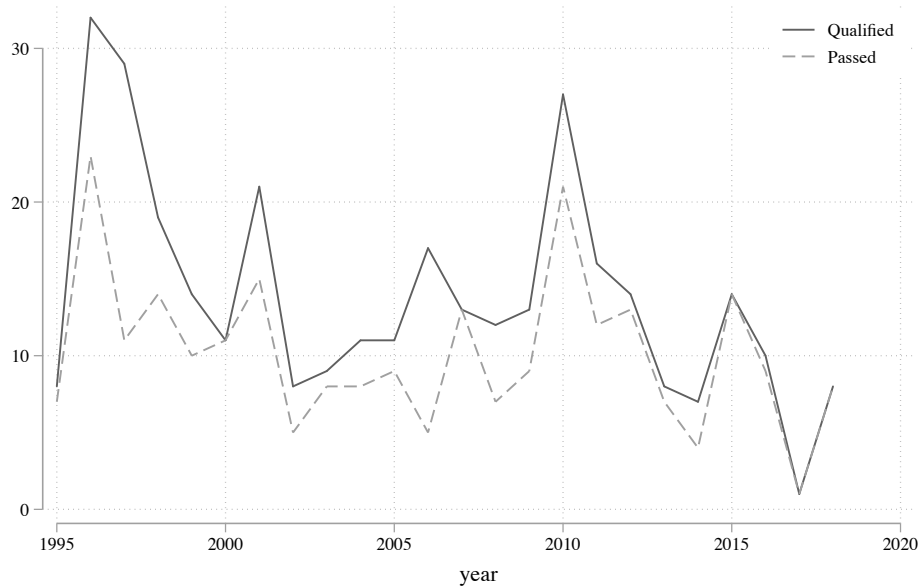
⁵Two of the five recalled state legislators were Republican state Assembly members removed by Republicans in heavily Republican districts, ostensibly for compromising with Democrats (Morton 2006). The most recent legislator, Democratic State Senator Josh Newman, lost a 2018 recall after winning his seat by 0.8% in 2016. The Republican-led recall protested his support for a bill that raised California’s gas tax. Source: <https://www.sfchronicle.com/politics/article/Recall-of-state-Sen-Newman-costs-Democrats-12971819.php>.

⁶Source: California Election Data Archive, available at <http://hdl.handle.net/10211.3/210187>.

⁷It turns out that Persky was not the first lower court judge recalled in California history. Three Los Angeles County judges were recalled in 1932 following a campaign against them by the California Bar Association (Smith, 1951). To our knowledge, this precedent was never cited in contemporary coverage of the Persky recall.

⁸https://www.courts.ca.gov/documents/California_Judicial_Branch.pdf.

Figure 1 Qualified and Successful Recalls of Local Officials in California, 1995-2018.



considerations. Since 1977, sentencing for most crimes operates according to a triad system, in which the judge is given the choice between upper, middle, and lower “base” terms. For example, Assault with a Deadly Weapon (§245(a)(1) of the California Penal Code) carries a base term of 2, 3, or 4 years in prison. Although there is a presumption in favor of the middle term in the absence of aggravating or mitigating factors, few sentences precisely match the three prescribed base terms, for three reasons. First, judges have discretion over whether the sentences for convictions on multiple counts run consecutively or concurrently. Second, judges can issue sentencing enhancements for aggravating factors such as gang or hate crimes, or prior convictions. Third, since 2011, judges have been granted discretion to issue suspended or split sentences for certain felonies.⁹ As is generally the case in the United

⁹Effective since 2015, many crimes that are neither sexual crimes, violent crimes nor serious crimes are also eligible for county jail sentencing (for terms of 16 months, 2 years, or 3 years).

States, the vast majority of cases are resolved via plea bargain. This generally includes a sentencing recommendation to the presiding judge, who retains ultimate discretion over the actual sentence imposed.

2.2 The Persky Recall

Our empirical analysis focuses on a shock to the salience of the recall threat to judges in California brought about by the campaign to recall Judge Aaron Persky from 2016 to 2018. The campaign was initiated in response to Judge Persky’s sentencing decision in a widely publicized sexual assault case. On January 28, 2015, Brock Turner, a white Stanford student athlete, sexually assaulted Chanel Miller,¹⁰ a visiting student, and was arrested. Five days later, Turner was indicted on two rape counts, two felony sexual assault counts, and one attempted rape count. The rape charges were later dropped, and in March 2016, Turner was convicted on the sexual assault and attempted rape charges.

Turner faced a maximum sentence of 14 years for these convictions, but on June 2, 2016, Judge Persky sentenced Turner to six months in prison and three months probation. The lenient sentence and Miller’s impact statement, published by BuzzFeed, sparked widespread national attention.¹¹ On June 6, 2016, Stanford Law School Professor Michele Dauber announced the formation of a committee and began the process of collecting signatures to recall Judge Persky. With 94,000 verified signatures collected, the Santa Clara Registrar certified the signature threshold had been met on January 24, 2018. Finally, Judge Persky was recalled (with 61.5% supporting removal) on June 5, 2018. According to the Palo Alto Daily Post, the campaign to remove Persky raised more than one million dollars.¹²

¹⁰While ordinarily we would strictly adhere to the norm of respecting the anonymity of sexual assault victims, Miller has specifically expressed a preference *not* to remain anonymous, both in public appearances and her memoir (aptly titled *Know My Name*).

¹¹<https://www.buzzfeednews.com/article/katiejmbaker/heres-the-powerful-letter-the-stanford-victim-read-to-her-ra>.

¹²<https://paldailypost.com/2018/05/27/recall-persky-campaign-raises-more-than-1-million>

Criticisms of the recall campaign were immediate and widespread. 95 California law professors signed an open letter in August 2017 opposing the recall petition. California mayors, state legislators, former Supreme Court justices, and hundreds of Superior Court judges supported the Retain Judge Persky Campaign.¹³ Critics were primarily concerned with judicial independence and impartiality (Santa Clara County Association, June 14, 2016; Law Professors Statement, August 15, 2017). Some critics also predicted an increase in judicial punitiveness, with disproportionate effects on minority defendants (Butler, July 11, 2016; Gersen, June 17, 2016; Woolf, June 24, 2016). These predictions were bolstered by the empirical literature, cited below, documenting how concerns with reelection induce trial judges to impose longer sentences; as well as the significant literature, also discussed below, documenting the disproportionate burden imposed by the criminal justice system on minority defendants.

2.3 Related Research

Electoral incentives. The current research contributes to a rich literature on the incentive effects of electoral institutions on the behavior of incumbents generally (see, especially, Besley and Case, 1995; Alt, Bueno de Mesquita, and Rose, 2011; Ferraz and Finan, 2011) and judges specifically (Besley and Abigail Payne, 2013; Brace and Hall, 1995; Huber and Gordon, 2004; Gordon and Huber, 2007; Lim, 2013; Matsusaka et al., 2010). One feature of judicial elections that makes them particularly noteworthy in the empirical analysis of electoral incentives is the nature of the informational environment in which they occur. Voters

By contrast, an attempt to recall an Orange County judge, Marc Kelly, in 2015, raised less than \$25,000 and did not achieve the required number of signatures (Source: <https://www.nbcnews.com/news/us-news/group-pushing-recall-effort-stanford-rape-case-judge-it-long-n590431>).

¹³The associated website, Voices Against Recall, has since been removed. An archived version is available here: <https://web.archive.org/web/20180423164925/http://www.voicesagainstrerecall.org/>.

often lack verifiable information to evaluate judicial performance, a problem further complicated by the fact that judges often face voters in retention elections (in which there are no challengers) and nonpartisan elections (in which voters lack clear cues such as party labels). As a result, voters may be highly responsive to well-publicized examples of apparent judicial “error,” as revealed by the media, organized interest groups, or challengers. As noted by Gordon and Huber (2007), the threat of such widely publicized judicial error is asymmetric, owing to differences in the availability of voter information. Instances of perceived judicial leniency may be publicized by victims’ rights groups, the media, or challengers, backed up by sensational instances of recidivism. By contrast, overpunishment and/or wrongful conviction are generally revealed only years after the fact, if they are at all, and then only after costly and lengthy investigations. This asymmetry in the likelihood of “bad news” for the judge creates an incentive akin to pandering (Canes-Wrone, Herron, and Shotts, 2001), such that an increase in a perceived threat to a judge’s electoral prospects will push that judge to become more *punitive*, rather than more responsive to underlying constituent preferences.

