SI Appendix for: No Evidence for Systematic Voter Fraud: A Guide To Statistical Claims About the 2020 Election

Andrew C. Eggers[a], Haritz Garro[b], and Justin Grimmer[c]

[a]Political Science. University of Chicago
[b]Democracy and Polarization Lab. Stanford University
[c]Corresponding Author. Democracy and Polarization Lab, Political Science, and Hoover Institution. Stanford University. 616 Jane Stanford Way, Stanford CA 94305. jgrimmer@stanford.edu.

September 21, 2021

A Identifying prominent statistical claims about the 2020 election

In the aftermath of the 2020 election, President Trump, his associates, and other skeptics advanced dozens of claims casting doubt on the election result. From this large set of claims, we sought to identify the most prominent *statistical* claims, by which we mean claims that rely on purportedly anomalous features of the official results. Such claims can be addressed by analysis of the official results that either re-examines the purported fact or assesses whether the fact is actually anomalous. Thus we steer clear of claims that require audits of submitted ballots, such as have been performed by election officials in Georgia, claims that rely on videos or other documentary evidence, and claims that relate to legal questions such as whether a state's electoral rules were constitutional. We lack any expertise to evaluate such claims.

To identify the set of prominent statistical claims, we constructed an initial list of claims by reviewing Trump's post-election speeches and social media posts. We then used a form of snowball sampling to expand the list: we looked for adjacent claims made in articles and social media posts mentioning claims on our original list; we also considered claims that arose when we shared our initial analysis on social media and on the Hoover Institution website. Although it would be impossible to address all vaguely statistical claims casting into doubt the 2020 election, we have done our best to focus on the most prominent ones. While we have decided what to include in part based on which claims are amenable to statistical analysis (i.e. claims based on publicly available data and with a comprehensible logic), we have not used the *results* of our statistical analysis to decide what to include (that is, we have not excluded claims on the basis that we cannot debunk them).

B More Votes than Voters

We attempted to identify as many potential objections to the 2020 election as possible and then we selected those claims that could be adjudicated with a statistical analysis. In this section, we discuss two additional claims about more votes than voters that were made nationally and in specific states.

More votes in PA, MI, and Nationally Some high-profile claims were simple and made obvious arithmetic errors. Perhaps the best example is the audacious claim that there were more votes than voters in the 2020 election. The original source of this claim seems to be a tweet from Bill Binney. His claim found a receptive audience among the 2020 election skeptics, including President Trump, who repeated the claim in a December 30th tweet.[1]

Binney's claim that there were more votes than voters comes from a confusion of different voting rates. The Washington Post published an article citing that 66.2% of the 239,247,182 people in the voting eligible population turned out to vote. Binney took this turnout figure, but applied it to the smaller group of 212,000,000 registered voters (at the time of his article being written). As a result, Binney computed:

$$\text{Binney's calculation:} \quad \overbrace{.662}^{\text{Turnout rate}} \times \overbrace{212 \text{ million}}^{\text{Registered voters}} = 140.344 \text{ million} \quad < \overbrace{158.254 \text{ million}}^{\text{Votes cast}}$$

Using this approach Binney estimates a number of voters (140 million) that is much lower than the reported number of votes cast (158 million).[2] Binney attributes the difference to fraud, asserting that this sort of evidence is hidden in plain sight.

In fact, the number of votes cast is the same as the number of voters; Binney's calculation is wrong. The problem is that the turnout rate he found in the Washington Post (.662) captures the proportion of *eligible* voters who voted, not the proportion of *registered* voters who voted. When the correct figure is used, the number of voters and the number of votes are equal:

$$\text{Correct calculation:} \quad \overbrace{.6615}^{\text{Turnout rate}} \times \overbrace{239.247 \text{ million}}^{\text{Eligible voters}} = 158.254 \text{ million} \quad = \overbrace{158.254 \text{ million}}^{\text{Votes cast}}$$

More votes than voters in PA, MI In addition to Binney's claim about the nationwide total, there were several allegations that turnout was impossibly high in particular states and localities, suggesting ballot box stuffing. 2020 election skeptics erroneously claimed there were more votes than voters in Pennsylvania, with essentially no evidence. In fact, claims about a too small number of votes occurred because state legislators making the claim failed to count provisional ballots in their vote total. In Michigan it was claimed several townships had greater than 100% turnout, but these claims are the result of an analyst confusing Michigan and Minnesota.[3] And claims of an implausibly large jump in turnout

[1] https://web.archive.org/web/20210108053918/https://twitter.com/realdonaldtrump/status/1344367336715857921

[2] Michael P. McDonald estimates that 158,254,139 million voters voted for the president office.

[3] See e.g. Aaron Blake, "The Trump campaign's much-hyped affidavit features a big, glaring error", *Washington Post*, November 21, 2020: https://www.washingtonpost.com/politics/2020/11/20/

in Wisconsin were, similar to Binney's claim above, based on using a different definition of the electorate for 2020 and previous elections. When we use comparable turnout figures, Wisconsin's turnout in 2020 is consistent with turnout in previous elections.[4]

C Biden's Bellwether Haul is Not Surprising

Figure 1 shows bellwether's odd-ratios from election specific regressions. This shows that bellwether counties are actually less predictive than other similar counties.

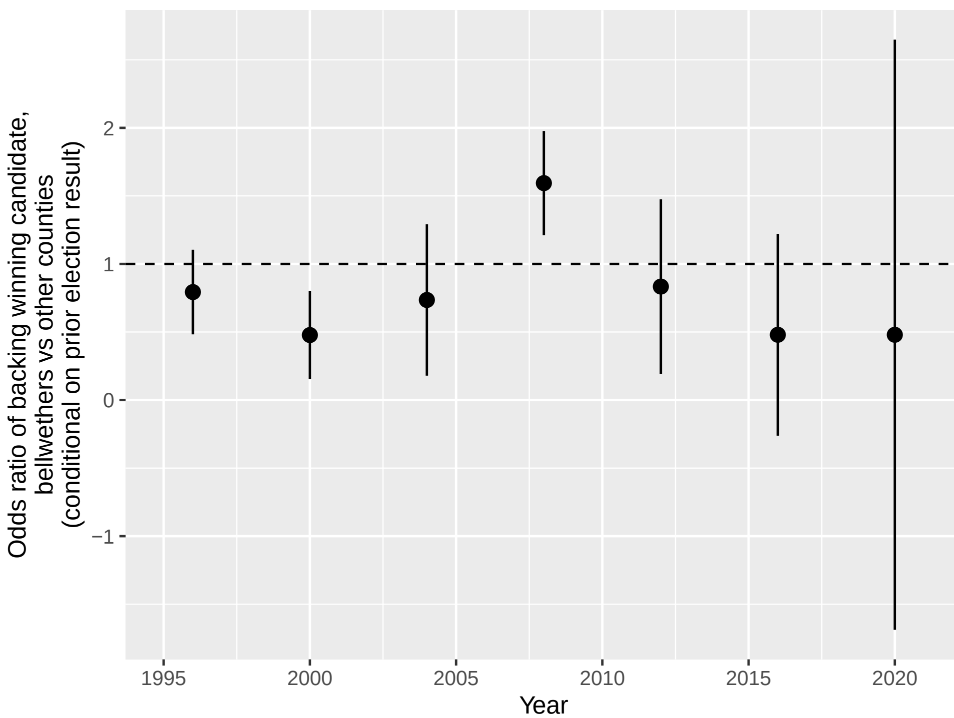


Figure 1: Difference in likelihood of supporting the winner (odds ratios) in bellwether counties vs other counties, conditional on support for winner's party in previous elections

In Table 1 we report similar model fits, but now using a linear probability model. Therefore, each coefficient can be interpreted as the difference in probability of a bellwether county selecting a winner. From this we can see that the probability of Biden winning a bellwether county, relative to other counties, was similar to historical performance of bellwethers.

Predicted Number of Biden Bellwether Wins We use a variety of methods to show that Biden's small number of bellwether county wins are the direct result of Trump winning the bellwether counties by a large margin. We perform this analysis in two stages.

trump-campaigns-much-hyped-affidavit-features-big-glaring-error/.

[4] Eric Litke, "Fact check: Wisconsin turnout in line with past elections, didn't jump 22% as claimed", *USA Today*, November 5, 2020 https://eu.usatoday.com/story/news/factcheck/2020/11/05/fact-check-wisconsin-voter-turnout-line-past-elections/6176028002/.

Table 1: Bellwether counties are no more predictive than other similar counties and often substantially less predictive. Each regression includes all bellwether counties as of the previous election and all counties whose absolute Democratic vote margin is lower than the largest absolute Democratic vote margin among the bellwether counties, ensuring that we include all counties whose two-party results were as close or closer than the results in bellwether counties.

