Supplementary Materials

The Direct Primary and the Incumbency Advantage in the U.S. House of Representatives

Michael Olson

Contents

A Personal Incumbency Advantage

A.1 Relationship Between Differences-in-Discontinuities Estimand and the Personal Incumbency Advantage

Fowler and Hall (2014) note that the RD estimand β can be expressed as

$$\beta = 2 \times \left(\text{Partisan Advantage} + \Pr(\text{Inc. Seeks Reelection}) \times \text{Personal Advantage}\right)$$

I express the Personal Advantage as a function of the other terms:

$$\text{Personal Advantage} = \frac{\beta - 2 \times \text{Partisan Advantage}}{2 \times \Pr(\text{Inc. Seeks Reelection})}$$

The change in the personal advantage due to primary adoption ΔPersonal can therefore be expressed, assuming that direct primary adoption is unconfounded, as

$$\Delta\text{Personal} = \left(\frac{\beta_p - 2 \times \text{Partisan Advantage}_p}{2 \times \Pr(\text{Inc. Seeks Reelection})_p}\right) - \left(\frac{\beta_c - 2 \times \text{Partisan Advantage}_c}{2 \times \Pr(\text{Inc. Seeks Reelection})_c}\right)$$

with the direct primary nominating regime noted with p and convention (non-primary) nominating systems with c. If $\text{Partisan Advantage}_p = \text{Partisan Advantage}_c = 0$, then this simplifies to

$$\Delta\text{Personal} = \left(\frac{\beta_p}{2 \times \Pr(\text{Inc. Seeks Reelection})_p}\right) - \left(\frac{\beta_c}{2 \times \Pr(\text{Inc. Seeks Reelection})_c}\right)$$

This simplification also holds if both Partisan Advantage and Pr(Inc. Seeks Reelection) are the same in convention and primary contexts. Each element of this expression is estimable from the data. β_p and β_c are the regression discontinuity estimates on the primary and non-primary samples, respectively. Pr(Inc. Seeks Reelection) is the probability that an incumbent who won the previous election with exactly 50% of the two-party vote will seek reelection. I discuss estimation of this quantity in the following section. Combining the estimates of β and Pr(Inc. Seeks

Reelection), both calculated using within-district and within-year demeaned outcomes to address confounding of direct primary adoption, produces an estimate of the change in the personal incumbency advantage due to direct primary adoption. Naturally, this estimator re-gains some concerns about endogenous reelection-seeking that the regression discontinuity estimator seeks to eliminate.

A.2 Estimating Probability that Incumbents Seek Reelection

To estimate the probability that a bare-winner incumbent seeks reelection, I use local linear regression in a procedure similar to the regression discontinuity estimation. I first demean the outcome, here an indicator for whether an incumbent seeks reelection in period $t+1$, by district and year; because the specific value of Pr(Inc. Seeks Reelection) is needed for each nominating regime (rather than a difference, as in the regression discontinuity estimation), I add back the mean district mean and the mean year mean. Using this demeaned and re-centered outcome, I then estimate a linear regression with a bandwidth and (triangular) kernel weights identical to the regression discontinuity estimation. The predicted value of this regression at *Incumbent Party Vote Share* = 0.5 is the estimate of the quantity of interest.[1] Put simply, this is the same procedure used to estimate the portion of an RD design above the discontinuity. Figure A.1 presents estimates of Pr(Inc. Seeks Reelection) for a variety of bandwidths for each of the primary and convention nominating regimes. These estimates are combined with the regression discontinuity incumbency advantage estimates to produce estimates of the change in the personal incumbency advantage (ΔPersonal) described above.

[1] This is a similar procedure to that used by Fowler and Hall (2014) to estimate the same quantity of interest, though they use OLS with a fourth-order polynomial of *Incumbent Party Vote Share*.