Recall Elections. To our knowledge, there exists no extant empirical research on the incentive effects of recall elections. The political science research on recall elections has instead focused on voter behavior in recall elections – see, e.g., Ho and Imai (2006); Segura and Fraga (2008); Masket (2011); Shaw, McKenzie, and Underwood (2005).¹⁴ One explanation for this lacuna might be that the most straightforward research designs available to researchers do not translate well to the recall setting. Because the institution of recall is not randomly assigned, comparing the behavior of officials in states with and without recall is likely to suffer from numerous unobservable confounders – including, *inter alia*, other variation in electoral

¹⁴Others have examined the effects of direct democracy on the availability of information about voter preferences. Kousser, Lewis, and Masket (2007) explore state Assembly members’ roll-call behavior immediately after the 2003 recall of CA Governor Gary Davis. They found that Democratic (and to a smaller extent Republican) Assembly members responded to shifts in voter ideology (and/or voter perceptions) signalled by the Governor’s recall.

institutions (such as the availability of referendum and recall). There are also issues characterizing variation in the “treatment” of officials within the same state because the timing and occurrence of recall attempts are random and idiosyncratic. A useful comparison in this regard is studies of the effect of electoral proximity in states with staggered electoral calendars (Huber and Gordon, 2004). Finally, studying changes in the behavior of an individual official subject to a recall effort will afford essentially no statistical power. More generally, a challenge to studying the effects of recall elections on official behavior is that the threat of recall will be “priced into” the behavior of the officials. Unanticipated shocks, should they occur, are likely to be exceptionally rare and highly localized.

Judicial bias and asymmetric burdens of criminal justice system African Americans face a six-fold greater rate of imprisonment than whites in the United States (Bronson and Carson, 2019). While noting potential racial differences in criminal behavior, a number of recent papers have highlighted the influence of disparities induced by judicial and prosecutorial discretion, even among defendants facing similar charges and of similar criminal backgrounds. Evidence from randomly assigned cases indicates that judges differ in the degree to which race influences their likelihood of incarceration (Abrams, Bertrand, and Mullainathan, 2012). In Kansas, retention pressures, discussed above, induce increased judicial punitiveness but only in cases involving black felons (Park, 2017). Capital punishment sentences involving white victims are significantly more likely to be overturned by appellate courts when the defendant is African American, providing evidence that lower courts discount the wrongful convictions of black defendants (Alesina and La Ferrara, 2014). Racially disparate judicial decision-making is in turn compounded by racial disparities in charging and plea bargaining (Rehavi and Starr, 2014), jury decision-making (Bayer, Hjalmarsson, and Anwar, 2012) and policing (Grogger and Ridgeway, 2006).

Defendant race is, of course, just one of many substantively irrelevant factors that may affect judicial behavior. Judges assign tougher sentences to randomly assigned cases im-

mediately after local university football teams experience unexpected defeats, an increase asymmetrically borne by African American defendants (Eren and Mocan, 2018). The sequential order of case hearings also increases judicial punitiveness, with judges discontinuously issuing lenient sentences after food breaks (Danziger, Levav, and Avnaim-Pesso, 2011). Judicial characteristics, including ideology, gender, and race, are strongly associated with judicial decision-making (Harris and Sen, 2019).

3 Data and Method

3.1 Data on Sentencing in California

Unlike in other states, at the time of writing there is no publicly accessible, centralized repository for sentencing data. To overcome this limitation, we scraped 458,099 criminal cases with hearing dates between January 2015 and December 2018, inclusive, from the websites of the seven superior courts that make these data available in one form or another: Alameda, Fresno, Napa, Sacramento, Santa Barbara, San Bernardino and Santa Cruz. After encountering problems with Alameda, our search ultimately produced a total of 19,798 cases encompassing 22,111 felony charges with initial sentencing dispositions in the remaining six courts.¹⁵ The sample counties represent 19% of California’s total incarcerated population. While we make no claims concerning how representative these counties are of the broader state, we at the same time have no reasons to believe that the effect of the recall should be larger or smaller in these counties than in counties for which data were not readily available.

Each charge is associated with a sentence length in days. For each offense code, we acquired base terms from the State of California Attorney General’s office operative for the period of our sample.¹⁶ Additional case information in our final dataset include the charge

¹⁵Sentences may be amended – for example, in cases of probation violations.

¹⁶<https://oag.ca.gov/law/code-tables>.

(410 unique offenses) and sentencing judge (157 unique judges). Of the 15,490 charges for which non-guilty plea/ plea bargaining status could be identified, 76% were plea bargained. We categorized crimes as nonviolent or violent based on offense codes from the California Attorney General: 75% of charges in the sample are classified as violent and 5% of the charges in the sample are classified as sex crimes.

To explore heterogeneity by race, we linked defendants in our data to publicly available arrest records sourced from county and municipal law enforcement agencies in California.¹⁷ We crawled 201,066 arrest records. Defendants were matched based on first name, middle name, last name, county of arrest and arrest date. Across the six counties, 12,844 defendants could be matched to arrest records, of which 11,184 defendants have race identified.

3.2 Empirical Approach

In the main part of our analysis, we look for sharp increases in judicial punitiveness immediately following key moments during the recall campaign. In particular, we consider two critical events: the announcement of the campaign itself, on June 6, 2016; and the recall election itself, on June 5, 2018.¹⁸ Our main specification is the following local linear estimator of a regression discontinuity in time (RDit; see Hausman and Rapson (2018)):

$$y_{ijt} = \beta_0 + \beta_1 \mathbb{I}(t > t_k) + \beta_2 f(t - t_k) + \varepsilon_{ijt} \quad (1)$$

Where t_k is the calendar date of a critical event k ; $y_{ijt} \equiv \min\{s/\bar{s}, 1\}$ is the normalized sentence of conviction i by judge j at time t (cf., Lim, 2013); and $f(\cdot)$ is smooth function of time. The normalization divides the sentence length in days s by the upper base term \bar{s} , creating a fractional measure of judicial discretion expressible in percentage terms and

¹⁷We scraped arrest records from this source: <https://www.localcrimenews.com/>.

¹⁸Another candidate is the date on which petition signatures were certified by the Santa Clara Registrar: January 24, 2018. Results from this event may be found in the Appendix.

comparable across different offenses. So, for example, a sentence of six months on an assault with a deadly weapon charge with an upper base term of four years would be coded as 0.125. The measure is censored at one so as not to be skewed by cases with unusual aggravating factors that increase the sentence above the upper base term. In point of fact, 96% of cases fall at or below the upper base term. That said, in robustness tests we use the uncensored measure as well as the raw sentence (in days) as outcome measures.

In addition to this unadjusted specification, we also present results throughout that adjust for a vector of judge- and offense-specific fixed effects. This entails discarding sentencing data from Sacramento County, whose data do not include judge identifiers. As is standard in RD designs, we weight observations using a triangular kernel. Standard errors for the unadjusted specifications are clustered at the county-statute level, and adjusted specifications at the judge-statute level. RD bandwidths are MSE-optimal (Calonico, Cattaneo, and Titiunik, 2014).

The identifying assumption of regression discontinuity designs is that treatment assignment is ignorable (conditional on covariates) sufficiently close to the cutoff (the critical event in the RDit setting). We examine threats to inference arising from shocks that vary discontinuously within the treatment windows. A sequence of placebo regressions for all dates in each calendar year alleviates the concern that the findings result from some confounding structural break (for instance, the ratification of two laws in September 2016 requiring mandatory sentences for sexual assault). To bolster further our claim that the recall events do not coincide with unrelated shocks to judicial decision-making, we examine contemporaneous sentencing patterns in the nearby state of Washington. To limit the effects of prosecutor responses to the recall campaign, we test whether our findings are robust to excluding charges filed after the recall events. Finally, we assess robustness to various specifications of the outcome and bandwidth.

Next, we examine whether any observed effects of key events on punitiveness are driven

by sentencing for sexual, non-sexual violent or nonviolent crimes. As the recall campaign centered around Judge Persky’s sentencing in a sex crime case, judges might have anticipated that voters would pay greater attention to perceived leniency on similar cases.

Third, we assess whether any increase in judicial punitiveness induced by the recall campaign placed a disproportionate burden on minority defendants, as anticipated by some of the campaign’s critics. It is important to note that there are two channels through which this might operate. First, it could be that judges are apprehensive that a racially biased electorate might react especially negatively to perceived leniency toward minority defendants. In that case, we would anticipate that the effect of the recall on sentencing would be larger for them than for white defendants. Call this the *direct racial burden* hypothesis.

