Appendix
What Explains Incumbent Success? Disentangling Selection on Party, Selection on Candidate Characteristics, and Office-Holding Benefits

Defining PM, CS, and OB within a Potential Outcomes Framework

Consider an election *i* where an incumbent seeks reelection. Define

$V_i(1)$ = the two-party vote share of the incumbent.

This quantity is observed. Also define

$V_i(2)$ = the two-party vote share that the same candidate would have received had she not been an incumbent, and

$V_i(3)$ = the two-party vote share that would have been received by the incumbent party had the incumbent not sought reelection.

$V_i(2)$ and $V_i(3)$ are unobserved, counterfactual quantities. The definition of $V_i(2)$ asks us to imagine the same election and the same candidate seeking office, but where the incumbent is not an incumbent and does not benefit from any of the benefits of office holding. The definition of $V_i(3)$ asks us to imagine the same election, but where the incumbent has not sought reelection and another candidate from the same party runs in her place.

With these counterfactuals defined, we can define our quantities of interest as functions of these quantities:

$OB_i = V_i(1) - V_i(2)$,

$CS_i = V_i(2) - V_i(3)$, and

$PM_i = V_i(3) - .5$.

In the paper, the terms *officeholder benefit*, *characteristic selection*, and *party match* refer to average quantities across many elections. Therefore, we can formally define these quantities as the average values of OB_i, CS_i, and PM_i, respectively, across many elections:

officeholder benefit = $E[OB_i]$,

characteristic selection = $E[CS_i]$, and

party match = $E[PM_i]$.

PM_i tells us how well a non-incumbent from the incumbent party would do in this election. In most cases, we'd expect $PM_i > 0$, because the previous electoral winner is more likely to have come from the preferred party of the electorate. CS_i tells us how much better then incumbent candidate will perform compared to whatever non-incumbent candidate would run in her place if she stepped down. If $CS_i > 0$, the incumbent is better than this counterfactual candidate from the same party, and we'd expect this to be true more often than not if high-quality candidates are more likely to win elections and become incumbents. Lastly, OB_i tells us how much better an incumbent candidate performs because she is the incumbent.

This framework also clarifies why party match, characteristic selection, and officeholder benefit are exhaustive and mutually exclusive. By definition, all of an incumbent's electoral success ($V_i(1) - .5$) is explained by a linear combination of these three quantities:

$$OB_i + CS_i + PM_i = [V_i(1) - V_i(2)] + [V_i(2) - V_i(3)] + [V_i(3) - .5] = V_i(1) - .5.$$

Additional Test of the Exclusion Restriction (No Partisan Incumbency Advantage)

In estimating officeholder benefit, I assume that there is no partisan incumbency advantage. In other words, I assume that the only way that a close election result could influence subsequent elections is through personal incumbency. In the main text, I discuss several previous studies which provide empirical support for this assumption. Fowler and Hall (2014) estimate the partisan incumbency advantage in state legislatures with term limits and find that it is substantively small and statistically indistinguishable from zero. Butler and Butler (2006) and

Broockman (2009) find that close election results do not influence the way that the electorate votes for other offices in subsequent elections. For example, the result of a close Senate race does not influence the way that a state votes in its other Senate seat 2 or 4 years later, and the result of a close U.S. House election does not change the way a district votes in future presidential elections. These results are reassuring for my exclusion assumption, because it appears that election results only influence future voting insofar as they change the personal incumbency status of candidates running in the same setting in the future.

Here, I provide one additional test of this exclusion restriction. If the assumption of no partisan incumbency advantage holds, we should see no effect of one election result on subsequent elections in settings where incumbents are not allowed to seek reelection. The only such settings within the U.S., to my knowledge, are gubernatorial elections in states with a one-term limit. Currently, Virginia is the only state holding its governor to a limit of one term, but 14 other states (AL, FL, GA, IN, KY, LA, MS, MO, NC, OK, PA, SC, TN, and WV) had a one-term limit at some point in time since 1950. Using data from Alt, Bueno de Mesquita, and Rose (2011) on term limit laws, I isolate the sample of gubernatorial elections in my previous data set where governors faced a one-term limit. In other words, incumbents were never allowed to seek reelection. With this sample of 100 gubernatorial elections, I implement Lee's RD design (2008) to estimate the effect of one election result on voting in the next gubernatorial election (analogous to the reduced form equation in the estimation of β_3). Again, if the exclusion restriction is valid, the effect of election results on future voting should be zero because incumbents never seek reelection.

Table A1 presents the results of this test. Unfortunately, the results are imprecise, because there are only 100 relevant elections, and many of them were lopsided. However, across multiple specifications, I find no evidence that election results influenced future elections when incumbents are not allowed to seek reelection. Column 1 shows the basic RD design with the running variable included as a fourth-order polynomial. Columns 2 and 3 add state and year fixed effects to this specification, respectively. Instead of utilizing an RD design, column 4 estimates the effect of election results with a differences-in-differences design, meaning that the running variable is not included, but state and year fixed effects account for time trends and partisan differences across states. Again, the estimates are imprecise, because of data limitations, but I find no evidence against my assumption of no partisan incumbency advantage. This result, combined with previous evidence from different, independent sources, provides compelling empirical evidence for my identifying assumptions.

Table A1. Testing the Exclusion Restriction using Governors with a One-Term Limit

	DV = Vote Share			
	(1)	(2)	(3)	(4)
Lagged Victory	−.065	−.026	−.114	.000
	(.052)	(.063)	(.118)	(.060)
f(Lagged Vote Share)	X	X	X	
Year Fixed Effects			X	X
State Fixed Effects		X		X
Observations	100	100	100	100

State-clustered standard errors in parentheses. None of the point estimates are statistically distinguishable from zero.

