No Evidence for 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]Democracy and Polarization Lab, Political Science, and Hoover Institution. Stanford University

February 3, 2021

Abstract

After the 2020 US presidential election Donald Trump refused to concede, alleging widespread and unparalleled voter fraud. Trump's supporters deployed several statistical claims that supposedly demonstrated that Joe Biden's electoral victory in some states, or his popular vote in the country, were fraudulently obtained. Reviewing the most prominent of these statistical claims, we conclude that none of them is even remotely convincing. The common logic behind these claims is that, if the election were fairly conducted, some feature of the observed 2020 election result would be unlikely or impossible. In each case, we find that the purportedly anomalous fact is either not a fact or not anomalous.

1 Introduction

Following the 2020 U.S. elections, President Trump and other Republicans questioned Biden's victory in public statements and lawsuits. Although Trump's legal challenges were unsuccessful, many of his supporters were apparently convinced by his claims that the election was stolen: a survey in December 2020 found that over 75% of Republican voters found merit in claims that millions of fraudulent ballots were cast, voting machines were manipulated, and thousands of votes were recorded for dead people.[1] Trump's efforts to overturn the election

[1]Jan Zilinsky, Jonathan Nagler and Joshua Tucker, "Which Republicans are most likely to think the election was stolen? Those who dislike Democrats and don't mind white nationalists", *Washington Post*, January 19, 2021: (Available here).

outcome culminated in a rally in Washington, DC, on January 6, 2021, which led some of his supporters to violently attack the U.S. Capitol in an attempt to "stop the steal".

In this paper, we consider several widely disseminated claims purporting to call the 2020 U.S. presidential election result into question. These claims were made in a range of formats including social media posts, expert witness testimony, and research papers. In many cases the logical and factual deficiencies of these claims have already been pointed out in disparate venues, but in others they have not. Our purpose in this paper is to address several of the most pervasive claims in one place using a common conceptual framework.

In Table 1 we briefly summarize the statistical claims we address in this paper.[2] We note that although these claims come from various sources and imply different types of wrongdoing, they follow a common logic: each one highlights an alleged feature of the 2020 election result that would purportedly be surprising or impossible if the election were free and fair. After carefully considering each claim, we conclude that each one fails in one of two ways. In some instances, accurate claims are made about the election results but the observed facts are not actually inconsistent with a free and fair election. In other instances, the supposedly surprising fact about the 2020 election result turns out to be incorrect on further inspection, whether because of an arithmetic error or because of insufficiently rigorous statistical analysis. Thus in each case a purportedly anomalous fact about the election turns out to be either not anomalous or not a fact. We are left with no evidence of anything out of the ordinary.

Our primary goal in assessing allegations of malfeasance in the 2020 election is to provide a resource for anyone who encounters these and other similar claims and either wonders whether they are true or seeks to better understand what is wrong with them. We view this goal as important given the proportion of Americans who reportedly doubt the results of the 2020 election. While a hard-core conspiracy theorist will not be convinced by what follows,

[2]Other claims have been addressed by others including journalists, expert witnesses, and public officials. In Appendix Table 5 we document a broader set of claims and include links to reputable rebuttals of those arguments.

Claim	Claim is correct, but fact is not surprising	Claim is wrong; once corrected, not surprising	Section
Biden won fewer counties than Trump	✓		3.1
Biden won one bellwether county	✓		3.2
Differences between:			
early- and late-counted votes	✓		3.3
2016 and 2020 state vote counts	✓		3.4
More votes than voters in US		✓	4.1
Dominion machines flipped votes to Biden		✓	4.2
Biden received disproportionate absentee votes in Fulton and Allegheny counties		✓	4.3
Unexplained high turnout in “suspicious” counties		✓	4.4

Table 1: Claims About the 2020 Election and Why They Are Wrong

many others may be interested in the basis for the claims made by those who advocated overturning the election result. If Trump's tactics are emulated by others, similar claims will undoubtedly arise in future elections, and perhaps our analysis will help in evaluating those claims.

As a secondary goal, we hope this paper will help some readers to better understand the logic of hypothesis testing in statistics. As we explain below, the claims we address have the same logical structure as null hypothesis significance testing in classical statistics: they report an alleged fact about the election outcome and they assert (perhaps implicitly) that this fact would be very surprising if the election were properly administered. We consider each claim in light of this logic. In some cases, we show that the alleged fact is true but is not surprising. In other cases, the alleged fact is not true and the corrected fact is not surprising. At the end of the paper we highlight some limits of statistical tests to detect election fraud that apply more broadly to null hypothesis significance testing. Briefly, a statistically anomalous election result can raise questions about the null hypothesis (here, that the election was free and fair) even as it provides little reason to believe any particular alternative hypothesis, such as that one side engaged in electoral fraud. Indeed, many of the supposed anomalies highlighted by the Trump campaign could just as easily be seen as evidence of fraud by the Trump campaign itself.[3]

It might be argued that the fraud claims we address are so flimsy, and made in such transparently bad faith, that they do not deserve a response. This may be true of some arguments, but others have the appearance of serious research and cannot be dismissed out of hand. At any rate, given the stakes involved we think it is better to err on the side of taking doubts seriously and making earnest efforts to address them.

2 The logic of statistical claims about election integrity

Doubts about the integrity of the 2020 presidential election were raised in a variety of forms: Facebook memes, tweets, media reports, research papers, and expert testimony. The logic of these claims was not always clearly stated, but one can recognize a standard statistical logic that applies across these claims. In this section we state that logic, which will help in our discussion of what is wrong with each claim.

In non-technical terms, these statistical claims consist of an observation about the election outcome (e.g. the number of counties carried by the election winner) and an assertion that this observation is inconsistent with a properly administered election. According to this logic, we have reason to doubt that the election was properly administered if we observe that something has happened that would be very unlikely to occur if the election had been properly administered.

This is the logic of hypothesis testing in classical statistics. Statistical claims about election fraud consist of a *test statistic* (for example, the number of counties won by the election winner), a *null hypothesis* (here, the hypothesis that the election was properly administered), and a *null distribution*, which describes how the test statistic would be distributed across

[3] The broader point is that statistical tests that cast doubt on the null hypothesis in any scientific study may not provide much evidence in favor of any particular alternative hypothesis, especially when no alternative hypothesis is *a priori* likely relative to the null and/or several alternative hypotheses that could have produced the observed data.

hypothetical properly administered elections. If we observe a value of the test statistic that is extreme relative to the null distribution (e.g. the number of counties carried by the winner being smaller than we would expect in 99% of properly administered elections), we may consider the test to have provided evidence against the null hypothesis that the election was properly administered.

Given this logic, there are two clear ways in which a statistical claim of electoral malfeasance can be refuted. First, it may be that the test statistic is computed incorrectly. This could be due to a simple error, such as when an analyst claims to have found that the total number of votes exceeded the number of people voting by several million, but is shown to have used the wrong figures in his calculations. It could also be more subtle, as when an analyst claims to have found an "all else equal" difference in voting patterns between two sets of counties but the difference is shown to evaporate when additional factors are controlled for.

Second, it may be that the analyst has correctly stated the test statistic but is incorrect in stating that this test statistic is extreme relative to a reasonable null distribution. It could be, for example, that the share of counties won by Biden is correctly reported but not at all surprising given recent electoral patterns in U.S. politics.

In what follows, we will address several statistical claims about the 2020 election by questioning either the reported value of the test statistic or the (often implied) null distribution against which it is compared. In all cases we find that the supposed anomaly is not an anomaly once we compare the properly computed test statistic to a reasonable null distribution. In other words, facts that are purportedly so surprising as to throw the 2020 election result into question are either not facts or not surprising.

3 Claims based on facts that are not actually anomalous

3.1 Number of counties won

Conservative radio talk show host and political activist Charlie Kirk tweeted on December 20, 2020, "Does anyone else have a hard time believing Joe Biden won a record-high number of votes despite winning a record-low number of counties?"[4] Later that day, he provided numbers to back up the claim, stating that Barack Obama won 69 million votes and 873 counties (in 2008) and Donald Trump won 74 million votes and 2,497 counties (in 2020), while Biden won 81 million votes and just 477 counties (also in 2020).[5] Other conservative activists had previously highlighted the same contrast between votes and counties on social media.[6] The intended conclusion was apparently that Biden's huge vote total must be fraudulent given how few counties he won.

As it happens, Kirk's tweet understates the number of counties Biden won: he won 537 counties, not 477. The basic point is correct, however: Biden won more votes than Trump or Obama while winning far fewer counties than Trump and somewhat fewer counties than Obama. Is this surprising?

The contrast between Biden in 2020 and Obama in 2008 loses some of its intended shock value when we convert vote totals to vote shares. In 2020, both the electorate and the turnout rate were larger than in 2008. Thus, Obama won fewer votes than Biden, but he won a larger vote share. It is not very surprising that Obama would win more counties than Biden given that he also won a larger share of the vote (52.9% vs 51.3%).