	Dependent variable:							
	County vote for Winner in:							
	(2020)	(2016)	(2012)	(2008)	(2004)	(2000)	(1996)	(1992)
Bellwether	−0.063	−0.111	−0.021	0.037	−0.0004	−0.063	0.019	−0.192
	(0.056)	(0.058)	(0.035)	(0.022)	(0.018)	(0.016)	(0.018)	(0.016)
4^{th} Polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	948	922	1,413	2,338	2,306	2,826	2,509	3,096
R^2	0.769	0.311	0.664	0.612	0.640	0.549	0.618	0.591

First, we calculate the probability of Biden winning a county, given its 2016 Democratic margin and the results of the 2020 election. This is perhaps easiest to see with a method that we call "manual bins" in the main text. Here, we divided the space of Democratic 2016 margin into bins and then calculated the proportion of counties Biden won in those bins. For example, in the 2016 election Trump won 8 bellwether counties by 15 percentage points or more and Biden won no counties where Trump had this margin in 2016. So, Biden's expected share of victories for these counties is 0%. Trump won 4 bellwether counties by 10 to 15 percentage points in 2016 and Biden won only 1.6% of these counties in 2020, Trump won 6 bellwether counties by 5-10 percentage points and Biden won 9.8% of these counties, and finally Trump won 1 bellwether county by 0-5 percentage points and Biden won 58.6% of those counties. We can then use these numbers to calculate Biden's expected number of bellwether county victories

$$\text{Bellwether Wins} \quad = \quad 8 \times 0 + 4 \times 0.016 + 6 \times 0.098 + 1 \times 0.586 = 1.238 \tag{1}$$

Of course, we might object to the process of forming these manual bins to calculate these probabilities. To ensure our findings are robust to different ways of relating the 2016 Democratic vote margin to Biden's probability of victory, we use several regression methods. This includes generalized additive models, a logistic regression, a gradient boosted decision tree (GBDT), and a neural network. As we show in the table in Figure

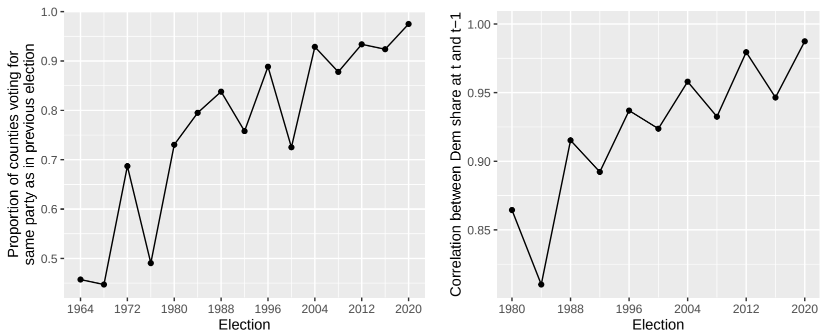
Biden's Poor Bellwether Performance Has Historical Precedent While President Trump and his allies have claimed that Biden's poor performance is without historical precedent, this is not true. Using data from 1960 we calculated the performance of bellwether counties at comparable intervals for prior elections. As it turns out, Biden's poor performance with bellwether counties mirror's Barack Obama's performance in 2008. The analogous bellwether counties in 2008 were counties that had been correct in every election since 1968. After the 2004 election there were 85 such counties and Obama won only 11 counties—an 87.1% decrease in the number of bellwether counties, similar to Biden's 94.7% decrease in counties.

Table 2: Biden's poor performance in Bellwether counties in 2020 has historical precedence: it is similar to Obama's performance in 2008

Year	% Δ in No. Bellwethers	Starting Year	Years Elapsed
2008	-0.80	1984	24
2020	-0.85	1996	24
2008	-0.72	1980	28
2020	-0.91	1992	28
2008	-0.85	1976	32
2020	-0.95	1988	32
2008	-0.85	1972	36
2020	-0.95	1984	36
2008	-0.87	1968	40
2020	-0.95	1980	40

Consistency of Election Results Our explanation for the poor performance of Biden in Bellwether counties is simple: Trump won those counties by a large margin in 2016 and there was a historically strong relationship between the 2016 and 2020 county level results. Figure 2 shows in two distinct ways that the serial correlation of county-level election results has increased over time, with 2020 showing the most serially correlated results in decades (and perhaps ever).

Figure 2: The left-hand plot shows the proportion of counties voting for the same candidate as in the previous election (higher in 2020 than in any other election shown). The right-hand plot shows the serial correlation of county-level Democratic vote shares (again, higher in 2020 than in any other election shown).



D Early and Late Differences in Cicchetti's Expert Report

Charles Cicchetti's other claim about statistically improbable results deals with comparisons of early- and late- counted votes. In the 2020 election, concerns around COVID-19 led a much larger share of citizens to vote by mail. In many states, Democrats were particularly likely to vote by mail—in part because Joe Biden encouraged his supporters to do so. In Pennsylvania, Georgia, and other states, election administrators were barred by law from counting these votes until election day. As a result, in many states Trump had a large lead in early counts but fell behind after mail-in ballots were counted. While this blue shift was expected, widely discussed and well documented [1],[5] this late shift in votes was regularly cited by Trump and his legal team as evidence of fraud.

So, in addition to testing the differences between 2016 and 2020, Charles Cicchetti also compared Joe Biden's support between early and late returns. In this analysis Cicchetti claims to test the null hypothesis that the early- and late-counted votes are random samples from the same population. This implies that Cicchetti would test whether Biden's vote share was the same in early- and late-counted votes. But instead, Cicchetti tests the null that Biden received the same *number* of votes from early- and late-counted votes. This is a perplexing choice: there were many more early-counted votes than late-counted votes, a fact that Cicchetti reports. As a result, even if Biden received the same share of votes in the early- and late-counted ballots, Cicchetti's test would produce a large test statistic.

Cicchetti assumes that every vote is an independent Bernoulli trial. This implies that the total number of early counted and late-counted votes for Biden are random variables that follow a Binomial Distribution and by the central limit theorem they will converge on a normal distribution. If T_{early} is the total number of early votes and $P_{\text{Biden,early}}$ is the proportion of early votes for Biden, then $B_{\text{early}} = T_{\text{early}} \times P_{\text{Biden,early}}$. Similarly, we can define the number of late votes for Biden as $B_{\text{late}} = T_{\text{late}} \times P_{\text{Biden,late}}$. By the central limit

[5] For example, see David A. Graham, "The 'Blue Shift' Will Decide the Election", *The Atlantic*, August 10, 2020 at https://www.theatlantic.com/ideas/archive/2020/08/brace-blue-shift/615097/.

theorem $B_{\text{early}} \sim \text{Normal}(T_{\text{early}} \times P_{\text{Biden,early}}, T_{\text{early}} \times P_{\text{Biden,early}} \times (1 - P_{\text{Biden,early}}))$, with the analogous normal distribution for late counted votes. Cicchetti's test statistic is then, test $= \frac{T_{\text{early}} \times P_{\text{Biden,early}} - T_{\text{late}} \times P_{\text{Biden,late}}}{\sqrt{T_{\text{early}} \times P_{\text{Biden,early}} \times (1-P_{\text{Biden,early}}) + T_{\text{late}} \times P_{\text{Biden,late}} \times (1-P_{\text{Biden,late}})}}$. This makes clear that even if the early- and late-counted votes had the exact same share of Biden votes $P_{\text{Biden,early}} = P_{\text{Biden,late}}$ Cicchetti would obtain a large test-statistic because of the massive differences in the number of votes in each category. In fact, it is easy to see that even in settings in which the Biden vote share is equal for both early and late votes, this test-statistic will depend on the number of early and late votes and on Biden's vote share. In short, this is a very poor test of the hypothesis Cicchetti sets out to test.

Thus, it is not surprising that he obtains a massive z-score of 1,891 which, Cicchetti notes, corresponds to an extremely small probability. If we test the more appropriate (but still wrong-headed) null, i.e. that Joe Biden had an equal vote share in early- and late-counted votes, we still obtain a large z-score of 282. This still indicates that there is a substantial difference between the early- and late-vote share for Biden.

It should be quite obvious that this difference is not, however, indicative of fraud in the election. The basic logic of Cicchetti's test is flawed: in a free and fair election there is no guarantee that a candidate's vote share will be the same in early- and late-counted ballots. If voters' preferences are correlated with how they vote, then systematic differences in vote share are likely to occur, even in a free and fair election. In Georgia and Pennsylvania, Democratic voters were more likely to use absentee ballots, and recently passed laws in both states forbid the counting of ballots until election day. As a result, there was a disproportionate number of votes from Democrats left to count at 3 a.m. on November 4th.

In Arizona, however, the correlation between when individuals cast their ballots and who they vote for was reversed: Democrats tended to cast ballots that were counted early, while Republicans tended to cast ballots that were counted later. So, when we apply Cicchetti's test in Arizona, we find systematic differences between early- and late-counted votes, with later votes favoring Trump. After election night, Joe Biden held a 93,016 vote lead in Arizona, with Biden receiving 51.7% of the two-party vote, and there were 604,375 Biden or Trump votes left to be counted. Among this group of late-counted votes, Biden received 43.2% of the two-party vote.[6] If we test the null that Biden received the same number of early- and late-counted votes we obtain an extremely large z-score of 1,263. If we instead test the null hypothesis that Biden's vote share was the same in the early and late votes, we obtain a more modest (though still exceedingly large) absolute z-score of 120.8.

Ballots are not submitted or counted at random. Rather, where ballots are cast and how they are cast affects when they are counted. Because location and method are both correlated with voters' preferences, we should (and did) expect shifts in vote totals or vote shares

[6] We use media reports to obtain the "early" vote total https://twitter.com/Politics_Polls/status/1324092133473689611?s=20, but we obtain identical numbers using tallies of late ballot counting from public sources https://alex.github.io/nyt-2020-election-scraper/all-state-changes.html.

E Evidence for P-Hacking in Dominion Report

The methods used by the original analysis are both unusual and opaque. The first step in the analysis appears to be a regression of Biden's 2020 county-level vote share on county-level predictors from the census. (The analysis does not indicate what predictors or model were used, and we were unable to replicate the results.) The second step in the analysis assesses how the prediction errors from the first regression relate to the type of voting machine used in the county. In the simplest version, the analysis compares the average residual across machine types; elsewhere the analysis regresses Biden's actual vote share in the county on the predicted vote share in the county separately for supposedly problematic machines and others.

