Appendix A: Data and coding procedures

Data source and processing

The federal corruption prosecution data come from the 2009 edition of the National Caseload Statistical Data (NCSD), an anonymized database that is regularly released by the Offices of the United States Attorneys at the Department of Justice under the Freedom of Information Act.[1] This dataset includes the universe of federal prosecution files and is effectively a snapshot of the DOJ database of cases (including cases filed and closed in previous years) as of the end of the 2009 fiscal year. We retain the non-appellate criminal cases within the fifty states that were categorized by DOJ as pertaining to state, local, or other public corruption (i.e., federal public corruption cases were excluded[2]). We excluded the 171 charges listed as "opened in error." In order to avoid double-counting charges that were either superseded by a new filing or included in another case, we used the record from the final case that included the defendant in question.[3] Legally, public corruption can range from a government employee stealing office supplies to embezzlement and bribery. To focus on cases that are high-profile enough to have the potential for political repercussions, we follow Gordon (2009, 551) and restrict our attention to the cases coded as national priorities or both national and district priorities. We therefore exclude 977 defendants who cases were coded as only district priorities, which Gordon (2009, 551) reports "are typically clerical workers," as well as the 2041 defendants whose cases were coded as neither a national or district priority or whose priority was undetermined.[4]

Defendant identification

Defendants were identified using the Public Access to Court Electronic Records (PACER) website (`www.pacer.gov`), a fee-based service provided by the federal courts to offer public

[1] Gordon (2009) uses data from the Transactional Records Access Clearinghouse and the Bureau of Justice Statistics but these secondary sources should be drawn from the raw data we accessed directly.

[2] It is of course possible that U.S. attorneys are also biased in deciding whether to prosecute cases of federal public corruption that could damage their own party. However, few members of the opposition party are presumably charged in such cases and we therefore do not examine them here.

[3] The record includes each defendant's full case history. The charges that were eventually superseded are visible in the later case and are still used when analyzing the initial charges filed against the defendant.

[4] We were concerned that some districts did not appear to use the national or national/district priority codes. As a validation step, we coded all 250 defendants from this group who were charged within 24 weeks of an election in a district that did not use the national or national/district priority codes for any public corruption defendants during an entire presidential administration (either Clinton or Bush). All but twenty of these 250 defendants were charged in New Jersey during the Bush years when then-U.S. attorney Chris Christie launched an unprecedented anti-corruption crusade (e.g., Sampson 2007). Of these, 80 were partisans and all were from New Jersey during the Bush administration. However, we observe no clear partisan patterns in case timing or severity around elections, which may reflect the intense scrutiny that Christie received due to allegations that his efforts were politically motivated (e.g., Conte 2012).

access to electronic court records. Research assistants initially searched PACER for cases in which the United States was a party that were filed within two calendar days of the case filing date provided in the Department of Justice (DOJ) data. If no matches were found, they expanded the window to four days on either side of the case filing date.

They then matched cases in the DOJ data to PACER using the case filing date, the number and type of charges against the defendant, the case closing date (if any), and the punishment (fine amount and/or months of probation/incarceration). Additional steps were taken to match defendants in the DOJ data to PACER records in multiple defendant cases, including using separate spreadsheets to record information from PACER on all defendants and then match them to the DOJ records.

Matches were allowed when minor discrepancies existed between the DOJ and PACER data if the PACER defendant data matched the DOJ defendant data on at least two identifying variables either and no other defendant in PACER did so. When too many discrepancies existed or a match could not be found, the identity of the defendant in the DOJ data was coded as missing. Minor date variation (e.g., five days or less) between the DOJ and PACER data was considered to reflect normal bureaucratic imprecision and delays in data entry. Charge/count variation between the DOJ and PACER data sometimes occurs and appears to reflect differences between charges filed (PACER) and those sustained (DOJ at least in some cases). The fact that a charge is listed in PACER but not DOJ is therefore relatively common.

A second research assistant blindly double-coded the most difficult cases, including multiple defendant cases and those for which defendants matched on two identifying variables, and resolved any discrepancies with the first coder and/or the authors to ensure that defendants were matched properly.