This leaves the contrast between Biden in 2020 and Trump in 2020. Is it surprising that

[4] https://twitter.com/charliekirk11/status/1340692425635979266

[5] Available Here

[6] https://eu.usatoday.com/story/news/factcheck/2020/12/09/fact-check-joe-biden-won-most-votes-ever-and-fewest-count
3865097001/

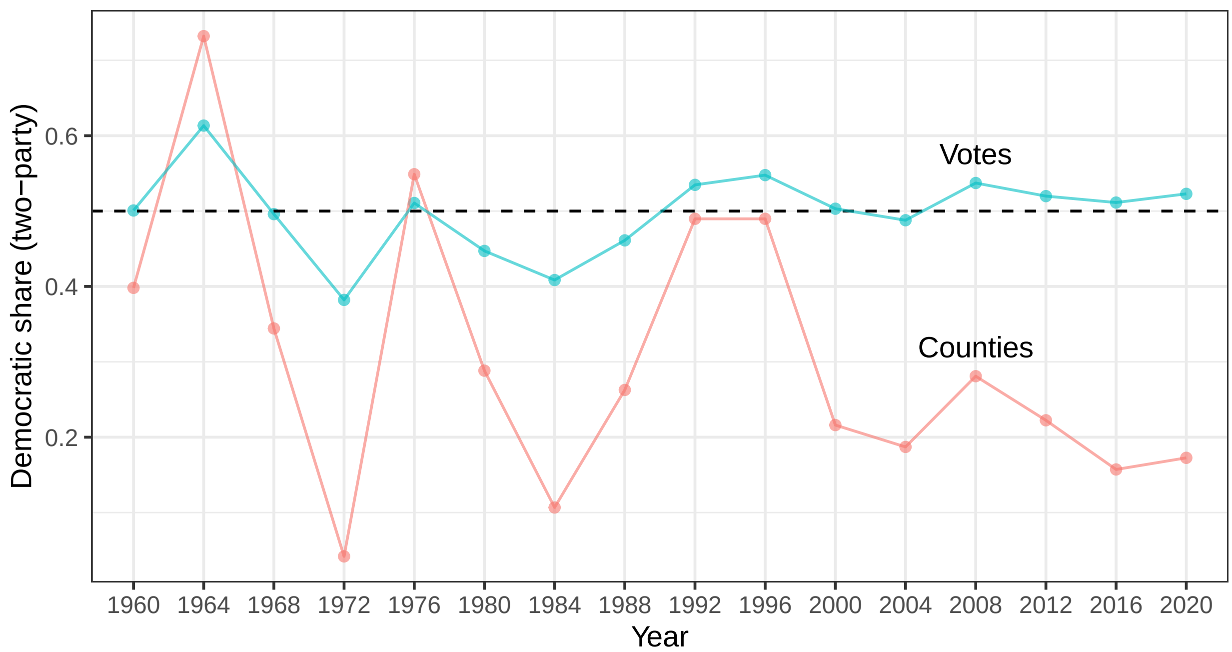


Figure 1: Share of votes and counties won by Democratic presidential candidates over time: in recent elections (including 2020), Democrats win national majorities while carrying a small proportion of counties

Biden would win a larger share of the national vote while winning under 20% of counties? Not in light of recent presidential elections.

Figure 1 shows the proportion of counties and votes carried by Democratic presidential candidates in each election since 1960. In every election other than 1964 and 1976, the Democratic candidate has won a smaller share of counties than votes. Since 2000, the gap between vote shares and county shares has been wide and stable: Democratic candidates tend to win around half of votes and around one-fifth of counties. The 2020 election fits this pattern exactly. Democratic presidential candidates win a small share of counties because their support is concentrated in more urban and populous counties, while Republican candidates tend to win small and rural counties. Compared to Hillary Clinton in 2016, Biden won a slightly larger share of the vote and a slightly larger share of counties.

Thus, the apparent discrepancy between Biden's popular vote haul and the share of counties he won is not anomalous in the least. It is typical of Democratic presidential candidates in recent elections.

3.2 Number of bellwether counties won

A related claim was made about Biden's performance in "bellwether" counties, which are counties where a majority of voters have supported the election winner in a number of previous elections. Of the 19 counties that voted for the eventual winner in every presidential election from 1980 to 2016, Biden defeated Trump in only one. Several commentators viewed this fact as a suspicious anomaly. As stated in *The Federalist*, "Amazingly, [Biden] managed to secure victory while also losing in almost every bellwether county across the country. No presidential candidate has been capable of such electoral jujitsu until now."[7] Trump later recited this fact in a rally in Georgia, stating that he "won 18 of 19 bellwether counties. You know what a bellwether county is? A big deal. So I won 18 of 19, a record."[8]

On what grounds might a reasonable person see the 2020 bellwether results as evidence of fraud or wrongdoing? Bellwether counties' perfect record of siding with the winner in past elections suggests that they might serve as a reliable gauge of sentiment in 2020 as well. If these 19 counties unerringly sided with the winner over the previous 10 elections, and they overwhelmingly sided with Trump in 2020, doesn't it suggest that (despite official reports to the contrary) Trump was the real winner in 2020?

This view relies heavily on the assumption that counties that have sided with the winner in the past (i.e. bellwether counties) reliably continue to do so in the future. In principle, this could be true—for example, if bellwether counties consistently mirror the demographic makeup of key swing states. In practice, it is not the case.

To assess whether counties that have sided with the winner in the past (i.e. "bellwethers") are more likely than other counties to side with the winner in the future, we analyzed each election since 1996.[9] We modeled a county's probability of correctly choosing the winner in a

[7] https://thefederalist.com/2020/11/23/5-more-ways-joe-biden-magically-outperformed-election-norms/

[8] https://www.washingtonpost.com/politics/2020/12/07/no-bidens-win-wasnt-statistically-impossible/

[9] Examining elections from 1940 to 1964, Tufte and Sun (1975) similarly asked whether bellwether counties identified on the basis of past results are more likely to side with the winner in future elections. They conclude that "the usual concept of a bellwether electoral district has no useful predictive properties" (p. 10) and that bellwether counties "are a curiosity and probably should be forgotten" (p. 17).

given election as a function of the Democratic margin in the county in the previous election and an indicator for whether the county had sided with the winner in each past election since 1980. Figure 2 shows the odds ratios comparing bellwethers and other counties; an odds ratio above 1 indicates that bellwethers were more likely to side with the winner of that year's election than other counties (conditional on the previous election result). In most elections from 1996 to 2020 we find an odds ratio not significantly different from one, suggesting that bellwethers' past performance is no indication of future success. In 2008 bellwethers track the winner slightly better than other counties, but in 2000 they do worse. Bellwether counties, by definition, have a track record of siding with the winner in the past, but this pedigree does not seem to result in any inherent tendency to track winners in the future. This in itself should not be too surprising: even if counties chose a presidential candidate at random in each election, there would be some counties that by chance would have consistently sided with the winner, despite having no special ability to pick winners in the future.

If counties with a history of siding with the winner do not reliably continue to do so, it should be less surprising that the 19 bellwether counties did not line up behind Biden in 2020. Still, what accounts for Trump winning almost all of these counties?

Figure 3 shows the Democratic margin in 2020 (vertical axis) and 2016 (horizontal axis) for every US county, with the 19 bellwethers highlighted in red. For a county to side with the winner in both 2016 and 2020, it would need to land in the upper left quadrant, labeled "Trump-Biden". Only 63 counties did so, of which only one was a bellwether. But the proportion of bellwether counties that sided with Biden in 2020 (1 in 19) appears similar to the corresponding proportion among other counties that narrowly supported Trump in 2016.. Indeed, Figure 3 suggests that the main reason why the 19 bellwether counties overwhelmingly voted for Trump in 2020 is that (by definition) they voted for Trump in 2016, and counties that voted for Trump in 2016 overwhelmingly voted for Trump in 2020. Once we are disabused of the notion that bellwether counties have any special ability to pick

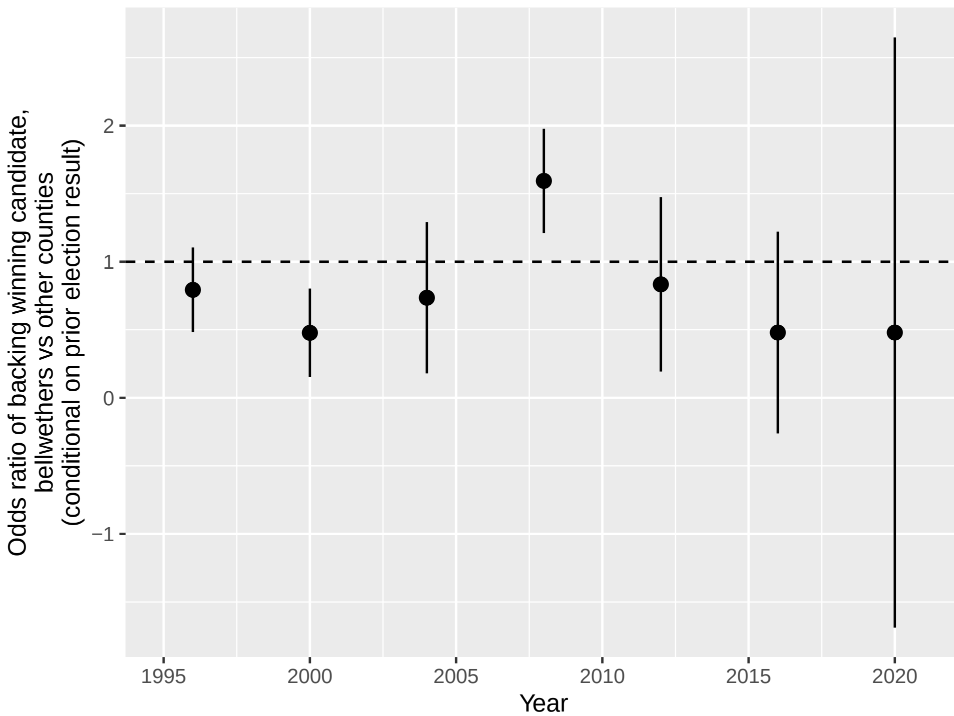


Figure 2: Difference in likelihood of supporting the winner (odds ratios) in bellwether counties vs other counties, conditional on support for winner's party in previous elections

future winners, the 2020 result among bellwethers is utterly unsurprising.

As Tufte and Sun (1975) established almost 50 years ago, bellwether counties are statistical accidents with no special powers to detect the will of the American people. Claiming that Biden's victory must have been illegitimate because bellwether counties voted for Trump is just as sound as claiming that Spain's victory in the 2008 UEFA Euro tournament was illegitimate because Paul the Octopus predicted Germany to win.

