

# A Silent Corrupting Force?

## Judicial Recall Elections and Criminal Sentencing

Sanford C. Gordon\*      Sidak Yntiso†

December 2019

### Abstract

Does the threat of recall effect the behavior of elected officials? In the United States, 39 states authorize recall elections at some level for elected officials. Nonetheless, the electoral incentives of recall elections are not well-understood. We explore how changes in the threat of recall salience alter the sentencing behavior of judges. In 2016, national news coverage following the conviction and sentencing of Brock Turner, a Stanford athlete, for sexual assault sparked a petition drive to recall the sentencing judge, Aaron Persky. Critics of Persky, pointing to the apparent leniency of the prescribed sentence, accused the judge of bias in favor of an affluent white male defendant. The recall drive was ultimately successful, with Persky losing his seat by a wide margin in 2018. Using data on almost 20 thousand sentences from Superior Courts in six California counties, we examine whether critical points in the recall campaign were accompanied by corresponding changes in judicial punitiveness. Our analysis suggests a discontinuous increase in sentencing stringency associated with the initial petition announcement, but not with either the petition's certification or the recall itself. We document heterogeneity in the magnitude of the effect across felony classes, but not defendant race. Our findings imply that judges updated their beliefs about the electoral consequences of their decisions in response to the possibility of a recall, and not the recall itself. They also point to the possibility that specific political actions may have unintended consequences reaching beyond their intended targets.

---

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

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

# 1 Introduction

On June 5, 2018, 61.5% of voters in Santa Clara County, CA took the highly unusual step of voting to recall Superior Court Judge Aaron Persky – that is, to remove from office prior to the expiration of his term. Two years earlier, Persky had achieved notoriety for imposing an unusually lenient sentence on an affluent, white Stanford student athlete convicted of two counts of sexual assault and one count of attempted rape. 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 affects 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 recall effort, even while condemning the leniency of the sentence that instigated it.

This paper addresses whether the recall effort exerted an incentive effect on the subsequent behavior of California judges, with particular focus on their sentencing decisions. While the immediate impetus is assessing the unintended consequences of the Persky recall itself, the research conducted herein speaks to broader questions of electoral accountability and judicial impartiality. Electoral incentives are a defining feature of democratic governance and have implications both for accountability and responsiveness. While the profession of judge is associated with the objective of the neutral application of law, judges are human and thus may be subject to numerous idiosyncratic factors that may make them more or less lenient. At the same time, lower court judges in 39 states in the U.S. face some sort of periodic review by voters, and an extensive research agenda has documented that variation in electoral incentives brought about by electoral proximity, institutions, and competitiveness account for differences in variation in the punitiveness of judges (Huber and Gordon, 2004; Gordon and Huber, 2007; Lim, 2013).

The specific effect of recall elections may contribute to the economy of incentives facing public officials, including judges. As we shall see, however, assessing the incentive effects of recall elections *in particular* is challenging under ordinary circumstances, owing to incommensurability problems with cross-sectional comparisons (across offices or states), difficulty in gauging exposure to treatment, and statistical power issues. Additionally, if public officials rationally anticipate the

consequences of a recall threat, they may take pains to avoid it, thus complicating attempts to connect recalls to official behavior.

To circumvent these difficulties, we examine the effect of three potential shocks to the salience of the recall threat brought about by the Persky recall initiative on the sentencing behavior of judges: the initial announcement of the recall petition; the official signature certification by the Santa Clara County Registrar; and the recall election itself. Specifically, using a regression discontinuity in time (RDit) approach (Hausman and Rapson, 2018), we first examine the effects of these shocks on sentencing in over 19,000 felony cases from 2016-2018 by 157 Superior Court judges in six California counties scraped from the web portals of those counties' superior courts.

In this paper, we examine the effect a shock to the salience of the recall threat brought by Judge Persky's recall election on sentencing in California courts.<sup>1</sup> We collected felony sentencing decisions issued by 157 Superior Court judges between January 2015- December 2018. Our data include charging information, defendant characteristics and initial sentencing dispositions for cases in Fresno, Napa, Santa Barbara, San Bernardino, and Santa Cruz counties.

Our baseline regression discontinuity in time specification implies 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. We find that the effect is driven by increases in sentencing on violent crimes not of a sexual nature, and not by sentencing on sex crimes or nonviolent crimes; and find no evidence of disparities in the effect of the petition announcement on sentencing by defendant race.

By contrast, we find no evidence that two other critical events associated with the Persky recall – the certification of petition signatures and the recall election itself – induced change in sentencing. These null findings suggest that these latter two events conveyed no additional information concerning the recall threat beyond that communicated to judges by the initial petition announcement.

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 five counties for which we have data.

---

<sup>1</sup>We wish to strongly caution against interpreting any empirical findings as speaking to the appropriateness or inappropriateness of the specific sentence that motivated Judge Persky's recall.

These findings contribute to a literature on recall elections (Dee 2007; Segura and Fraga 2008; Alvarez and Kiewiet 2009; Ho and Imai 2006*b*; Masket 2011; Shaw, McKenzie and Underwood 2005) as well as the effects of direct democracy writ large (Matsusaka 2005; Matsusaka et al. 2010). We document evidence of strong incentive effects of direct democracy on the behavior of elected officials. The findings also shed light on the mechanisms through which these incentive effects operate (Matsusaka 2014). In addition to a direct influence, direct democracy can have an indirect influence on elected officials either by inducing incentives for preemptive action (Gerber 1996; Matsusaka and McCarty 2001) or by revealing new policy-relevant information. Our event-level strategy allow us to identify an independent indirect effect of the recall process that is at least as large as the direct effect.