Due to a lack of case summary information in PACER, it was not possible to identify defendants in the following districts: California Central, Indiana South, Louisiana Middle, Nevada, New York East, Oregon, Texas West, and Virginia West. A lack of case summaries also precluded defendant identification for cases filed between December 16, 1993 and July 20, 1995 in Maryland.

After matching the defendant, research assistants copied and pasted a series of fields from the PACER case summary into the data.

Defendant partisanship

Research assistants searched for the defendant in Lexis-Nexis Academic, Google, Google News, Proquest, and the list of federal candidates compiled by Open Secrets. When possible, they identified each defendant's job title or position, city, county, state, and the level of government in which they worked: federal government, state executive branch/bureaucracy,

state legislative, local government, private/nonprofit, relative/personal relationship with accused, or a military or postal worker (excluded from federal category).

The research assistants also coded the public partisanship of each defendant and any supervisor, associate, or ally of the defendant who was mentioned in news accounts or official documents about the case using the same data sources used to identify the defendant. When possible, partisan codings were corroborated using data from Gordon (2009). It is important to note that partisan codings do *not* reflect party registration or other private behavior by defendant or their associates. Individuals were only coded as partisans if they were publicly identified as members of the Democratic or Republican party in news accounts or public documents or as associates of prominent partisans.

The data employed in the analysis above classifies as partisans both defendants in public corruption cases who were publicly identified with one of the major parties as well as defendants with ties to prominent partisan figures.

Defendant data: Match rate and reliability

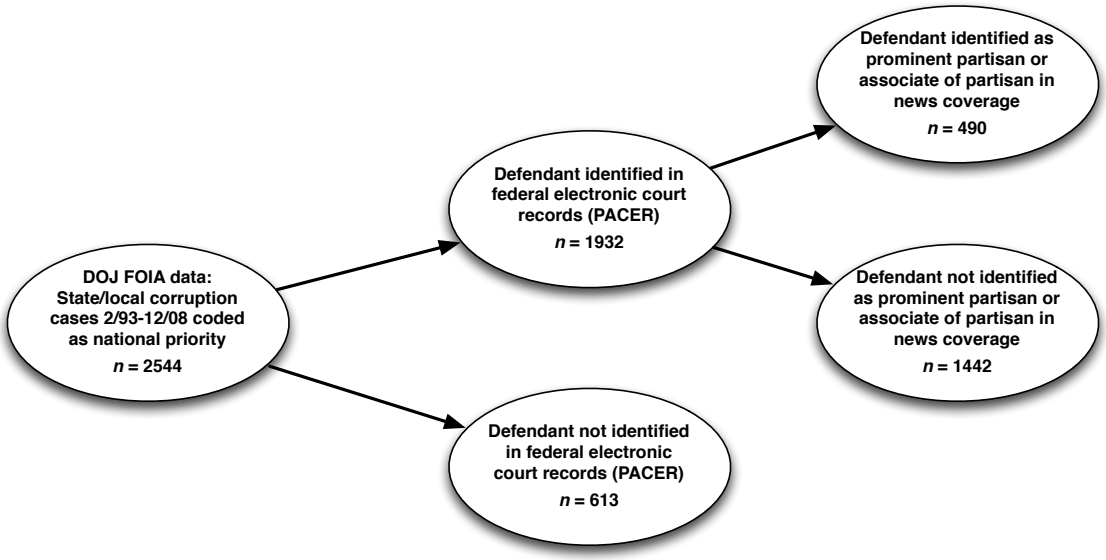
We were able to identify 1932 of the 2545 qualifying defendants (76%) spread across 1177 cases (out of 1334 total). Of the 1932 identified defendants, 490 (25%) were publicly identified with one of the major parties (353 Democrat, 137 Republican) either individually (153), as an associate of a publicly-identified partisan (314), or as both a publicly-identified partisan and an associate of a partisan (23). We illustrate the steps in this process of identifying partisan defendants from the set of qualifying public corruption prosecutions in Figure A1.