3.3 Differences between early- and late-counted vote results

Other claims about statistically improbable results deal 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

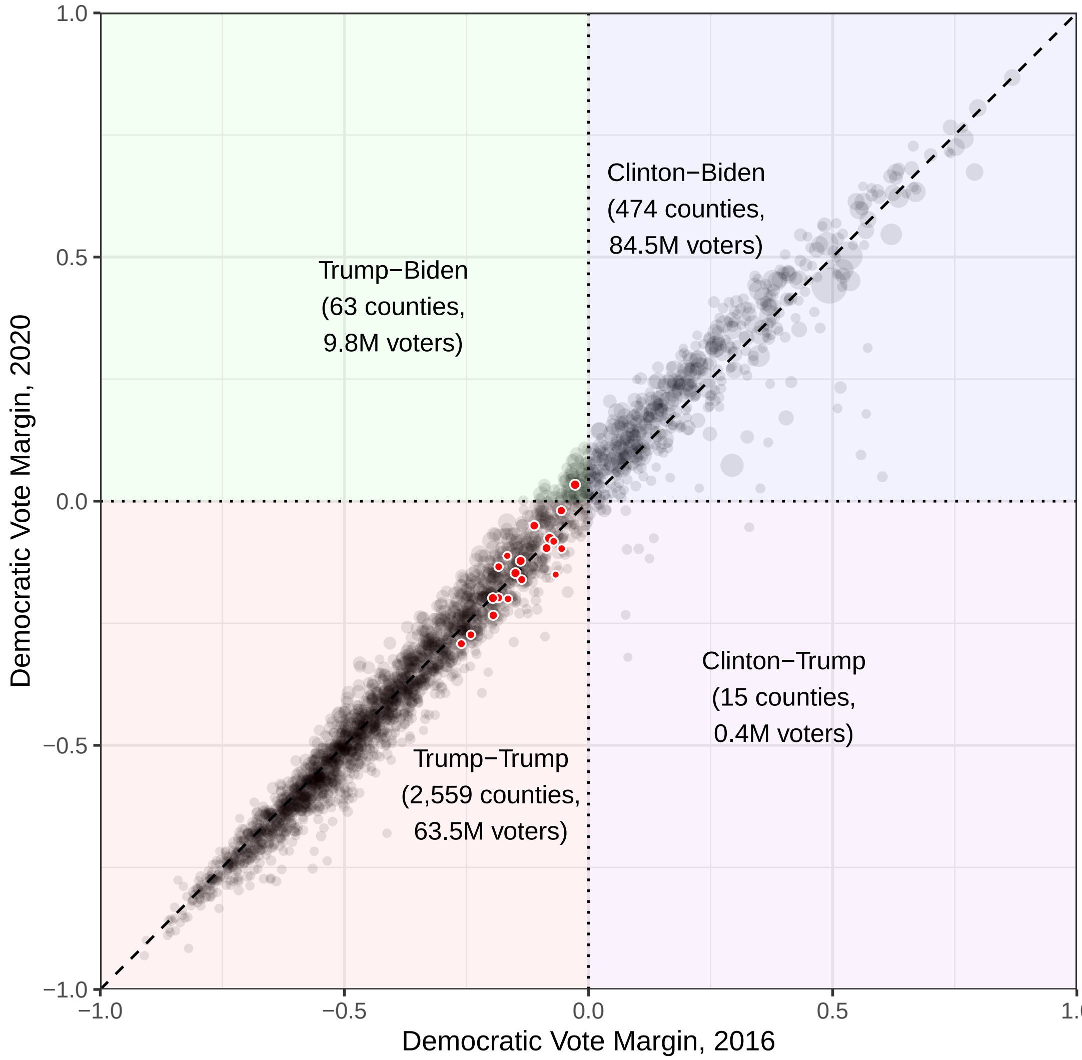


Figure 3: Democratic vote margin in 2016 (horizontal axis) and 2020 (vertical axis) by county: support in most counties did not shift much, and "bellwethers" (colored red and green) were no exception

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 (Li, Hyun and Alvarez, 2020),[10] this late shift in votes was regularly cited by Trump and his legal team as evidence of fraud.

In fact, Trump advocates used a statistical analysis to argue that there was a "one-in-a-quadrillion" chance of this shift occurring. This claim comes from an expert report attached to a Supreme Court lawsuit Texas Attorney General Ken Paxton filed against the state of Pennsylvania. In that report, Paxton claims that the expert, Charles Cicchetti, calculated a "one-in-a-quadrillion" chance of Biden winning given Trump's lead in early-counted votes in Georgia and three other swing states. Cicchetti concludes his report arguing that "In my opinion, the outcome of Biden winning in all these four states is so statistically improbable, that it is not possible to dismiss fraud and biased changes in the ways ballots were processed, validated, and tabulated."

Cicchetti, however, never actually calculated the probability of Biden's victory. Instead, Cicchetti intends 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.[11]

[10]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/.

[11]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 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

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 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.[12] 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

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.

[12] 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.

obtain a more modest (though still exceedingly large) absolute z-score of 120.8.

To be clear, the fact that we would reject Cicchetti's null hypothesis in Arizona is not evidence that fraud also occurred in that state; rather, it demonstrates that the premise of Cicchetti's test—that early- and late-counted votes should be the same in a properly administered election—is plainly false. 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. Indeed, given these expectations, it would have been more surprising if we had *not* seen differences in Biden's support between early- and late-counted ballots. The differences in early and late ballots that Cicchetti highlights is thus not at all anomalous and provides no evidence of fraud.

3.4 Differences between 2016 and 2020

Cicchetti's expert report also claims that the changes from 2016 to 2020 are suspiciously large and, therefore, the results of the 2020 election deserve full scrutiny. Again focusing on vote counts rather than vote shares, Cicchetti tests the null hypothesis that Joe Biden's vote count in 2020 was the same as Hillary Clinton's in 2016 in Georgia and other swing states. Under the null hypothesis that the underlying count of Democratic votes in Georgia is the same in 2016 and 2020, Cicchetti calculates a z-score of 396.3. He then dramatically asserts that this corresponds to "a chance in 1 in almost an infinite number of outcomes". He goes on to argue that "the statistical differences are so great, this raises important questions about changes in how ballots were accepted in 2020".

Just like Cicchetti's test of early- and late-counted ballots, Cicchetti's comparison of the 2016 and 2020 election results is based on a deeply flawed premise. He certainly provides convincing evidence that the 2016 and 2020 elections were not statistically identical twins. He seems to think that this implies something was amiss about the 2020 result. But of course knowing that 2016 and 2020 were different elections has no bearing on whether the 2020 election was fraudulent. We *know* that the two elections are different: they were conducted at different times, in different electorates, with different candidates. The differences he highlights between 2020 and 2016 are full consistent with both elections being held in a free and fair manner.

One way to illustrate the misguided premise behind Cicchetti's analysis is to conduct the same test for other pairs of elections. Because an election held in one year is different from the election held four years earlier, we should not be surprised to find that Cicchetti's test rejects the null hypothesis of no difference. Indeed, this is what find. Conducting Cicchetti's test over pairs of proximate presidential elections within each state, we find a large z-statistic in almost every election in every state. We present those statistics in Figure 4. Across 1,498 within-state election comparisons we find only 10 instances in which Cicchetti's test produces a p-value greater than 0.05. 98.9% of all comparisons produce absolute z-score with accompanying p-values that are less than 0.001, and 98.3% of all comparisons produce a p-value less than 0.0001. In 5.7% of cases, the change in states' vote total produces a t-statistic that is larger than the "suspicious" z-score in Georgia (396) that Cicchetti claimed had a "1 in almost-infinite" chance of occurring.

Thus Cicchetti detects real differences that occur from 2016 to 2020, but these differences occur in any election. The observed facts are utterly unsurprising and provide no evidence whatsoever of electoral fraud in 2020.

4 Claims based on facts that are not actually facts

4.1 More votes than voters in the US

While some claims made about the 2020 election used apparently sophisticated statistical techniques, other 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.