We also provide novel evidence on the impact of electoral incentives on judges. Evidence from contested judicial elections indicates electoral cycles in criminal sentencing severity consistent with models wherein retention incentives increase electoral accountability (Huber and Gordon 2004; Gordon and Huber 2007; Berdejó and Yuchtman 2013; Lim 2013). In these models, judges trade-off personal preferences about sentencing against the negative electoral response to lenient sentencing. However, beyond variation in electoral competition (Gordon and Huber 2007; Dippel and Poyker 2019), few scholars have investigated the factors underlying judges’ perceptions of the electoral response. We provide evidence that judges appear to rationally update perceptions in response to threats of electoral sanction.

## 2 Background

### 2.1 Electoral Incentives and Incumbent Behavior

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 empirically is the nature of the informational environment in which they occur. Voters often lack verifiable information to evaluate judicial performance, a

problem further complicated by the fact that judges often face retention elections (in which there are no challengers), nonpartisan elections (in which voters lack clear cues such as party labels) and/or uncompetitive elections. 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 “headline risk” is asymmetric, owing to the information potentially available to voters. 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.

## 2.2 Recall Elections

**Assessing the Incentive Effects of Recall Elections.** 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 as 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.<sup>2</sup> 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.<sup>3</sup>

Since 1995, the earliest year for which we have 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

---

<sup>2</sup>Source: Complete List of Recall Attempts, California Secretary of State. Available at <https://www.sos.ca.gov/elections/recalls/complete-list-recall-attempts/>.

<sup>3</sup>Two of the five recalled state legislators were Republican state Assembly members removed by Republicans in heavily Republican districts, ostensibly for comprising 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>

successful.<sup>4</sup> Figure 1 displays qualified attempts and successes over time; as is evident from the figure, every year voters recall some local officials in California. That being said, the data suggest that no superior court judge had ever been recalled until the Persky case.

To our knowledge, there exists no extant empirical research on the incentive effects of recall elections.<sup>5</sup> 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 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.

**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 Stanford student athlete, sexually assaulted Chanel Miller,<sup>6</sup> 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

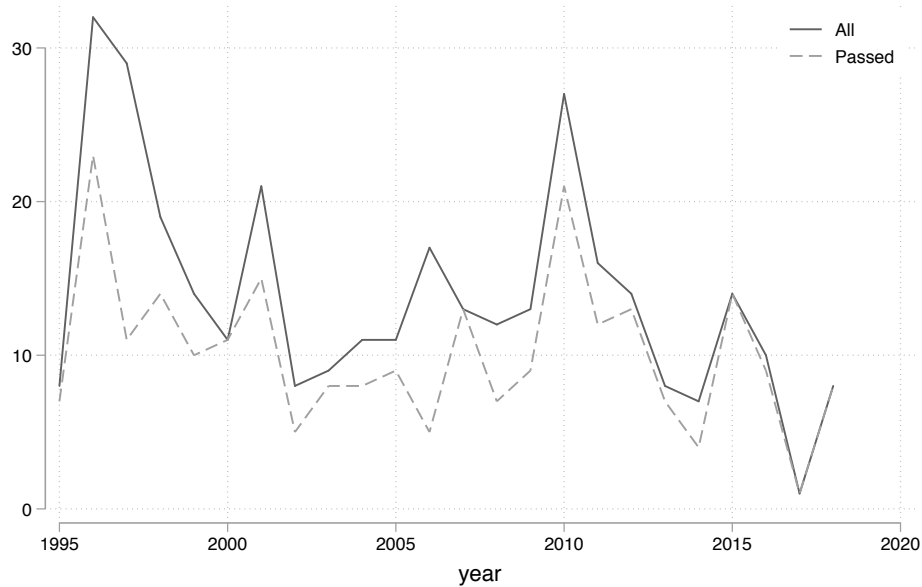
---

<sup>4</sup>Source: California Election Data Archive, available at <http://hdl.handle.net/10211.3/210187>.

<sup>5</sup>The political science research on recall elections has focused on voter behavior in recall elections – see, e.g., Ho and Imai (2006a); Segura and Fraga (2008); Masket (2011); Shaw, McKenzie and Underwood (2005).

<sup>6</sup>While 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*).

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



and Miller’s impact statement, published by BuzzFeed, sparked widespread national attention.<sup>7</sup> On June 6, 2016, Stanford Law School Professor Michele Dauber announced the formation 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.

Criticisms of the recall campaign were immediate and widespread. 95 California law professors signed an open letter to Dauber in August 2017 opposing the recall petition. Californian mayors, state legislators, former Supreme Court justices, and hundreds of Superior Court judges supported the Retain Judge Persky Campaign.<sup>8</sup> Critics were primarily concerned with judicial independence and impartiality (California Public Defenders 2016; Law Professors Statement 2017). Some critics also predicted an increase in judicial punitiveness, with disproportionate effects on minority defendants (Butler 2016, Pauly 2017, Gersen 2016). These predictions were bolstered by the empirical literature, cited above, documenting how concerns with reelection induce trial judges to increase the likelihood of and average length of incarceration; as well as the significant literature on the dis-

<sup>7</sup><https://www.buzzfeednews.com/article/katiejmbaker/heres-the-powerful-letter-the-stanford-victim-read-to-her-ra>

<sup>8</sup>The Voices Against Recall website has since been removed. An archived version is available here: <https://web.archive.org/web/20180423164925/http://www.voicesagainstrerecall.org/>

proportionate burden imposed by the criminal justice system on minority defendants (e.g., Grogger and Ridgeway, 2006; Bayer, Hjalmarsson and Anwar, 2012; Park, 2014; Alesina and La Ferrara, 2014; Abrams, Bertrand and Mullainathan, 2012).