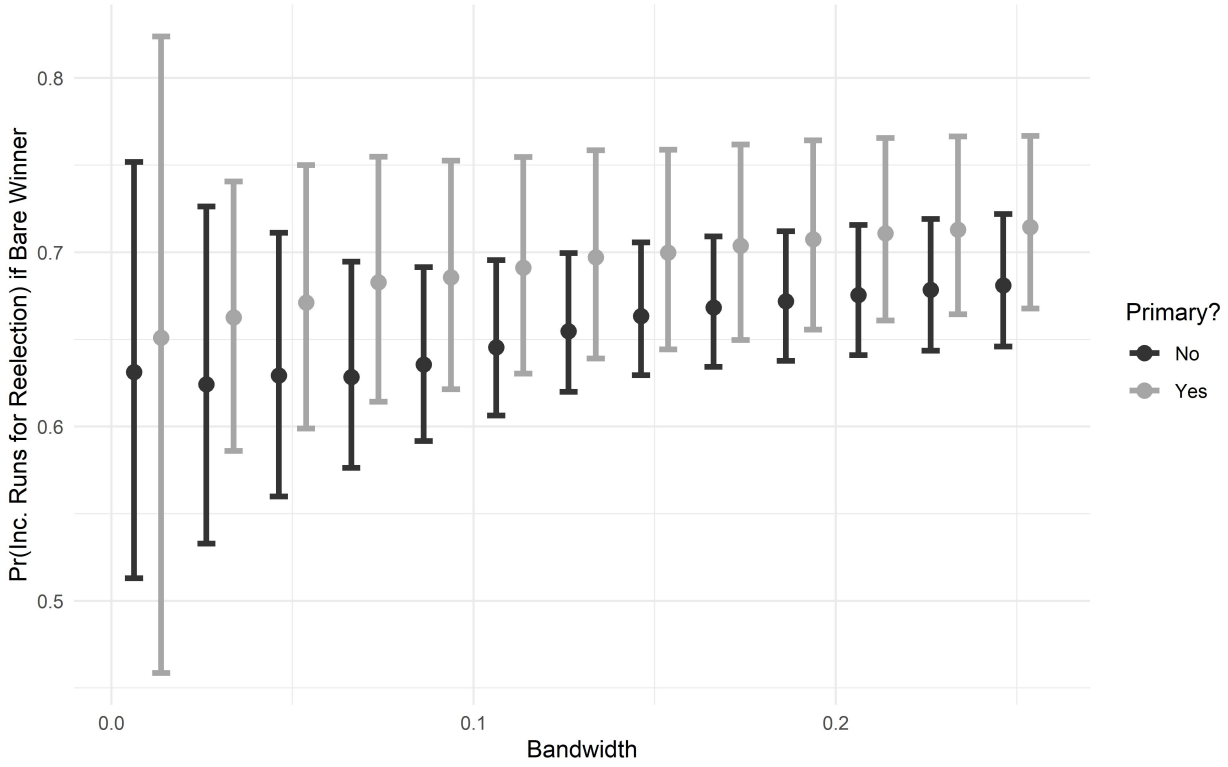

FIGURE A.1: Probability of Bare-Winner Incumbents Seeking Reelection

B Data Appendix

B.1 Primary Adoption Dates

TABLE B.1: Primary Adoption Years

State	Year	In Sample
Alabama	1902	
Arizona	1909	✓
Arkansas	1900	
California	1909	✓
Colorado	1910	✓
Connecticut	1955	✓
Delaware	1969	✓
Florida	1902	
Georgia	1898	
Idaho	1909	✓
Illinois	1908	✓
Indiana	1918	✓
Iowa	1907	✓
Kansas	1908	✓
Kentucky	1912	✓
Louisiana	1904	
Maine	1911	✓
Maryland	1910	✓
Massachusetts	1911	✓
Michigan	1909	✓
Minnesota	1901	✓
Mississippi	1902	
Missouri	1907	✓
Montana	1912	✓
Nebraska	1907	✓
Nevada	1909	✓
New Hampshire	1909	✓
New Jersey	1911	✓
New Mexico	1939	✓
New York	1913	✓
North Carolina	1915	
North Dakota	1907	✓
Ohio	1913	✓
Oklahoma	1908	✓
Oregon	1906	✓
Pennsylvania	1907	✓
Rhode Island	1947	✓
South Carolina	1892	
South Dakota	1907	✓
Tennessee	1909	
Texas	1905	
Utah	1937	✓
Vermont	1915	✓
Virginia	1905	
Washington	1907	✓
West Virginia	1915	✓
Wisconsin	1906	✓
Wyoming	1911	✓

Note: Information drawn from Hirano and Snyder (2019), Table 2.A. States of the former Confederacy are excluded from the sample for the in-text analyses.