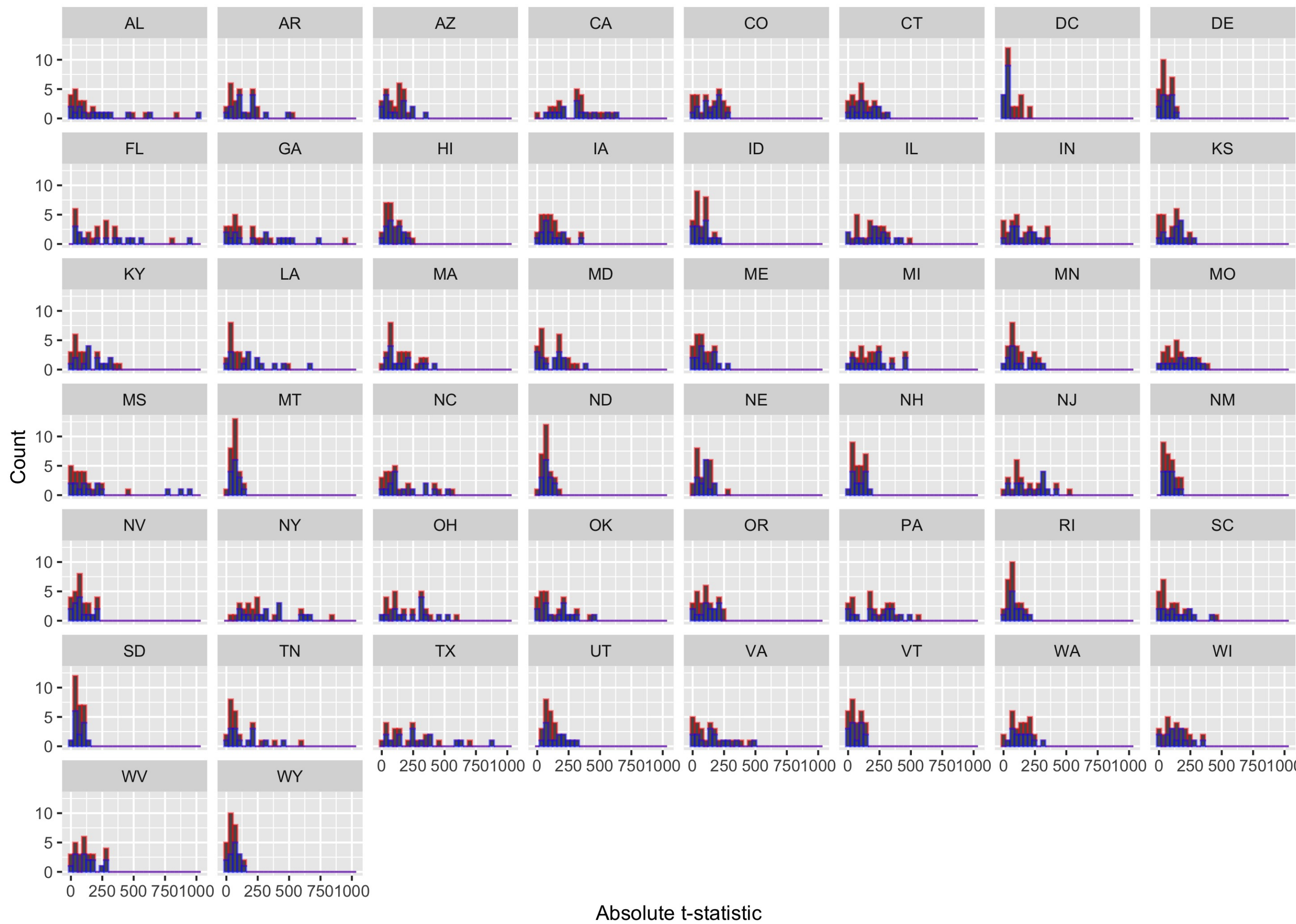


Figure 4: Cicchetti's analysis produces large absolute t-statistics in every state across elections since 1960

His claim found a receptive audience among the 2020 election skeptics, including President Trump, who repeated the claim in a December 30th tweet.[13]

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. 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).[14] 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:

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

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. As we describe in Table 5, 2020 election skeptics erroneously claimed there were more votes than voters in Pennsylvania, with essentially no evidence. In Michigan it was claimed several townships had greater than 100% turnout, but these claims did not stand up

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

[14]Michael P. McDonald, from the United States Election Project estimates that 158,254,139 million voters voted for the president office.

to scrutiny either.[15] And claims of an implausibly large jump in turnout 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.[16]

Claims of turnout over 100% were damaging in part because they are so easily understood. They are also easily debunked: after basic errors are corrected, nothing surprising is taking place. We now turn to claims that are not so straightforwardly debunked.

4.2 Dominion voting machines do not increase Trump vote share

Trump's legal team claimed at various points after the election that voting machines run by Dominion Voting Systems switched votes from Trump to Biden. Trump lawyers Rudy Giuliani and Sydney Powell argued for a global conspiracy that undermined democracy everywhere Dominion was present. In late December, an anonymous analysis[17] was widely circulated on social media claiming to show that Biden outperformed expectations in counties that used Dominion voting machines. The right-wing news outlet *The Epoch Times* reported that the analysis showed Biden outperformed expectations in 78% of the counties that use Dominion or Hart voting machines and that the analysis "also indicates that Biden consistently received 5.6 percent more votes in those counties than he should have."[18]

Inferring that a particular set of voting machines caused Biden to receive more votes is difficult, because machines are not randomly assigned to counties. It turns out that the analysis behind these claims failed to address some of the obvious differences between counties using different types of machines. Once this problem is addressed, we find no

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

[16] 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/.

[17] The report is available here.

[18] "Joe Biden Appears to Outperform in Counties Using Dominion or HART Voting Machines: Data Analyst".

significant difference in voting outcomes between counties using Dominion machines and other counties. There is, thus, no evidence that Dominion machines created votes for Biden; what was presented as an anomalous fact turns out not to be a fact.

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.[19] 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.[20]

Given these problems with the original analysis, we carry out own analysis to check for evidence that Dominion machines switched votes from Trump to Biden. In column 1 of Table 2 we show the results of a bivariate regression of Biden's share in 2020 on an indicator for whether the county used a Dominion machine, finding a very slight and statistically insignificant difference. In column 2 we adjust for Clinton's share of the vote in 2016, which strongly predicts the 2020 outcome (note the R^2 of .964); the Dominion coefficient becomes very slightly negative, though again it is not significant. In column 3 we add a fixed effect for being in a state in which Dominion machines were used and in column 4 we add a fixed effect for each state; in both cases we find coefficients that are statistically significant in the *negative* (i.e. pro-Trump) direction, although very small in magnitude. In short, we find absolutely no evidence that Biden outperformed expectations in counties where Dominion machines were used.

In the appendix we report additional tests probing the robustness of our finding. We check different ways of coding whether a county uses Dominion machines (based on the US Election Assistance Commission as in Table 2 or through our own hand coding of counties in swing states on Table 7)[21]; controlling for census covariates (Table 9) or Biden's predicted

[19] 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.

[20] 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.

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

Table 2: 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.

	Dependent variable:			
	Biden Vote Share, 2020			
	(1)	(2)	(3)	(4)
Dominion Machines	0.007	−0.002	−0.009	−0.006
	(0.010)	(0.002)	(0.002)	(0.003)
Clinton Share of Vote, 2016		1.032	1.029	1.011
		(0.004)	(0.004)	(0.004)
Observations	3,111	3,111	3,111	3,110
R^2	0.0002	0.964	0.965	0.975
Dominion-State Fixed Effects			✓	
State Fixed Effects				✓

performance (with predictions generated by a random forest regression) rather than Clinton's 2016 share Table 11; comparing Dominion BMD machines and Hart machines against others as in the original report rather than all Dominion machines (Table 10); conducting the analysis with vote margins rather than vote shares (Table 8). 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.

In summary, we fail to find any compelling evidence for an effect of Dominion vote machines on Biden's vote share. The ostensible proof of fraud presented in the original report appears to be the result of "fishing" for evidence. The actual differences in Biden's performance between counties with and without Dominion machines are easily explained by other factors that vary across counties. Thus, the supposed fact that Biden curiously outperformed expectations in counties using Dominion machines turns out not to be a fact at all.

4.3 Absentee vote counting in Pennsylvania and Georgia did not produce suspicious extra ballots for Biden

Another focus of the Trump team's accusations was the processing of absentee ballots in key states that Biden narrowly won. Among other claims, they alleged that Fulton County, Georgia, and Allegheny County, Pennsylvania, were major centers of voter fraud in the 2020 election. Most of these allegations relied upon hearsay affidavits or debunked videos purportedly showing voters stuffing ballots. But in a paper posted in late December 2020, Lott (2020) claims to provide statistical evidence that irregularities in the absentee vote counting procedure in Fulton County and Allegheny County suppressed votes for Trump and bolstered Biden's vote count. The paper received immediate and widespread attention. Peter Navarro, Assistant to the President and Director of the Office of Trade and Manufacturing Policy, touted the claim as solid evidence of fraud. President Trump tweeted out a link to the paper.

Lott's claims, however, are entirely baseless. Our reanalysis of Lott (2020)'s data shows that Lott's claims about absentee voting in GA and PA depend on an entirely arbitrary decision about how counties are entered in the data set: the conclusion is reversed when an alternative and equally justified data entry rule is used. 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. Because this error is more subtle than those we have discussed so far, we will more thoroughly diagnose how the error occurred and our strategy for avoiding the issue.[22]

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

[22] 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.

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.[23] 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.[24] 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

[23] 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.

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

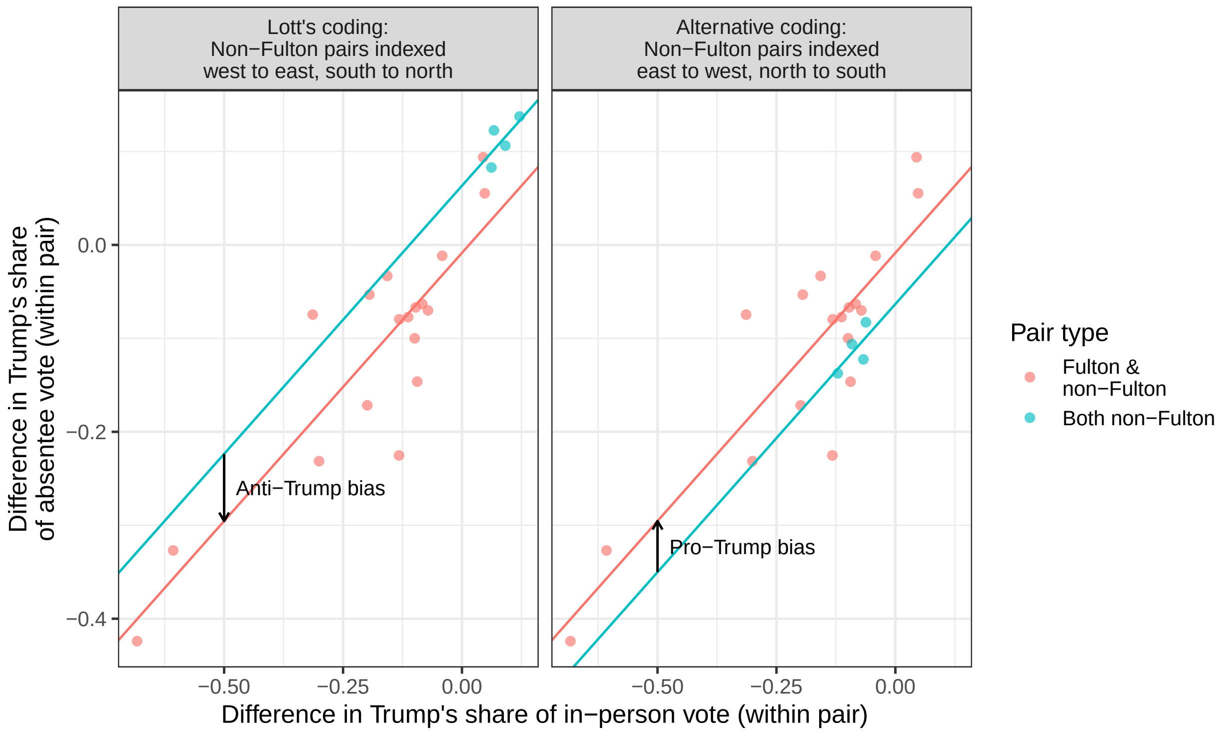
pair. Thus, β_0 measures the within-pair difference in Trump's share of the absentee vote among pairs that don't involve a suspect county (adjusting for the within-pair difference in Trump's in-person share). The key coefficient is δ, which compares the adjusted difference in Trump's share of the absentee vote within pairs involving the suspect county against the corresponding adjusted difference within pairs not involving the suspect county. The 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),[25] 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.[26]

[25] 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.

