Online supplementary appendix

A Identification

Table A1: χ^2 Test of Outcomes in Close Elections

Observed	**Losses**	**Wins**
Percent	48.8	51.2
Number	(536)	(563)

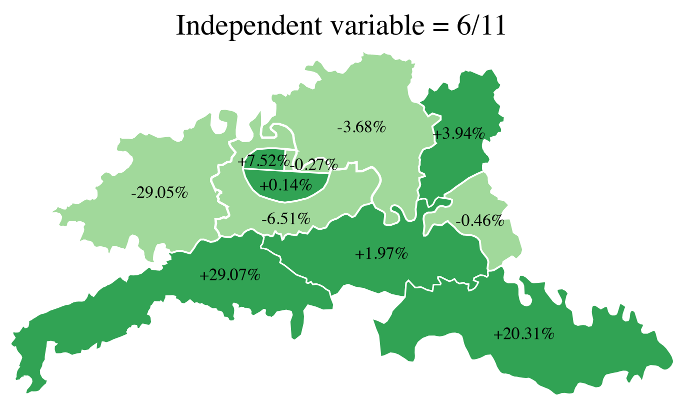
$\chi^2_1 = 0.66$, $P = 0.42$

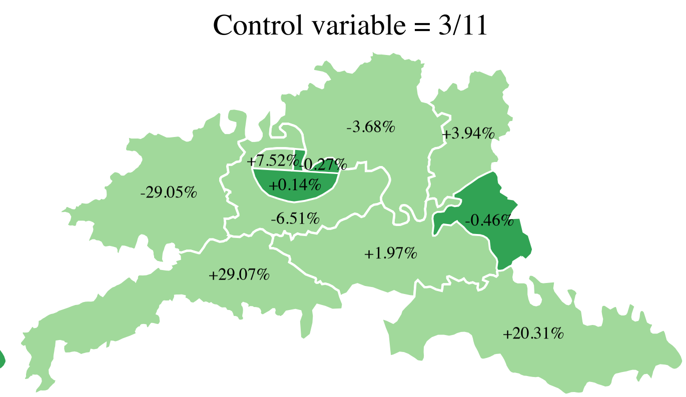
Table A2: Correlation of Close Elections over Time

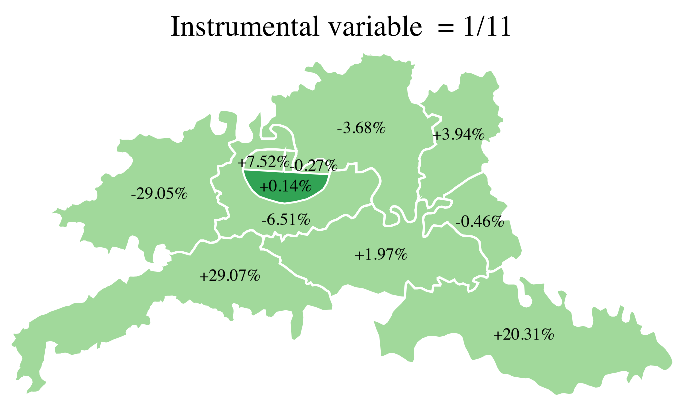
	(T)	$(T-1)$	$(T-2)$
(T)			
$(T-1)$	0.03		
$(T-2)$	0.02	0.03	
$(T-3)$	0.00	0.01	0.04

Pearson correlation coefficients. N is 2556 for $(T-1)$, 2246 for $(T-2)$, and 1939 for $(T-3)$.

Figure A1: Maps illustrating construction of right-hand-side electoral variables

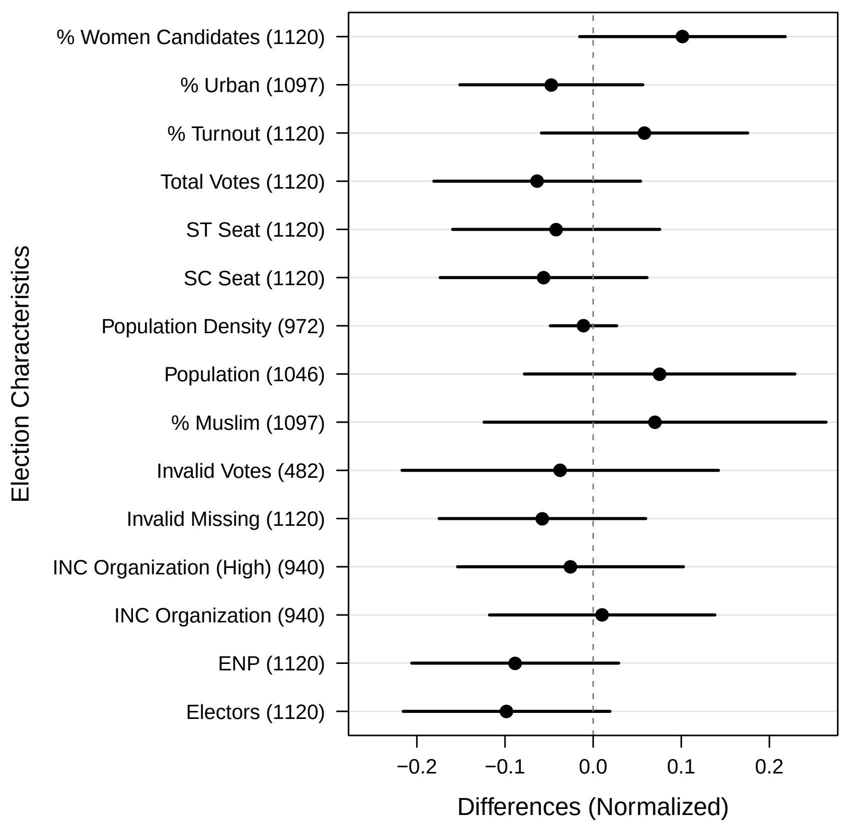






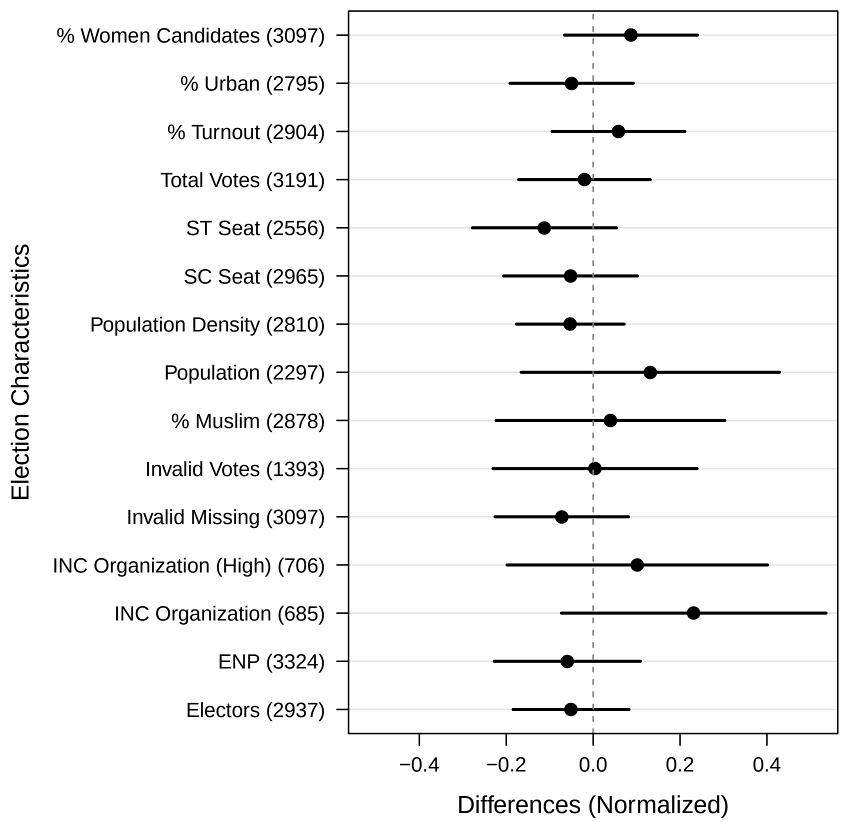
Notes: Map displays state assembly elections in Agra District, Uttar Pradesh, in 1985. Numbers represent the Congress candidate's margin of victory or loss in each MLA constituency. The number of dark shaded areas represents the numerator of the variable, while the denominator is the total number of MLA constituencies in the district. This shows a bandwidth of 1%.