### 2.3 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.<sup>9</sup> 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 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, for a variety of reasons, few sentences precisely match the three prescribed base terms. 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.<sup>10</sup> As is generally the case in the United States, the vast majority of cases are resolved via plea bargain (subject to the approval of the presiding judge).

---

<sup>9</sup>[https://www.courts.ca.gov/documents/California\\_Judicial\\_Branch.pdf](https://www.courts.ca.gov/documents/California_Judicial_Branch.pdf)

<sup>10</sup>Effective 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, 3 years).



## 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 handful of California 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 19,846 cases encompassing 22,111 felony charges with initial sentencing dispositions in the remaining six courts.<sup>11</sup> 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.<sup>12</sup> Additional case information in our final dataset include the charge (410 unique offenses) and sentencing judge (157 unique judges). Combining information on defendant first and last name with data on county prison demographics from the California Sentencing Institute<sup>13</sup> permits us to calculate Bayesian posterior probabilities of defendant race (Voicu, 2018). Of the 13,033 charges for which non-guilty plea/ plea bargaining status could be identified, 72% were plea bargained. We categorized crimes as nonviolent or violent based on offense codes from the California Attorney General: 73% of charges in the sample are classified as violent.

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.<sup>14</sup> Defendants were matched on first name, last name, county of arrest and arrest date. Across the six counties, 9,000 defendants could be linked across both data-sets.

---

<sup>11</sup>Sentences may be amended – for example, in cases of probation violations.

<sup>12</sup><https://oag.ca.gov/law/code-tables>

<sup>13</sup><http://casi.cjcj.org/>.

<sup>14</sup>We scraped arrest records from this source: <https://www.localcrimenews.com/>

### 3.2 Empirical Approach

In the main part of our analysis, we look for sharp increases in judicial punitiveness immediately following key events during the recall campaign. In particular, we consider three dates in the recall campaign: the announcement of campaign itself, on June 6, 2016; the signature certification by the Santa Clara Registrar, on January 24, 2018; and the recall election itself, on June 5, 2018. 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 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 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. As is standard in RD designs, we weight using a triangular kernel. Finally, we cluster standard errors at the judge and offense code to account for within-judge and within-charge correlation of sentences. 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 alleviate 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). We

also examine case characteristics to ensure that there are not coincident potential shifts in the composition of cases heard.

Next, we explore the heterogeneity of observed effects. In particular, we examine whether any observed unconditional effects are driven by violent or nonviolent crimes.<sup>15</sup> We also examine whether effects differ conditioning on the posterior probability of a defendant is nonwhite.

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 effects of any 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 Main Results: Instantaneous Effects of Critical Events

### 4.1 Graphical Evidence

Before proceeding to our local linear estimation, our first step in assessing the effect of the three 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 45 days before and after each event.<sup>16</sup> A local polynomial is fit separately on either side of the event date.

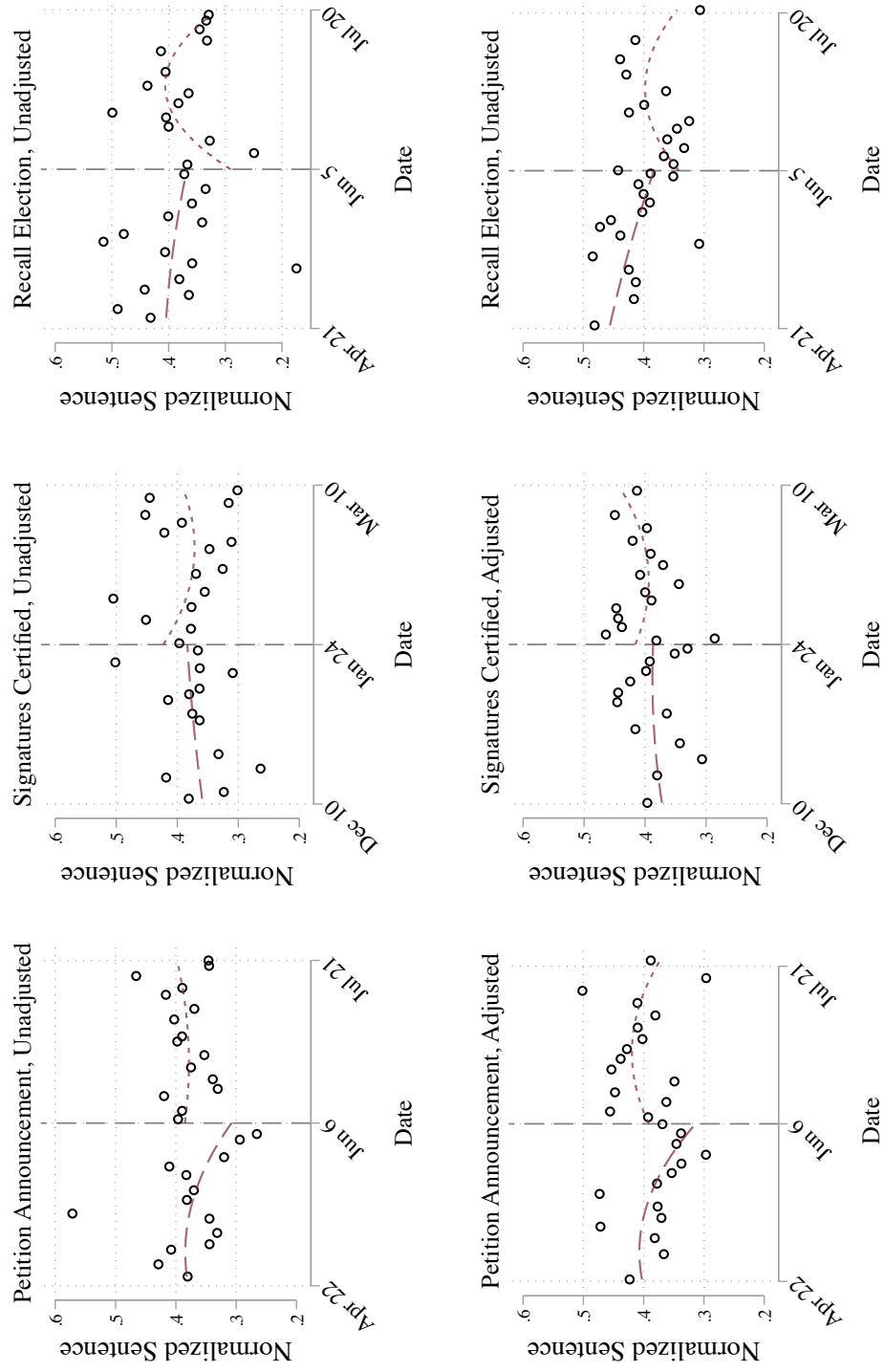
The three panels in the top row present plots of the raw data. We observe a large, discontinuous increase in average normalized sentence transitioning from immediately prior to, to immediately following the June 6, 2016 petition announcement. This increase corresponds to an increase of approximately 10 percentage points, reflecting a proportionate increase of approximately 29 percent. Turning to January 24, 2018, the date on which the county election board certified the signatures