[26] 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

Figure 5: Evidence for fraud in Fulton County, GA, is reversed if arbitrary coding rule is reversed



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 5 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 5 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

is not Lott's design.

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.

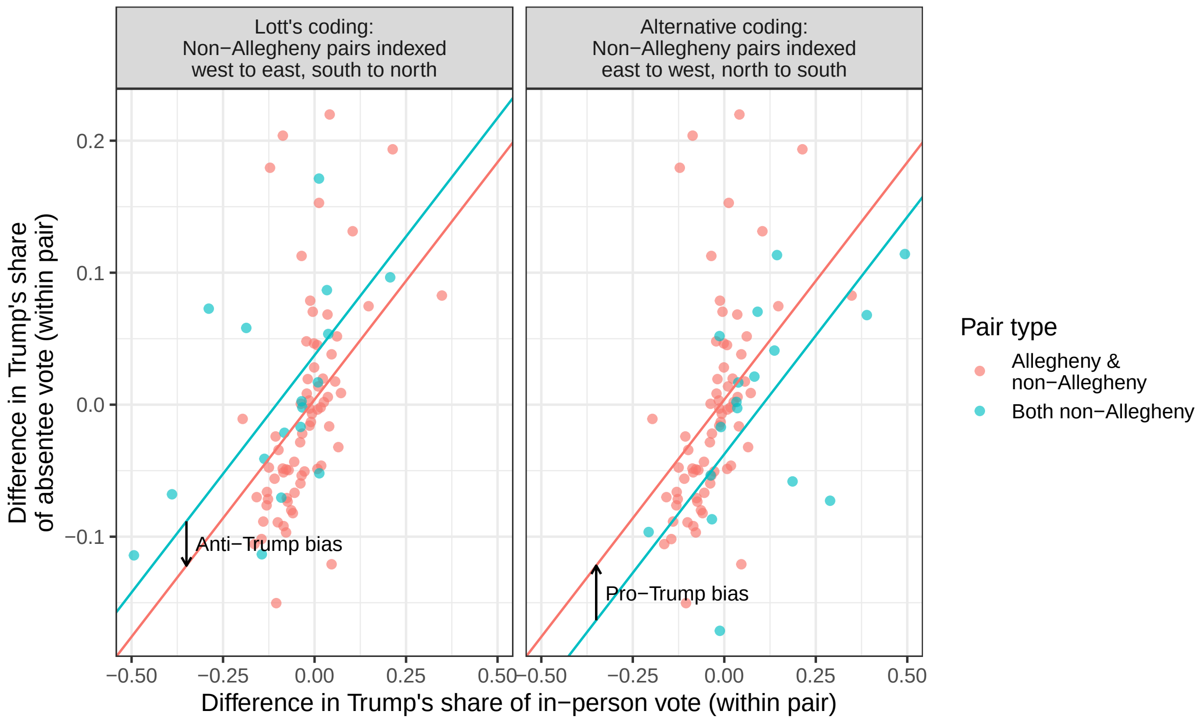
In the right panel of Figure 5, we show that the conclusion is reversed when we reverse Lott's arbitrary coding rule: instead of subtracting east from west and north from south in computing differences for non-Fulton precinct pairs, we subtract west from east and south from north. The scatterplot looks 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 12 (Appendix) reports coefficient estimates and standard errors for both sets of analyses depicted in Figure 5. 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 6 and Table 13 (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 7 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 evidence

Figure 6: Evidence for fraud in Allegheny County, PA, is reversed if the arbitrary coding rule is reversed

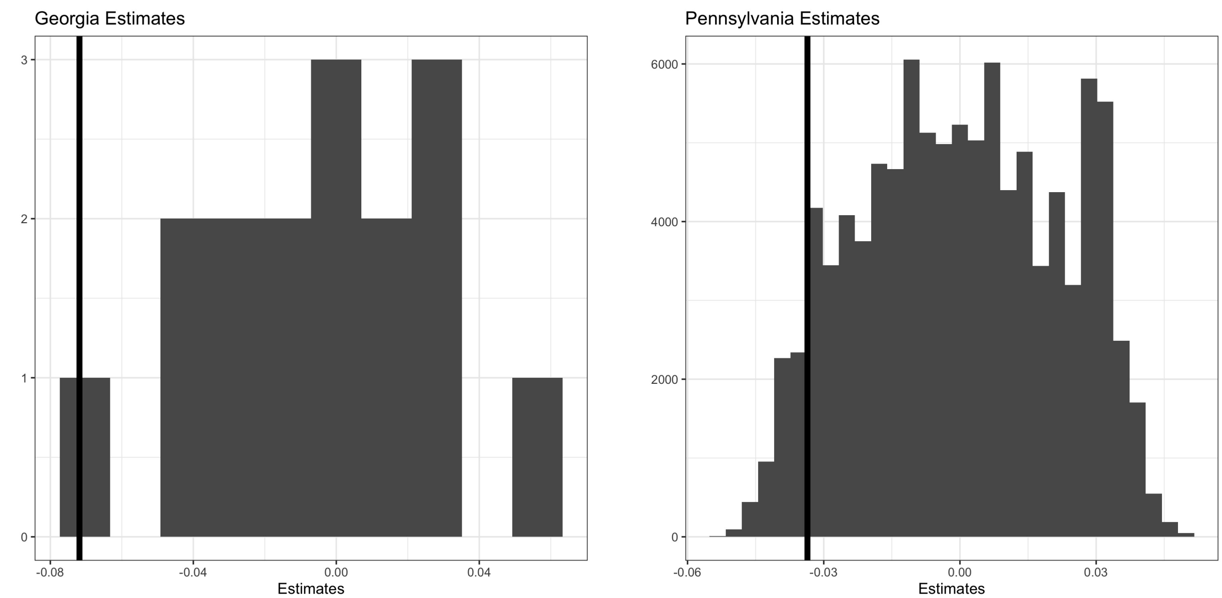


for anti-Trump fraud in Allegheny County.[27] In the Fulton County analysis, there are four non-Fulton precinct pairs and thus $2^4 = 16$ possible rules for computing differences within non-Fulton pairs. The left panel of Figure 7 shows the histogram of the key coefficient across these sixteen possible 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 7 shows the distribution of estimates for a random sample (with replacement) of 100,000 of these rules,[28] with the actual estimate

[27]In personal communication, Lott said the ordering of precincts followed a rule in a prior American Economics Review paper. We believe that is Bronars and Lott (1998).

[28]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 7: Evidence for fraud in Georgia and Pennsylvania depends on arbitrary coding rules; Lott's estimates are outliers in the distribution of estimates



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.

To more effectively achieve Lott's objective of comparing voting patterns across county boundaries, we reanalyze Lott's data using a more standard specification that does not suffer from these problems. Rather than using within-pair differences as Lott does, we employ a simple fixed-effects model. The regression equation for this model can be written as

$$\text{Absentee}_i = \beta_1 \text{InPerson}_i + \delta \text{SuspectCounty}_i + \sum_{k=1}^{K} \alpha_k I(\text{pair}_i = k) + \epsilon_i \tag{1}$$

where Absentee_i and InPerson_i denote Trump's share of the absentee and in-person vote (respectively) in precinct i, SuspectCounty_i indicates whether precinct i is located in a "suspect" county (Fulton or Allegheny, depending on the state being analyzed), and each precinct is identified with one of K precinct pairs indexed by k, with α_k denoting the fixed effect for pair k. The regression thus asks whether Fulton or Allegheny county precincts have lower absentee support for Trump than would be expected controlling for their in-person support for Trump and any factors (observable or unobservable) that are common to paired precincts. Precinct pairs that do not involve a suspect county contribute to estimating the coefficient β_1 but do not otherwise contribute to the estimation of the key coefficient δ. Crucially, no arbitrary coding decisions are necessary.[29]

Once corrected, the basis for Lott's (2020) claims of fraud disappears: the actual difference in Trump's absentee support between the key counties and neighboring counties is fully consistent with the null hypothesis that absentee ballots were handled correctly in both counties. We report the results of the fixed-effect analyses for Georgia in Table 3 below. In column 1, we simply regress Trump's share of the absentee vote on Trump's share of the in-person vote and a dummy for Fulton County; in column 2 we add precinct-pair fixed effects as in equation 1, essentially allowing the intercept to vary across Lott's precinct pairs; in column 3 we instead use county-pair fixed effects, with one intercept for Fulton-Coweta pairs, another for Carroll-Coweta pairs, etc. None of these specifications shows a substantively or statistically significant difference between Trump's share of the absentee vote in Fulton County precincts and other precincts. Similarly, Table 4 shows that when properly conducted, there is no evidence of differences in absentee votes in Allegheny County, Pennsylvania.