Careful reading reveals that the headline claim in the report does not correspond to the analysis in the report. That claim is, "In counties using Dominion BMD voting machines, candidate Biden appears to have consistently received 5.6% more votes than he should have received." There are two sets of analysis claiming to show a 5.6% over-performance (a Chi-square automatic interaction detection (CHAID) analysis and a regression of observed on predicted), but neither analysis actually compares counties using Dominion machines to other counties: in both cases counties using one set of Dominion machines (Democracy Suite 5.5, or BMD) are combined with counties using Hart machines; this combined group of counties is then compared against others. Thus the headline claim that mentions only Dominion BMD is not correct: no analysis in the report finds a 5.6% over-performance for these machines separate from Hart machines.

The decision to focus on this set of counties raises another red flag about the analysis. The report offers no justification for analyzing Dominion BMD machines separately from Dominion D-Suite 4.14, or for lumping Dominion BMD machines with Hart machines. The reason they focus on these machines appears to be that the counties using these machines had the highest average residuals. But having defined the set of potentially problematic counties on the basis of the residuals, one cannot then *test* whether the residuals are especially high for this set of counties. The practice of choosing which tests to run on the basis of the results is known in social science as "fishing", and it is known to produce unreliable findings.[7] Following the critique we made in the previous section, it is not surprising that a researcher would find that Biden outperforms expectations in counties using Dominion BMD and Hart machines given that the researcher *chose* those machines because of Biden's performance in those counties.[8]

E.1 Robustness of Dominion Results

We check different ways of coding whether a county uses Dominion machines (based on the US Election Assistance Commission as in Table **??** or through our own hand coding of

[7] Social science research usually is presented in a way that makes fishing harder to detect, e.g. by providing some apparently principled justification for what was actually an *ex post* choice of specification.

[8] Put differently, the researcher's statistical software may state that the chance of finding an association as strong as the observed one is, say, .05, but given that the researcher was willing to fish for a significant association the true probability may have been .5 or higher.

counties in swing states on Table 3)[9]; controlling for census covariates (Table 5) or Biden's predicted performance (with predictions generated by a random forest regression) rather than Clinton's 2016 share Table 7; comparing Dominion BMD machines and Hart machines against others as in the original report rather than all Dominion machines (Table 6); conducting the analysis with vote margins rather than vote shares (Table 4). In some cases we find a pro-Biden difference between Dominion counties and others, but as soon as we control for the most obvious covariates the difference goes away or even changes sign. These null results are similar to null results on the use machines in 2004, but in that election the fraud was alleged against Republicans [2].

Table 3: Using an alternative coding of the presence of Dominion voting machines we continue to find no evidence that Dominion voting machines caused an increase in support for Biden. We code the presence of Dominion voting machines using Secretary of State websites in swing states (Arizona, Colorado, Florida, Georgia, Michigan, Minnesota, North Carolina, Nevada, Pennsylvania, Wisconsin, Florida, Texas).

	Dependent variable:			
	Biden Vote Share			
	(1)	(2)	(3)	(4)
Dominion Machines	0.030	0.013	−0.003	−0.006
	(0.010)	(0.002)	(0.003)	(0.004)
Clinton Share of Vote, 2016		0.994	0.979	0.967
		(0.008)	(0.007)	(0.007)
Observations	984	984	984	984
R^2	0.009	0.946	0.952	0.957
Dominion-State Fixed Effect			✓	
State Fixed Effects				✓

E.2 No Evidence Dominion or ES & S Transfers Votes

In a viral video posted to the moderation-light video site *Rumble*, Edward Solomon argues that the operators of voting machines in PA were involved in a surreptitious effort to steal votes from Trump and allocate them to Biden [3]. While Solomon describes an effort using Dominion machines, he later corrected himself because ES & S machines are used in the Philadelphia area. Solomon uses the live updates to vote counts from the New York times to argue that there is evidence that perpetrators had a target ratio of Trump to Biden votes in place to swing the election to Biden. He then alleges that he uncovers evidence of this ratio being "transferred" to precincts, as election officials attempted to subtly distribute votes to

[9] We obtained data on county voting machines for swing states from state official webpages.

Table 4: Dominion Voting Systems Did not Cause an Increase in Biden Votes. This table uses data from all states and the coding of Dominion voting systems from the US Election Assistance Commission, using Democratic margin as the DV.

	Dependent variable:			
	Democratic Margin, 2020			
	(1)	(2)	(3)	(4)
Dominion Machine	0.009	0.009	−0.001	−0.008
	(0.020)	(0.003)	(0.004)	(0.005)
Democratic Margin, 2016		1.030	1.028	1.007
		(0.003)	(0.003)	(0.003)
Observations	3,111	3,111	3,111	3,110
R^2	0.0001	0.975	0.975	0.980
Dominion State-Fixed Effects			✓	
State Fixed Effects				✓

Biden. In a 50 minute live video using an excel spreadsheet, he identifies ratios involving prime numbers and then asserts this “shouldn’t happen” because the distribution of the greatest common denominators are “too weird”.

Solomon does not, however, provide a statistical justification for his claim that the prime number ratios are in fact “weird.” Before taking on the statistical argument, we clarify two important inaccuracies in the Solomon story. The Solomon story has been popular with 2020 skeptics who have described this as conclusive evidence of invalid behavior. But in the rearticulation of Solomon’s analysis it is common to hear the incorrect perception that Solomon shows the repeated use of an *exact* ratio of Trump to Biden votes in the Philadelphia area. This is not what he shows: Solomon computes the greatest common denominator (GCD) and then forms the ratio of Trump to Biden votes. This will, by definition, be a whole number, but the actual ratio of Trump to Biden votes will have a remainder. So,in actuality, different precincts have different ratios, as we would expect with any data set. Beyond this very basic statistical misunderstanding, there is also the basic technical facts of the ES & S machines make it clear they couldn’t carry out the alleged fraud. The ES & S machines were not linked to each other in a way that would have enabled the transmission of the votes or the ability to reweight the votes. What’s more, the vote totals were in line with historical results for the Philadelphia area.

Specifically, Solomon does the following. He takes the vote counts for Trump and Biden in a precinct from a live update. He then calculates the greatest common denominator between the two numbers. He then divides the Trump and Biden total by this greatest common denominator and then uses these whole numbers to form the ratio between Trump and Biden. He then sorts his spreadsheet to identify ratios that he finds “odd”, including

Table 5: No evidence Dominion causes an increase in Biden vote share or margin when we condition on census covariates. We use the ACS to identify characteristics of counties and then adjust for those characteristics using both margin and vote share. The null effect is found for the coding of the treatment from the UEAC and from our own hand coding.

	Dependent variable:					
	Democratic Margin, 2020					Biden Share
	(1)	(2)	(3)	(4)	(5)	(6)
Dominion (Hand)	0.065	0.020	0.026	0.001		
	(0.017)	(0.003)	(0.004)	(0.005)		
Dominion (UEAC)					0.001	−0.0002
					(0.004)	(0.002)
Dem Mar., 2016		0.962	0.969	0.974	0.991	
		(0.006)	(0.006)	(0.008)	(0.004)	
Dem Share, 2016						1.008
						(0.005)
Log(Population)	0.070	0.008	0.008	0.012	0.013	0.007
	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)
% Female	0.003	0.0002	−0.0001	−0.0002	0.001	0.0005
	(0.003)	(0.001)	(0.001)	(0.001)	(0.0003)	(0.0002)
% Black	0.008	0.0001	−0.00001	−0.00004	0.0001	−0.0002
	(0.001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
% Asian	0.033	0.005	0.005	0.004	−0.001	−0.001
	(0.006)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0002)
% Hispanic/Latino	0.002	−0.001	−0.002	−0.002	−0.002	−0.001
	(0.0005)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)
Median HH Income	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
% Over 65	0.006	0.0002	0.0002	0.0003	−0.0003	−0.0005
	(0.002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0001)
Observations	985	985	985	984	3,110	3,110
R^2	0.389	0.979	0.979	0.982	0.987	0.986
Dominion-State			✓			
State				✓	✓	✓

Table 6: No Evidence Dominion/Hart machines increase in Biden's turnout, replicating coding from original paper. We construct the independent variable as a more recent Dominion machine or Hart machine present in the county. We then compare those counties to any other county that has a voting machine. While there is a bivariate relationship between the Dominion/Hart machines and Biden's performance, this a precisely estimated zero once we adjust for Clinton's performance in 2016 and State-fixed effects.

	Dependent variable:							
	Biden Vote Share				Biden Vote Margin			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dominion 5.5, Hart	0.032	−0.0003	−0.001	0.002	0.059	−0.001	0.010	0.005
	(0.008)	(0.011)	(0.002)	(0.002)	(0.016)	(0.022)	(0.003)	(0.004)
Clinton Share			1.025	1.028				
			(0.005)	(0.005)				
Clinton Margin							1.029	1.022
							(0.004)	(0.004)
Observations	1,720	1,720	1,720	1,720	1,720	1,720	1,720	1,720
R^2	0.009	0.258	0.961	0.972	0.008	0.254	0.972	0.977
State fixed Effects		✓		✓		✓		✓

Table 7: Dominion machines do not cause an increase in vote share for Biden. Here, we use the Dominion coding from the UEAC and adjust for the covariates using predictions from a random forest regression, rather than Clinton's performance in the prior election. Again, we find no significant difference in favor of Biden's vote share or vote margin.