Reweighting Procedures to Address Concerns about Heterogeneity

In Table A2, I presented pooled estimates (across all offices) implementing different reweighting schemes designed to make my local average estimates of β_2 and β_3 more comparable. As explained in the main text, β_2 is estimated through a differences-in-differences regression that implicitly puts more weight on the units that see more variation in incumbency, and β_3 is estimated with a fuzzy RD design that implicitly puts more weight on close elections. If the magnitudes of characteristic selection or officeholder benefit differ between the entire sample and the sample where we see more variation in incumbency, then my estimate of party match will be biased, and if the magnitude of officeholder benefit differs between settings with more close elections and those with more variation in incumbency, then my estimate of characteristic selection will be biased. Below, I discuss how I allow for certain kinds of heterogeneity by reweighting the data, and I show that the results are largely unchanged when relaxing different homogeneity assumptions, suggesting that these concerns have little implication for my main results.

Assume temporarily that the only relevant heterogeneity occurs across states. In other words, assume that the magnitudes of characteristic selection and officeholder benefit are homogenous within state but could differ across states.[1] We could reweight the three estimation steps so that they put the same relative weight on each state, allowing for comparable estimates. I accomplish this by reweighting the second and third estimation steps so that they put approximately the same weight on each state as the first estimation step where all elections are weighted equally. The fixed effects regression which estimates β_2 implicitly puts more weight on

[1] Heterogeneity in party match is irrelevant, because estimates of party match are not used to estimate either of the other quantities. Therefore, while we can imagine variation in party match across settings, it is not relevant for estimating characteristic selection or officeholder benefit without bias.

states that experience more variance in incumbency, so I reweight this regression by the inverse of the variance of incumbency in each state. The RD design which estimates β_3 implicitly puts more weight on states that have more close elections, so I weight by the inverse of the proportion of close elections in the state.[2] Replicating my empirical strategy with the addition of these weights, I can obtain new estimates of party match, characteristic selection, and officeholder benefit that allow for heterogeneity across states. Table A2 shows that this approach produces results that are nearly identical to those from unweighted approach.

Instead of worrying about heterogeneity across states, we might worry about heterogeneity across time, because, for example, incumbency advantages are thought to have increased over time (e.g., Cox and Katz 1996; Gelman and King 1990; Ansolabehere and Snyder 2002–and see Figure 1). Now, temporarily assume that the values of characteristic selection and officeholder benefit are homogenous within a given year but could vary across years. Under this assumption, we can reweight each observation according to the variation of incumbency in each year and the prevalence of close elections in each year (in the same way that we previously reweighted by state). This approach produces estimates that are robust to heterogeneity across time—as opposed to the previous results which were robust to heterogeneity across states. Again, the results are shown in Table A2 and they are virtually identical to the original, unweighted results. Therefore, whatever heterogeneity exists over time, it appears to hold no consequences for my estimates.

Lastly, I conduct a similar reweighting procedure for each state-decade, this time assuming homogeneity within each state-decade but allowing for heterogeneity across state-

[2] In calculating the proportion of close elections, I arbitrarily call an election *close* if the two-party vote share fell between 45 and 55 percent, although the results of this exercise are robust to different cutoffs.

decades. Again, the results in Table A2 are similar to the unweighted results, but the estimate of characteristic selection is smaller and the estimate of officeholder benefit it larger. However, characteristic selection is still statistically significant and not statistically distinguishable from the unweighted estimate. If there is heterogeneity in characteristic selection and officeholder benefit across states, time, or state-decades, it does not appear to meaningfully influence my estimates. Therefore, concerns about different local averages do not appear to pose a problem for my results.

I should point out that party match and characteristic selection cannot be estimated without imposing some kind of homogeneity assumption (e.g., within states, within years, within state-decades, etc.). As demonstrated by the re-weighting procedures we can relax the homogeneity assumption and allow for certain kinds of heterogeneity. However, we'll always have to make some kind of homogeneity assumption (e.g., characteristic selection and officeholder benefit are the same, on average, for some subset of election). Suppose, for example, that I recovered the implicit regression weights for each observation in the differences-in-differences analysis and reweighted each observation by its inverse weight. Then, I would simply recover β_1 from the cross-sectional regression and learn nothing about party match or characteristic selection. Similarly, if I conducted the RD analysis while reweighting by the inverse closeness of an election, I would again recover something similar to β_1 and learn nothing new. Conversely, if I tried to estimate equations 1 or 2 using only close elections, I would remove party match and characteristic selection and learn only about officeholder benefit.

Table A2. Pooled Results under Different Re-weighting Schemes

	PM	CS	OB
Unweighted	**.112**	**.021**	**.044**
	[.109,.115]	[.018,.024]	[.041,.047]
Re-weighted by State	**.112**	**.019**	**.046**
	[.108,.115]	[.015,.023]	[.043,.051]
Re-weighted by Year	**.110**	**.021**	**.046**
	[.107,.113]	[.018,.025]	[.042,.049]
Re-weighted by State-Decade	**.114**	**.012**	**.051**
	[.111,.118]	[.006,.017]	[.046,.057]

Block bootstrapped 95% confidence intervals in brackets.
The table presents pooled results under different re-weighting schemes designed to increase the comparability of different local averages, thereby increasing the validity of the comparisons of coefficients. Results are similar across all reweighting schemes, suggesting that concerns about different local average effects do not significantly alter the results.

Separate Estimates by Office and Decade

To be thorough, Figure A1 shows separate estimates of PM, CS, and OB for each office and each decade where data is available. In most settings, CS and OB increased over time, while PM was relatively constant, especially in non-southern states.

Figure A1. Variation over Time and Across Settings