B.2 Descriptive Statistics

Descriptive statistics presented in Table B.2 include information for state-level covariates drawn from the U.S. Census (U.S. Census Bureau 1890 to 1930) and used in the OLS panel estimation strategy discussed below.

TABLE B.2: Summary Statistics

Statistic	Mean	Median	St. Dev.	Min	Max	N
Democratic Wins	0.29	0	0.45	0	1	6,222
Democratic Vote Share	0.44	0.45	0.14	0.00	1.00	5,015
Direct Primary	0.53	1	0.50	0	1	6,222
ln(Population) (State)	14.96	15.01	0.89	10.65	16.35	6,222
Percent Black (State)	0.03	0.02	0.03	0.001	0.21	6,222
Percent Other Race (State)	0.004	0.001	0.01	0.0000	0.16	6,222
Percent Foreignborn (State)	0.17	0.17	0.08	0.01	0.45	6,222
Percent Urban over 2,500 (State)	0.55	0.55	0.20	0.00	0.92	6,222
Percent Urban over 25,000 (State)	0.38	0.39	0.20	0.00	0.74	6,222
Civil Service Reform	0.35	0	0.48	0	1	6,222
Party Block Ballot	0.60	1	0.49	0	1	6,222
Office Block Ballot	0.33	0	0.47	0	1	6,222

Note: Sample is non-southern U.S. House elections, 1890 to 1930.

C Robustness Checks

C.1 Lagged Outcome Estimates

TABLE C.1: Incumbency Advantage with and without Direct Primary, Lagged Outcome

	Coefficient	SE	Bandwidth	Tot. Obs.	Used Obs.
Direct Primary	-0.001	0.008	0.084	2124	959
No Direct Primary	0.002	0.007	0.062	1353	649

Note: Point estimates and standard errors are "conventional" estimates from local linear regression implemented with the rdrobust function in R. Standard errors are clustered on state. Bandwidths are chosen using the rdbwselect function. The running variable is the two-party Democratic vote share, and the outcome is the lagged two-party Democratic vote share. Sample is non-southern U.S. House elections, 1890 to 1930.