	Dependent variable:			
	Biden Share		Biden Margin	
	(1)	(2)	(3)	(4)
Dominion	−0.007	−0.015	−0.020	−0.032
	(0.005)	(0.008)	(0.009)	(0.015)
Predicted Biden Share	1.099	1.094		
	(0.015)	(0.015)		
Predicted Biden Margin			1.091	1.087
			(0.015)	(0.015)
Observations	1,720	1,720	1,720	1,720
R^2	0.760	0.820	0.753	0.814
State Fixed Effects		✓		✓

ratios involving prime numbers. He is particularly concerned that ratios he views as "odd" are found before and after an update, but found in two different precincts. When an "odd" or "suspicious" ratio appears in two different precincts, Solomon calls that a "transfer". He never, however, explains why he thinks these ratios are odd, other than the ratios are personally surprising to him.

To assess whether Solomon's analysis could possibly uncover evidence of fraud, we conduct a simulation and see that the "transfers" Solomon identified are actually quite common in a fraud-free election. This is because Solomon's analysis is a classic example of how fishing for a finding and lack of a null distribution to compute "anomalous" results can lead an argument astray. Solomon fishes for results when he identifies "weird" ratios without first defining what is "weird". And a null distribution that corresponded with the rate these "weird" ratios would occur in a free and fair election would enable him to compute a p-value, to at least provide a sense of his uncertainty.

We used Solomon's data to show that there does not seem to be anything particularly suspicious in the counts or ratios in the live results for the Philadelphia area. In fact, we can replicate the Solomon-like analysis with our simulated data, which—by definition–randomly generates the results. Specifically, we randomly generated results from 3,110 precincts 4 times. We then identified a "suspicious" ratio: 40 Trump Votes to 197 Biden votes in the third update in our simulation. As it turns out, we find four precincts that "inherited" (in Solomon's words) this same ratio, even though the simulation is purely random. So even though Solomon says "this shouldn't happen" we in fact find several instances of patterns like this happening.

In short, it is difficult to reason about the distribution of ratios in data. A formal statistical analysis is necessary to know if any of the patterns Solomon identified are truly anomalous. This formal analysis would be useful because it forces the analyst to define what they are testing, why they think it is weird, and how it doesn't correspond with the results we would find under reasonable models of how data are generated.

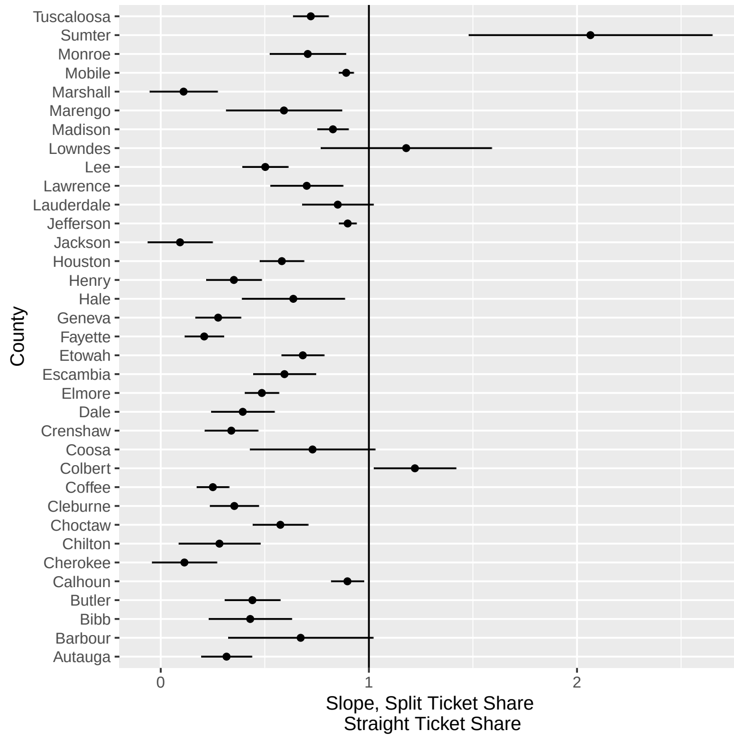


Figure 3: Across Alabama counties the slope from a regression of McCain's split-ticket share on his straight-ticket share is less than 1, consistent with a regression to the mean explanation for the "suspicious" pattern Ayyudarai identifiess.

F Alabama Counties Show Negative Relationship Between Split-Ticket Shares and Straight-Ticket Shares

We provide a theoretical argument to explain why the negative relationship Shiva Ayyudrai identifies should be expected. We also demonstrated that in several counties in Alabama that there was a negative and linear relationship between McCain's split-ticket vote share in 2008 and his straight-ticket vote share. In Figure 3 we formally test the relationship between straight-ticket and split-ticket vote shares. Within each county we performed a linear regression of McCain's precinct-level split-ticket share and his straight-ticket share. Figure 3 provides the estimated slope coefficient and the 95% confidence intervals.

Figure 3 shows that in 32 of the 35 counties the regression of McCain split-ticket share on his straight-ticket share has a coefficient less than 1 and in 29 of those counties we reject the null of equality to 1. This implies that in the vast majority of counties we find the negative relationship between split-ticket and straight-ticket voting shares that Ayyudarai argued was evidence for fraud.

G Demonstration of Lott's (2020) Specification Error

When we replace Lott's unusual specification with a more standard approach that does not depend on arbitrary coding rules, we find absolutely no evidence for fraud in either Fulton County or Allegheny County. In this section we characterize the error in Lott's original analysis.[10]

Lott (2020) claims that a comparison of adjacent election precincts in Georgia and Pennsylvania supports the Trump campaign's allegations that the 2020 presidential election was "stolen" through fraud. In Lott (2020)'s abstract, he estimates that fraud in Fulton County contributed 11,350 votes to Biden (which would account for nearly all of Joe Biden's margin of victory in Georgia) and fraud in Allegheny County contributed about 55,270 votes to Biden's victory in Pennsylvania (which would account for around 2/3 of Biden's margin in Pennsylvania). To make this claim about absentee ballots, Lott intends to tests the null hypothesis that, after controlling for all relevant factors, there is no average difference in Trump's absentee support as we move from precincts in Fulton County to adjacent precincts in bordering Republican counties. To eliminate some of these alternative explanations for differences in Trump's absentee support between "suspect" counties and neighboring counties, Lott (2020) focuses on precincts that lie along county borders. Specifically, he forms pairs of precincts that lie along a boundary separating a suspect county (i.e. one where Republicans have alleged that fraud took place) and an adjacent county where Trump won a majority of the vote and no fraud allegations have been made.[11] Lott (2020) also forms pairs of precincts that lie along the boundary between two of these Republican counties, which serve as a kind of control group for the other pairs. Lott (2020) then conducts his analysis using within-pair *differences* in each variable: he regresses the difference in Trump's share of the absentee vote between the two precincts on the difference in Trump's share of the in-person vote between the two precincts and an indicator for whether the pair contains a precinct in a suspect county.[12] That is, his basic regression equation is

$$\left(\text{Absentee}_i - \text{Absentee}_j\right) = \beta_0 + \beta_1\left(\text{InPerson}_i - \text{InPerson}_j\right) + \delta\text{SuspectCounty}_i + u_{ij},$$

where Absentee_i is Trump's share of the absentee vote in precinct i, InPerson_i is Trump's share of the in-person vote in precinct i, SuspectCounty_i indicates whether precinct i is located in a "suspect" county, and i and j are adjacent precincts that Lott assigns to a pair. Thus, β_0 measures the within-pair difference in Trump's share of the absentee vote among pairs that don't involve a suspect county (adjusting for the within-pair difference in Trump's in-person share). The key coefficient is δ, which compares the adjusted difference in Trump's share of the absentee vote within pairs involving the suspect county against the corresponding adjusted difference within pairs not involving the suspect county. The

[10] After we posted an original version of this analysis, Lott retracted this analysis. Nevertheless, because the original claims were so widely viewed and disseminated, we think it is essential to explain the logic of why Lott was wrong.

[11] Lott (2020) provides no justification for not comparing Fulton and Allegheny counties (or others where fraud was alleged) with surrounding counties carried by Biden. By ruling out these comparisons, Lott severely restricts his sample size and likely excludes the most similar comparisons.

[12] In some specifications he also includes differences in various race-and-gender groups between the two precincts.

underlying logic seems to be that fraud is the likely explanation if there is a bigger drop in Trump's share of the absentee vote when we cross from, for example, Coweta County to Fulton County than when we cross from Coweta County to Carroll County, which are two Republican counties where no fraud has been alleged.