Testing the direct burden hypothesis is subtle, as the following examples demonstrate. Suppose, hypothetically, that judges tend to sentence both white and minority defendants at the midpoint of their discretionary sentencing range on the normalized scale described above. A direct burden would be corroborated if normalized sentences increased more for minority than white defendants following the petition announcement or recall election. Suppose, on the other hand, that racial disparities in discretionary sentencing were *already present* prior to the electoral shock, so that white defendants were sentenced at the lower, and minority defendants at the upper, ends of judges’ discretionary sentencing range. In that case, and even in the presence of the posited underlying mechanism, we might observe a (weakly) larger effect of the electoral shock for white than minority defendants. To assess this possibility, we examine whether race-based disparities in discretionary sentencing were in evidence immediately prior to the electoral shocks. Note that this is a purely descriptive exercise aimed at clarifying the mechanism – we lack a strong identification strategy for assessing the causal effect of race on sentencing discretion directly (clearly a critical task, but one beyond the scope of the current paper).

Suppose that we uncover no evidence for the direct burden hypothesis. This does not

rule out the possibility that minority defendants are hit harder by the consequences of an electoral shock brought about by the recall drive. This is because it could be that white and minority defendants are charged with crimes that vary in their severity. Suppose judges increase their discretionary sentencing in an apparently race-neutral way – e.g., from 0.5 to 0.6 on the normalized sentencing scale – but that minority defendants tend to be charged with crimes with higher statutory maximum penalties. Then the electoral shock will mechanically lead to a higher cumulative sentencing load for minorities. Call this the *indirect racial burden* hypothesis. To assess the indirect burden hypothesis, we examine whether minority defendants in our sample are charged with crimes having higher maximum penalties.

The RDit approach identifies a local average treatment effect (LATE) at the precise moment of the critical event in question. In the final part of our analysis, we compute aggregate effects, which require estimating longer-term consequences of electorally-induced shifts in judicial behavior. Doing so requires more stringent identifying assumptions than those necessary to identify the LATE. Accordingly, rather than committing ourselves to one set of assumptions, we adopt four separate approaches: (1) assuming the estimated LATE persists as an average treatment effect in a window of time after the announcement; (2) a fully parametric approach that attributes any post-announcement time trends to decay or growth in the effect of the announcement itself; (3) a linear reweighting estimator; and (4) a propensity score estimator (the latter two approaches suggested by Angrist and Rokkanen (2015)).

4 Empirical Results

4.1 Main Results: Instantaneous Effects of Critical Events

Graphical Evidence. Before proceeding to our local linear estimation, our first step in assessing the effect of the events described above is graphical. Figure 2 illustrates the main

effects at the core of the paper, presenting binned averages of normalized sentence length (sentence length as a fraction of maximum sentence, top-coded at one), within the 90 day window surrounding each event.¹⁹ Quadratic curves and local polynomial smoothers are fit separately on either side of the event date.

The two panels in the top row present plots of data unadjusted for covariates. We observe a large, discontinuous increase in average normalized sentence transitioning from immediately prior to, to immediately following the June 6, 2016 petition announcement. This increase corresponds to an increase of approximately 10 percentage points on the normalized sentencing scale, reflecting a proportionate increase of approximately 29 percent over pre-announcement levels. Turning to June 5, 2018, the date of the recall election itself, we observe no change in average sentence length from before to after that date.

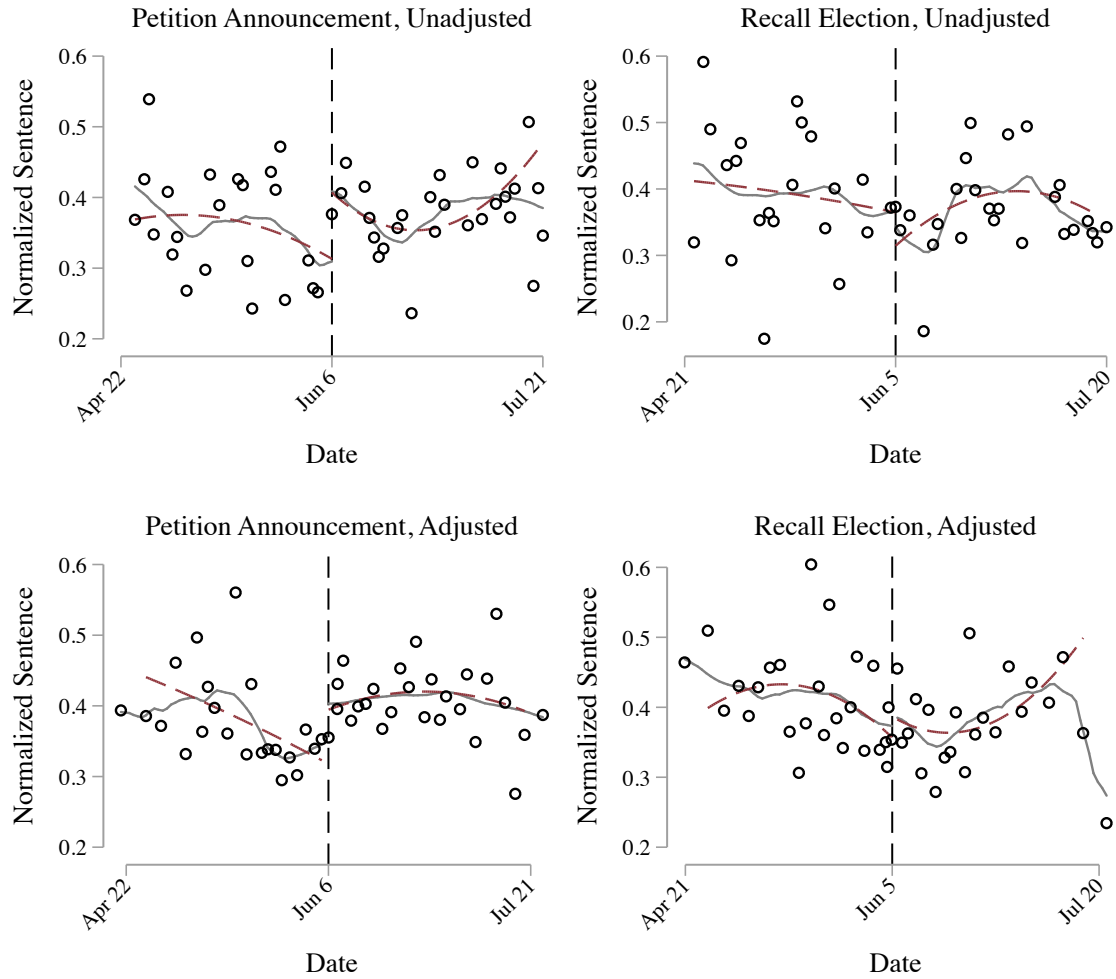
Plots in the bottom row depict binned means residualized using judge- and offense-specific fixed effects. The graphical analysis of the petition announcement adjusting for the fixed effects reveals a similar pattern to that in the unadjusted panel: an increase of around eight percentage points. Using the adjusted estimates, we again no apparent difference before and after the election itself.

Local Linear Regression Results To interrogate the preliminary inferences suggested by the graphical analysis in a more rigorous fashion, we next present local linear regression estimates of the local average treatment effect (LATE), β_1 from equation (1). The LATE estimates appear in Table 1. The running variable (time) is restricted to MSE-optimal bandwidths reported in the table, and the data are weighted using a triangular kernel. Following the recommendation of Gelman and Imbens (2019), we report results from a local linear specification rather than estimating higher-order polynomials (which are susceptible to over-fitting).

Estimates in the table corroborate the results from the graphical analysis. We estimate

¹⁹The ± 45 day window approximates the MSE-optimal bandwidth; see below.

Figure 2 Effect on Sentencing of Critical Events in Persky Recall: Graphical Analysis



Tokens in each panel depict average normalized sentence length in equally-sized bins; quadratic curve (maroon) and local polynomial smoother (gray) fit separately on each side of the event under consideration.

Table 1 Effect on Sentencing of Critical Events in Persky Recall: RD Estimates

	Petition Announced		Recall Election	
RD estimate	0.089 (0.041)	0.101 (0.039)	-0.034 (0.049)	-0.009 (0.051)
Left-side Intercept	0.296 (0.028)	–	0.37 (0.041)	–
Bandwidth	36.8	44.3	41.9	42.1
Judge fixed effects	N	Y	N	Y
Statute fixed effects	N	Y	N	Y
Effective observations	1,268	1,336	1,440	1,278

Estimates employ triangular kernel. Specifications with judge fixed effects exclude Sacramento County, which does not report judge identifiers. Standard errors clustered at the county-statute level for specifications with no fixed effects, and at the judge-statute level for specifications with fixed effects.

a large, statistically significant effect of the June 6 petition announcement: unadjusted (first column), the estimated effect is 8.9 percentage points on the normalized sentencing scale; adjusted for judge- and offense-specific fixed effects (second column), the estimate increases to 10.1 points. To give a sense of the substantive significance of these estimates, immediately prior to the announcement, the estimated average normalized sentence length (the left-side intercept in the Table) was 0.30; hence, these effects correspond to an immediate proportionate increase of 30.1 to 34.1 percent. Both RD estimates easily surpass conventional thresholds for statistical significance.