Figure A2: Balance Test: T-Test for difference in means between Congress wins and losses by less than 1 percent



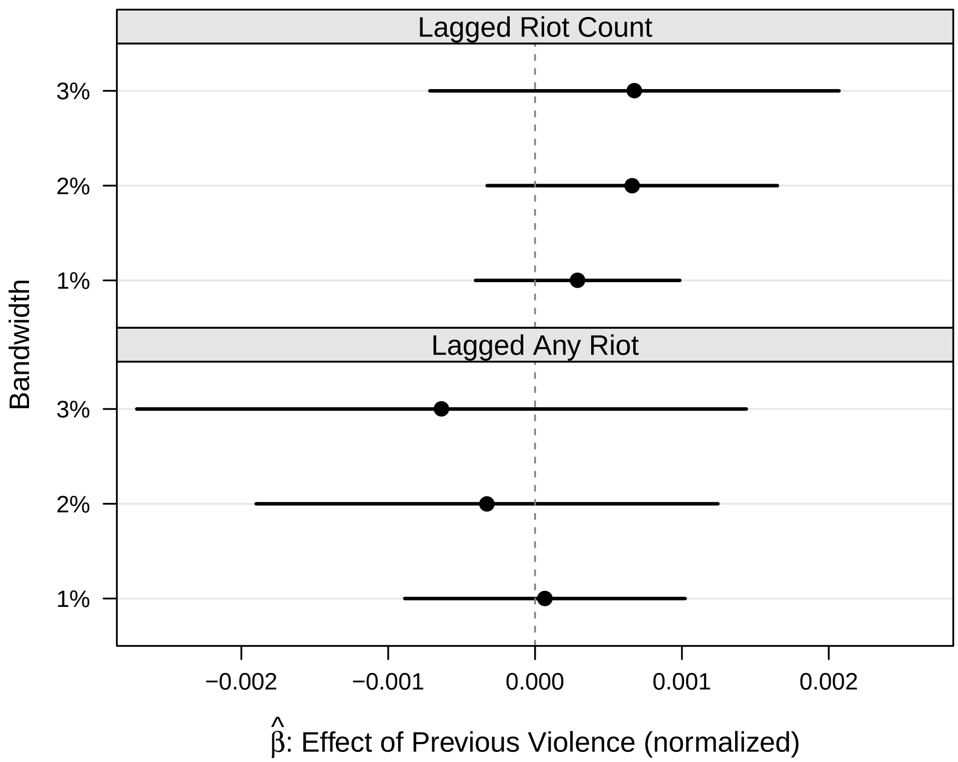
Notes: Results from a two-sided t-test of the difference in means between close elections won and lost by Congress candidates across several election characteristics. Confidence intervals are based on robust standard errors clustered at the district level. N for each test is in parentheses next to the variable being tested.

Figure A3: Balance Test: Local Linear Regression Discontinuity



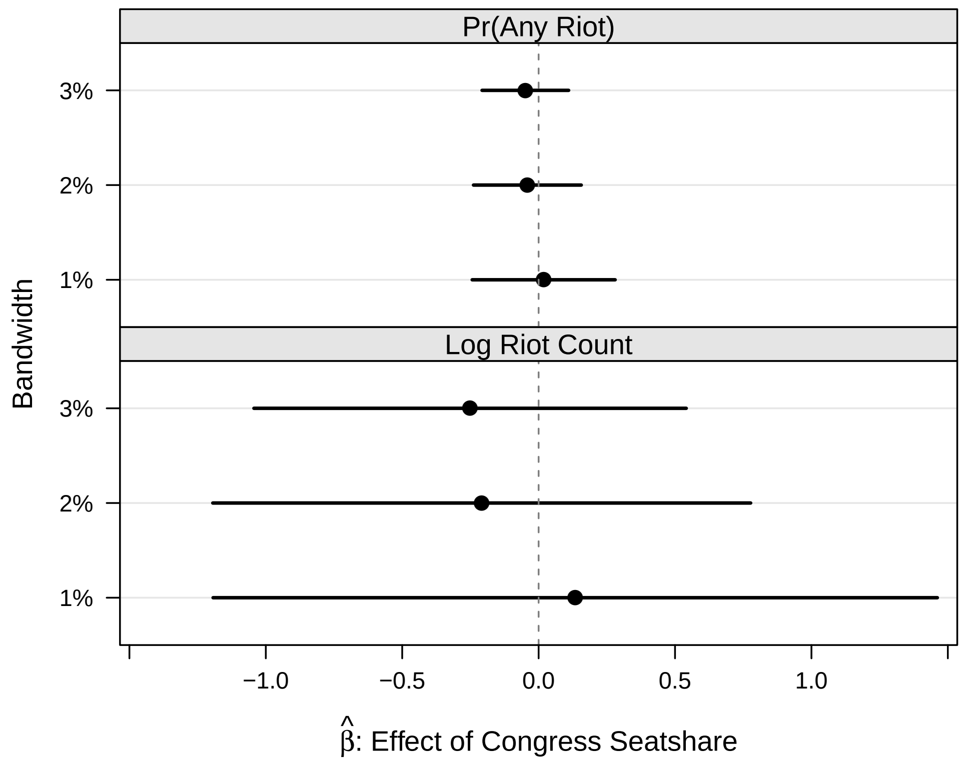
Notes: Results from a local linear regression to estimate differences at the discontinuity between Congress candidates winning and losing election across several election characteristics. Bandwidths are estimated using optimal bandwidth selection suggested by Imbens and Kalyanaraman. While bandwidths differ for each outcome, they are all between two and three percent. Confidence intervals are based on robust standard errors. N for each test is in parentheses next to the variable being tested.

Figure A4: Randomization test—estimates showing that violence at time $t-1$ does not predict our instrument at time t



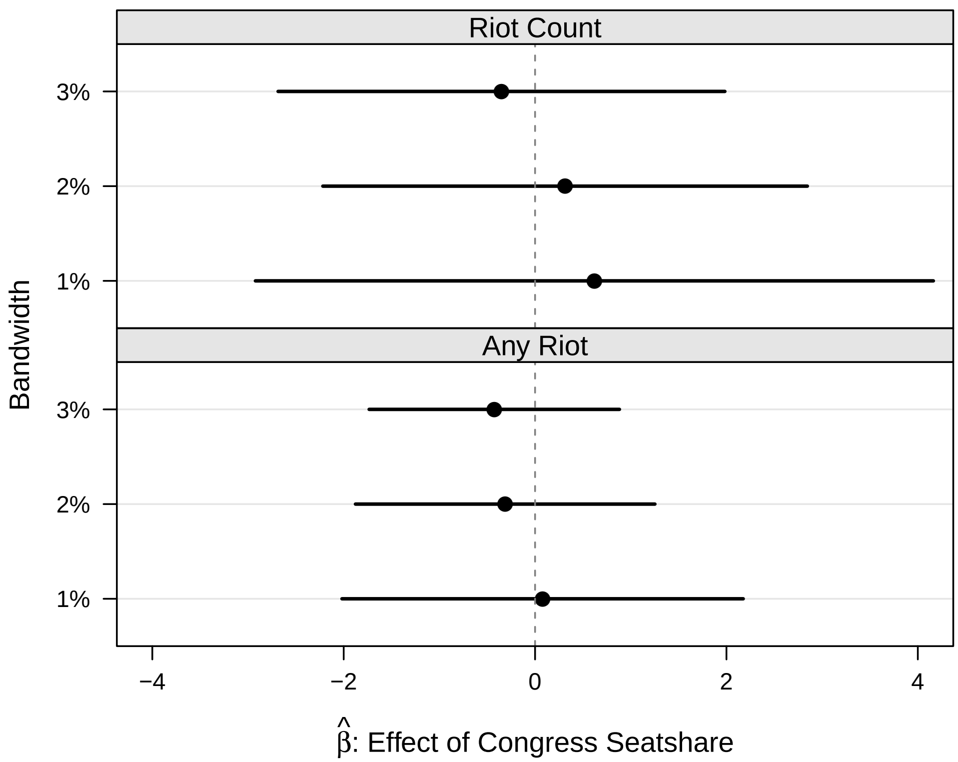
Notes: OLS regressions of the instrument ($CongCloseWin$) in election years t on violence measures in the election cycle preceding the election ($t-1$). Regressions include $CongCloseProp$ as a saturated control. Bars represent 95% confidence intervals based on robust standard errors clustered the district level.