Even if we stipulate that focusing on adjacent precincts eliminates all between-county differences in true absentee support for Trump (conditional on Trump's in-person support),[13] Lott (2020)'s design suffers from a fatal flaw. As noted, Lott (2020)'s design measures a difference between two differences: is the drop in Trump's share of the absentee vote larger when we cross the Fulton County border into Republican counties than when we cross the border of one Republican county into another Republican county? The problem arises in measuring the second drop: there is no clear rule for determining the order of the difference. For example, should we record the change in Trump's absentee vote share as we move from Carroll to Coweta, or as we move from Coweta to Carroll? Neither county is "suspect", so either approach could be justified. Lott (2020, footnote 13) chooses one rule (subtracting east from west and north from south) but the opposite rule or indeed any rule would be equally justified. This arbitrariness is a symptom of the underlying lack of compelling logic behind this aspect of the design: there is no clear reason to benchmark the difference in voting patterns across the key county boundary against the corresponding difference across another boundary.[14]

As it turns out, Lott (2020)'s evidence for fraud in Fulton County, GA, and Allegheny County, PA, relies entirely on this arbitrary coding rule: if a different but equally valid rule is used, we reach the opposite conclusion from Lott (2020). Figure 4 illustrates the point for Fulton County. In both panels, each red dot corresponds to a pair of precincts lying on opposite sides of the Fulton County boundary; each blue dot corresponds to a pair of precincts lying on opposite sides of the boundary between two nearby Republican counties. The vertical axis shows the difference in Trump's share of the absentee vote within the precinct pair; the horizontal axis shows the difference in Trump's share of the in-person vote within the precinct pair.

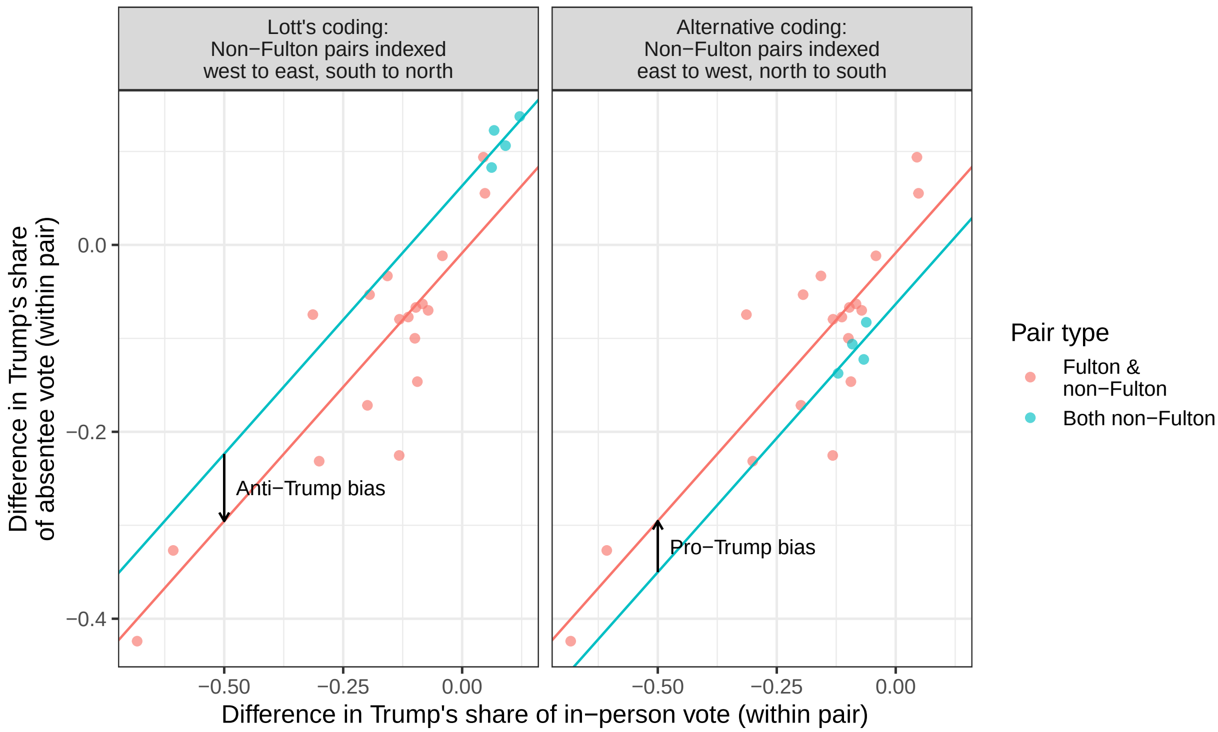
The left panel of Figure 4 shows the analysis using Lott (2020)'s coding: for pairs including a Fulton County precinct, the Trump share for the non-Fulton County precinct is subtracted from the Trump share for the Fulton County precinct; for pairs not including a Fulton County precinct, Lott (2020) uses the arbitrary rule noted above. This coding results in what Lott interprets as evidence for anti-Trump bias in Fulton County. Conditional on the difference in Trump's in-person vote share within a precinct pair, the difference in Trump's absentee vote share is lower in precinct pairs involving Fulton County than in other precinct pairs.

In the right panel of Figure 4, we show that the conclusion is reversed when we reverse Lott's arbitrary coding rule: instead of subtracting east from west and north from south in

[13]This is doubtful. For example, Trump won just 9.6% of the in-person vote in a precinct in Fulton County (FA01B) that is adjacent to a precinct in Coweta County where Trump won 78% of the in-person vote (Fischer Road). It seems unlikely that precincts that differ so markedly in voting outcomes would be similar in e.g. voters' propensity to vote in person vs. absentee conditional on their vote choice.

[14]One could imagine a better design that compared the *magnitude* (i.e. absolute value) of differences across suspect boundaries and other boundaries. In this case, the ordering of precinct pairs would not matter. This is not Lott's design.

Figure 4: Evidence for fraud in Fulton County, GA, is reversed if arbitrary coding rule is reversed



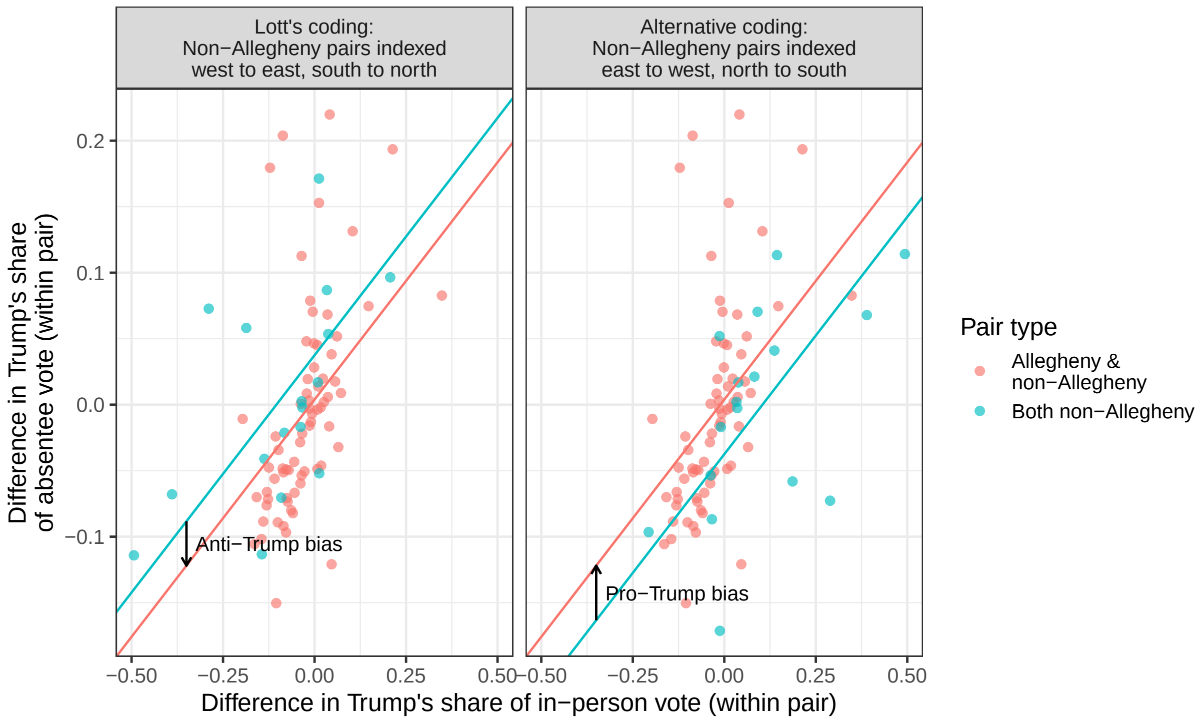
computing differences for non-Fulton precinct pairs, we subtract west from east and south from north. The scatterplot looks identical to the left panel except that the four blue dots (representing non-Fulton precinct pairs) are reflected through the origin. Notably, this small change reverses the conclusion: by Lott (2020)'s logic we now have evidence of pro-Trump bias in Fulton County.

Table 8 (Appendix) reports coefficient estimates and standard errors for both sets of analyses depicted in Figure 4. The evidence of pro-Trump fraud with the alternative coding rule has a similar absolute t-statistic ($t = 1.67$) as Lott's evidence of anti-Trump fraud with the original coding rule ($t = 1.89$).

The Pennsylvania results also depend on Lott's arbitrary coding rule, as we show in the same manner in Figure 5 and Table 9 (Appendix). Lott (2020) concludes from his analysis that anti-Trump fraud took place in Allegheny County. But, if we apply a different but equally valid coding rule, we find (by the same logic) stronger evidence for *pro-Trump* fraud in Allegheny County: the positive coefficient we obtain with the alternative coding rule is both larger in magnitude and more significant than the negative coefficient Lott reports.