The third and fourth columns of the table display the analogous estimates for the recall election date. In contrast to the announcement estimates, the estimated effect, whether adjusted or unadjusted for the judge and offense-specific fixed effects, is small, negative, and statistically indistinguishable from zero.

4.2 Robustness of the Announcement Finding

Placebo tests for temporal confounding. Our main analysis implies that the announcement of the recall petition caused a substantial and immediate increase in the length of felony sentences in California. One threat to inference is the possibility that other events may have been taking place around the time of the announcement. One event that is particularly relevant for our analysis is the 2016 California primary, which took place on June 7. A second event is Persky’s actual sentence of Brock Turner on June 2.

With respect to the primary, there are two immediate responses. First, a Superior Court judge who faced a challenger in 2016 did so initially in a top-two primary, and would only need to face the voters in the general election upon placing in the top two but receiving less than 50% of the vote. Owing to California’s unusual electoral rules, the vast majority of judges would thus see the sway of electoral incentives *diminish* following a contested primary or remain roughly constant following an uncontested one. The anticipated behavioral response (given the prior research cited above) would be a reduction in average sentence length; hence, the overall effect would be to bias the above results *downward*. In point of fact, only one incumbent judge in our sample (in San Bernardino County) faced a primary challenge, and she did not hand down a sentence in the sample period.

With respect to the Brock Turner sentence, it is less clear what the direction of the bias might be. It is conceivable that judges, anticipating the electoral backlash from outrage over the sentence, might ratchet up sentencing in their courtrooms in response, and that this anticipation is what our main results are capturing. This would confirm the power of anticipated electoral threat, but complicate our efforts to make inferences about the specific effect of the petition announcement. On the other hand, perhaps the Turner sentence signaled the acceptability of unusual sentences. In the first account, our main estimates are biased upward; in the second; biased downward.

Another threat to inference with which to contend is that there may be numerous struc-

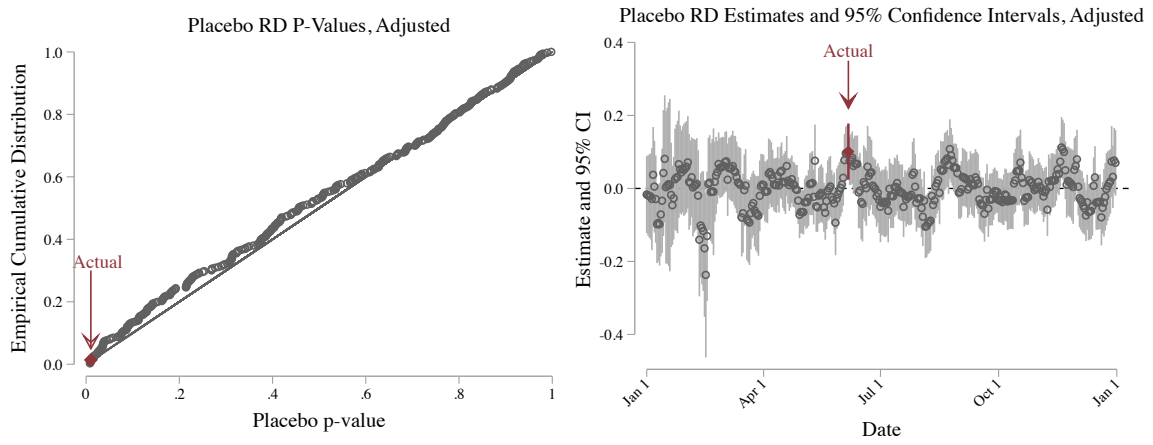
tural breaks throughout the calendar year that affect sentencing, some associated with the explicitly political stimuli discussed above and others associated with, *inter alia*, changes in sentencing guidelines, news accounts of prison overcrowding (or litigation on that issue), or shifts in prosecutorial behavior. The relevant question then becomes whether the shift associated with the June 6, 2016 cutoff was particularly unusual relative to other candidate breakpoints (including the Turner sentence).

To answer this question, we conducted a sequence of placebo tests, using every day of calendar year 2016 as a breakpoint in (MSE-bandwidth optimal) regression discontinuity analyses including judge- and charge-fixed effects.²⁰ Figure 3 displays the results. The left panel displays the empirical cumulative distribution of p-values for the placebo tests, along with the June 6 p-value (labeled “Actual”). The figure shows that the estimated June 6 p-value is lower than 98.9% of its placebo analogs. The right panel displays all 365 placebo estimates plus their associated 95% confidence intervals, along with the June 6 estimate and its confidence interval (again labeled “Actual”). There are two things to specifically note about the right figure. First, the only larger RD estimates are for August 24-25. It is possible that this large effect is random noise, although a review of contemporary news accounts of that week uncovered coverage of a report that felony arrests had plummeted 28.5% since California voters approved Proposition 47, a measure aimed at lowering criminal sentences by reclassifying certain felonies as misdemeanors Thompson (2016). Second, looking at the neighborhood around June 6, we see that the estimate for that specific date is larger than any of the surrounding placebo estimates – including June 7 and June 2. In fact, the placebo estimate for the date of the Brock Turner sentence is statistically indistinguishable from zero.

Washington as a placebo state. To further assuage concerns that the petition announcement coincided with an unrelated shock to judicial decision-making, we assess shifts

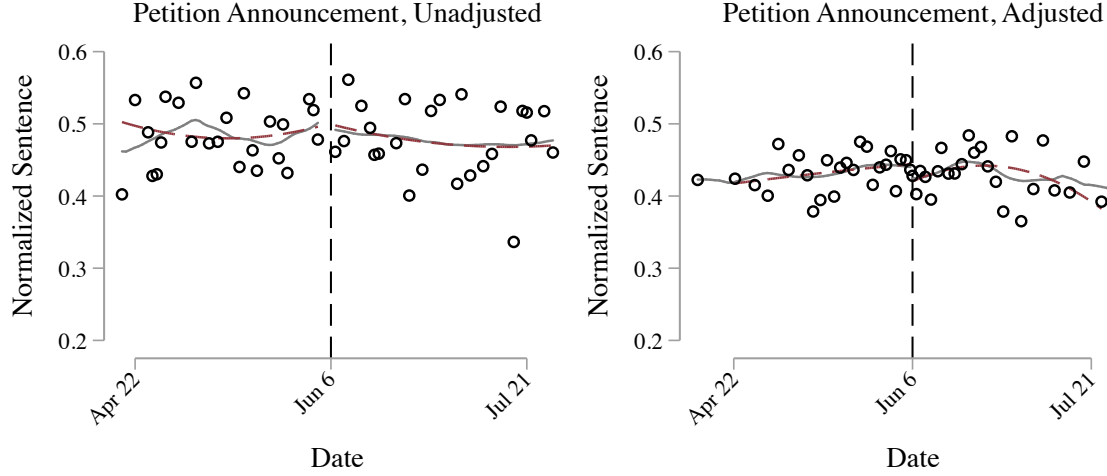
²⁰Placebo tests using unadjusted RD-estimates produce substantively identical conclusions.

Figure 3 Placebo Tests for Main Effect of Petition Announcement



The left panel displays the empirical cumulative distribution of estimated placebo p-values (in gray), with the actual petition announcement p-value overlaid in maroon. The right panel displays placebo RD estimates and associated 95% confidence intervals (in gray), with the actual estimate and confidence interval in maroon.