---

<sup>15</sup>We lack sufficient data to disaggregate further, e.g., to sexual assault cases.

<sup>16</sup>The  $\pm 45$  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 displays average normalized sentence length in equally-sized bins; local polynomial smoothers fit separately on each side of the event under consideration.*

on the petition, thus ensuring that the issue would go to the voters, we see effectively no change in average sentence length immediately before and immediately after that date. Finally, we observe a slight decrease in sentencing associated with the recall election itself on June 5, 2018.

Plots in the bottom row depict binned means residualized by 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 observe a slight uptick in sentencing for the signature certification date (of approximately four percentage points), and no difference before and after the election itself.

## 4.2 Local Linear Regression Results

To explore the preliminary inferences suggested in the graphical analysis in a more rigorous fashion, we next present local linear regression estimates of the local average treatment effect,  $\beta_1$  from equation (1). These 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 suggestion of Gelman and Imbens (2019), we report results from a local linear specification rather than estimating higher-order polynomials.

Estimates in the table corroborate the results from the graphical analysis. We estimate a large, statistically significant effect of the June 6 petition announcement: unadjusted, the estimated effect is 9.2 percentage points; adjusted for judge- and offense-specific fixed effects, the estimate increases to 11.0%. To give a sense of the substantive significance of these estimates, immediately prior to the announcement, the estimated average normalized sentence length was 0.31; hence, these effects correspond to an immediate proportionate increase of 29.7 to 35.5 percent. Both estimates easily surpass conventional thresholds for statistical significance.

By contrast, for neither of the other two critical dates, whether adjusted or unadjusted for the judge and offense-specific fixed effects, are estimates statistically distinguishable from zero. As noted above, this is consistent with an account in which judges fully internalized the novel threat of recall at the time the petition was announced.

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

	Petition Announced		Signatures Certified		Recall Election	
RD estimate	0.092 (0.038)	0.11 (0.039)	-0.045 (0.042)	0.014 (0.045)	-0.053 (0.05)	-0.02 (0.051)
Bandwidth	44.2	44.2	44.5	44.5	41.5	41.5
Defendant demographics	N	Y	N	Y	N	Y
Judge fixed effects	N	Y	N	Y	N	Y
Offense fixed effects	N	Y	N	Y	N	Y
Effective observations	1,365	1,365	1,324	1,324	1,281	1,281

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

### 4.3 Robustness of Announcement Finding

**Temporal confounding.** Our main analysis causally suggests 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 structural 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 analysis including judge- and charge-fixed effects.<sup>17</sup> Figure 3 displays the results. The left panel displays the empirical cumulative distribution of p-values for the placebo tests (in blue), along with the June 6 p-value (in red). The figure shows that the estimated June 6 p-value is lower than 98.6% 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 (in red). There are two things to specifically note about the 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).