We can further highlight the dependence of Lott's results on arbitrary coding decisions by exploring the universe of possible fraud estimates that Lott could have reported with equally justified alternative coding rules. In Figure 6 we show that, among the possible rules that could be used, any alternative rule would have produced weaker apparent evidence for anti-Trump fraud in Fulton County and almost any rule would have produced weaker

Figure 5: Evidence for fraud in Allegheny County, PA, is reversed if the arbitrary coding rule is reversed



evidence for anti-Trump fraud in Allegheny County.[15] In the Fulton County analysis, there are four non-Fulton precinct pairs and thus $2^4 = 16$ possible rules for computing differences within non-Fulton pairs. The left panel of Figure 6 shows the histogram of the key coefficient across these sixteen possible rules, with a vertical line highlighting the estimate for the rule Lott used. Among the sixteen possible rules, Lott's rule produces the strongest apparent evidence of anti-Trump fraud; six possible rules produce apparent evidence of pro-Trump fraud. In the Pennsylvania analysis we have seventeen non-implicated precinct pairs, allowing for over 130,000 possible coding rules. The right panel of Figure 6 shows the distribution of estimates for a random sample (with replacement) of 100,000 of these rules,[16] with the actual estimate again shown with a vertical line. The distribution is centered around zero, with roughly as many rules producing apparent evidence of pro-Trump and anti-Trump fraud; Lott's rule again happens to produce among the strongest apparent evidence of anti-Trump fraud. Figure 7 shows that we get nearly identically shaped histograms if we focus on the t-statistics–with t-statistics that would be significant at conventional levels for either pro-Trump or pro-Biden bias.

[15]It appears that the ordering of precincts follows a rule in a prior American Economics Review paper. We believe that is Bronars and Lott (1998).

[16]To explore the space of changes to the difference order, we first sample the number of difference orders to change from a Uniform(1, 16). Once this number is obtained, we then randomly sample the specific units that will have the difference order changed. This explores the space, but does not provide a sampling distribution that gives an equal probability to each rearrangement, because our sampling method is biased towards either too few or too many rearrangements.

Figure 6: Evidence for fraud in Georgia and Pennsylvania depends on arbitrary coding rules; Lott's estimates are outliers in the distribution of estimates

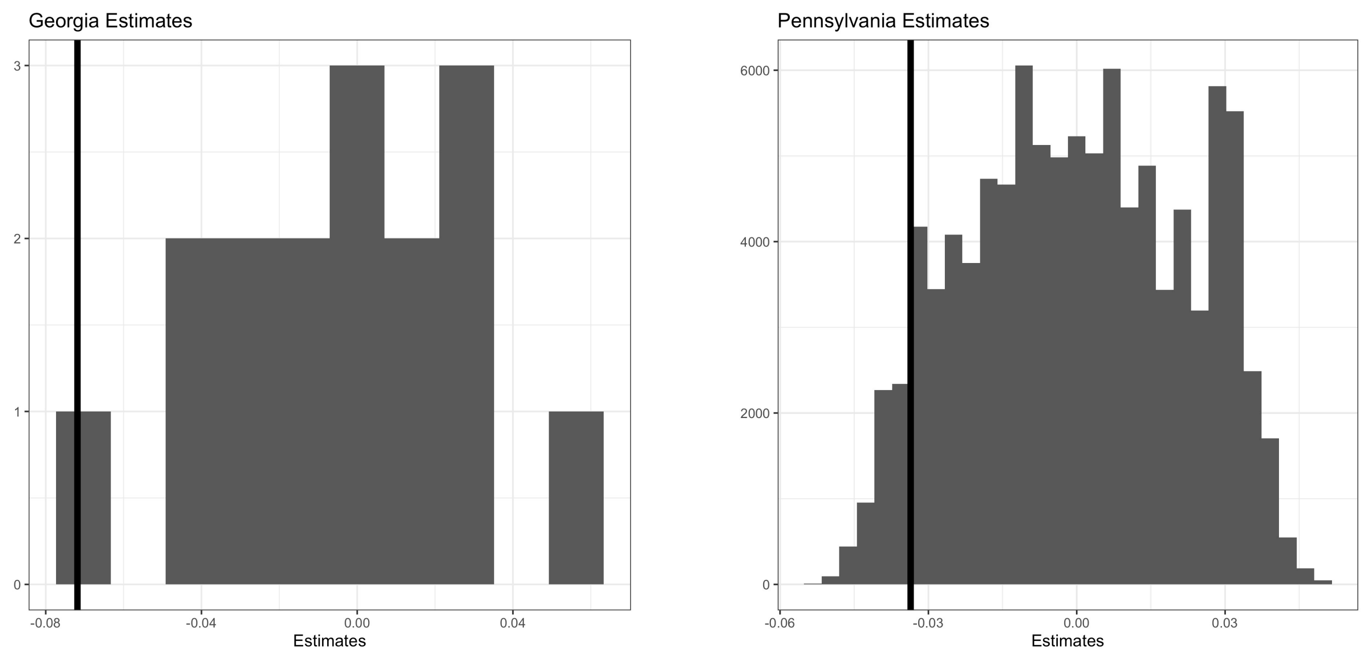
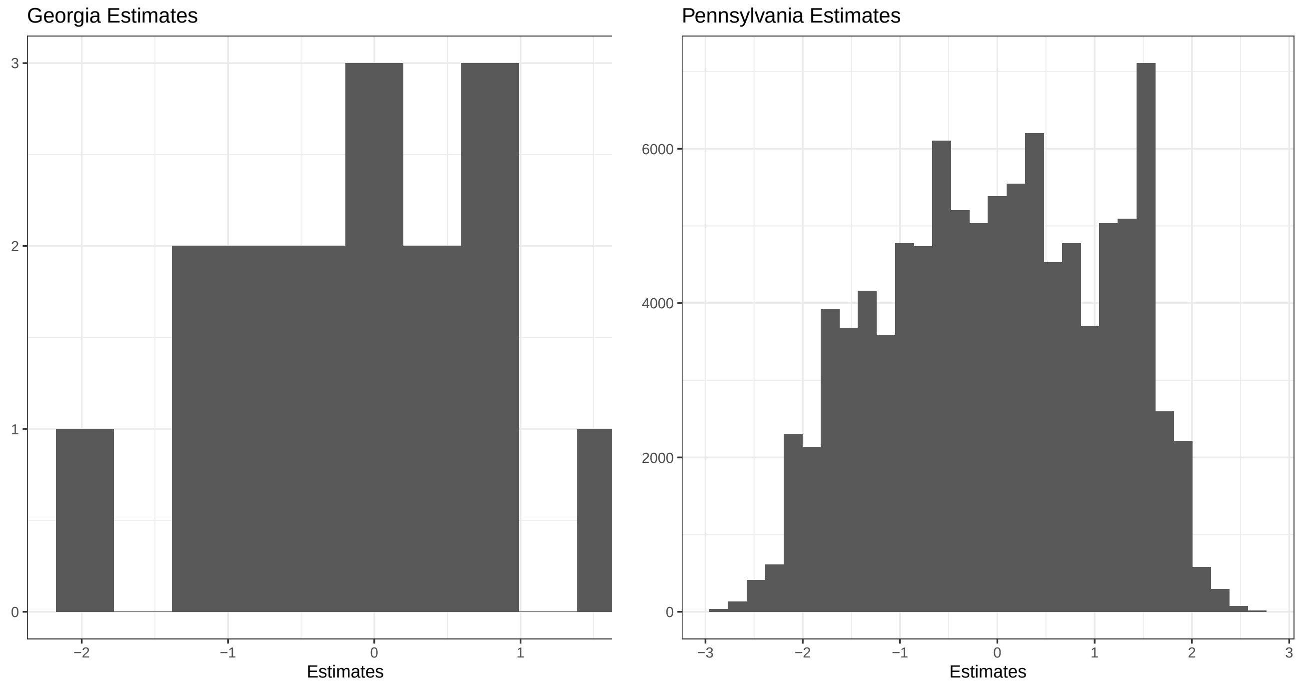


Figure 7: Evidence for fraud in Georgia and Pennsylvania depends on arbitrary coding rules; Lott's estimates are outliers in the distribution of estimates



G.1 Additional Results for Lott's (2020) Absentee Vote Analysis

Table 8: Lott's Conclusions Are Reversed if the Arbitrary Ordering of Precinct Differences is Reversed (Georgia)

	Dependent variable:	
	Difference, Trump Absentee (Lott (2020), Table 2)	
	(1)	(2)
Difference, Trump In-Person Vote	0.574	0.574
	(0.073)	(0.073)
Fulton County	−0.072	0.055
	(0.038)	(0.033)
Observations	22	22
Reverse Coding		✓

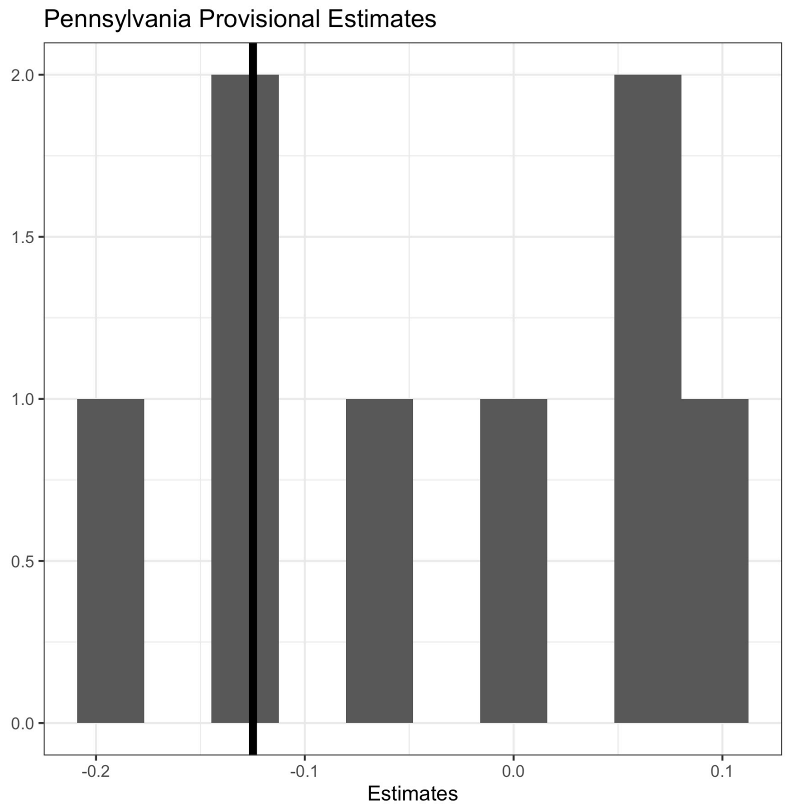
Table 9: Lott's Conclusions Are Reversed if the Arbitrary Ordering of Precinct Differences is Reversed (Pennsylvania)