Figure A5: Placebo test—estimates showing that the instrument cannot predict "pre-treatment" violence outcomes



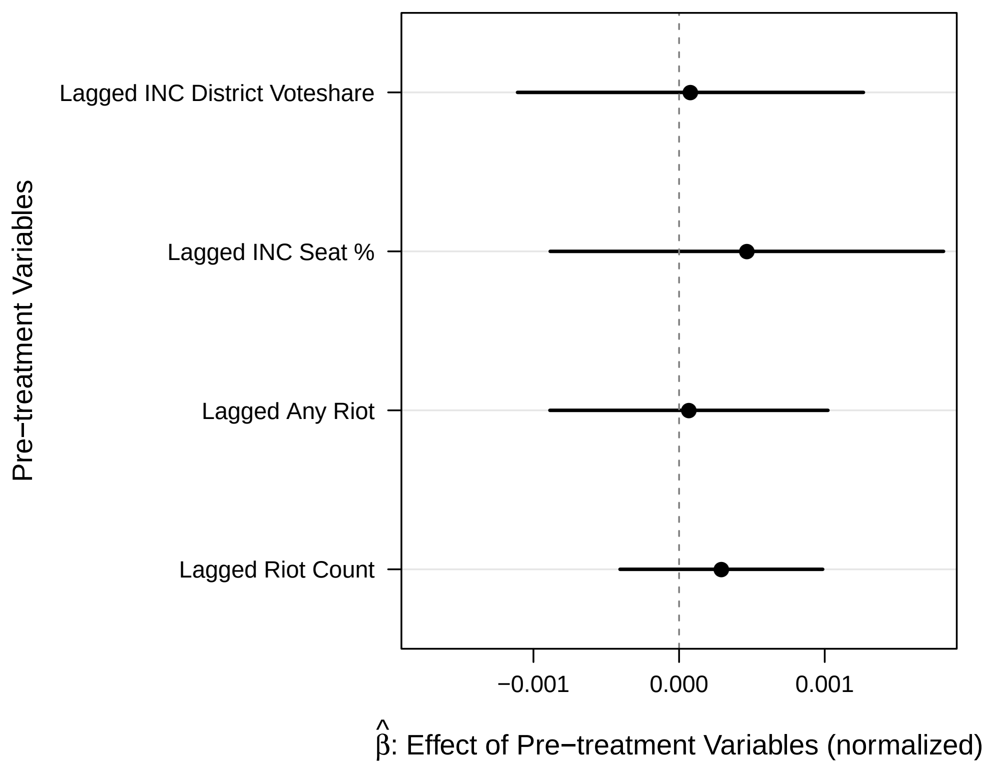
Notes: Results from IVLS regressions of logged and binary riot outcomes at $t-1$ (previous election cycle) on $CongSeatShare$ at time t. Bars represent 95% confidence intervals using robust standard errors clustered at the district level.

Figure A6: Placebo test—estimates from the reduced-form maximum likelihood models showing that the instrument cannot predict "pre-treatment" violence outcomes



Notes: This figure reproduces the previous figure using reduced-form negative binomial and probit regressions of riot outcomes at $t - 1$ (previous election cycle) on $CongSeatShare$ at time t. Bars represent 95% confidence intervals based on robust standard errors clustered at the district level.

Figure A7: Randomization test—estimates showing that violence and INC performance at time $t-1$ does not predict our instrument at time t



Notes: OLS regressions of the instrument ($CongCloseWin$) in election years t on normalized violence and INC performance measures in the election cycle preceding the election $(t-1)$. Regressions include $CongCloseProp$ as a saturated control. Bars represent 95% confidence intervals based on robust standard errors clustered the district level. N is 2556 for all regressions.

B Data

Included states. The states included in the analysis are: Andhra Pradesh, Arunachal Pradesh, Assam, Bihar, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal. Other states—all of which, with the exception of Jammu and Kashmir, are extremely small—were omitted because data were unavailable.

Creating the district panel. Our analyses required compiling a variety of data and aggregating them to create a panel dataset for constant geographic units across time. This is necessary as Indian administrative district boundaries have changed periodically. In 1961, for example, there were 331 districts; by 2011 there were 640.

Changes to administrative district boundaries took two forms. The vast majority were "simple" splits in which one district was cleanly divided into two or more districts. In other cases, new districts were the result of "complex" splits: the new district's territory was formed out of multiple existing districts. Our raw annual data on riots are recorded using the district boundaries as they existed at the time the riots took place. Our goal was to aggregate these data back to 1961 districts.

We define the original unsplit districts as "parents" and the new districts as "children." To match parents to children, we used Appendix 1 to Table A-1 from the General Population Tables (Part II-A) of the 2001, 1991, 1981, and 1971 Indian censuses. These tables record all districts extant in the year of the census. When a new district has been created, the table indicates the parent district or districts out of which it was carved. For each census round, we identify the changes that took place.

Our dependent variables are count data. In the simplest case, district boundaries are unchanged across census years. When the children districts are the result of a *simple* split, aggregating backwards is straightforward: since there is only one parent district, we simply sum up the counts of all its children. For complex splits, the procedure is more involved.

In such cases, we take a weighted sum of the counts from the children districts. Using the Census tables, we calculate what proportion of the territory in a child district j came from each parent district i and define this as the weight W_{ij}. We compute some count variable X for parent district i by taking the weighted portion of X from each child district j. That is, we sum over the product of each X_j and W_{ij} as follows:

$$X_i = \sum_{1}^{j} X_j W_{ij} \tag{A}$$

More precisely, we use weights calculated from each census to bring districts back first from 2001 to 1991, then from 1991 to 1981, from 1981 to 1971, and finally from 1971 to 1961.[1]

We further had to map state legislative (MLA) constituencies onto the 1961 administrative districts in order to create our right-hand-side electoral variables: $CongSeatShare$, $CongCloseWin$, and $CongCloseProp$. Like administrative districts, the boundaries of state legislative constituencies changed over time. Throughout, however, these constituencies remained *perfect subsets* of administrative districts.

We used the reports of the Delimitation Commission of India from 1961, 1971, and 1976 to assign each legislative constituency at election time t to the administrative district to which it belonged, also at time t. (After 1976, legislative districts were not redrawn until 2008, easing the process for this period.) If these administrative districts had gone unchanged since 1961, then no further work was needed—the MLA constituency was already matched to the correct 1961 district. If the constituency had ended up in child district produced by a *simple* split, then we simply reallocated this seat to the original parent district. In cases where an MLA constituency belonged to a child district produced via a *complex* split, we used tehsil and village information contained within the Delimitation Reports, as well as district maps, to manually assign the constituency to the correct 1961 district. In this manner, we were able to accurately assign all MLA constituencies between 1962 and 2008 to 1961 administrative

[1] Equation A is easily generalized for simple changes. When the district remains unchanged, $i = j$. When there is a simple split, each $W_{ij} = 1$.

districts.

District Muslim population. To measure the proportion of the population in a district that was Muslim, we used reports from the 1961, 1971, and 1981 censuses. These data included the total population for a district and the total number of Muslims in a district. Applying the same procedure for reconstructing 1961 district boundaries, we added up total population and total Muslim population for 1961 districts. We thereby calculate the proportion of the district that was the Muslim.

Congress state governments. We used secondary historical sources to compile a list of all parties that formed state governments in India between 1961 and 2008. This list included the party of the Chief Minister as well as any other parties in coalition governments. In our analyses, we used this data to create a dummy variable indicating whether the Chief Minister was from the Congress Party in a given state-year.

Riots. As mentioned in the paper, we use the Wilkinson-Varshney database of Hindu-Muslim riots from 1950 to 1995. We append these with data collected by Mitra and Ray (2014), bringing the panel to 2000. In cases where these data did not report the district in which the riots occurred, we used the state and locality of the riot to find the district.