As a validation step, we merged our data with the replication files from Gordon (2009) and resolved any unintended discrepancies in defendant identification, party affiliation, or position among partisan defendants. After this step, we matched 94% of his defendants, including 99.4% of the partisans (one defendant appears to be omitted from our DOJ data). The sentencing data corresponds almost perfectly between datasets as well (97% on incarceration and 98% on both probation and fines among matching defendants). Finally, our coding matches Gordon's very closely on party identification (90% of those defendants who match across datasets) and public/private sector positions (95% of matching defendants).[5]

Table A1 compares the distribution of the summary statistics presented in Table 1 between those defendants we were able to identify with those that we could not. We find that the defendants whom we could not identify faced fewer charges and counts and were found guilty less often and of less severe crimes. In particular, the mean sentence for non-identified defendants was only eight months compared for 22 months among those whom we could identify (the median sentences were 0 months and 6 months, respectively), sug-

[5]The remaining differences appear to reflect slight variations in coding rules and procedures.

Figure A1: Partisan defendant identification procedure



Sample: All federal criminal cases targeting state and local public corruption filed by U.S. attorneys between February 1993 and December 2008 and coded as national priorities or national and local priorities.

gesting that the defendants we identified were those who were incriminated in more severe corruption cases.

Election timing

For each case, we calculated the electoral distance variable to the closest election before or after the case filing (i.e., the minimum absolute value), which is the one we expect to be most salient. Since most state elections coincide with federal elections, this variable measures the number of weeks until or since the closest federal election except for a subset of cases in the five states with off-year electoral cycles (Kentucky, Louisiana, Mississippi, New Jersey, and Virginia). For those five states, the closest election was a state gubernatorial or legislative election for 153 of 200 defendants. The resulting electoral distance variable ranges from -365 (a case filed on November 8, 1999 in West Virginia South — approximately one year before the 2000 federal elections) to 366 (two cases, including one filed in Missouri East on November 4, 1993—one year and one day after the 1992 federal elections). As described in the main text, we round our electoral distance variable down to the nearest complete week from the election. This week variable ranges from -52 to 52. Cases filed less than 7 days before and after the from the election were classified as 0.5 and -0.5, respectively.

We also construct placebo election dates on the first Tuesday of November in off-years for

Table A1: Summary statistics

	Non-identified		Identified	
	Mean	SD	Mean	SD
Charge characteristics				
Number of distinct charges	1.670	[1.357]	2.129	[1.399]
Total counts	4.042	[17.64]	5.768	[15.35]
Statutory max: most serious charge (months)	158.9	[76.79]	161.8	[80.79]
Case resolution				
Guilty of any charge	0.615	[0.487]	0.867	[0.339]
Number distinct charges pled guilty	0.473	[0.789]	0.971	[0.876]
Number counts pled guilty	1.302	[12.99]	2.382	[6.220]
Statutory max: most serious guilty plea (months)	113.4	[108.2]	161.9	[105.7]
Months of incarceration	8.388	[29.86]	22.27	[47.06]
No plea agreement	0.269	[0.444]	0.153	[0.360]
Sentencing departure	0.0326	[0.178]	0.0782	[0.268]
Timing				
Weeks from case received to filed	46.15	[57.58]	45.47	[62.53]
Number of defendants	613		1932	