In short, when we reanalyze Lott (2020)'s data with a more sensible fixed effects specification, we find no evidence of differences in voting patterns between precincts in Fulton County or Allegheny County and adjacent precincts in Republican-leaning counties. If such differences existed they would hardly be convincing evidence of fraud, given possible differences between precincts located in different counties that are served by different school systems. But we find no such differences, undermining the basis for Lott (2020)'s claims.[30]

[29] In an updated version of the paper, Lott uses this model by continuing with the within-county difference approach but instead forcing the intercept to be zero. This is mathematically equivalent to our fixed-effects approach, though he seems unaware of that fact.

[30] In the Appendix we also replicate and extend Lott's analysis of provisional ballots in Pennsylvania. As with his analysis of absentee voting, his conclusions about provisional ballots depend on the arbitrary coding of non-Allegheny precinct pairs (Figures 12 and 13) and fixed effects estimation shows a substantively small difference in Biden's share of the provisional vote in Allegheny precincts and other precincts (Tables 14 and 15).

Table 3: A Fixed Effects Specification Shows Nothing Suspicious in Fulton County, GA

	Dependent variable:		
	Trump Share Absentee		
	(1)	(2)	(3)
Trump Share, In-Person	0.760	0.606	0.654
	(0.049)	(0.077)	(0.056)
Fulton County	0.019	−0.003	0.006
	(0.019)	(0.020)	(0.018)
Observations	44	44	44
Precinct-Pair Fixed Effects		✓	
County-Pair Fixed Effects			✓

Table 4: A Fixed Effects Specification Shows Nothing Suspicious in Allegheny County, PA

	Dependent variable:		
	Trump Share, Absentee		
	(1)	(2)	(3)
Trump Share, In-Person	0.511	0.307	0.442
	(0.042)	(0.066)	(0.048)
Allegheny County	0.003	0.003	0.006
	(0.008)	(0.009)	(0.009)
Observations	174	174	174
Precinct-Pair Fixed Effects		✓	
County-Pair Fixed Effects			✓

4.4 Turnout rates in counties where Republicans alleged fraud were not unusually high

Lott (2020) provides a second analysis that he claims demonstrates evidence for voter fraud. After arguing that electoral fraud can result in inflated turnout rates through a variety of mechanisms, he claims to show that 2020 turnout rates were higher than one would otherwise expect in a set of counties where Republicans have alleged that fraud took place. Lott argues that there was an "unexplained increase in voter turnout" in the key counties of between 1.26 and 2.42 percent, which Lott says is equivalent to 150,000 to 289,000 votes in those states. Lott concludes that this is evidence consistent with fraud.

While Lott's analysis of absentee voting results focused on narrow comparisons precincts located on county boundaries, his turnout analysis is based on county-level turnout figures for hundreds of counties across nine battleground states. Specifically, Lott checks whether turnout in the 2020 election was higher than would be expected (given previous turnout, political leaning, and local demographics) in counties where, according to Republican lawsuits filed after the election, fraud may have taken place. Lott identifies 19 counties across six swing states where Republicans made fraud allegations.[31] Lott (2020) compares turnout in these counties to turnout in other counties in the same six states plus all counties in three other swing states (Florida, Ohio, and North Carolina). He argues that, if turnout is higher in these counties than would be expected given covariates, it would be evidence of fraud.

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,

[31] Lott identifies the following "suspicious" counties—Georgia: Fulton, DeKalb; Pennsylvania: Allegheny, Centre, Chester, Delaware, Montgomery, Northampton, Philadelphia; Arizona: Apache, Coconino, Maricopa, Navajo; Michigan: Wayne; Nevada: Clark, Washoe; Wisconsin: Dane.

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?[32] Highly anomalous turnout figures could provide evidence of a problem, but a percentage point or two of unexplained turnout has other more plausible explanations and could not on its own establish fraud.

Nevertheless, given the possible implications of such a serious claim, we investigate the issue to see if Lott (2020) has shown a genuinely unexplained anomaly in the counties where Republicans have alleged that fraud took place. We assembled an original dataset that would allow us to assess Lott (2020)'s claims beyond his chosen set of states, if necessary.[33]. We provide our county-average turnout rates by state in the appendix. We note that our estimates of turnout are lower than Lott (2020)'s average turnout rates, but closer to official statistics.

Visually comparing turnout in 2020 to turnout in 2016 for counties in the six states where Lott alleged fraud, we find nothing remarkable about the turnout rate in the suspicious counties. In Figure 8 we plot turnout in 2020 against turnout in 2016 for counties in the six states with counties that Lott codes as having alleged fraud; we do this separately by state, with counties where fraud was alleged colored red and a linear regression line superimposed.[34] On a simple visual inspection, there is nothing puzzling about 2020 turnout in the highlighted

[32]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.

[33]We use turnout rates for the county citizen voting-age population. For total votes, we use Dave Leip's county-level vote results for 2020 and 2016. For the number of voting-aged citizens we use the five-year ACS from 2019 and 2015. This follows best practice from Michael McDonald Available here (McDonald, N.d.)

[34]The regression line is drawn based on the non-suspect counties.

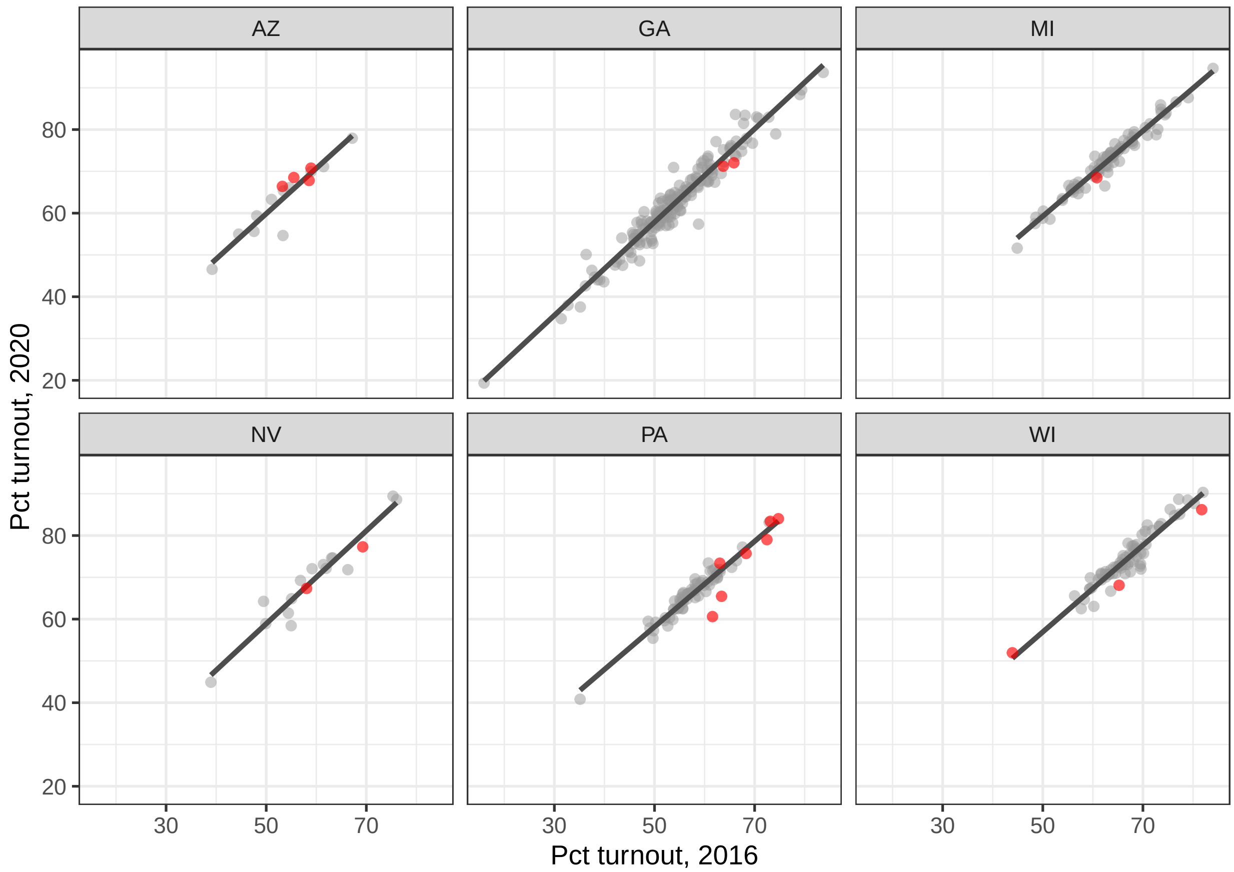


Figure 8: “Suspicious” counties (in red) are not remarkable relative to other counties in their state

counties. In fact, turnout seems to have been lower on average in these counties than in other counties in the same state, conditional on prior turnout. In light of this observation, Lott (2020)’s finding is puzzling: why would he conclude that turnout is suspiciously high in these counties, given the information in this figure?