	Dependent variable:	
	Difference, Trump Absentee (Lott (2020), Table 5)	
	(1)	(2)
Difference, Trump In-Person Vote	0.359	0.359
	(0.069)	(0.069)
Allegheny County	−0.034	0.041
	(0.019)	(0.020)
Observations	87	87
Reverse Coding		✓

Table 10: Pennsylvania Provisional Ballot Results

	Dependent variable:			
	Difference, Trump Provisional (Lott (2020), Table 6)	Trump Provisional Vote		
	(1)	(2)	(3)	(4)
Difference, Trump In-Person Vote	1.038			
	(0.558)			
Trump, In-Person Vote		0.729	1.055	0.690
		(0.222)	(0.552)	(0.257)
Allegheny County	−0.125	−0.004	−0.036	−0.047
	(0.141)	(0.036)	(0.044)	(0.048)
Observations	34	120	120	120
Precinct-Pair Fixed Effects			✓	
County-Pair Fixed Effects				✓

Figure 8: Distribution of Estimates for Alternative Precinct Differencing Orders, Pennsylvania Provisional Ballots



H Selection Issues in Lott's (2020) Turnout Analysis

Before digging deeper into Lott (2020)'s turnout analysis, we emphasize that we dispute the premise of Lott (2020)'s analysis; that is, we do not believe that even a robust finding of slightly higher than expected turnout in a set of counties Republicans targeted in post-election lawsuits would constitute convincing evidence of electoral fraud. The differences Lott claims to have found are small (1-2 percentage points), and in the absence of fraud, turnout is not perfectly explained by the covariates that Lott (2020) uses: a particularly energetic local mobilization campaign (on either side) or an especially effective down-ballot candidate could affect turnout by these amounts. Perhaps more to the point, Lott (2020) looks for unexplained turnout in places Republicans chose to target in post-election lawsuits. We do not know how Republicans chose which counties to target, but it seems plausible that they targeted counties based on district characteristics that are related to turnout (but not modeled by Lott (2020)) or even based on observed results (including turnout). This creates a thorny selection problem: was fraud the cause of high turnout, or was high turnout the cause of allegations of fraud?[17] Highly anomalous turnout figures could provide evidence of a problem, but a percentage point or two of unexplained turnout has other more plausible

[17] Thus we could see Lott and Republican legal teams as engaged in a joint fishing expedition similar to the one we describe above in the Dominion analysis.

Figure 9: Distribution of Estimates for Alternative Precinct Differencing Orders, Share of Biden Ballots from Pennsylvania Provisional Ballots

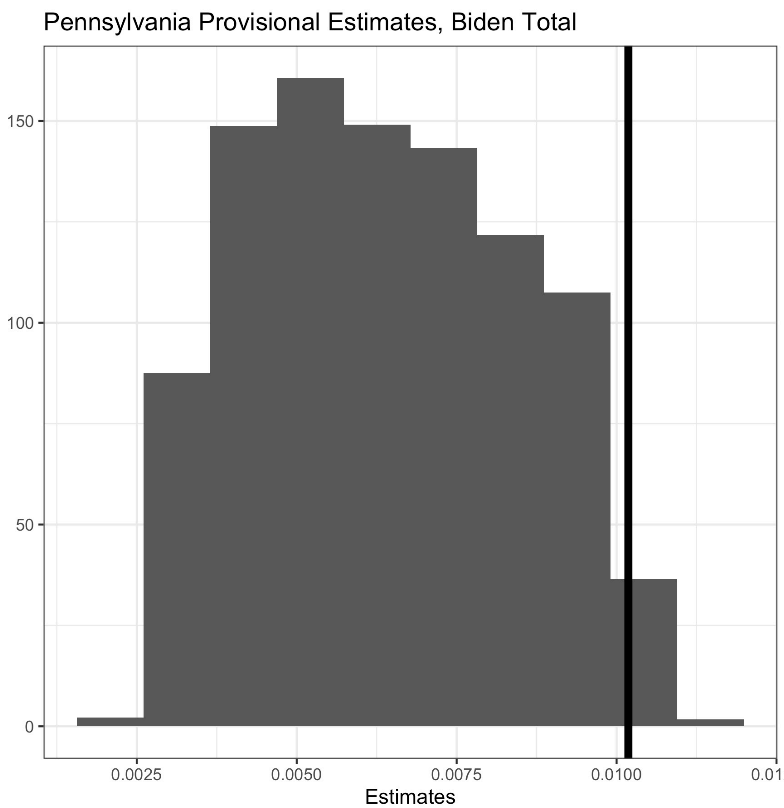


Table 11: Pennsylvania Provisional Ballot Results, Total Ballots

	Dependent variable:			
	Difference, Biden Share of Votes From Provisional Ballots (Lott (2020), Table 7a)	Biden Share of Votes From Provisional Ballots		
	(1)	(2)	(3)	(4)
Difference, Share of Trump Vote from Provisional Ballots	0.364 (0.105)			
Share of Trump Vote from Provisional Ballots		0.371 (0.078)	0.385 (0.103)	0.342 (0.082)
Allegheny County	0.010 (0.004)	0.007 (0.002)	0.007 (0.002)	0.007 (0.002)
Observations	87	174	174	174
Precinct-Pair Fixed Effects			✓	
County-Pair Fixed Effects				✓

explanations and could not on its own establish fraud.

H.1 Additional Evidence that Lott's Turnout Model Overstates the Turnout Differences in "Suspicious" Counties

To highlight the deficiency of Lott's approach, we undertake a falsification test. To reiterate, the fundamental problem with Lott's analysis is that it compares "suspect" counties in states that experienced large turnout increases against a pooled control group comprising of non-suspect counties in states that experienced large turnout increases and all counties in states that experienced smaller turnout increases. Given this flaw, we should find similar evidence of fraud if we replace Lott's coding of "suspect" counties with a random set of counties in the same states. To investigate this, we repeatedly draw a random set of counties from the states where Republicans alleged fraud, designate these counties (counterfactually) as "suspect", and conduct the same analyses reported in Figure **??**.[18] If Lott (2020)'s design is valid, the coefficient on "suspect county" should be significant in only 5% of random draws. We expected otherwise: by including states with lower turnout increases in the control group (without including state fixed effects or otherwise accounting for cross-state turnout differences), Lott (2020)'s analysis builds in a bias toward finding "inexplicably" high turnout increases in counties where Republicans have alleged fraud.

[18] In a state where n counties had allegations of fraud, we randomly draw n counties to be the pseudo-suspect counties.