C Explanation of simulations

To simulate the expected count of riots for our entire sample (Figure D10), we estimate the reduced-form equation of our instrumental variables design using negative binomial regression and a 1% bandwidth. We then generate new copies of the data for several counter-factual scenarios in which Congress won close elections with different probabilities: 0, 0.2, 0.4, 0.6, 0.8, and 1. Next, we generate 1,000 clustered bootstrap simulations. For each bootstrap simulation, we estimate the vectors of coefficients, $\tilde{\boldsymbol{\beta}}$, then calculate predicted values, $\tilde{y}$ for the actual data and for 250 counterfactual datasets for each victory probability,[2] and transform these into the expected counts by taking $e^{\tilde{y}}$. Finally, for each scenario, we take the sum over all observations in the data, giving us the expected number of riots. By using bootstrap simulation, we are able to estimate the uncertainty of predictions from our model.

[2] This is necessary, since we have to randomly sample which close elections are won and lost, given a probability of victory.

D Supplements to the main analysis

Table D1: Descriptive Statistics

Variables	Mean	SD	N	Min	Max
Number of riots	0.348	1.509	2871	0	47.351
Number of riot casualties	9.170	73.535	2871	0	2386.000
Number of riot days	0.704	3.806	2871	0	88.351
Any Riot	0.160	0.367	2871	0	1.000
% Congress seats	0.436	0.319	2871	0	1.000
% Congress close wins	0.018	0.052	2871	0	1.000
% Congress close elections	0.036	0.072	2871	0	1.000
Number of Seats	11.044	6.529	2871	1	55.000
% Muslim	0.100	0.091	2799	0	0.614
Congress chief minister	0.556	0.497	2863	0	1.000
% Congress vote	0.362	0.133	2870	0	0.963
% Turnout	0.572	0.119	2871	0	0.890

Table D2: First Stage F-Test

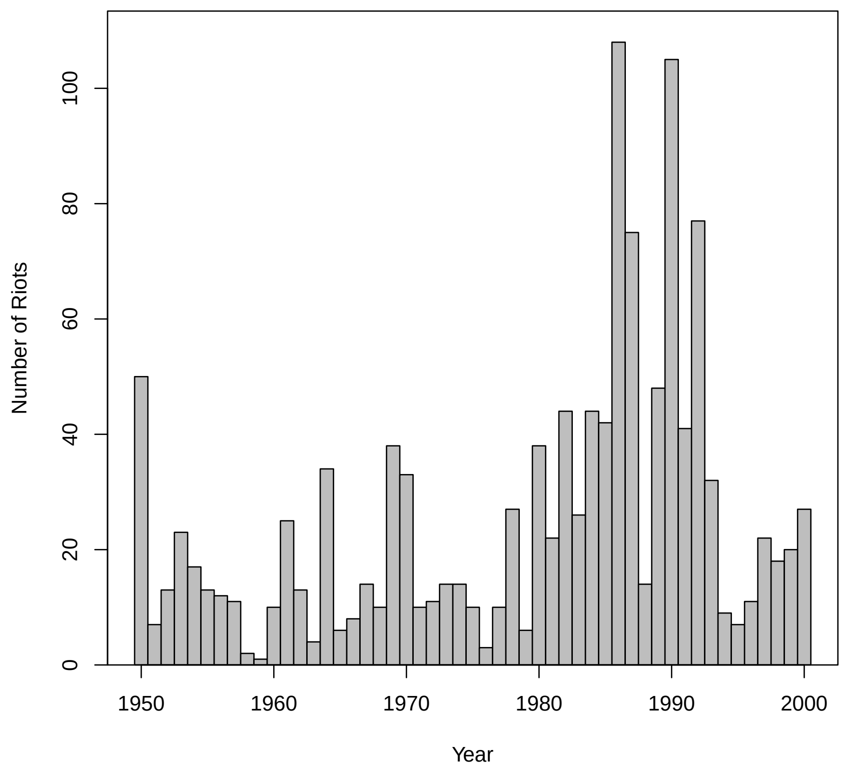
Bandwidth	F-Statistic	p
1%	55.50	0.00

Table D3: Correlates of Close Elections

	Prop. Close Elections	Any Close Election	% Muslim	% Urban
Prop. Close Elections				
Any Close Election	0.98			
% Muslim	0.05	0.07		
% Urban	0.04	0.05	0.18	
Population Density	0.12	0.15	0.41	0.11

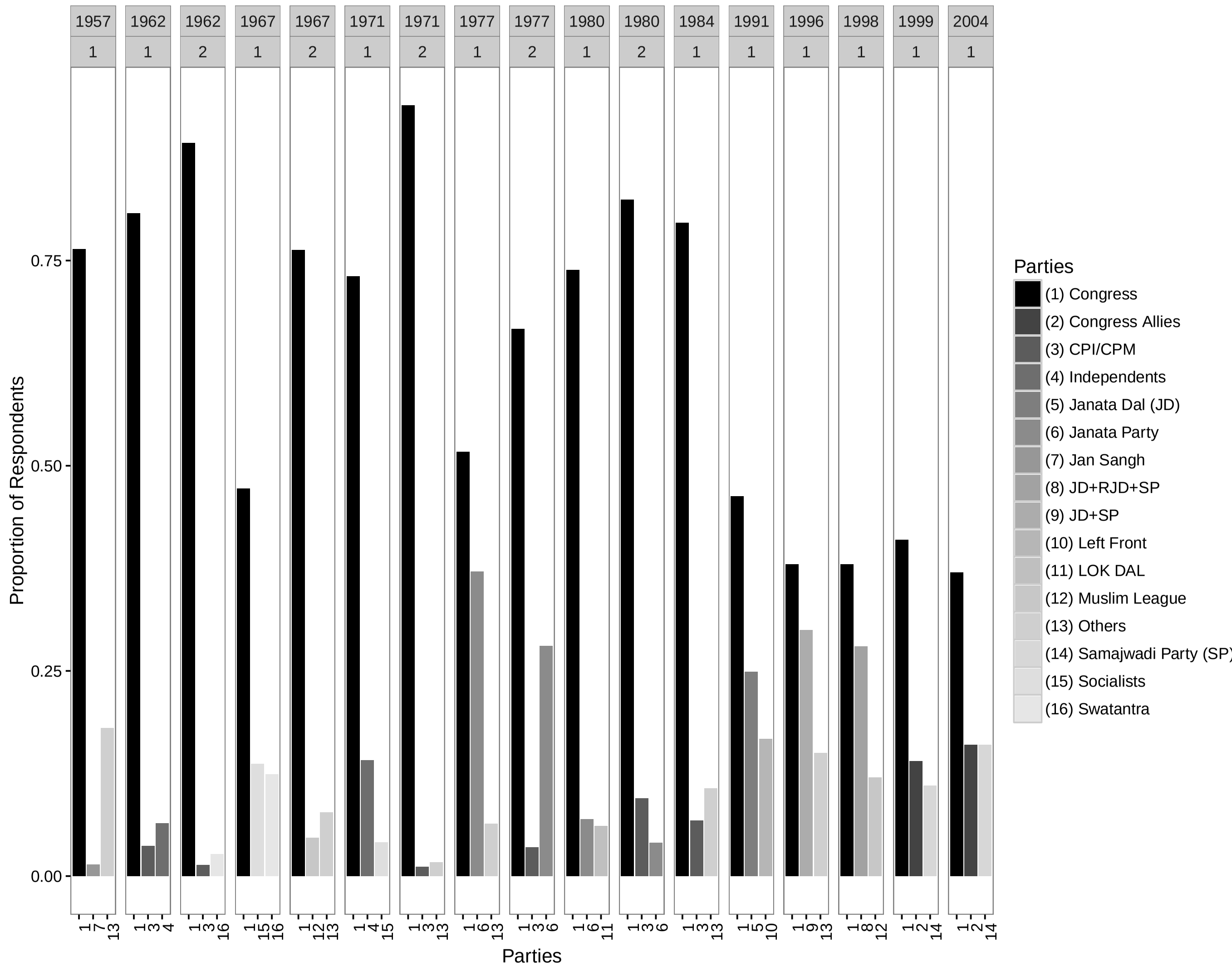
Spearman rank correlation coefficients. N is 2787 for % Muslim and Urban and 2436 for Population Density.