The answer is that Lott’s conclusions are driven by the inclusion of states that have lower turnout increases and no “suspicious” counties—namely Florida, North Carolina, and Ohio. Figure 9 shows that, conditional on turnout in 2016, turnout in these three states was lower than turnout in the six states that contain a suspicious county in Lott’s analysis. This is relevant because Lott (2020)’s analysis compares changes in turnout in suspicious counties with changes in turnout in all other counties, so these smaller increases in turnout rates across states will be conflated with the suspicious county indicator in his analysis. The

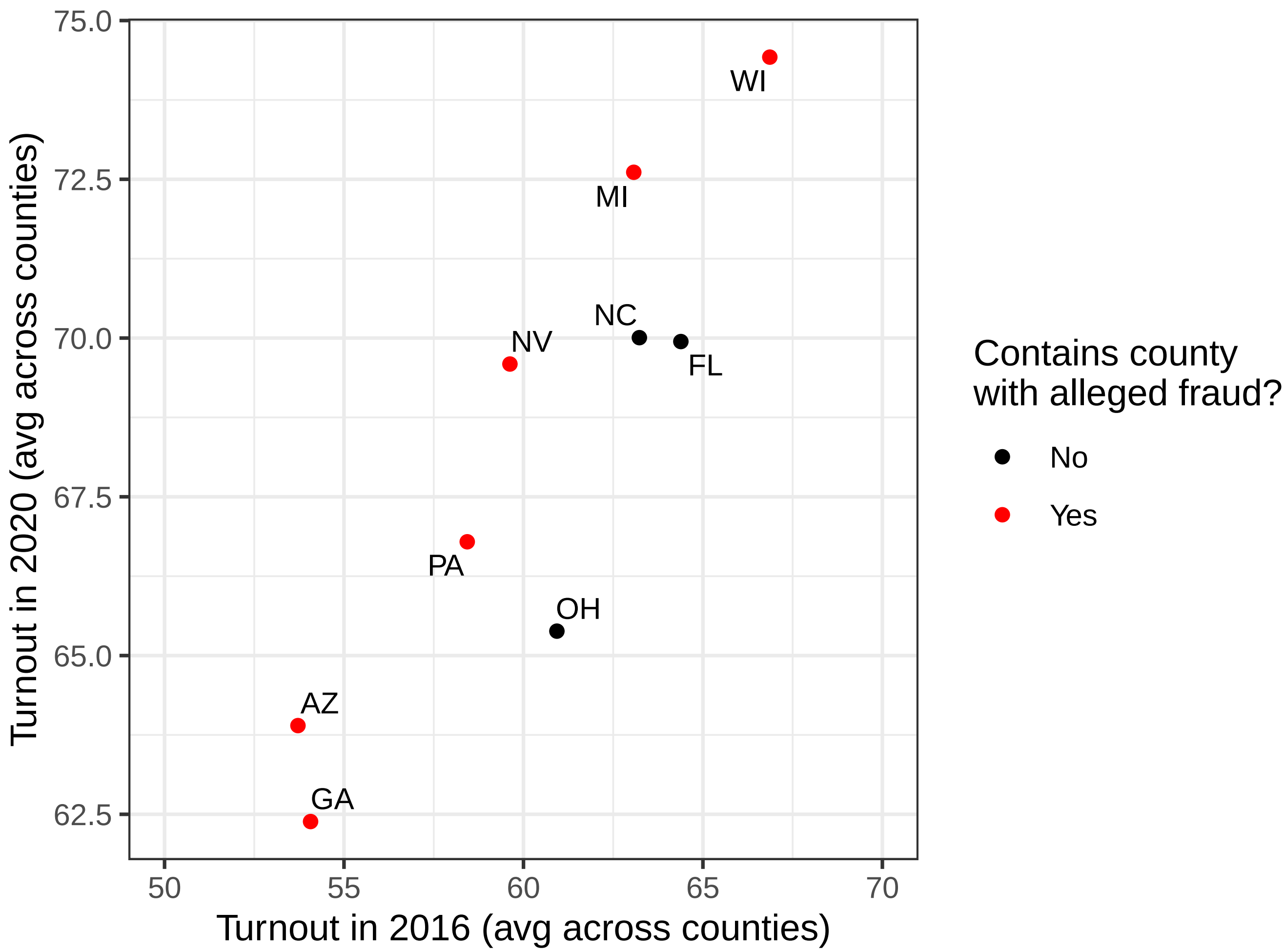


Figure 9: Swing states without suspicious counties had smaller average turnout increases, which drives Lott's (2020) results

smaller the turnout increase in these three "non-suspect" states, the more turnout in the suspect counties will appear to be suspiciously high, even if the changes in turnout in these suspect counties are unremarkable relative to the changes in turnout in other counties in their own state.

Figure 10 shows that, once we address the level differences across states, Lott's (2020) estimates of the turnout differences in suspicious counties go to zero and become null. We examine all four of Lott's (2020) models (organized on the vertical axis) and present the estimate of the average difference in turnout rates for suspicious counties. The circle/purple estimates of suspicious county turnout depict the estimates using the four specifications for

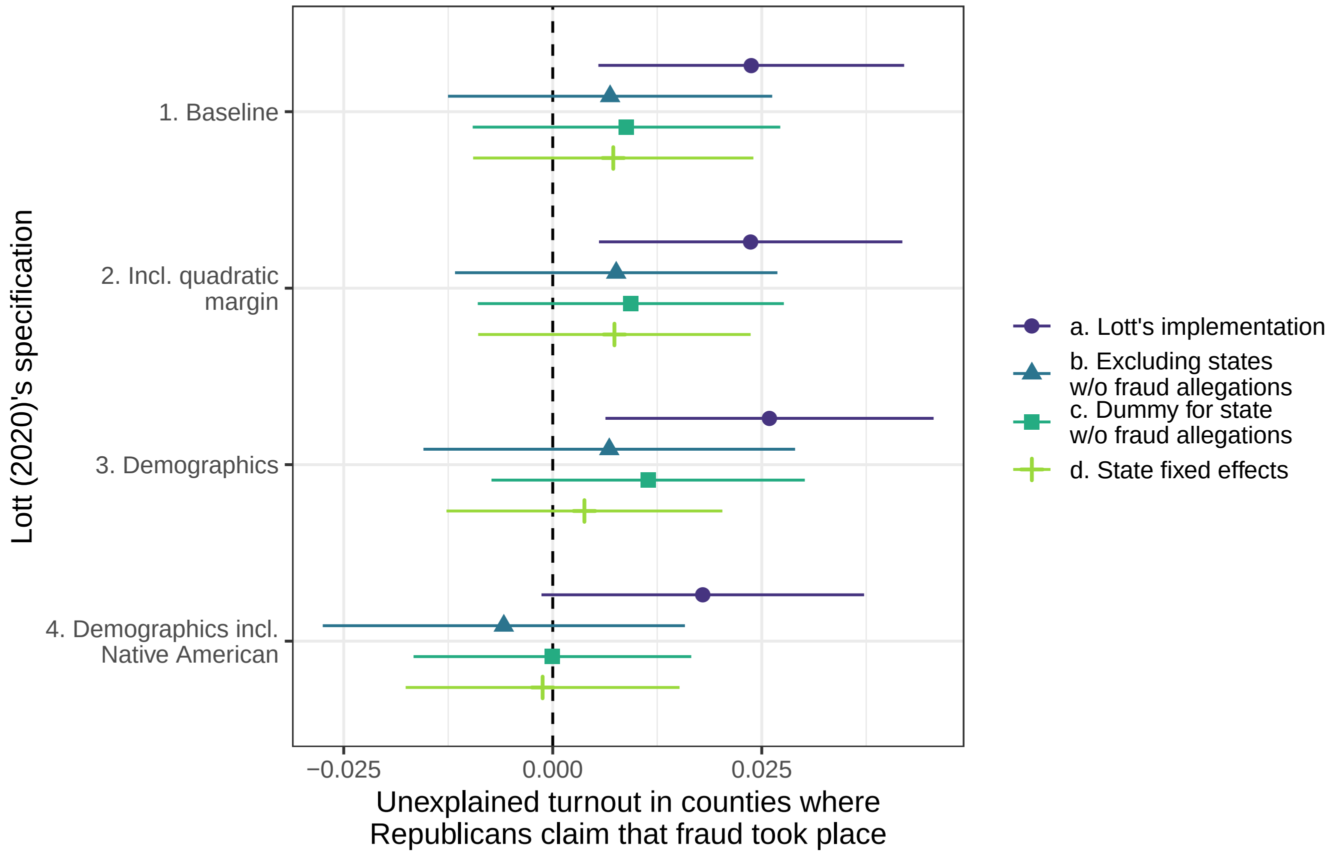


Figure 10: Lott's (2020) estimates of suspicious county differences in turnout are zero and null once we address state-level differences.

which Lott (2020) presents results in his Table 10. The triangle/dark-green estimates depict our estimates when we exclude Florida, Ohio, and North Carolina— three states in which no fraud was alleged. Across models, the difference in suspicious counties is close to zero and—in the case of model 4—the estimate is negative. The square/light-green estimates are from a model where we merely include an indicator for a state that has suspicious counties. Again, this reduces the estimate to null. Finally, the last estimates (plus/lime-green) include state-level fixed effects. Across models, this gives a close to zero and null difference for suspicious counties. Thus, simply by focusing only on states where at least one county had alleged fraud (i.e. swing states that Biden won) or allowing that state-wide turnout trends may differ across states or groups of states, we are able to explain what Lott (2020) claimed was unexplained turnout in counties where Republicans had claimed fraud.

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 10.[35] 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.