Figure 4 Effect on Sentencing of Petition Announcement in Washington State: Placebo Test

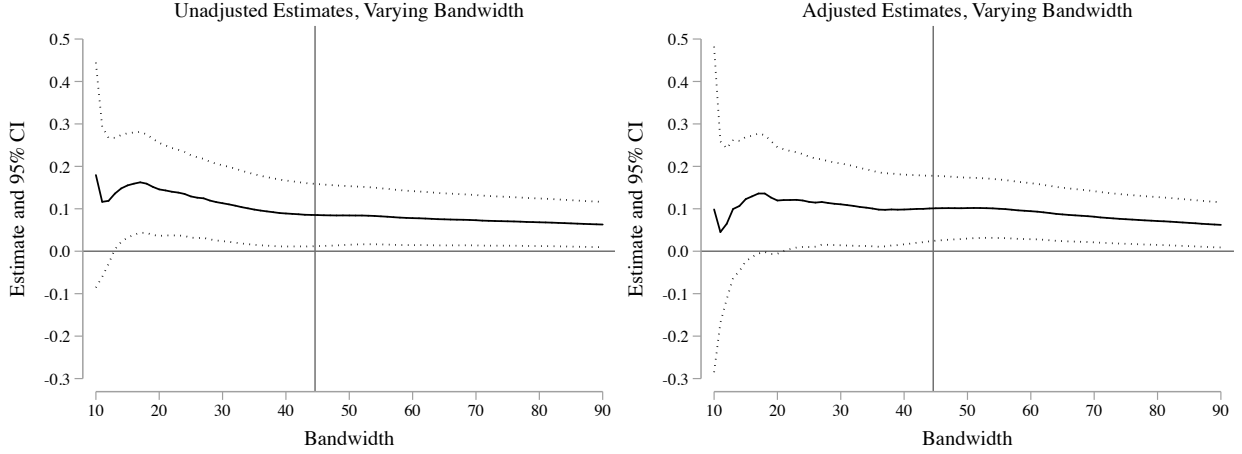


See notes in Figure 2.

in judicial punitiveness in nearby Washington state. Like judges in California, judges in Washington face nonpartisan elections (four year terms) and have broad discretion to issue sentences within the appropriate sentencing guidelines. However, unlike California, the Washington state constitution does not allow for the recall of judges. Using data sourced from the Washington State Department of Corrections, we extracted the sentencing judge, charge and sentence length associated with 135,984 charges. In Figure 4, we present binned averages of the normalized sentence within 45 days of the petition announcement date. Neither the unadjusted averages nor the averages adjusted on judge and charge fixed effects significantly change after the petition announcement date. Local linear regression results confirm the null finding.

Tests for bandwidth artifacts. While the estimates above employ a principled means of selecting the optimal bandwidth, we wish to make sure that the significance of our results is not overly dependent on the breadth of the interval employed in the analysis. Accordingly, we re-ran our analysis of the effect of the petition announcement for different bandwidths, ranging from one week to 90 days. Results of this exercise appear in Figure 5. For very

Figure 5 RD Estimates Varying Bandwidth



As in the main analysis, estimates employ triangular kernel, with standard errors clustered at the judge-charge level. The solid line denotes the MSE-optimal bandwidth.

short bandwidths, of course, the sample size declines dramatically, substantially diminishing the precision of the estimates. However, past around a two-week bandwidth for both adjusted and unadjusted specifications, our main results are robust to a wide range of different bandwidths, up to values that can scarcely be labeled “local” given the August shocks to sentencing that we document above.

Tests for charge adjustment. An additional threat to inference concerns strategic adjustment by prosecutors to changes in the electoral environment for judges brought about by the Persky recall effort. Prosecutors have enormous discretion when deciding what charges to bring against a defendant alleged to have committed a specific criminal act. Suppose that prosecutors, anticipating an increase in judicial punitiveness, responded by becoming more aggressive in the charges they file. Empirically, the effect would be to bias the observed sentencing effect toward zero, because the electorally-motivated increase in judicial punitiveness would be divided between the charge inflation and (observed) sentencing conditional on the charge inflation. While the anticipated direction of bias should strengthen our assessment

of the robustness of the findings concerning the petition announcement, it may weaken that for the null findings on the recall election dates.

One way to obviate these concerns is to restrict attention to cases that were filed *before* the critical dates in question: for example, in assessing the effect of the petition announcement, delete any cases that were filed after the announcement itself, and thus before any opportunity for prosecutorial adjustment to any novel information conveyed by the event. Table A.2 in the Appendix displays the revised estimates restricting the data in this way. Owing to the relatively short (45 day) bandwidth and the typical duration of criminal justice proceedings (the median time from charge date to sentencing is 109 days, with 76% of cases longer than 45 days), restricting the data in this way has minimal impact on the estimated RD estimates.

Alternative measures of the outcome. Finally, we consider whether our estimates are influenced by the choice of outcome variable. Table A.3 in the Appendix replicates the main analysis in Table 1 using the same normalization but without top-coding at one. This operationalization will pick up increases in judicial punitiveness that result from, e.g., decisions to let sentences for multiple charges run consecutively instead of concurrently. Using this alternative coding leads to slight changes in coefficient magnitudes, but reproduces the main results: substantial, statistically significant increases in sentencing associated with the petition announcement, and no significant change associated with the recall election.

Table A.4 in the Appendix uses the raw sentence (in days) rather than the normalized measure. Here, the offense-specific fixed effects are particularly important, as they pick up mean sentence length for specific charges. Using the non-normalized time scale as the outcome, the substantive import of the findings remains unchanged, with our fixed effects estimates suggesting that the petition announcement had an average (within-charge, within-judge) effect of 141 days additional incarceration.

Table 2 Heterogeneous Effects of the Petition Announcement: RD Estimates by Crime Type

	Sex Crimes		Other Violent Crimes		Nonviolent Crimes	
RD estimate	-0.059 (0.141)	0.531 (0.301)	0.129 (0.047)	0.148 (0.05)	-0.049 (0.06)	-0.036 (0.082)
Left-side intercept	0.288 (0.133)	–	0.269 (0.029)	–	0.396 (0.044)	–
Bandwidth	60.9	62.1	38.1	38.8	72.2	54.6
Judge fixed effects	N	Y	N	Y	N	Y
Statute fixed effects	N	Y	N	Y	N	Y
Effective observations	81	55	962	829	655	390

See notes in Table 1 for estimation details.

4.3 Effects by Type of Crime

A natural question to consider in assessing the above result concerning the petition announcement is the extent to which it is driven by increases in sentencing for different crimes. To the extent that the precipitating event for the Persky recall was a lenient sentence for a violent sex crime, we wish to consider whether the incentive effect of the petition was confined to sex crimes. Accordingly, we partition the sample of felon cases into sex crimes, non-sexual violent crimes, and nonviolent crimes, and run the local linear regression estimator (unadjusted and adjusted for judge and offense fixed effects) separately for each of the three categories. Results appear in Table 2.

Turning first to the analysis of sex crimes (the first and second columns), one immediately notes the very small sample size despite the 50% increase in the bandwidth relative to the main analysis. This contributes to marked imprecision in the estimated coefficients: these estimates switch signs depending on specification, and are nowhere close to statistical significance in either. Our estimates for nonviolent crimes (fifth and sixth columns) are more precisely estimated given the larger number of cases with which to conduct the analysis; however, in neither specification is the effect of the petition announcement significantly

different from zero.

Focusing on violent crimes not of a sexual nature, we observe large and precisely estimated LATEs for both specifications, suggesting an approximately 13 to 15 percentage point immediate increase in sentencing associated with the petition announcement. Relative to a baseline normalized sentence for non-sexual violent crimes of 0.27 immediately prior to the announcement, this represents an immediate 48-55 percent proportionate increase in punitiveness. In other words, the effects described in the main analysis appear to be driven by increases in sentences for non-sexual violent crimes. This is consistent with the prediction by critics of the recall effort that any resulting increases in sentencing stringency would not be confined to sex crimes.

4.4 The Recall Petition and Disproportionate Burden by Race

We next assess the argument made by critics of the recall effort: that notwithstanding the aim of sanctioning a judge for imposing a lenient sentence for a white defendant, any increase in judicial punitiveness driven by the recall itself would likely be disproportionately borne by minority defendants. As discussed above, doing so requires adjudicating between the direct and indirect racial burden hypotheses.

The direct racial burden hypothesis. As described above, two patterns in the data would be consistent with the direct burden mechanism: (1) the immediate effect of the petition announcement on normalized sentences being more severe for minority defendants than white defendants; OR (2) the immediate effect of the petition announcement on normalized sentences being weakly more severe for white than for minority defendants, but higher normalized sentences for minority defendants prior to the announcement.

To assess the presence of absence of the first pattern, the top half of Table 3 displays local linear RD estimates of the petition announcement reported separately for black, Hispanic, and white defendants. The first thing to note is that relative to our main analysis, the

effective sample size is considerably smaller, owing to the difficulty matching arrest and sentencing records. This contributes to a lack of precision in our estimates. We find a 12 to 24 percentage point increase in normalized sentences for African Americans and a 9 to 12 percentage point increase for Hispanics. These effects are not robust to changes in specification. We can reject the null hypothesis of no effect for African Americans in the specification with judge and statute fixed effects, but not in the specification without fixed effects. The reverse is true for Hispanic defendants. The corresponding effect for Whites is 16 to 28 percentage points and exceeds conventional thresholds of statistical significance for both specifications.