Figure D1: Hindu-Muslim riots in India by year, 1950–2000



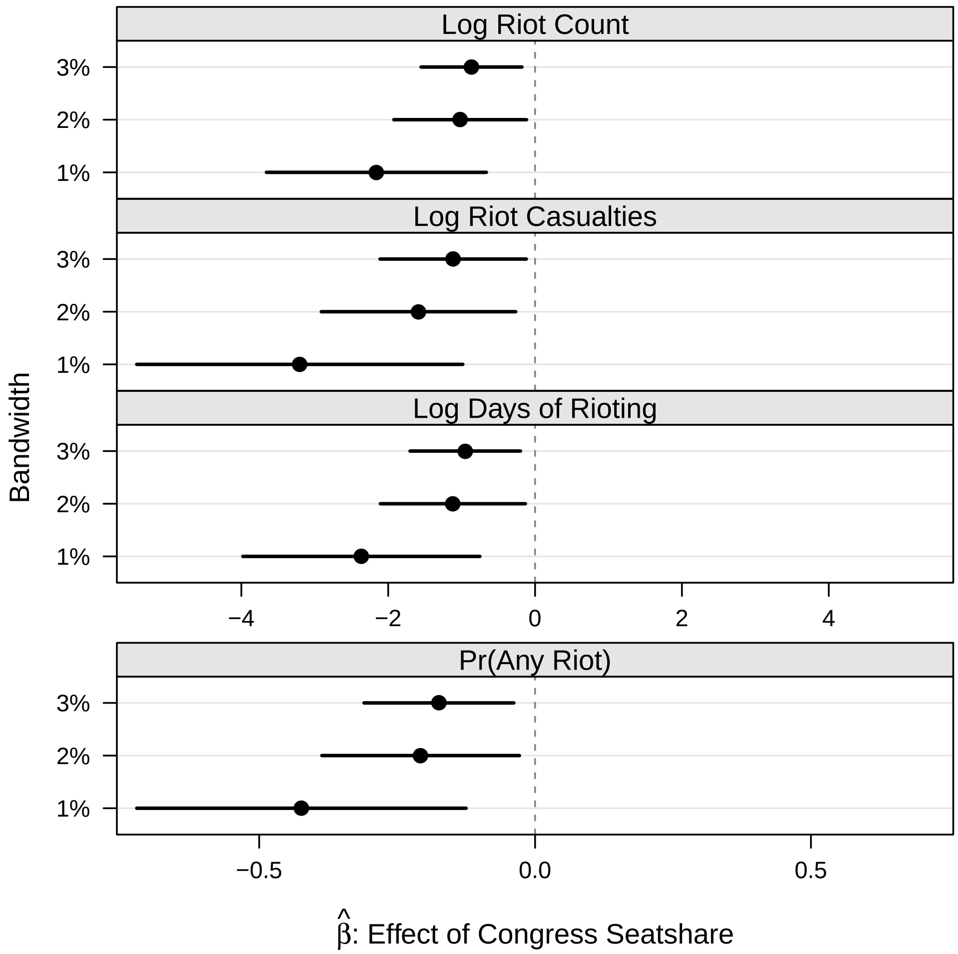
Notes: Data come from the *Varshney-Wilkinson Dataset on Hindu-Muslim Violence in India* and an extension of it to 2000 by Mitra and Ray (2014).

Figure D2: Percentage of Muslim votes going to various parties in Indian elections, 1957–2004



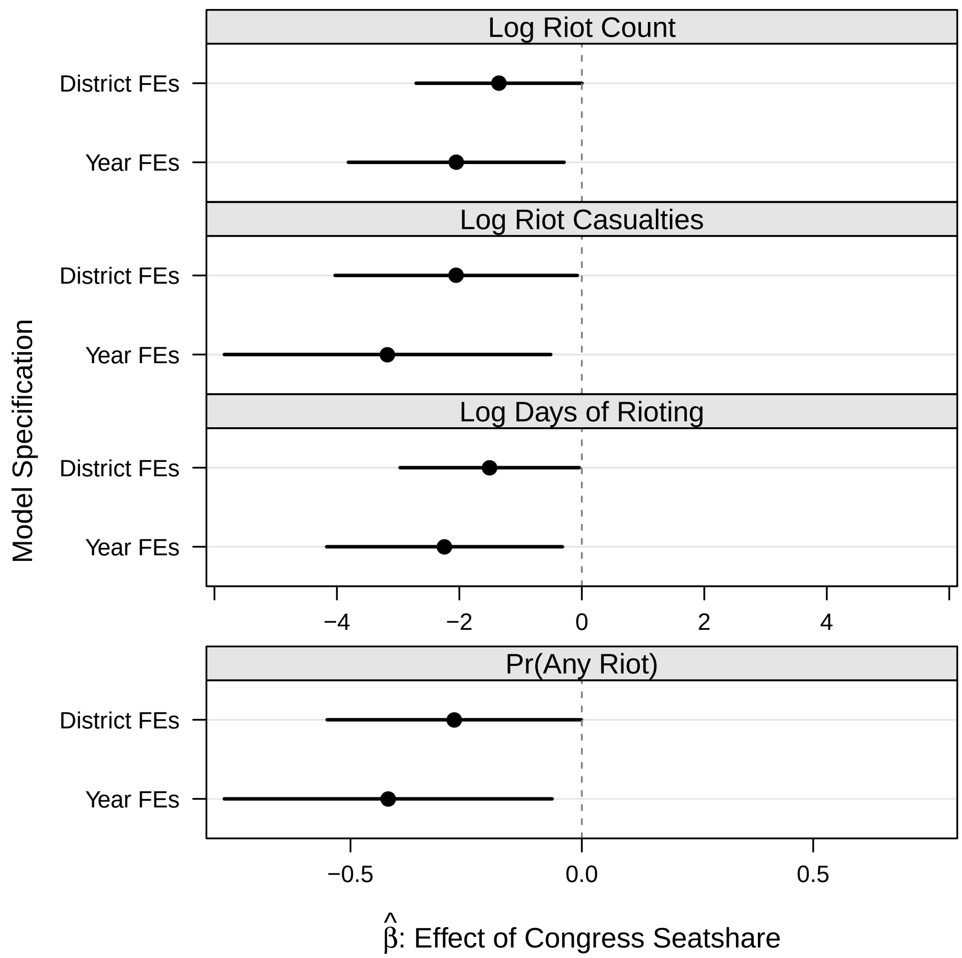
Notes: Data based on self-reports of Muslim voting found in various surveys conducted by the Centre for the Study of Developing Societies and newly compiled by the authors. Surveys sometimes report Muslim vote shares for formal (i.e. pre-election) coalitions of parties; this explains why some parties in the legend are grouped together. For election years when we have two surveys, we provide separate plots for each survey.

Figure D3: Instrumental variables estimates of the effect of *CongSeatShare* on riot outcomes, multiple bandwidths