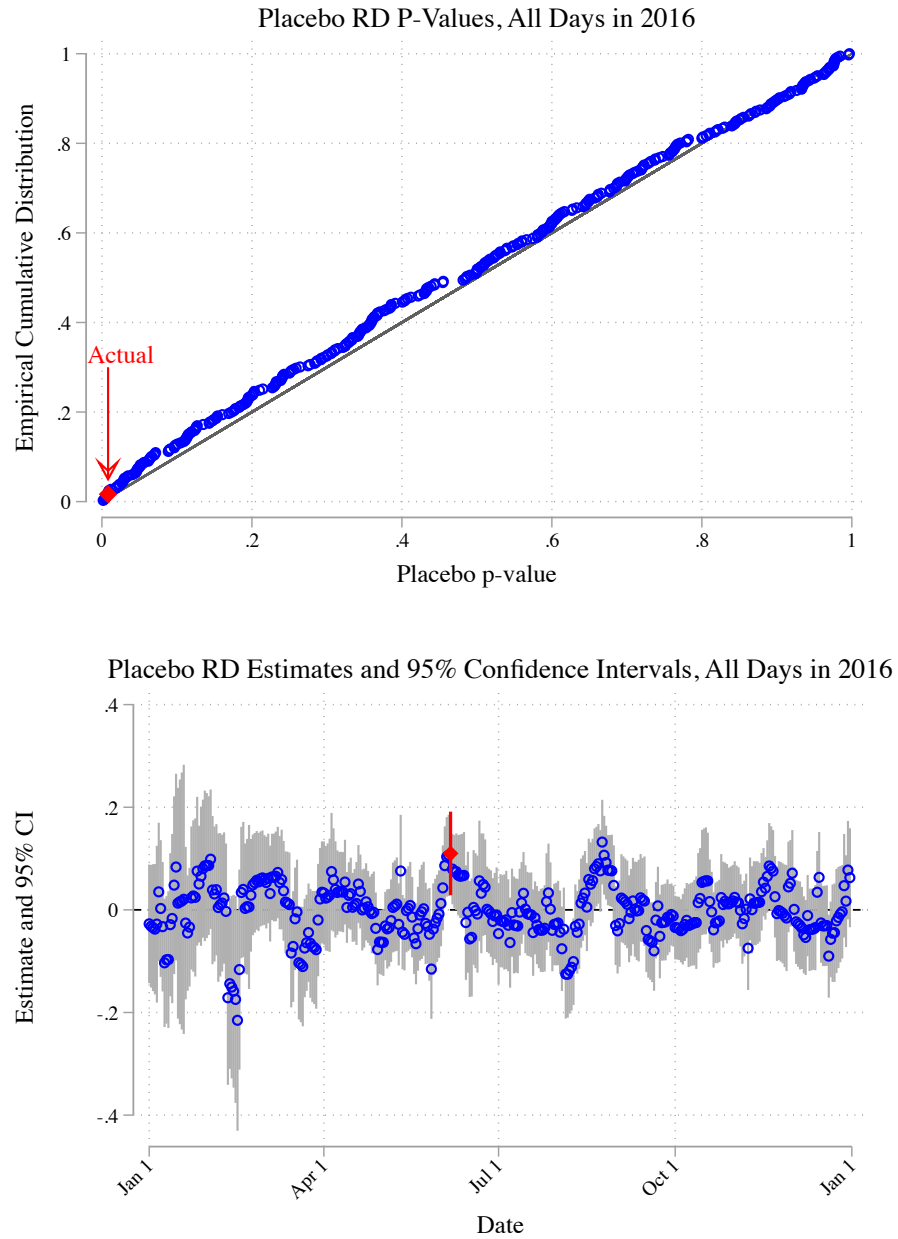
Second, looking at the neighborhood around June 6, we see that the estimate for that specific date is larger than any of the surrounding placebo estimates – including June 7 and June 2. In fact, the placebo estimate for the date of the Brock Turner sentence is statistically indistinguishable from zero.

**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.

---

<sup>17</sup>Placebo tests using unadjusted RD-estimates produce substantively identical conclusions.

Figure 3 Placebo Tests for Main Effect of Petition Announcement



*The top panel displays the empirical cumulative distribution of estimated placebo p-values (in blue), with the actual petition announcement p-value overlaid in red. The bottom panel displays placebo RD estimates and associated 95% confidence intervals (in blue), with the actual estimate and confidence interval in red.*