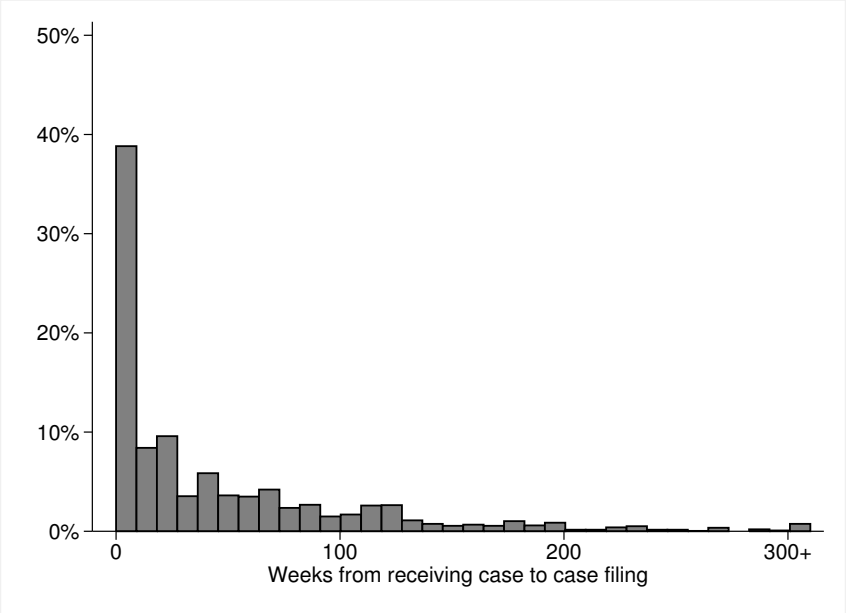
Sample: All federal criminal cases targeting state and local public corruption filed by U.S. attorneys between February 1993 and December 2008 and coded as national or national/local priorities. Charge severity measures were calculated using the approach developed in Rehavi and Starr (2012), which estimates the maximum potential sentence under the law for every criminal charge used by the Department of Justice. Weeks to file were calculated from the date the case was received to the date on which charges were filed (the 25 cases in which defendants were charged before the case was received due to a pre-arrest indictment are coded as 0; none were partisans). Number of defendants represents totals in the data; individual cell sample sizes vary slightly due to missing data. See Appendix A for further details.

partisan defendants charged with public corruption in the 45 states that hold state elections on the federal election calendar (excludes Kentucky, Louisiana, Mississippi, New Jersey, and Virginia) and estimate the number of weeks to the closest placebo election for these defendants. This measure is constructed analogously to the main electoral distance measure.

Weeks to file

We calculate the number of days elapsed from the date the case was recorded as being received by DOJ to the date that the prosecutor filed charges. A histogram of this measure, which is rounded to the nearest complete week, appears below (the 19 cases in which more than 300 weeks elapsed are collapsed in the rightmost bin).

Figure A2: Time to case filings



Sample: All federal criminal cases targeting state and local public corruption filed by U.S. attorneys between February 1993 and December 2008 and coded as national priorities or national and local priorities.

Charge severity

In both government databases and court documents, criminal charges are recorded using the exact section of the U.S. Code that the defendant is accused of violating. For example, a charge of 18:1347A refers to Title (18), Section (1347), Subsection (A) of the U.S. Code. When the relevant code has numerous subsections and paragraphs, the exact reference will be indicated in the charge by an additional series of lower case letters and numbers enclosed in parentheses. Lastly, the category (F for felony or M for misdemeanor) indicates whether the individual was charged with the felony or misdemeanor version of the offense when both exist for that crime. The severity of the individual charges filed against a defendant, the lead charge in the defendant's case, and the individual charges sustained against each defendant were quantified by matching each charge to the the charge severity measures developed by Rehavi and Starr (2012), which provide the maximum potential sentence under the law for every criminal charge used by DOJ since 2000 (i.e., the statutory maximum; see the data appendix in Rehavi and Starr 2012 for a detailed description).

All charges filed/sustained

The NCSD includes detailed information on every charge ever filed against a defendant (including those that were dropped or superseded). Using the charge severity matrix from Rehavi and Starr (2012), we calculated the maximum potential sentence among all charges filed against each defendant as well as the potential maximum among all charges that were sustained. Because most federal sentences are served concurrently, this measure calculates the maximum severity of the charges filed against each defendant as well as the maximum sustained charge.