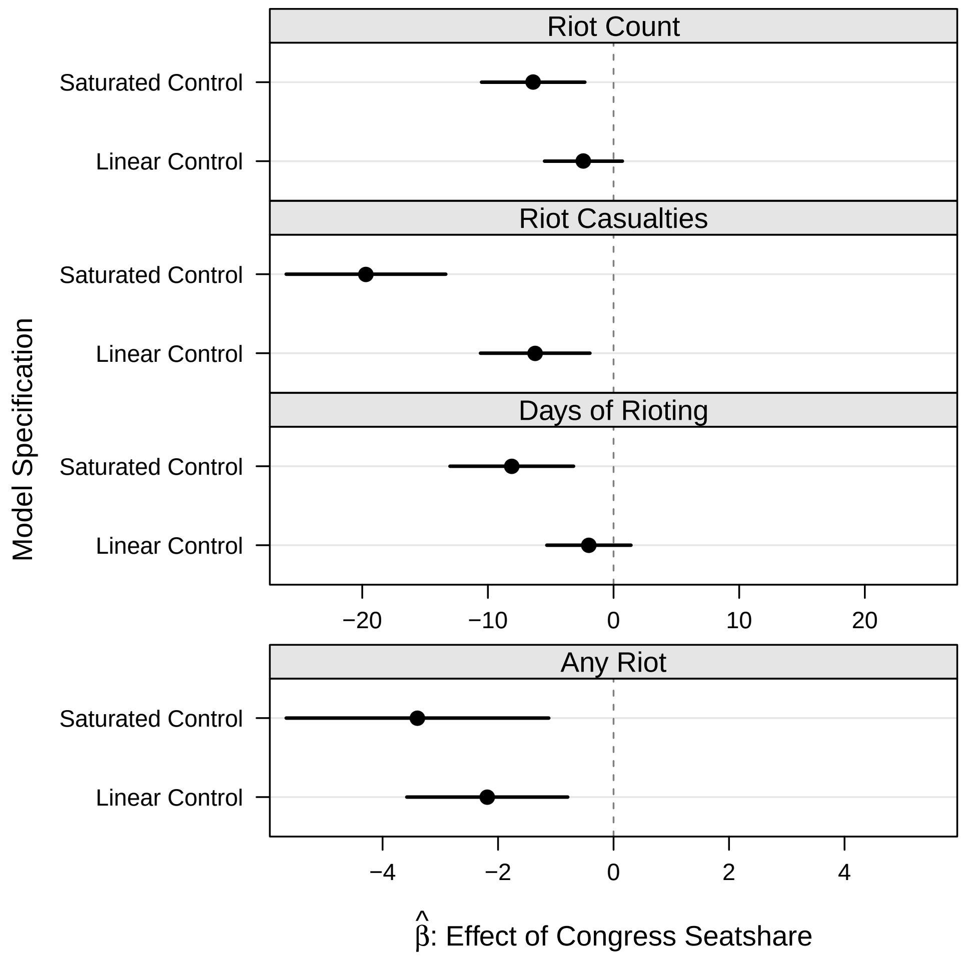
Notes: This figure presents coefficient estimates from IVLS regressions of logged or binary riot outcomes on *CongSeatShare*, using the approach described in the Data and Identification section, across multiple bandwidths. Bars represent 95% confidence intervals using robust standard errors clustered at the district level. Bandwidth refers to the margin of victory used to define a close election. N for all regressions is 2871, across 315 districts. The number of close elections for each bandwidth is 1099, 2212, and 3331 for 1%, 2%, and 3%, respectively.

Figure D4: Instrumental variables estimates of the effect of $CongSeatShare$ on riot outcomes, with district and year fixed effects



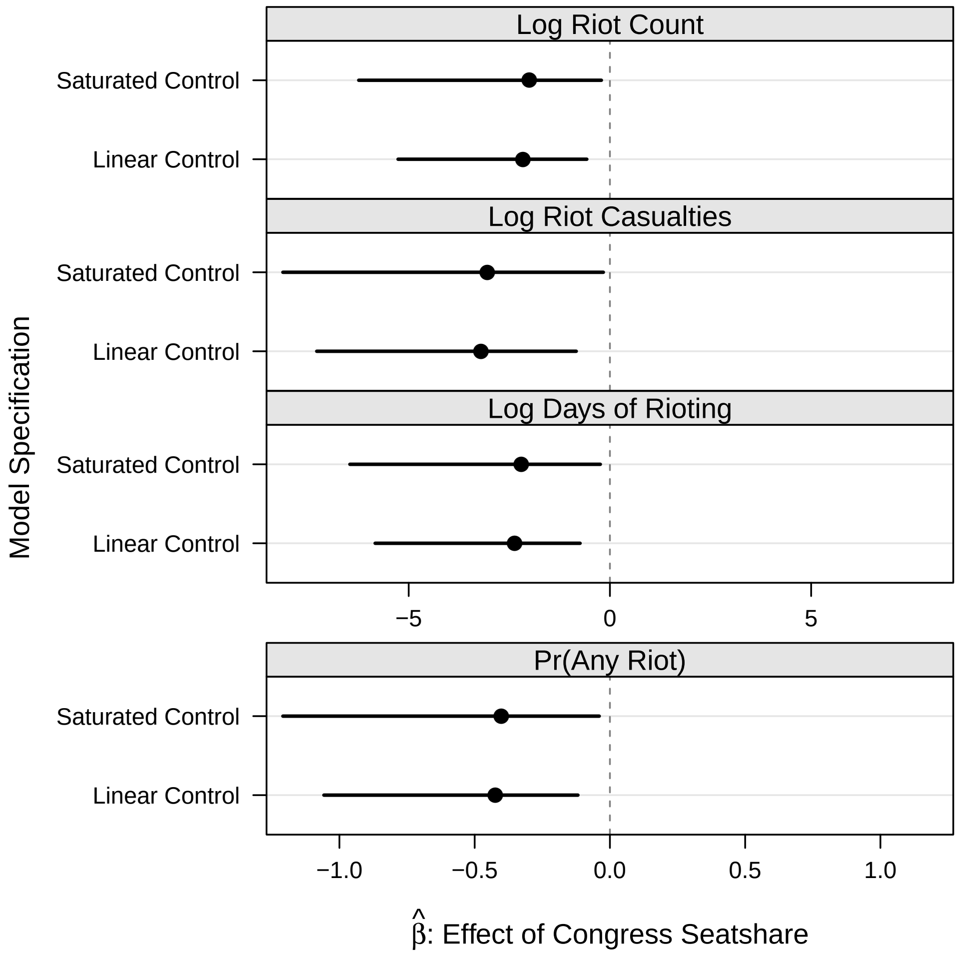
This figure presents coefficient estimates from IVLS regressions of logged or binary riot outcomes on $CongSeatShare$, using the approach described in the Data and Identification section, with district or year fixed effects. Bars represent 95% confidence intervals using robust standard errors clustered at the district level. N for all regressions is 2871, across 315 districts.

Figure D5: Maximum likelihood estimates of reduced-form equation



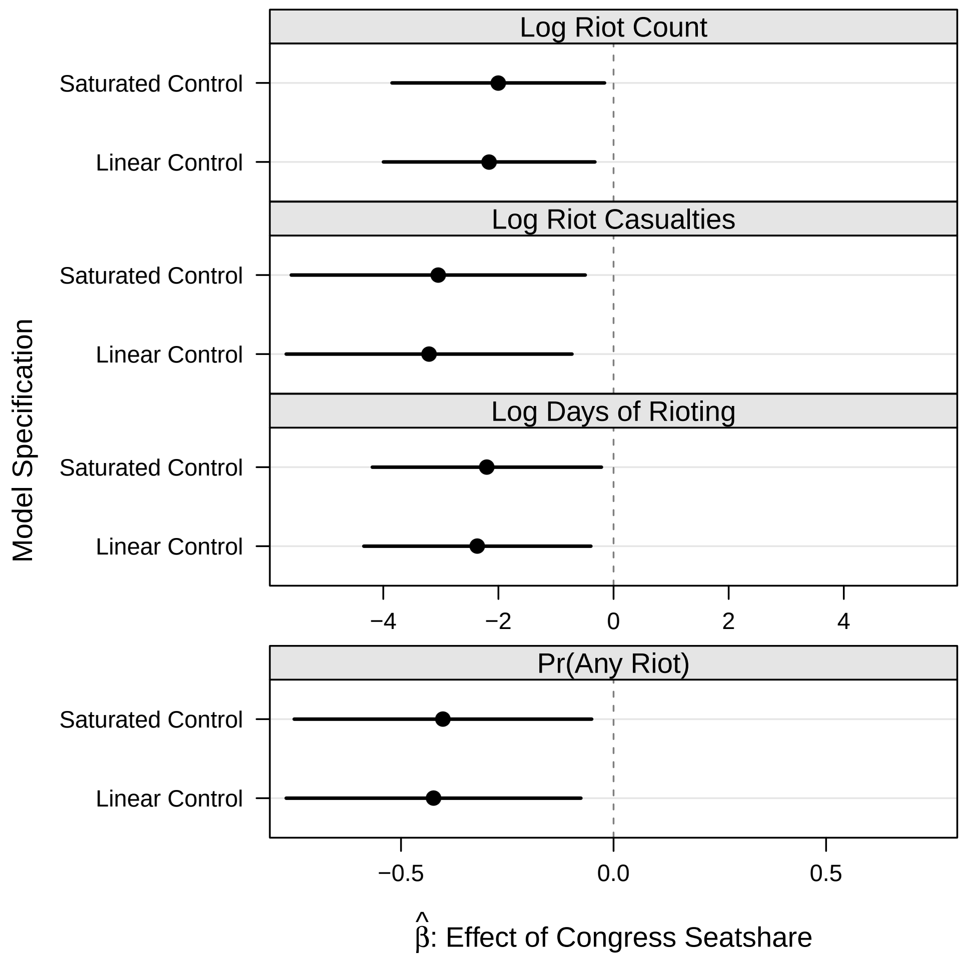
This figure presents estimates from negative binomial (top three panels) and probit (bottom panel) regressions of the reduced-form equation. That is, unlogged riot outcomes regressed on $CongCloseWin$ and $CongCloseProp$. Bars represent 95% confidence intervals derived from robust standard errors clustered at the district level. The results are in line with those shown in Figure 2, although specifications using a linear control are weaker and sometimes drop out of conventional significance. N for all regressions is 2871, across 315 districts.