We next assess whether the RD estimates displayed in the top half of the table are statistically different *from each other*. To conduct these tests, we estimate interactive specifications that pool black, white, and Hispanic defendants and estimates the three RD estimates in the same model, employing a common bandwidth rather than defendant race-specific bandwidths.²¹ These appear in the middle portion of the Table. For no comparison can we reject the null hypothesis that the race-specific RD estimates are identical. The null finding persists both in the specifications with and without judge and statute fixed effects. The conjunction of these findings rules out the first pattern that would constitute evidence for the direct burden hypothesis.

Next, we investigate the presence or absence of the second pattern: a null finding for racial differences in the RD estimates – documented above – *and* descriptive evidence of racial disparities in normalized sentences prior to the petition announcement. To assess pre-treatment racial disparities in discretionary sentencing, we compare the conditional expected normalized sentence immediately prior to the petition announcement, as captured by the left-side intercepts from the RD estimation.²² Statistical tests of differences by defendant race

²¹Comparing the coefficients obtained from the split sample approach yields substantively identical results, but assumes zero covariance among the RD estimates.

²²Owing to the difficulty of interpreting this intercept in the twoway fixed effect setting,

Table 3 Assessing Direct Racial Burden from the Petition Announcement: RD Estimates by Race of Defendant

	Black		Hispanic		White	
RD estimate	0.121 (0.104)	0.235 (0.128)	0.12 (0.059)	0.088 (0.074)	0.163 (0.083)	0.28 (0.119)
Left-side intercept	0.331 (0.07)	–	0.297 (0.039)	–	0.25 (0.067)	–
Bandwidth	53.1	58.4	46.7	44.1	44.9	56.7
Judge fixed effects	N	Y	N	Y	N	Y
Statute fixed effects	N	Y	N	Y	N	Y
Effective observations	257	205	534	473	296	293
Hypothesis Tests of Equality of RD Estimates						
$H_0 : RD_{Black} = RD_{White}$	0.041 (0.12)	0.077 (0.188)				
$H_0 : RD_{Black} = RD_{Hispanic}$	0.001 (0.11)	0.23 (0.15)				
$H_0 : RD_{Hispanic} = RD_{White}$	0.039 (0.094)	0.153 (0.153)				
Hypothesis Tests of Equality of Intercepts						
$H_0 : LSI_{Black} = LSI_{White}$	0.071 (0.102)	–				
$H_0 : LSI_{Black} = LSI_{Hispanic}$	0.045 (0.075)	–				
$H_0 : LSI_{Hispanic} = LSI_{White}$	0.045 (0.075)	–				
Judge, statute fixed effects	N	Y				

See notes in Table 1 for estimation details. Hypothesis tests of equality of estimates derived from combined specification that uses optimal bandwidth (47.68 days) for pooled data. LSI is the left-side intercept, i.e., the value of the regression function estimated using data prior to the petition announcement at the date of the announcement.

appear at the bottom of Table 3. For none of the three comparisons can we reject the null hypothesis that normalized sentences do not differ by race.

The indirect racial burden hypothesis. As discussed above, the fact that minority and white defendants receive comparable normalized sentences does not imply the absence of disproportionate burden on minorities. Even if the observed effect of the petition announcement on the use of judicial discretion (as captured by the normalized sentence) was identical across racial categories, this would still place a disproportionate burden on minorities if they are systematically charged with more severe crimes. To assess this indirect racial burden hypothesis, we considered whether, as a descriptive matter, this was the case. Table 4 displays estimates from a regression of a case’s *statutory maximum sentence* (in days) – a measure of severity – on indicator variables for race (Black and Hispanic, the omitted category is White)²³, adjusting in some specifications for judge-specific fixed effects. (As the primary instrument for manipulating charging severity is the choice of offense itself, we omit statute-specific fixed effects for this portion of our analysis.)

Our analysis is consistent with the indirect burden hypothesis: In our full sample (2015-2019) and restricting attention to the post-petition time period, African American and Hispanic defendants are systematically charged for more severe crimes than their white counterparts. In the full sample, African American defendants were charged for crimes whose maximum sentences were 78.6 to 85.3 days longer than white defendants, while Hispanic defendants were charged with crimes averaging 31.3 to 55 days longer than white defendants.

What is the implied magnitude of the indirect burden for African American defendants specifically? Recall from our main analysis that the average normalized sentence was 0.31 immediately prior to the petition announcement, and approximately 0.4 (depending on specification) after. If the average maximum sentence was 1,276 days for a white defendant, and

this portion of the analysis is confined to the unadjusted estimates.

²³Only a tiny fraction of defendants in the sample are identified as Asian or Native American in the arrest data.

Table 4 Assessing Indirect Racial Burden: Racial Disparities in Crime Severity, as Measured by Statutory Maximum Penalties

	Pre-Announcement		Post-Announcement		Full Sample	
Black	82.6 (19.9)	51.7 (55.6)	85.9 (27.5)	82.8 (17.1)	85.3 (25.1)	78.6 (15.9)
Hispanic	58.7 (40.7)	11.4 (20.8)	54.4 (23.4)	31.9 (16.0)	55.0 (25.3)	31.3 (15.7)
Intercept	1284.7 (25.2)	1325.4 (14.1)	1273.8 (10.9)	1293.4 (9.7)	1275.6 (11.3)	1296.6 (9.5)
Judge fixed effects	N	Y	N	Y	N	Y
N	1,942	1,260	10,328	8,159	12,270	9,419

Excluded category is white defendants. Standard errors clustered at the county or judge level.

1,361 days for a black defendant, then the average effective increase in incarceration would be 115 days for the former and 123 days for the latter – a difference of approximately 7%.

4.5 Aggregate Effects

An advantage of the regression discontinuity in time approach is that it precisely identifies a local average treatment effect at the time of the critical event under consideration under relatively weak assumptions. However, insofar as effects are only identified at the boundary, interpreting their broader substantive implications requires additional assumptions. In the current application, the most relevant consideration – both in terms of cost to defendants and cost to the state of California – is a counterfactual one: how does the shift in judicial behavior following the petition announcement translate into additional days, months, or years of additional prison time? Rather than commit ourselves to one set of assumptions, in this section we adopt four alternative approaches. Likewise, rather than extrapolate over a prolonged period of time (in which, per our placebo tests above, a sequence of additional factors not pertaining to the Persky recall may have affected judicial punitiveness), we restrict

ourselves to the 45 day window following the petition announcement (with the 45 day length approximating the optimal bandwidth from the RD estimates above).

The first approach is to assume that the identified local average treatment effect is the average treatment effect over the 45 day window. This approach assumes no decay or growth in the effect of the announcement on sentencing considerations. We proceed by summing the product of our LATE estimates, expressed as a percentages of a case’s statutory maximum, and a case’s statutory maximum itself, over all cases in the 45 day window.²⁴ We report results using the unadjusted LATE estimate and the estimate adjusted for judge- and offense-specific fixed effects.

The second approach is to estimate a fully parametric regression model that adjusts for time trends before and after the announcement, and use the predicted values from that model to estimate the aggregate effect over the 45 day window. This approach may capture growth or decay in the effect over the interval following the announcement; however, it may also erroneously attribute factors unrelated to the announcement to the announcement itself. In order to protect against the possibility that downward pre-announcement trends might artificially inflate anticipated sentencing differences, we constrain the trend to zero when predicting counterfactual sentences.