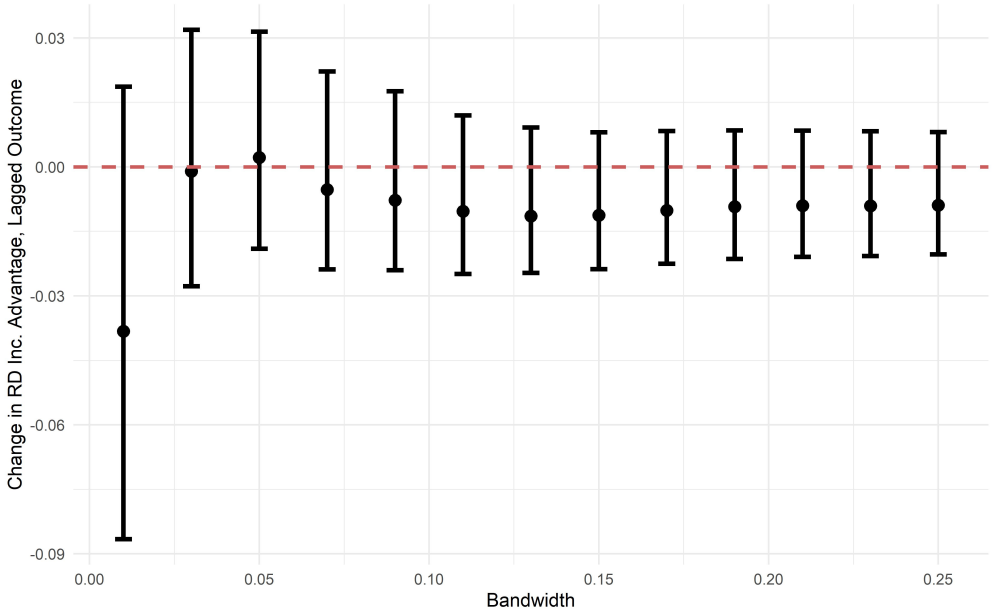

FIGURE C.1: Differences-in-Discontinuities Estimates, Lagged Outcome

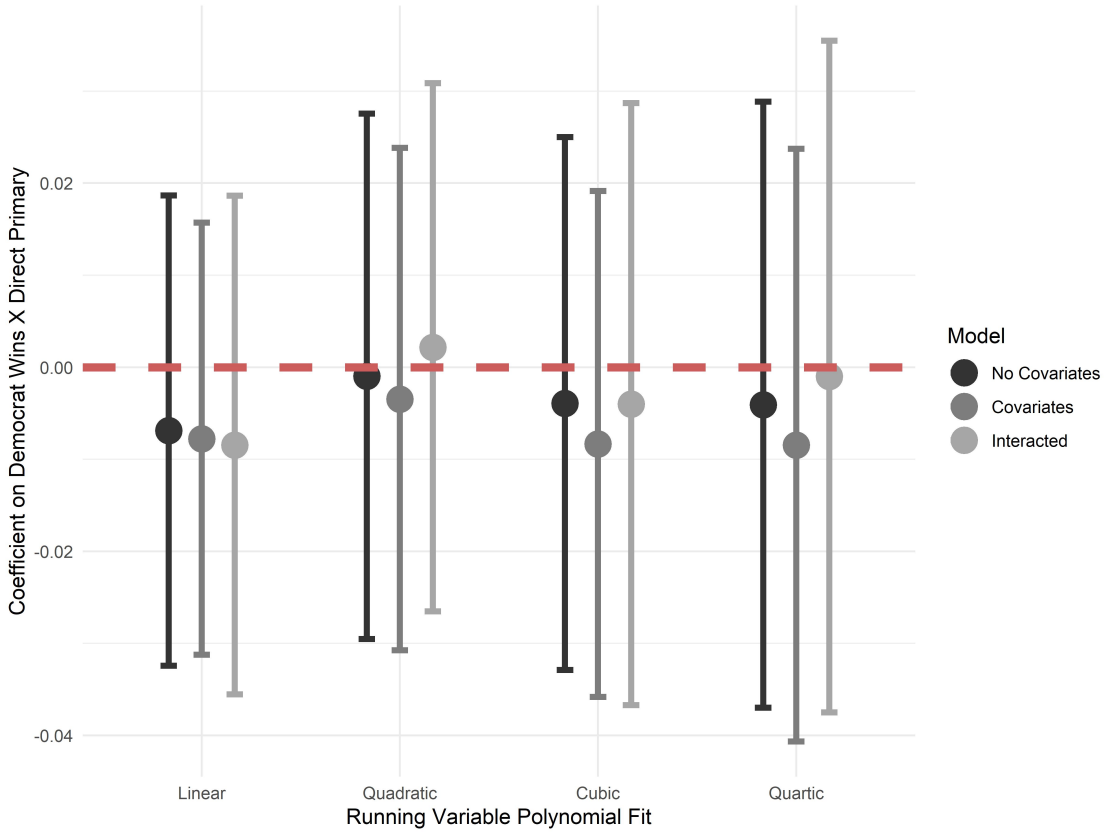

FIGURE C.2: Fixed Effects Regression Estimates, Lagged Outcome

figure presents estimates of the interaction between *DemWin* and *DirectPrimary* from the above fixed effects specification. The "no covariates" model includes no time varying covariates. The "covariates" specification additively includes the control variables described above. The "interacted" specification includes those same covariates, but also interacts institutional covariates for civil service reform and secret ballot usage with *DemWin*. 95% confidence intervals are based on state-clustered standard errors. sample is non-southern u.s. house elections, 1890 to 1930.