Figure D6: Instrumental variables estimates of the effect of $CongSeatShare$ on riot outcomes, with standard errors clustered by state



This figure presents coefficient estimates from IVLS regressions of logged or binary riot outcomes on $CongSeatShare$, using the approach described in the Data and Identification section. Bars represent 95% confidence intervals using robust standard errors clustered at the state level using cluster bootstrapping and percentile confidence intervals. Because the bootstrapped distribution is asymmetric, the confidence intervals are asymmetric around the point estimates. N for all regressions is 2871, across 315 districts.

Figure D7: Instrumental variables estimates of the effect of $CongSeatShare$ on riot outcomes, with standard errors clustered by state-election year



This figure presents coefficient estimates from IVLS regressions of logged or binary riot outcomes on $CongSeatShare$, using the approach described in the Data and Identification section. Bars represent 95% confidence intervals using robust standard errors clustered at the state-election year level. N for all regressions is 2871, across 315 districts.

Figure D8: Proportion of elections contested closely by INC by election years.

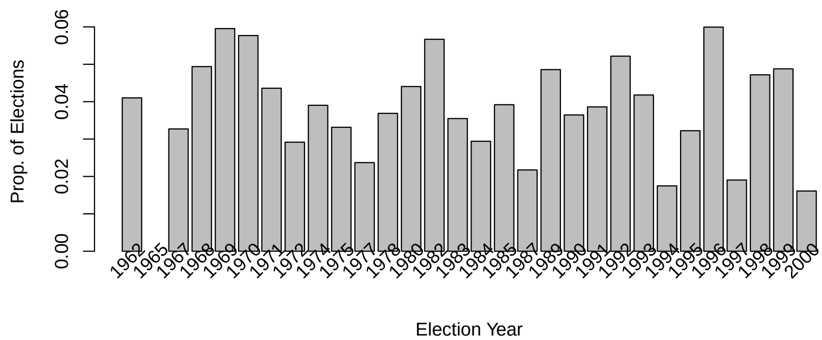


Figure D9: Proportion of elections contested closely by INC by states.

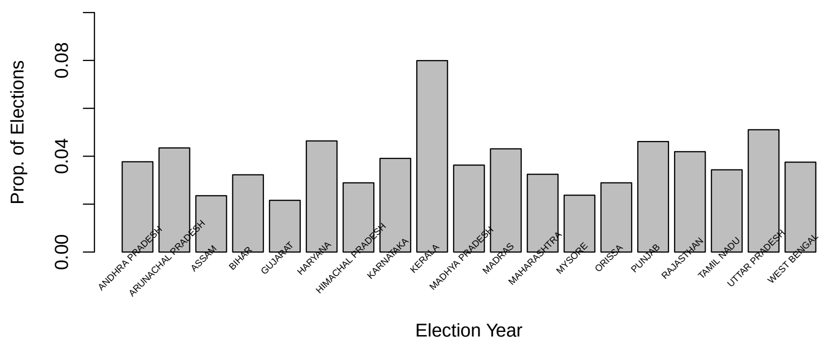
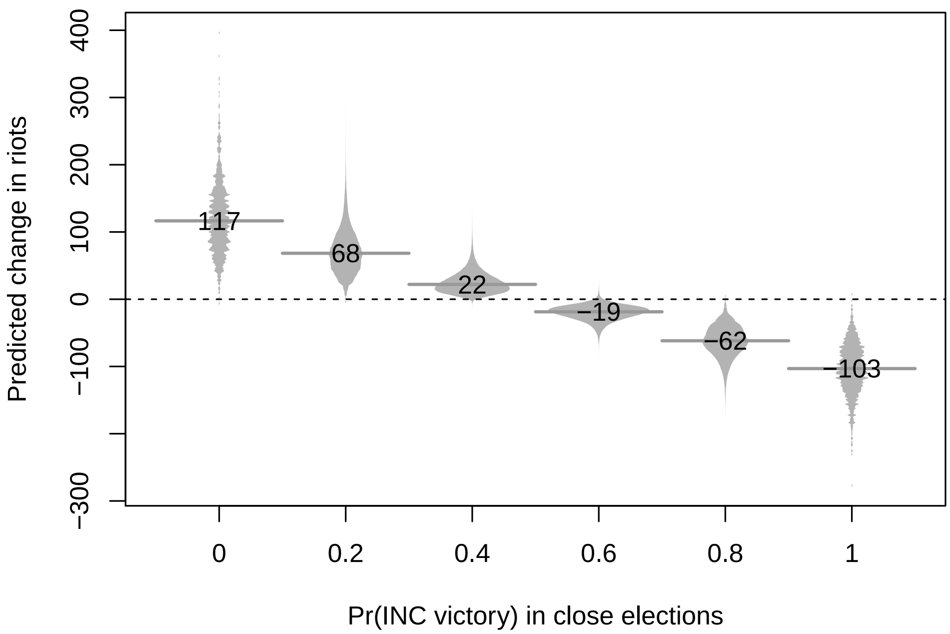
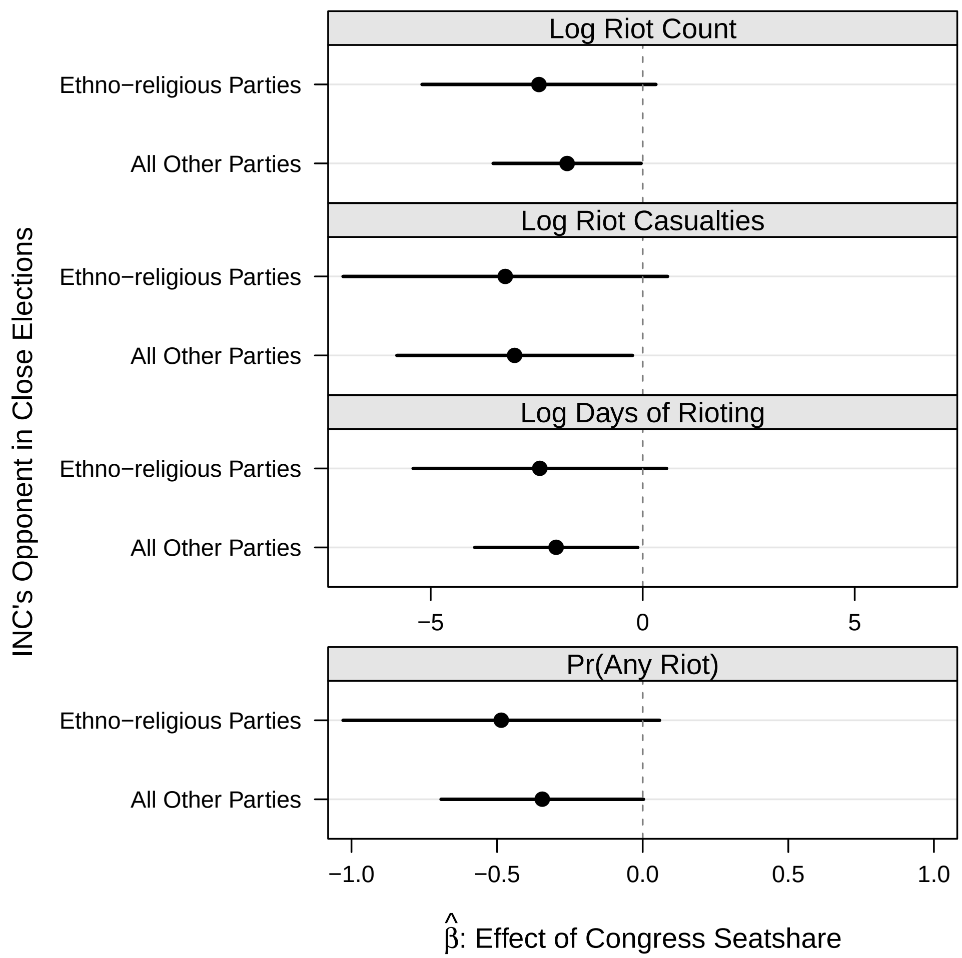