The third and fourth approaches employ estimators recommended by Angrist and Rokkanen (2015), which rely on a critical feature of the regression discontinuity design: that failure to control for the running variable (time) is the only source of omitted variables bias. Their approach to estimating average treatment effects away from the boundary is to test whether, conditioning on covariates, a relationship exists between the outcome and running variables; and if not, to estimate average treatment effects for an interval using either a linear reweighting or propensity score estimator (see Angrist and Rokkanen (2015) for details). We adopt

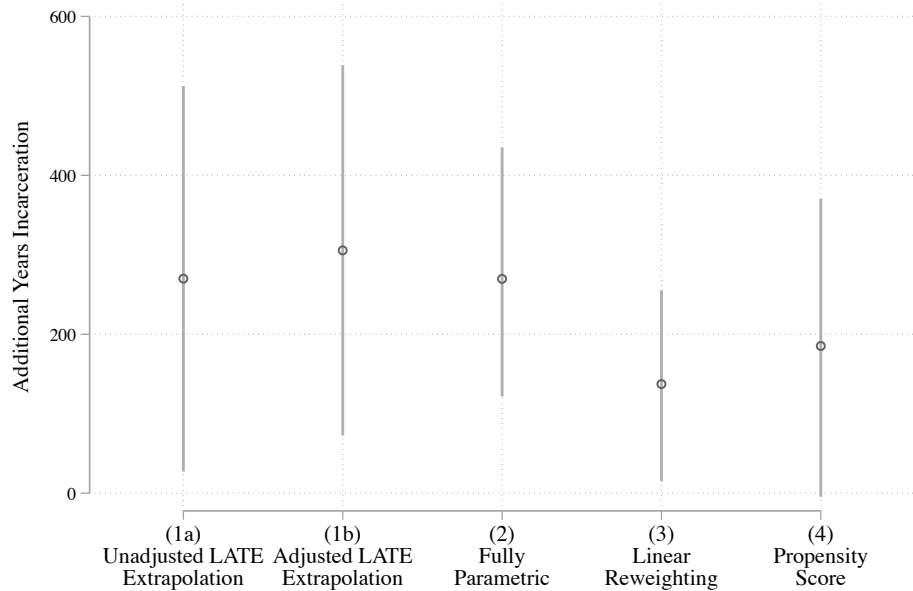
²⁴Using the results from LATE estimates using non-normalized sentence length as the outcome variable yields slightly larger results; insofar as these may be in artifact of variation in sentencing enhancements, we report the more conservative results.

both approaches, restricting attention to the 20 day period before the announcement for control observations given a downward linear time trend in evidence using a wider pre-treatment window.

Figure 6 displays the estimated additional incarceration (in years) for the counties in our sample using the approaches described above. Depending on the approach, point estimates suggest total effects of between 152 and 316 years additional incarceration associated with the announcement. The larger figure comes from the fully parametric model; note that if we did not implement the constraint described above, an estimated downward pre-announcement trend would have increased the predicted additional incarceration further, to 423 years. (The observed downward trend justifies the adjustment employed in the estimators based on the conditional independence assumption.) At the same time, the smaller estimates discard observations for which we lack covariate overlap pre- and post-treatment (e.g., sentences from the same judge both before and after the announcement), and are likely to be biased downward.

While the human cost of this estimate on defendants is difficult to calculate without very strong assumptions, a far easier calculation is the total cost to the state: In 2016-17, the average annual cost of incarceration in the California Department of Corrections was \$71 thousand per inmate. Using the most conservative 152 year estimate, our analysis suggests a total cost to the six counties in our sample of \$10.8 million. Note also that defendants from the counties in our sample make up just 12% of the incarcerated population in the state. Under fairly restrictive assumptions (most importantly, that the distribution of charges and the effect of the petition announcement are both uniform across the state) a back-of-the-envelope calculation using the most conservative estimate suggests that the total effect statewide is 1,266 years, reflecting a total cost to the state of \$89.9 million. If the effect of the petition announcement persisted longer than the 45 day window under consideration, actual costs could be considerably higher.

Figure 6 Estimating Aggregate Effects of the Petition Announcement: 45 Day Window



Confidence intervals for (2), (3), and (4) derived using nonparametric bootstrap.

5 Discussion

In 2016, a Superior Court judge in California had not been recalled from office by voters in 84 years. But in the summer of that year, days after a Santa Clara County judge handed down a widely-publicized lenient sentence to an affluent, white defendant for a sexual assault and attempted rape conviction, an unanticipated recall campaign against that judge raised the threat of potential electoral sanctions for other judges.

The research presented in this paper documents the far-reaching consequences of that threat for the criminal justice system. Using data from six California counties, we observe large, instantaneous increases in judicial punitiveness immediately following the announcement of the recall campaign, which are most readily apparent in sentencing for non-sexual violent crime. While we uncover no evidence that these instantaneous effects were significantly different for minority and white defendants, the fact that the former tend to be

charged with more severe crimes implies that the increased salience of the electoral threat may have indirectly produced a disproportionate burden on black and Hispanic defendants.

The broader import of these findings – for our understanding of the criminal justice system in the United States and our understanding of electoral accountability – is twofold. First, they underscore the fact that even political campaigns targeting individual officeholders may have broad, unintended consequences. This is because such campaigns do not operate in a vacuum, and thus may alter the expectations of other officeholders that they themselves might be subject to such campaigns. The fact that we document no observable effects of the eventual recall election itself is consistent with this shift in beliefs to a “new normal” in the political environment of sitting trial judges, about which the ultimate (and widely-anticipated) electoral outcome conveyed no additional information. And critically, although the defendant in the precipitating case was white, and the crimes for which he was convicted were sex crimes, the use of the recall tool cannot be restricted to similar cases. And as such, neither can any anticipatory responses to that threat by judges in their courtrooms.

Second, the research presented contributes to our understanding of the electoral incentives of public officials. We provide the first empirical evidence that the threat of recall affects the behavior of incumbent officials. In the current context, we provide evidence that an exogenous shock to judges’ beliefs in the risk of recall affected their sentencing decisions. We document a substantial and immediate increase in sentencing severity following the highly-publicized announcement of a recall campaign, and calculate aggregate effects of that increase on the order of 150 years of additional incarceration for around 600 defendants in the 45 day period following the announcement. Insofar as we restrict our attention to a narrow window of time and only six counties, these estimates likely substantially underestimate the broader effects of this change in the behavior of these officials.

Finally, our analysis provides a roadmap for studying non-standard electoral institutions whose structure does not lend itself to standard research designs that exploit, *inter alia*,

proximity to the next election, cross-sectional institutional variation, or term limits. This is particularly valuable for an institution such as the recall, which, although widespread, is poorly understood. In the same vein, understanding the scope of incentive effects of recall efforts that vary in their intensity, and the political responses of incumbent officials,²⁵ is an important topic for future research.

²⁵Consider for example, a group of Californian judges who launched the Judicial Fairness Coalition shortly after the Persky recall (<https://www.caljudges.org/CommFairness.asp>), in part to provide resources for judicial officers facing potential recall threats.

References

- Abrams, David S, Marianne Bertrand, and Sendhil Mullainathan. 2012. “Do judges vary in their treatment of race?” *The Journal of Legal Studies* 41 (2): 347–383.
- Alesina, Alberto, and Eliana La Ferrara. 2014. “A Test of Racial Bias in Capital Sentencing.” *American Economic Review* 104 (11): 3397–3433.
- Alt, James, Ethan Bueno de Mesquita, and Shanna Rose. 2011. “Disentangling accountability and competence in elections: Evidence from US term limits.” *The Journal of Politics* 73 (1): 171–186.
- Angrist, Joshua D., and Miikka Rokkanen. 2015. “Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff.” *Journal of the American Statistical Association* 110 (512): 1331–1344.
- Bayer, Patrick, Randi Hjalmarsson, and Shamena Anwar. 2012. “The Impact of Jury Race in Criminal Trials.” *The Quarterly Journal of Economics* 127 (2): 1017–1055.
- Besley, Timothy, and A Abigail Payne. 2013. “Implementation of anti-discrimination policy: does judicial selection matter?” *American Law and Economics Review* 15 (1): 212–251.
- Besley, Timothy, and Anne Case. 1995. “Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits.” *The Quarterly Journal of Economics* 110 (3): 769–798.
- Brace, Paul, and Melinda Gann Hall. 1995. “Studying courts comparatively: The view from the American states.” *Political Research Quarterly* 48 (1): 5–29.
- Bronson, Jennifer, and E Ann Carson. 2019. “Prisoners in 2017.” *Age* 500: 400.

Butler, Paul. July 11, 2016. “Judicial Recall Will Inevitably Lead to Harsher Sentences.”.

URL: <https://www.nytimes.com/roomfordebate/2016/06/08/should-an-unpopular-sentence-in-the-stanford-rape-case-cost-a-judge-his-job/judicial-recall-will-inevitably-lead-to-harsher-sentences>

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. “Robust nonparametric confidence intervals for regression-discontinuity designs.” *Econometrica* 82 (6): 2295–2326.

Canes-Wrone, Brandice, Michael C. Herron, and Kenneth W. Shotts. 2001. “Leadership and Pandering: A Theory of Executive Policymaking.” *American Journal of Political Science* 45 (3): 532–550.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. 2011. “Extraneous factors in judicial decisions.” *Proceedings of the National Academy of Sciences* 108 (17): 6889–6892.

Eren, Ozkan, and Naci Mocan. 2018. “Emotional Judges and Unlucky Juveniles.” *American Economic Journal: Applied Economics* 10 (3): 171–205.

Ferraz, Claudio, and Frederico Finan. 2011. “Electoral accountability and corruption: Evidence from the audits of local governments.” *American Economic Review* 101 (4): 1274–1311.

Gelman, Andrew, and Guido Imbens. 2019. “Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs.” *Journal of Business & Economic Statistics* 37 (3): 447–456.