C.2 Linear Trends

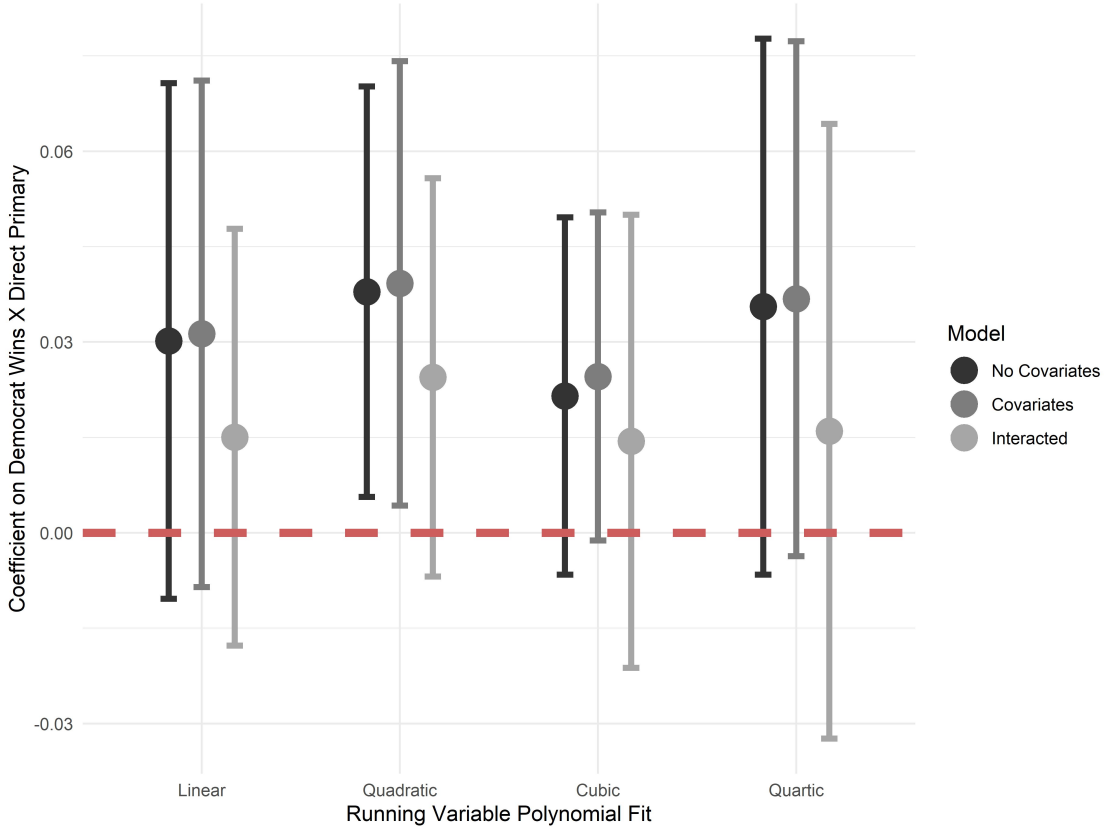

FIGURE C.3: Differences-in-Discontinuities Design, Linear District Trends

FIGURE PRESENTS RESULTS ANALOGOUS TO THOSE FROM FIGURE 4 IN THE TEXT, BUT USING MODELS THAT ADDITIONALLY INCLUDE A DISTRICT-SPECIFIC LINEAR TIME TREND. SAMPLE IS NON-SOUTHERN U.S. HOUSE ELECTIONS, 1890 TO 1930.