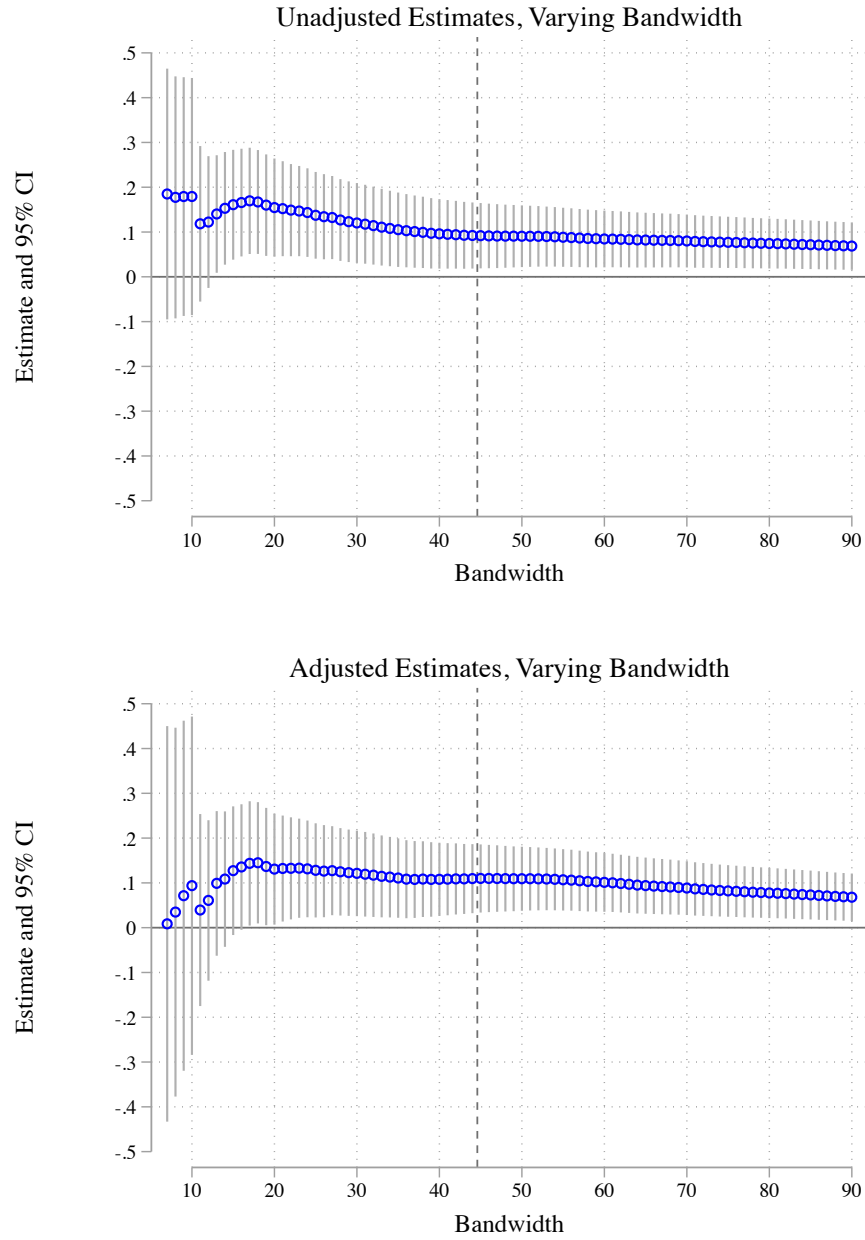
Results of this exercise appear in Figure 4. For very short bandwidths, of course, the sample size declines dramatically, substantially diminishing the precision of the estimates. However, past around a two-week bandwidth for both adjusted and unadjusted specifications, our main results are robust to 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.

**Charge adjustment.** An additional threat to influence 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 signature certification and 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.1 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.2 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

Figure 4 RD Estimates Varying Bandwidth



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

significant increases in sentencing associated with the petition announcement, and no significant change associated with either the signature certification

Tables A.3 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.

## 5 Heterogeneous Effects

### 5.1 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 what 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 local linear estimator (unadjusted and adjusted for judge and offense fixed effects) separately for each of the three categories. Results appear in Table ??.

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, each suggesting an approximately 16 percentage point immediate increase in sentencing associated with the petition announcement. Relative to a baseline normalized sentence for non-sexual violent crimes of 0.26 immediately prior to the announcement, this

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

	Sex Crimes		Other Violent Crimes		Nonviolent Crimes	
RD estimate	-0.036 (0.172)	0.348 (0.337)	0.155 (0.045)	0.159 (0.05)	-0.032 (0.076)	-0.038 (0.082)
Bandwidth	66	66	39.2	39.2	54.9	54.9
Defendant demographics	N	Y	N	Y	N	Y
Judge fixed effects	N	Y	N	Y	N	Y
Statute fixed effects	N	Y	N	Y	N	Y
Effective observations	57	57	880	880	390	390

*See notes in Table 1 for estimation details.*

represents an immediate 59 percent proportionate increase in punitiveness.