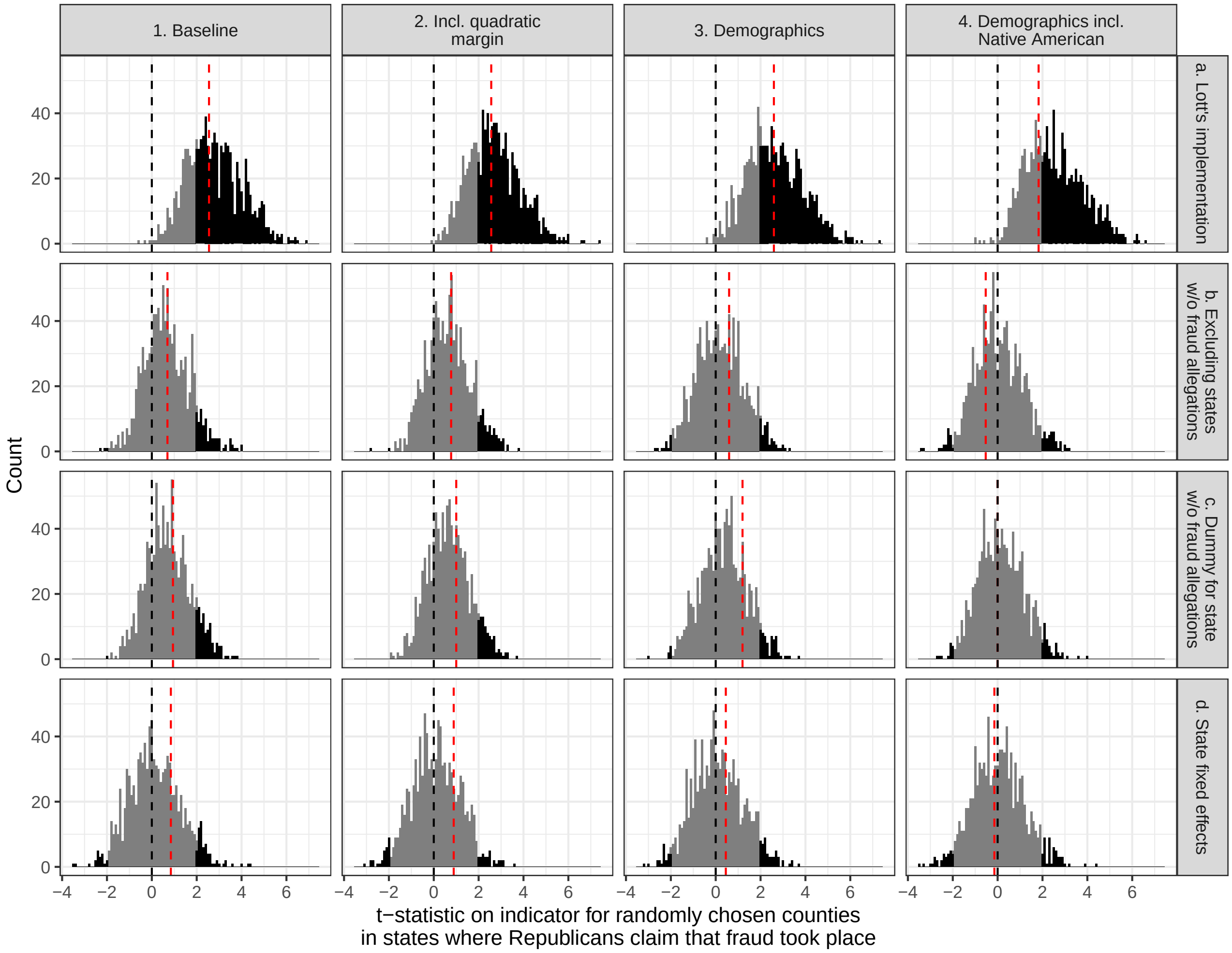


Figure 10: If "suspicious" counties were chosen at random rather than identified from Republican allegations as in Lott (2020), Lott (2020)'s test would usually find evidence of "fraud"; our improved specifications would not

Figure 10 shows the distribution of t-statistics across 1000 random reshufflings. The top row shows Lott (2020)'s specifications: the estimate from the true coding of suspect counties is statistically significant in each specification (as shown by the vertical red line at or above 2), but this t-statistic is actually typical of the distribution of t-statistics across random reshufflings (shown in the histogram). Across Lott (2020)'s specifications, the proportion of random reshufflings that produce a significant "effect" (the false discovery rate, or type I error, shown by the dark region of the histograms) is between .6 and .75. In fact, the t-statistic is larger on average when we randomly select counties than when we use the counties in which Republicans actually alleged fraud (according to Lott (2020))).

The next three rows of Figure 10 show the same exercise conducted for the alternative specifications we used in Figure **??** above. False discovery rates are near .05, suggesting that adjusting for differences in turnout across states renders Lott (2020)'s tests statistically valid.

I Ayyadurai's Claim Of Election Fraud in the 2020 Massachusetts Senate Primary is Baseless

As a candidate in the Republican primary in Massachussets, Shiva Ayyadurai lost his race by over 20 percentage points. Since his race, Ayyadurai has produced several viral internet videos alleging that a weighting feature in Dominion machines was used to decrease his vote total and increase his opponent's vote total. We address these claims here because they relate indirectly to claims made about voting machines in the presidential election.

In a lawsuit filed against the Massachusetts Secretary of State, Ayyadurai alleges that his vote total in machine-counted precincts was multiplied by 2/3, while his opponent's (Kevin O'Connor) vote total was multiplied by 1.2. To support this claim, Ayyadurai creates a histogram of his vote counts for precincts in Suffolk County. Focusing on precincts where he received 22 or fewer votes, he observes that there were many precincts in which he received an odd number of votes. He observes that this could be the result of his vote totals being multiplied by 2/3 and rounded, which would in fact disproportionately yield odd numbers: 1, 2, 4, 5, 7, 8 would all be converted to odd numbers (respectively: 1, 1, 3, 3, 5, and 5) for example. He claims to have performed better in counties that used paper ballots, where this weighting could not be applied. Ayyadaurai claims that, after the alleged weighting issue is addressed, he would have won.

Though we could make many critiques of these claims,[19] we focus on two. First, Ayyadurai neglects to assess whether the allegedly odd pattern he finds in Suffolk County appears anywhere else in Massachusetts. We show in Figure 11 that it does not. For each county in Massachusetts, we show the number precincts (horizontal axis) and the proportion of precinct vote totals for a given candidate that are even numbers. Suffolk County, the one Ayyadurai highlights, has the lowest share of even vote totals; other counties have about the same number of evens and odds, and many are almost as tilted toward an excess of *even* ballots as Suffolk was toward odd ballots. (The outlier is a rural county with so few precincts that departures from even-odd balance are not surprising.) The corresponding numbers for O'Connor, shown in the plot below, also look unremarkable. Thus even if we were to concede that the pattern in Suffolk County is unusual, there is no evidence that the pattern obtains beyond Suffolk County, which fatally undermines Ayyadurai's allegations.

Second, Ayyadurai's claim that he performed better in places with paper ballots is false. Table 12 shows that there is no difference between Ayyadurai's vote total in precincts that use paper ballots or machines. It is true that there is a bivariate relationship between paper ballots and Ayyadurai's returns; this occurs because Ayyadurai tended to perform better in more rural and small precincts, where paper ballots are also more likely to be used.[20] But

[19] While not crucial to our conclusions, we note several issues with this analysis. First, Ayyadurai groups together 0 and 1 in his histogram, for reasons that are unclear. Second, he never establishes a logic for multiplying by 2/3. In fact, even if multiplication occurred, it would be impossible to retrieve the ratio because he would fail to observe the true count in some precinct. Ayyadurai also makes inaccurate claims about the distribution of voting machines in Massachusetts. Every county uses machines, but many counties also use paper ballots. In short, even setting aside the problems we diagnose below, Ayyudurai's analysis cannot possibly demonstrate what he claims that it does.

[20] Note, this bivariate difference is far below the change we would expect to see if Ayyadurai's allegations were true.

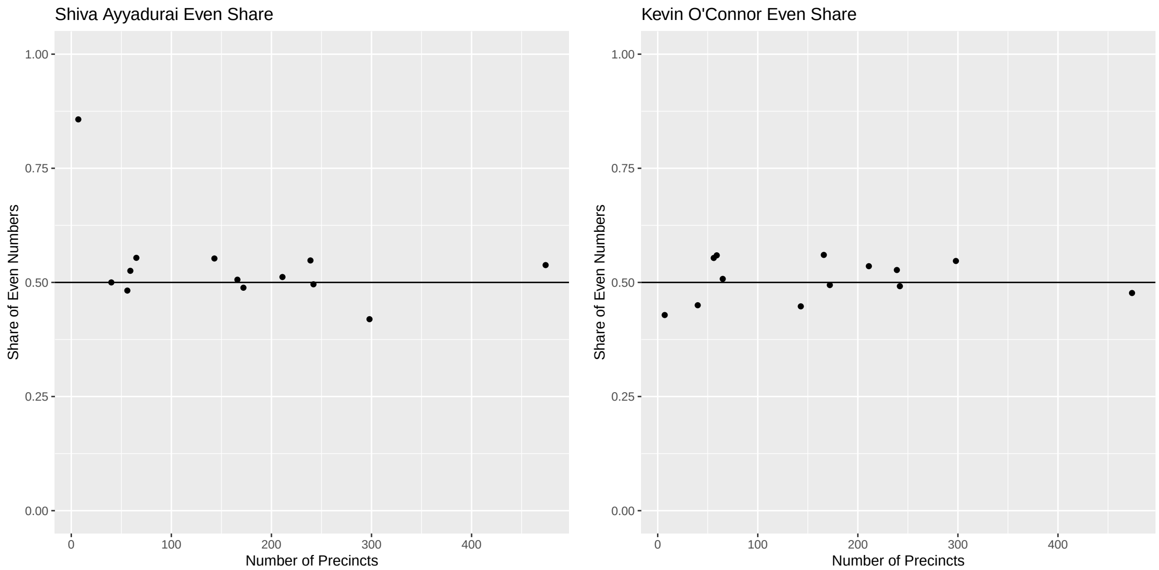


Figure 11: Ayyadurai's rate of even numbers is equivalent to his opponent's

this relationship disappears once we include county fixed-effects to account for local voting preferences. There is thus no evidence that electronic voting machines suppressed support for Ayyadurai in his election attempt.

Table 12: Ayyadurai does no better in locations with paper ballots.

	Dependent variable:			
	Ayyadurai Margin		Ayyadurai Share	
	(1)	(2)	(3)	(4)
Paper Ballot	0.068	0.004	0.046	0.010
	(0.025)	(0.027)	(0.013)	(0.014)
Observations	2,169	2,169	2,169	2,169
County Fixed Effects		✓		✓

References

[1] Yimeng Li, Michelle Hyun, and R Michael Alvarez. Why do election results change after election day? the" blue shift" in california elections. 2020.

[2] Michael C Herron and Jonathan Wand. Assessing partisan bias in voting technology: The case of the 2004 new hampshire recount. *Electoral Studies*, 26(2):247–261, 2007.

[3] Edward Solomon. Smoking gun: Ess transferring vote ratios between precincts in pa. 2020. https://rumble.com/vbas2t-smoking-gun-dominion-transferring-vote-ratios-between-precincts-in-pa.-by-e.html.