C.3 Parallel Trends

TABLE C.2: Parallel Trends Test

	Dependent variable:			
	Democratic Vote Share			
	(1)	(2)	(3)	(4)
Democrat Wins × Primary	0.040	−0.000	−0.012	0.039
	(0.033)	(0.040)	(0.045)	(0.043)
Democrat Wins × Primary (Lagged)	0.036	0.044	0.029	0.010
	(0.027)	(0.027)	(0.037)	(0.028)
Democrat Wins × Primary (Lead)	−0.041*	−0.014	0.010	−0.016
	(0.016)	(0.024)	(0.029)	(0.033)
Running Variable Poly. Order	1	2	3	4
District Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Covariate Control	✓	✓	✓	✓
Observations	3,426	3,426	3,426	3,426
Adjusted R^2	0.718	0.721	0.722	0.722

Note: Entries are linear regression coefficients with standard errors clustered on states in parentheses. All models include a polynomial fit of the running variable, interacted with an indicator for "Democrat Wins" and an indicator for the direct primary. Sample is non-southern U.S. House elections, 1890 to 1930. *p<0.05 (two-tailed test).

C.4 Regional Robustness

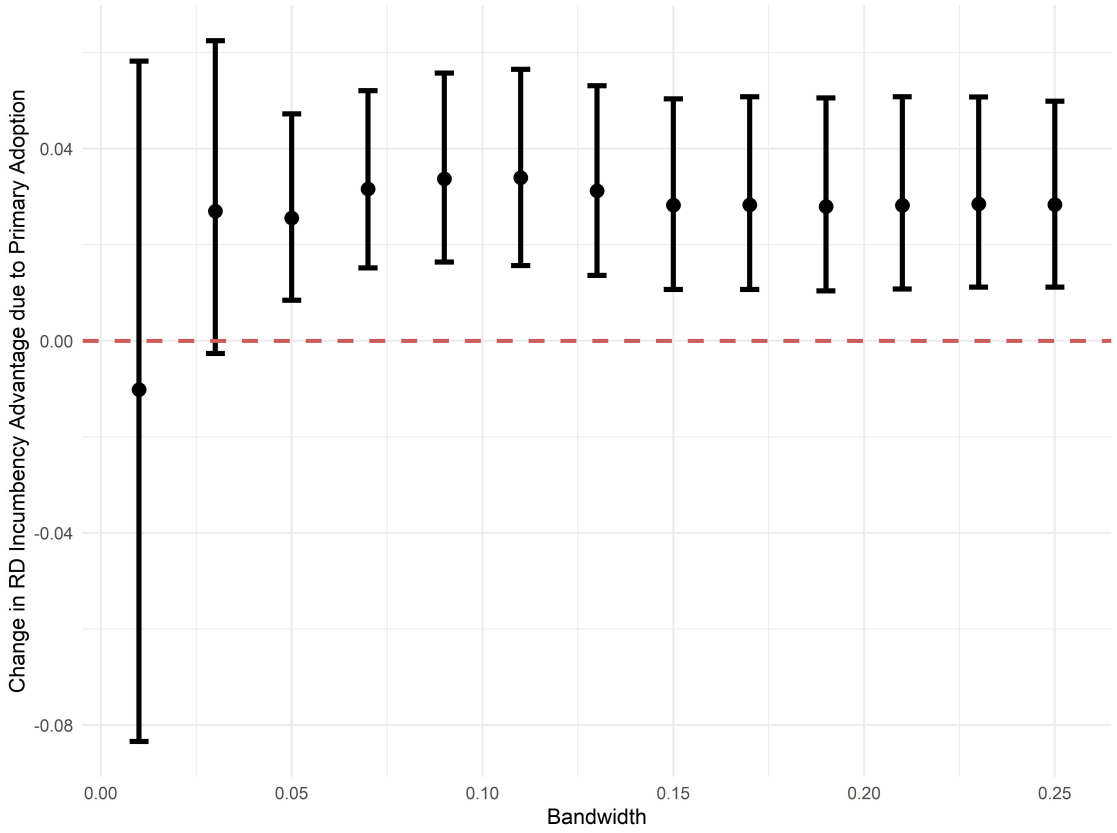

FIGURE C.4: Differences-in-Discontinuities, With South

figure presents results analogous to those from figure 5 in the text, but using a full sample of states, without omitting the south. 95% confidence intervals are based on state-blocked bootstrap. sample is u.s. house elections, 1890 to 1930.