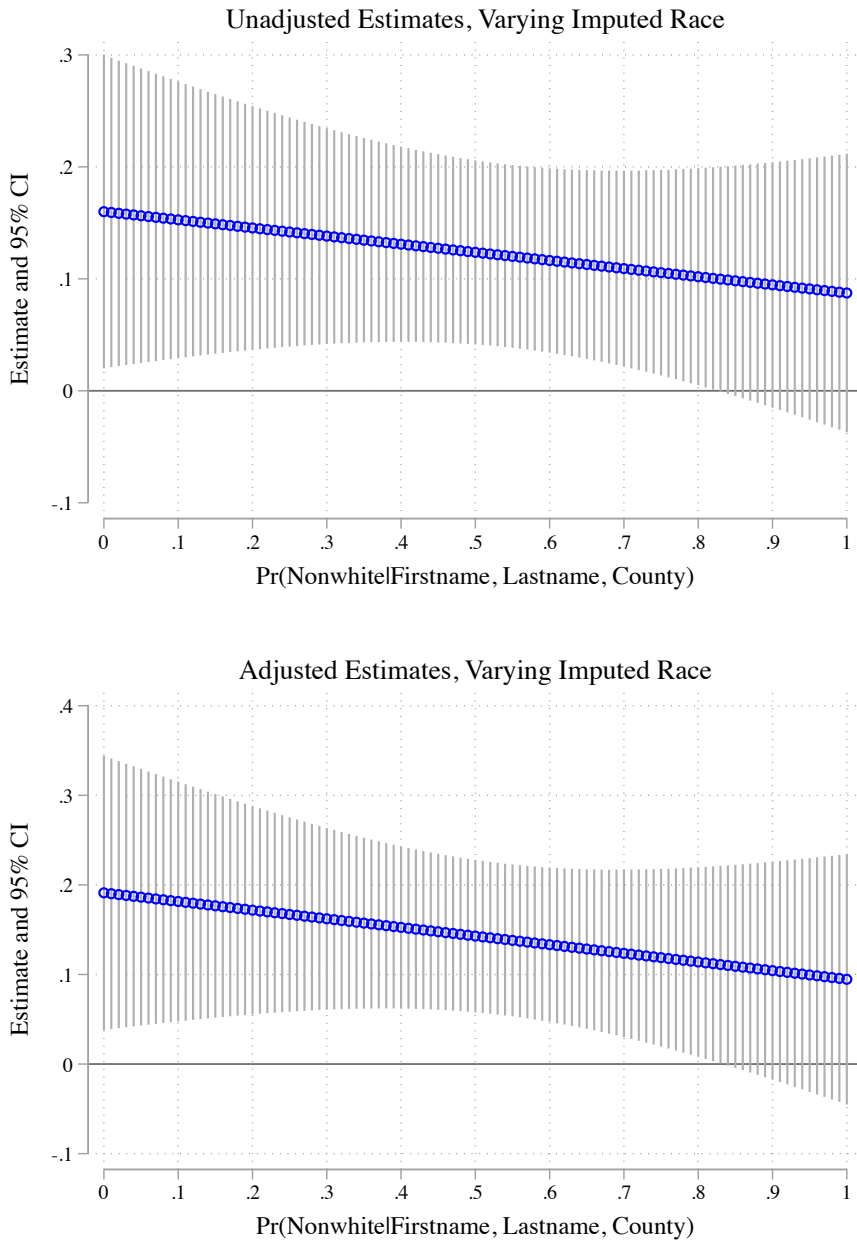
## 5.2 Effects by Race of Defendant

As noted above, critics of the recall campaign anticipated that, notwithstanding the aim of sanctioning a lenient sentence given to a white defendant, any increase in judicial punitiveness driven by the recall itself would likely be disproportionately borne by minority defendants. Unfortunately, we lack direct access to racial demographic data for individual defendants. However, as described above, we are able to calculate a measure of the posterior probability that a defendant is nonwhite, given the defendant’s first and last name and the county in which they were sentenced.

To determine whether the treatment effect of the petition announcement varies according to defendant race, we re-ran our main analysis, interacting the running variable and announcement indicator with the posterior probability that the defendant is nonwhite. Because models including triple-interaction effects are difficult to interpret, we plot the conditional LATE (unadjusted and adjusted for judge and offense effects) as a function of  $\Pr(\text{nonwhite} | \text{firstname}, \text{lastname}, \text{county})$  in Figure 5.

Surprisingly, the figure suggests that the effect of the petition announcement is, if anything, slightly larger for white than nonwhite defendants. However, for no two value of the posterior probability the defendant nonwhite can we reject the null that the estimate is the same. Hence, the available evidence suggests that the effect of an increase in judicial punitiveness brought about by the petition announcement does not vary by race for the marginal defendant. That being said, note that because minority citizens are disproportionately represented in the ranks of criminal defendants, even a treatment effect that was perfectly constant across defendants would dispropor-

Figure 5 LATE Estimates for Petition Announcement by Probability the Defendant is Nonwhite



tionately effect minority communities.

## 6 Cumulative 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.<sup>18</sup> 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 cumulative 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

---

<sup>18</sup>Using 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.

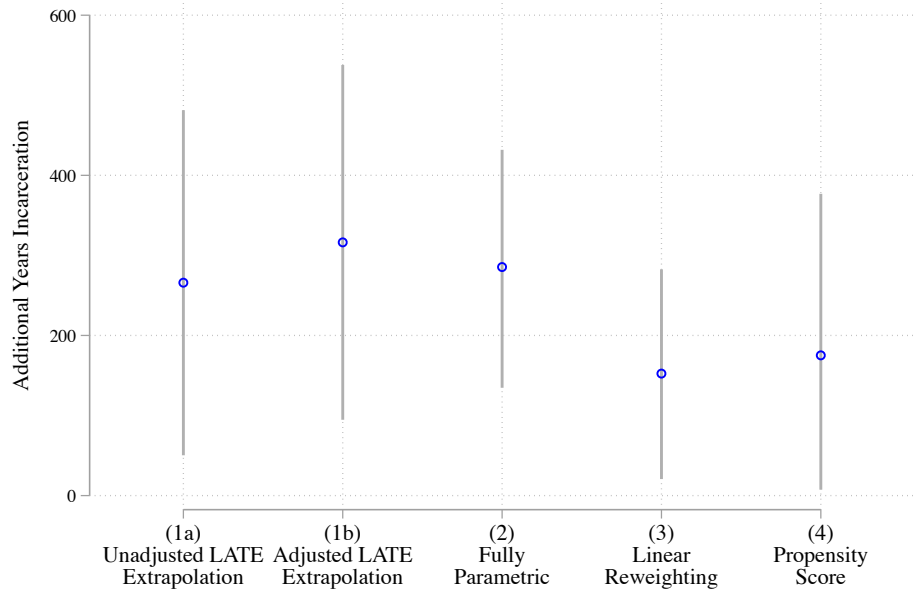
differences, we constrain the trend to zero when predicting counterfactual sentences.

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 in for an interval using either a linear reweighting or propensity score estimator (see Angrist and Rokkanen (2015) for details). We adopt 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 five 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

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



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

consideration, actual costs could be considerably higher.

## 7 Discussion

The research presented in this paper suggests that the campaign to recall Judge Aaron Persky shifted the beliefs of sitting Superior Court Judges about the potential electoral consequences of their sentencing decisions. We document a substantial and immediate increase in sentencing following the announcement of the campaign, and calculate cumulative 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 five counties, these estimates likely substantially underestimate the broader effects of this change in judicial behavior.