Lead charge

The NCSD data include the prosecutor's determination of the primary charge in the case, which is known as the lead charge. While it is often the most serious charge in the case, it is not always the most serious final charge. However, while cases and charges do evolve over time, the lead charge is not typically dropped (even in plea bargaining).[6] The lead charge is recorded at the case level in the NCSD data, making it a good indicator of the severity of the alleged criminal conduct in the case, but not necessarily of the severity of each individual defendant's alleged conduct in a case with multiple defendants. In our sample, the lead charge was only listed among each defendant's individually enumerated charges for 67% of defendants.

Another challenge is that the lead charge in the NCSD data often lacks the same level of statutory detail as the individually listed charges. We assessed the severity of lead charges in these cases using the following steps:

- The most notable omission is whether the charge was filed as a felony or a misdemeanor and the specific sub-section under which the defendant was charged. This omitted information was filled in from individual charge data for all defendants where this was possible. These cases were then assigned the relevant statutory maximum sentence from Rehavi and Starr (2012).

- For some of the remaining defendants, the missing sub-section and category were irrelevant for the statutory maximum sentence (i.e., there was either no relevant subsection in the statute or all subsections had the same maximum sentence) and they were accordingly assigned the statutory maximum sentence as well.

- Finally, some of the remaining defendants did not have charge type information, but had charges that only ever appeared in our data as one charge type (i.e., they were

[6] From 2003 onwards, the so-called "Ashcroft Memo" required line prosecutors to get special approval for charge reductions. Consistent with this requirement, Rehavi and Starr (2012) find that the initial lead charge filed was identical to the final charge in about 85% of federal criminal cases sentenced between 2007 and 2009.

always charged as felonies or always as misdemeanors). We thus calculated their charge severity assuming the case was filed under the category as every other case in the data.

After taking all of these steps, we were able to assess the maximum severity of the lead charge and charges filed against 2458 of the 2545 defendants in our data (97%).

Appendix B: Robustness checks

Table B2: Post-election change in probability of same-party case

	Window around election (weeks)			
	24	20	16	12
Local linear regression				
Election discontinuity	0.40*	0.43*	0.48*	0.49*
	(0.19)	(0.19)	(0.19)	(0.19)
LLR 200% optimal bandwidth	9.55	9.25	9.02	8.90
Flexible polynomial RD (logit)				
Election discontinuity	0.59*	0.68**	0.61*	0.45
	(0.26)	(0.25)	(0.29)	(0.45)
N	251	208	152	113

+, *, and ** denote significance at the 10%, 5% and 1% levels, respectively. Local linear regression estimated in Stata 11 using `rd` (Nichols 2011) with 200% of bandwidth calculated using the approach in Imbens and Kalyanaraman (2012). Flexible polynomial estimator includes third order polynomials estimated using logistic regression. Standard errors in parentheses (clustered by criminal case for logit models).

Sample consists of all federal criminal cases targeting state and local public corruption filed by U.S. attorneys during the February 1993–December 2008 period and coded as national or national and local priorities in which the defendants were publicly identified as a member of a major party or a prominent associate of a well-known partisan. For each case, we calculated the number of weeks from the date the case was filed to the closest election before or after at the federal or state level. See Appendix A for further details.

References

Conte, Michaelangelo. 2012. "Former Jersey City pol says corruption charges against him part of effort to elect Christie governor." *The Jersey Journal*, January 26, 2012. Downloaded Sepetember 13, 2013 from `http://www.nj.com/hudson/index.ssf/2012/01/former_jersey_city_pol_says_co.html`.

Gordon, Sanford C. 2009. "Assessing partisan bias in federal public corruption prosecutions." *American Political Science Review* 103 (4): 534–554.

Imbens, G., and K. Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3): 933–959.

Nichols, Austin. 2011. "rd 2.0: Revised Stata module for regression discontinuity estimation. rd 2.0: Revised Stata module for regression discontinuity estimation." `http://ideas.repec.org/c/boc/bocode/s456888.html`.

Rehavi, M., and S. Starr. 2012. "Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences." University of Michigan Law and Economics, Empirical Legal Studies Paper No. 12-002.

Sampson, Peter J. 2007. "Corruption on the run in N.J." *The Hackensack Record*, September 23, 2007.