C.5 Sample Robustness

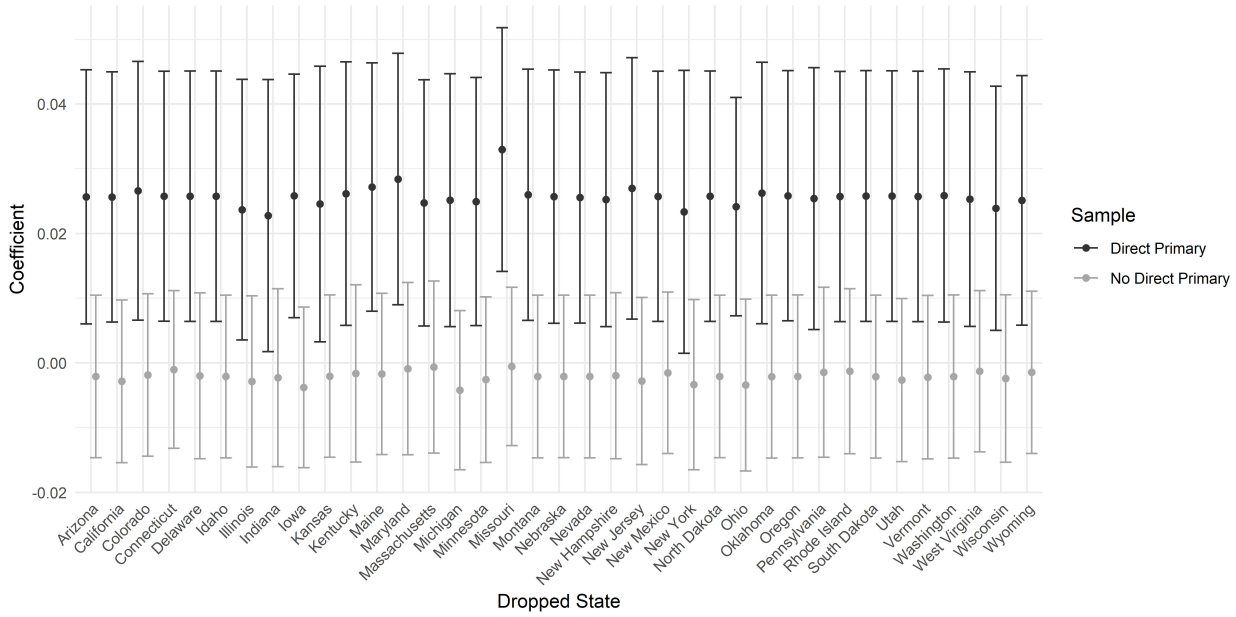

FIGURE C.5: Results, Iteratively Dropping States

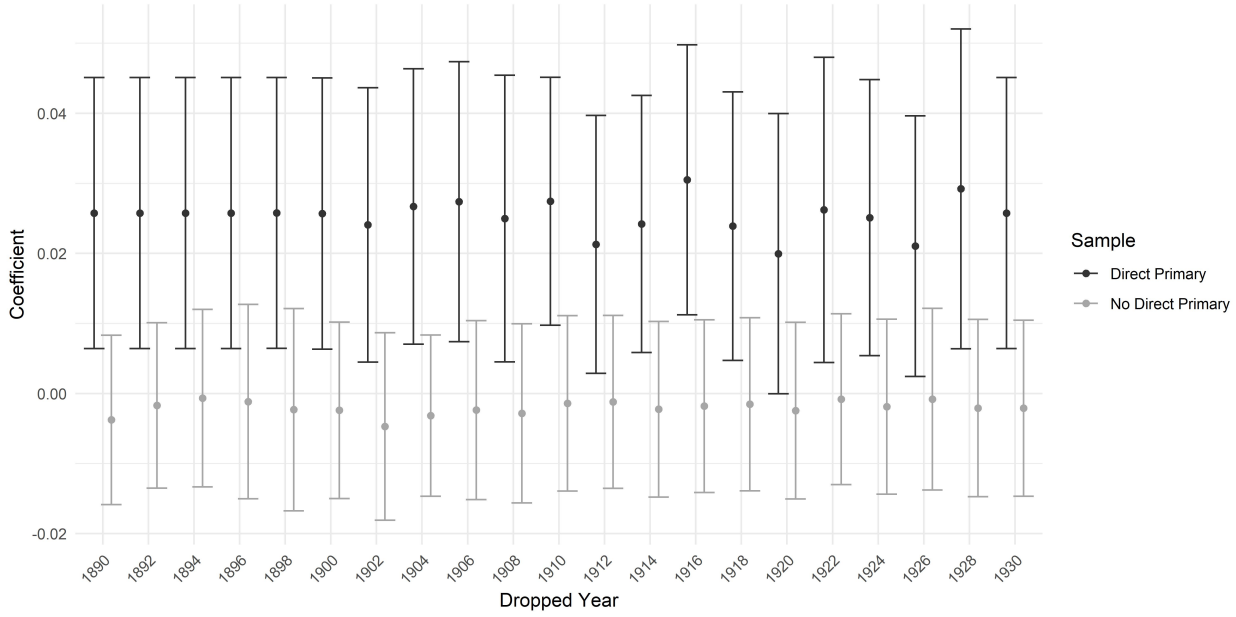

FIGURE C.6: Results, Iteratively Dropping Years

estimates in Figures C.5 and C.6 are based on optimal bandwidths calculated via rdbwselect. sample is u.s. house elections outside the u.s. south, 1890 to 1930.

C.6 Panel Estimates

TABLE C.3: Panel Models to Estimate Personal Incumbency Advantage

	Dependent variable:			
	Democratic Two-Party Vote Share			
	(1)	(2)	(3)	(4)
Democratic Incumbent × Primary	0.024*	0.021*	0.027*	0.019*
	(0.007)	(0.006)	(0.008)	(0.006)
Democratic Winner (t-1) × Primary	−0.009	0.001	−0.009	0.0005
	(0.005)	(0.005)	(0.008)	(0.006)
Incumbent Running × Primary	0.017*	0.028*	0.009	0.009
	(0.006)	(0.007)	(0.007)	(0.007)
Primary	−0.005	−0.038*		
	(0.015)	(0.010)		
Democratic Incumbent	0.015*	0.025*	0.015*	0.024*
	(0.004)	(0.004)	(0.004)	(0.004)
Democratic Winner (t-1)	−0.005	0.003	−0.002	0.008*
	(0.003)	(0.004)	(0.004)	(0.004)
Incumbent Running	−0.007	−0.008	−0.002	−0.002
	(0.004)	(0.004)	(0.005)	(0.006)
Dem. Two-Party Vote Share (t-1)		0.592*		0.515*
		(0.030)		(0.035)
District Fixed Effects	✓		✓	
Time Fixed Effects	Year	Year	State-Year	State-Year
Observations	3,925	3,477	3,925	3,477

Note: Entries are linear regression coefficients with standard errors clustered on states in parentheses. Sample is non-southern U.S. House elections, 1890 to 1930. *p<0.05 (two-tailed test).

References

Fowler, Anthony, and Andrew B. Hall. 2014. "Disentangling the Personal and Partisan Incumbency Advantages: Evidence from Close Elections and Term Limits." *Quarterly Journal of Political Science* 9 (4): 501–531.

Hirano, Shigeo, and James M. Snyder, Jr. 2019. *Primary Elections in the United States.* Cambridge University Press.

U.S. Census Bureau. 1890 to 1930. "State-Level Demographic Data, U.S. Decennial Census." *Social Explorer* .