Gersen, Jeannie Suk. June 17, 2016. “The Unintended Consequences of the Stanford Rape-Case Recall.”.

URL: <https://www.newyorker.com/news/news-desk/the-unintended-consequences-of-the-stanford-rape-case-recall>

- Gordon, Sanford C., and Gregory A. Huber. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2: 107–138.
- Grogger, Jeffrey, and Greg Ridgeway. 2006. "Testing for racial profiling in traffic stops from behind a veil of darkness." *Journal of the American Statistical Association* 101 (475): 878–887.
- Harris, Allison P, and Maya Sen. 2019. "Bias and judging." *Annual Review of Political Science* .
- Hausman, Catherine, and David S Rapson. 2018. "Regression discontinuity in time: Considerations for empirical applications." *Annual Review of Resource Economics* 10: 533–552.
- Ho, Daniel E, and Kosuke Imai. 2006. "Randomization Inference With Natural Experiments." *Journal of the American Statistical Association* 101 (475): 888–900.
- Huber, Gregory A., and Sanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind When It Runs for Office?" *American Journal of Political Science* 48: 247–263.
- Kousser, Thad, Jeffrey B Lewis, and Seth E Masket. 2007. "Ideological adaptation? The survival instinct of threatened legislators." *The Journal of Politics* 69 (3): 828–843.
- Law Professors Statement. August 15, 2017. "Law Professors' Statement for the Independence of the Judiciary and Against the Recall of Santa Clara County Superior Court Judge Aaron Persky."
- URL:** https://www.sfchronicle.com/file/232/4/2324-89_names_Law_Professors_Statement.pdf
- Lim, Claire S. H. 2013. "Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges." *American Economic Review* 103 (4): 1360–97.

- Masket, Seth E. 2011. "The Circus That Wasn't: The Republican Party's Quest for Order in California's 2003 Gubernatorial Recall Election." *State Politics & Policy Quarterly* 11 (2): 123–147.
- Matsusaka, John G et al. 2010. "Popular control of public policy: A quantitative approach." *Quarterly Journal of Political Science* 5 (2): 133–167.
- Morton, Rebecca. 2006. *Analyzing Elections*. WW Norton.
- Park, Kyung H. 2017. "The Impact of Judicial Elections in the Sentencing of Black Crime." *Journal of Human Resources* 52 (4): 998–1031.
- Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122: 1320–1354.
- Santa Clara County Association. June 14, 2016. "SCCBA Statement on Judicial Independence." .
URL: <https://sccba.site-ym.com/blogpost/1133925/249782/SCCBA-Statement-on-Judicial-Independence>
- Segura, Gary M, and Luis R Fraga. 2008. "Race and the recall: Racial and ethnic polarization in the California recall election." *American Journal of Political Science* 52 (2): 421–435.
- Shaw, Daron, Mark J McKenzie, and Jeffrey Underwood. 2005. "Strategic voting in the California recall election." *American Politics Research* 33 (2): 216–245.
- Smith, Malcolm. 1951. "The California Method of Selecting Judges." *Stanford Law Review* 3 (4): 571–600.
- Thompson, Don. 2016. "2014 Crime Measure Triggers Fewer Arrests." *San Bernardino Sun* p. A11.

Woolf, Nicky. June 24, 2016. “Stanford sexual assault: public defenders support judge in open letter.”.

URL: *<https://www.theguardian.com/us-news/2016/jun/24/stanford-sexual-assault-public-defenders-support-judge>*

Appendix: Additional Results

Table A.1 Descriptive Statistics

	All		Pre-&Post-Petition Announcement		Pre-&Post-Recall Election	
	Mean	SD	Mean	SD	Mean	SD
Sentencing Characteristics						
Sentence Length days	537.856	578.611	512.08	516.668	567.125	640.788
Uncensored Normalized Sentence	0.415	0.409	0.402	0.374	0.430	0.445
Normalized Sentence	0.388	0.336	0.384	0.331	0.392	0.342
Charge Characteristics						
Sex Crime	0.061	0.239	0.031	0.174	0.095	0.293
Violent non-sex Crime	0.716	0.451	0.739	0.439	0.689	0.463
Non-violent Crime	0.223	0.417	0.230	0.421	0.216	0.412
Defendant Characteristics						
Black	0.211	0.408	0.196	0.397	0.228	0.420
Hispanic	0.460	0.499	0.483	0.500	0.434	0.496
White	0.291	0.454	0.282	0.450	0.302	0.459
Male	0.859	0.348	0.854	0.354	0.864	0.342
Age	35.877	10.746	36.128	10.392	35.591	11.133
Num. Charges	22,111	22,111	1,595	1,595	1,567	1,567
Num. Cases	19,798	19,798	1,445	1,445	1,368	1,368
Num. Defendants	18,370	18,370	1,343	1,343	1,271	1,271

Columns 3-4 report statistics for the sample restricted to within 45 days of the petition announcement date. Columns 5-6 report statistics for the sample restricted to within 45 days of the recall election date.

Table A.2 Replication of Main Analysis Restricted to Pre-Breakpoint Cases

	Petition Announced		Recall Election	
RD estimate	0.091 (0.042)	0.107 (0.041)	0.238 (0.236)	-0.686 (0.446)
Bandwidth	36.4	42.3	69.1	63.1
Judge fixed effects	N	Y	N	Y
Statute fixed effects	N	Y	N	Y
Effective observations	1,150	1,167	113	103

Estimates employ triangular kernel. Standard errors clustered at the judge-charge level.

Table A.3 Replication of Main Analysis Using Uncensored Normalized Sentences as Outcome

	Petition Announced		Recall Election	
RD estimate	0.088 (0.043)	0.11 (0.041)	-0.047 (0.058)	-0.036 (0.059)
Bandwidth	45.3	48	40.1	39.4
Judge fixed effects	N	Y	N	Y
Statute fixed effects	N	Y	N	Y
Effective observations	1,613	1,444	1,375	1,166

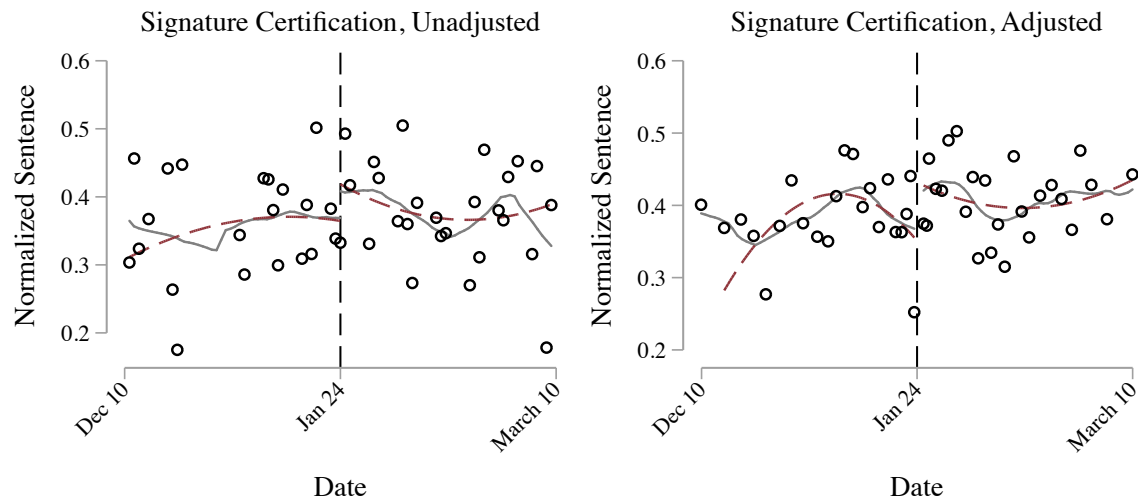
Estimates employ triangular kernel. Standard errors clustered at the judge-charge level.

Table A.4 Replication of Main Analysis Using Non-Normalized Sentence Length as Outcome

	Petition Announced		Recall Election	
RD estimate	119.539 (58.02)	121.83 (54.626)	-97.056 (84.497)	-47.804 (78.868)
Bandwidth	38.3	41.6	41.3	38.9
Judge fixed effects	N	Y	N	Y
Statute fixed effects	N	Y	N	Y
Effective observations	1,371	1,237	1,450	1,158

Estimates employ triangular kernel. Standard errors clustered at the judge-charge level.

Figure A.1 Effect on Sentencing of Signature Certification Date in Persky Recall: Graphical Analysis



Tokens in each panel depict average normalized displays average normalized sentence length in equally-sized bins; local polynomial smoothers fit separately on each side of the event under consideration.