Figure D10: Simulated difference in riots when Congress wins all close elections compared to its actual performance



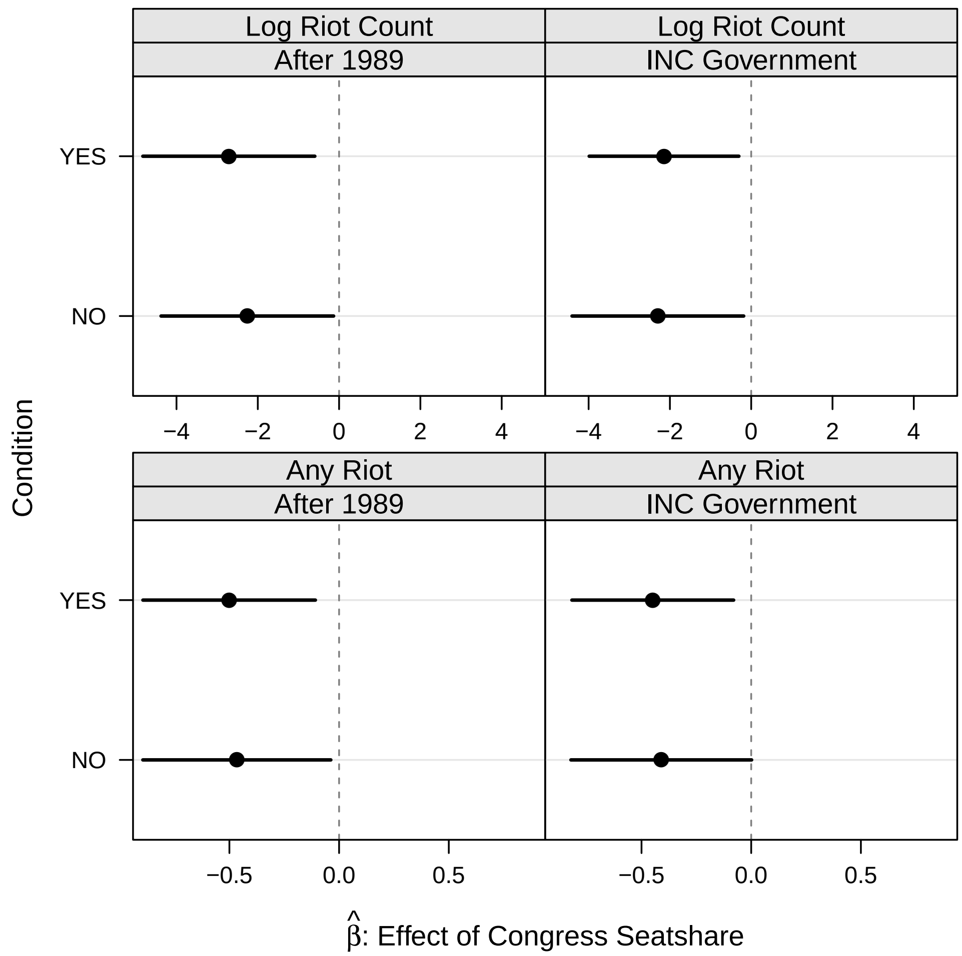
Notes: This figure plots the simulated predictions of how many fewer riots would have occurred if Congress had won close election with probabilities of 0, 0.2, 0.4, 0.6, 0.8, and 1, compared to its actual performance. The predictions are based on 1,000 clustered bootstrapped replications of negative binomial regression estimates of the reduced-form equation from our instrumental variables design, using a 1% bandwidth (2 simulations drop due to failure for the model to converge). The figure reports the median change in riots in the middle of each simulated distribution. See Appendix C for further explanation.

Figure D11: Effects of $CongSeatShare$, by opposition party type



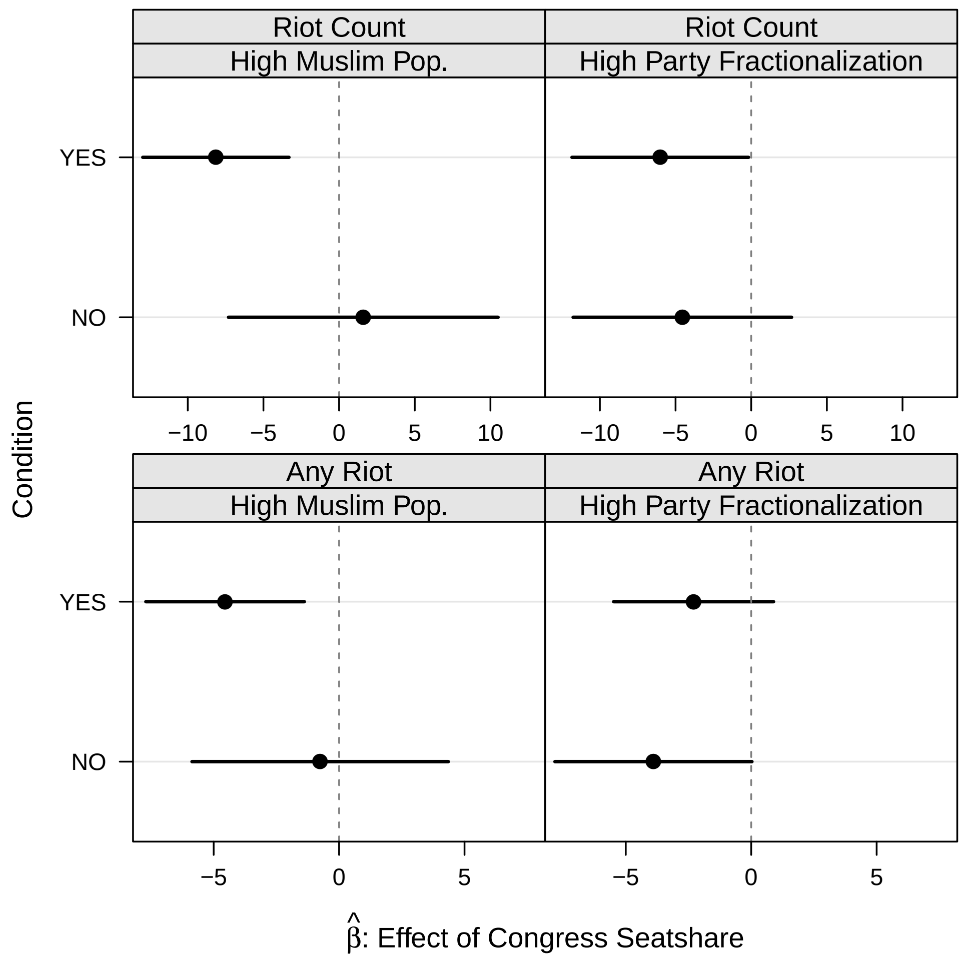
Notes: This figure presents coefficient estimates from IVLS regressions of logged or binary riot outcomes on $CongSeatShare$, using the approach described in the Data and Identification section. Bars represent 95% confidence intervals using robust standard errors clustered at the district level. $CongCloseProp$ uses a saturated specification. "Ethnic parties" refers to the BJS/BJP and the Shiv Sena, parties which mobilize along the Hindu-Muslim ethnic divide. "All Other Parties" refers to all parties other than the ethnic parties and the INC. N for all regressions is 2871, across 315 districts.

Figure D12: Effects of $CongSeatShare$, by time-period and state-government incumbency



Notes: This figure presents coefficient estimates from IVLS regressions of logged and binary riot outcomes on $CongSeatShare$, using the approach described in the Data and identification section. Bars represent 95% confidence intervals using robust standard errors clustered at the district level. N for Pre- and Post-1989 are 1953 and 918, respectively. N for Congress government and opposition are 1593 and 1270, respectively.

Figure D13: Reduced-form negative binomial estimates of heterogeneous effects



This figure presents estimates from negative binomial regressions of the reduced-form equation. That is, unlogged riot outcomes are regressed on $CongCloseWin$ and $CongCloseProp$. Bars represent 95% confidence intervals derived from robust standard errors clustered at the district level. The subgroups used to demonstrate the heterogeneous effects are described in the main paper. N for high and low Muslim population are 1427 and 1372, respectively. N for high and low party fractionalization are both 1397.