At the same time, we document no observable effects of two other critical events that occurred during the recall campaign: the certification of petition signatures by the Santa Clara Registrar and the recall election itself. To the extent that the initial petition announcement reflected a “new



normal” in the political environment of sitting trial judges, the absence of an observed effect for either of these two events is consistent with two possibilities: either judges anticipated the signature threshold being achieved, and Persky’s recall, so the two events conveyed no new information; or the events did not reflect the change in norms signaled by the announcement of the campaign itself.

Our paper sheds light on what might be called the “indirect effects of direct democracy.” The Persky recall was aimed at sanctioning a specific judge for a specific sentence perceived as unjust. But once this remedy is contemplated, it is impossible to control the circumstances under which it may be used to sanction judges for their leniency in the future. Judges in California appear to have anticipated this possibility, as indicated by the fact that the effect of the petition announcement was not confined to white defendants or sexual assault cases.

Finally, our paper provides evidence that recall elections can affect the behavior of elected officials, and so provides an additional contribution to the literature on electoral incentives of public officials more generally.

## 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.
- Alvarez, R Michael and D Roderick Kiewiet. 2009. “Rationality and rationalistic choice in the california recall.” *British Journal of Political Science* 39(2):267–290.
- 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.
- URL:** <https://doi.org/10.1080/01621459.2015.1012259>

- 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.
- Berdej , Carlos and Noam Yuchtman. 2013. “Crime, punishment, and politics: an analysis of political cycles in criminal sentencing.” *Review of Economics and Statistics* 95(3):741–756.
- 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.
- Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2014. “Robust nonparametric confidence intervals for regression-discontinuity designs.” *Econometrica* 82(6):2295–2326.
- Dee, Thomas S. 2007. “Technology and voter intent: Evidence from the California recall election.” *The Review of Economics and Statistics* 89(4):674–683.
- Dippel, Christian and Michael Poyker. 2019. How Common are Electoral Cycles in Criminal Sentencing? Technical report National Bureau of Economic Research.
- 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.  
**URL:** <https://doi.org/10.1080/07350015.2017.1366909>
- Gerber, Elisabeth R. 1996. “Legislative response to the threat of popular initiatives.” *American Journal of Political Science* 40:99–128.
- 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.
- 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. 2006a. "Randomization Inference With Natural Experiments." *Journal of the American Statistical Association* 101(475):888–900.  
**URL:** <https://doi.org/10.1198/0162145050000001258>
- Ho, Daniel E and Kosuke Imai. 2006b. "Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election." *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–.
- 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. 2005. "Direct democracy works." *Journal of Economic Perspectives* 19(2):185–206.
- Matsusaka, John G. 2014. "Disentangling the direct and indirect effects of the initiative process." *Public Choice* 160(3-4):345–366.
- Matsusaka, John G and Nolan M McCarty. 2001. "Political resource allocation: Benefits and costs of voter initiatives." *Journal of Law, Economics, and Organization* 17(2):413–448.
- 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. 2014. "Do Judges Have Tastes for Racial Discrimination? Evidence from Trial Judges."
- 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.
- Thompson, Don. 2016. "2014 Crime Measure Triggers Fewer Arrests."
- Voicu, Ioan. 2018. "Using First Name Information to Improve Race and Ethnicity Classification." *Statistics and Public Policy* 5(1):1–13.
- URL:** <https://doi.org/10.1080/2330443X.2018.1427012>

## 8 Appendix: Additional Results

Table A.1 Replication of Main Analysis Restricted to Pre-Breakpoint Cases

	Petition Announced		Signatures Certified		Recall Election	
RD estimate	0.096 (0.039)	0.116 (0.041)	-0.047 (0.043)	0.005 (0.046)	-0.058 (0.052)	-0.018 (0.055)
Bandwidth	42.4	42.4	43.6	43.6	38.8	38.8
Defendant demographics	N	Y	N	Y	N	Y
Judge fixed effects	N	Y	N	Y	N	Y
Statute fixed effects	N	Y	N	Y	N	Y
Effective observations	1,193	1,193	1,229	1,229	1,146	1,146

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

Table A.2 Replication of Main Analysis Using Uncensored Normalized Sentences as Outcome

	Petition Announced		Signatures Certified		Recall Election	
RD estimate	0.105 (0.042)	0.119 (0.04)	-0.054 (0.05)	0.008 (0.056)	-0.068 (0.059)	-0.045 (0.058)
Bandwidth	47.5	47.5	45.4	45.4	38.8	38.8
Defendant demographics	N	Y	N	Y	N	Y
Judge fixed effects	N	Y	N	Y	N	Y
Statute fixed effects	N	Y	N	Y	N	Y
Effective observations	1,452	1,452	1,324	1,324	1,186	1,186

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

Table A.3 Replication of Main Analysis Using Non-Normalized Sentence Length as Outcome

	Petition Announced		Signatures Certified		Recall Election	
RD estimate	166.161 (60.784)	141.2 (55.987)	-139.899 (75.857)	7.788 (82.533)	-138.075 (91.199)	-64.797 (79.531)
Bandwidth	41.4	41.4	39.6	39.6	39	39
Defendant demographics	N	Y	N	Y	N	Y
Judge fixed effects	N	Y	N	Y	N	Y
Statute fixed effects	N	Y	N	Y	N	Y
Effective observations	1,268	1,268	1,134	1,134	1,206	1,206
<i>Estimates employ triangular kernel. Standard errors clustered at the judge-charge level.</i>						