Figure 11 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 11 show the same exercise conducted for the alternative specifications we used in Figure 10 above. False discovery rates are near .05, suggesting that adjusting for differences in turnout across states renders Lott (2020)'s tests statistically

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

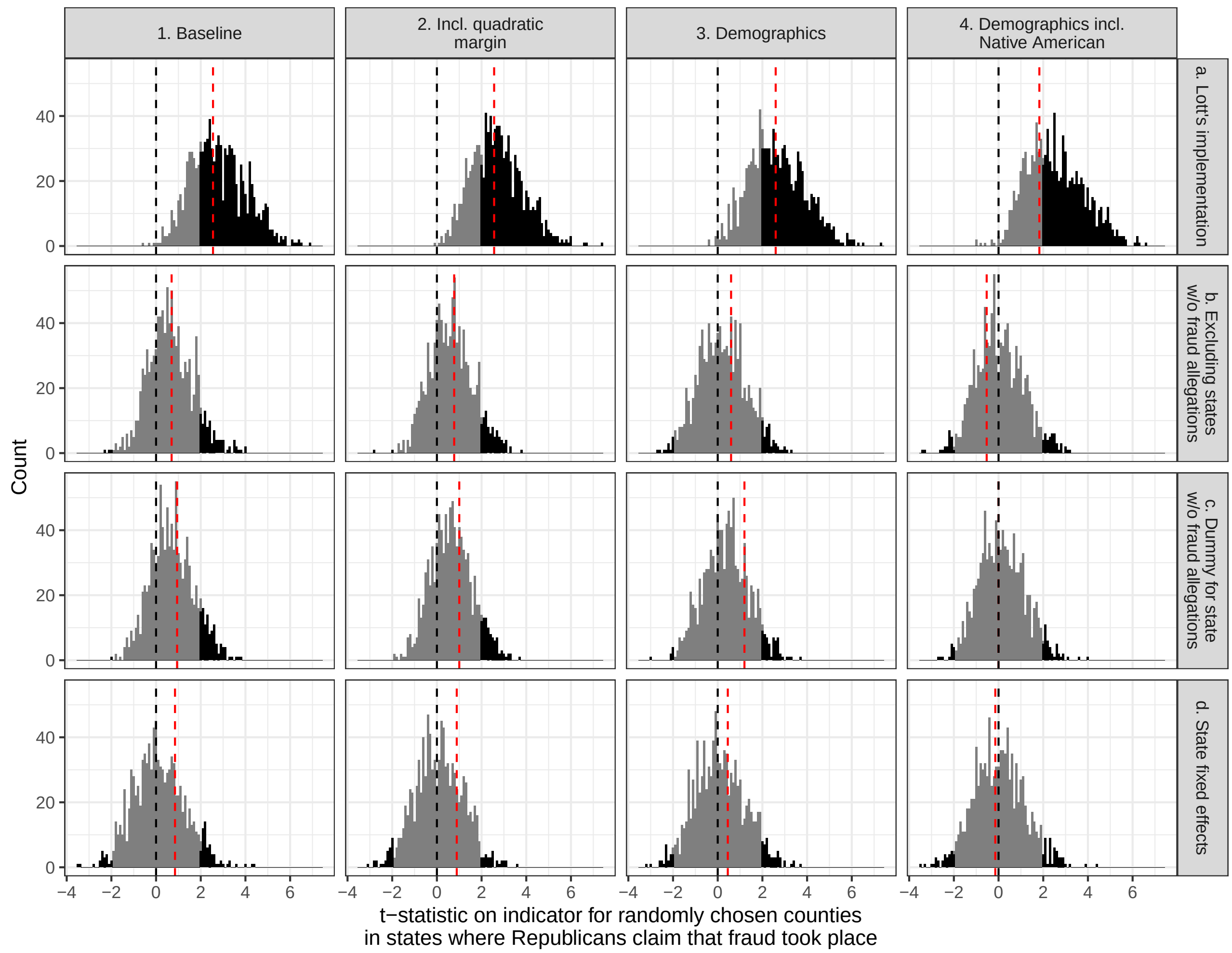


Figure 11: If "suspicious" counties were chosen at random rather than identified from Republican allegations as in Lott (2020), Lott (2020)'s test would usually find evidence of "fraud"; our improved specifications would not

valid.

Altogether we find no turnout differences in counties that Lott (2020) labels suspicious. The differences he documents occur because lower-turnout states where there were no suspicious counties were included in the analysis. But even if we had found lower turnout in suspicious counties, that does not necessarily imply that there was fraud, because there are many other explanations for high turnout. We now turn to discussing this limitation of statistical tests for electoral fraud in more general terms.

5 What do Statistical Tests Teach Us About Electoral Fraud?

We have so far shown that the Trump campaign's claims about the 2020 election fail in two ways: some claims are simply not true, and those that are true are not inconsistent with elections that are free of fraud. But, more generally, when we evaluate statistical claims made about elections, we have to be mindful of what the tests could possibly imply. These limitations are well understood, because they are well-known features of classical hypothesis testing. Highlighting their importance in this setting may help to illuminate the meaning and limits of common statistical practice.

The first limitation is that, even when a statistical test reveals a pattern that would be unlikely under the null hypothesis, it may not give us much reason to believe that some other hypothesis is more likely than the null hypothesis. Fundamentally, the reason is that an outcome that is statistically improbable *could* happen, and knowing that a pattern is improbable under the null does not tell us whether it is more probable under some alternative. In the context of the 2020 election, even if we believe that a properly administered election would be unlikely to produce some feature of the observed results, we may not have reason to believe that the election was *not* properly administered, both because that feature may still be unlikely under some other scenario and because improper administration may itself be very unlikely.

Consider a simple example to illustrate this point. It would be surprising to flip a fair coin 21 times and get the same result every time.[36] Nonetheless, we may still believe it more likely that the coin is fair than that its movements are controlled by distant magnets: 21 identical flips may also be unlikely under the magnetic-control hypothesis, and the magnetic-control hypothesis may also be *a priori* unlikely given what we know about the technical challenges of controlling coins with magnets and the prevalence of sinister coin-controlling physicists.

In more technical terms, our belief in the null hypothesis relative to some other hypothesis after observing a test statistic should depend not only on the probability of observing a value of the test statistic under the null (which is the focus of classical hypothesis testing), but also on the probability of observing the same test statistic under the alternative hypothesis and the prior probability of the null and alternative hypotheses.[37] Applied to the 2020 election, any truly anomalous feature of the election result could be explained either as a fluke or as the result of large-scale election malfeasance; in the absence of further evidence, it may be reasonable to prefer the "fluke" interpretation on the grounds that otherwise undetected large-scale malfeasance is unlikely and/or does not make the observed results any more likely.

The second limitation is that, even when a statistical test casts doubt on the null hypothesis, it may not give us much reason to prefer one alternative hypothesis rather than another. In the context of election malfeasance, we may observe an anomaly that makes us doubt that the election was properly administered but still have no idea which side is more likely to have engaged in malfeasance. Suppose, for example, that we find that counties administered by Democrats had higher reported turnout than counties administered by Republicans in a key state, controlling for all relevant factors, and that this difference was large enough to make us suspect that some electoral fraud occurred. Who was responsible for the fraud? Was it Democratic election administrators stuffing ballot boxes, or Republican election ad-

[36] The probability of this occurring is about 1 in one-million.

[37] This follows from Bayes Rule.

ministrators throwing out valid ballots? The statistical test provides no information about this; other evidence would be necessary to determine responsibility. Thus, even when the data makes us suspect that the null hypothesis is false and that another hypothesis is more likely, it may not clarify which alternative hypothesis is more plausible.

6 Conclusion

Even though the 2020 election is over and Donald Trump's attempt to overturn the results failed, the effects of the claims will reverberate for years. A large segment of the public remains skeptical that Biden won the election legitimately,[38] and Republican state lawmakers are taking steps to alter voting access in the name of preventing fraud.[39] The Trump campaign delivered a blueprint for losing candidates to undermine support for the winner or even steal the election. It seems unlikely that he will be the last to try these tactics.

We have closely examined what appear to be the main pieces of statistical evidence of fraud in the 2020 election. For each of these claims, we find that what is purported to be an anomalous fact about the election result is either not a fact or not anomalous. In many cases the alleged fact, if shown to withstand scrutiny, would hardly constitute convincing evidence that Biden was elected due to fraud: a modest advantage to Biden in counties that chose to use Dominion machines, for example, could be explained by chance, by factors not accounted for in statistical models, or indeed by pro-Trump fraud undertaken using *other* voting machines. As it happens, the evidence we examine either fails to withstand scrutiny (like the Dominion results) or is utterly unsurprising given common sense (like Cicchetti's observation that the 2020 election differed from the 2016 election) or familiarity with patterns in recent U.S. elections (like the bellwether counties claim).

[38] A poll from January 8-11 2021 indicated that 72% of likely Republican voters and 42% of independents said they continue to question the election results. Li Zhou, "About half of Republicans don't think Joe Biden should be sworn in as president", January 11, 2021, *Vox*, (Available here).

[39] See for example Gabby Birenbaum, "State GOPs have already introduced dozens of bills restricting voting access in 2021", *Vox* January 29, 2021: (Available here).

In some cases, members of the public who are confronted with a statistical claim of election fraud can apply the approach we took in this paper: first ask whether the allegedly anomalous fact is a fact; if so, ask whether it is anomalous. In many cases, assessing the validity and unexpectedness of an allegedly anomalous fact requires some statistical sophistication and even original data analysis. For these cases, we think academics (and data journalists and others with appropriate skills) have an important role to play. To safeguard future election results, it will be essential to have elections experts ready to evaluate claims made about whether an election is free and fair. And we think that social media organizations can do more to broadcast these evidence-based claims rather than merely flagging questionable assertions as disputed or asserting that the election was free and fair.

Rebuilding trust in American elections requires that we fairly evaluate claims about their failures and communicate those claims to a skeptical public. This paper is an effort in that direction.

References

Li, Yimeng, Michelle Hyun and R Michael Alvarez. 2020. "Why Do Election Results Change After Election Day? The" Blue Shift" in California Elections.".

Lott, John R. 2020. "A Simple Test for the extent of Vote Fraud with Absentee Ballots in the 2020 Presidential Election: Georgia and Pennsylvania Data." *Available at SSRN 3756988* .

McDonald, Michael. N.d. "I want congressional, state legislative district, or county VEP turnout rates.". Accessed on 1/15/2021.
URL: *http://www.electproject.org/home/voter-turnout/faq/congress*

Tufte, Edward R and Richard A Sun. 1975. "Are there bellwether electoral districts?" *Public Opinion Quarterly* pp. 1–18.

Table 5: Catalogue of Trump Election Claims

Claim	Source	Refutation
More votes than voters in MI	Expert Witness: Ramsland	USA Today
More votes than voters in PA	PA State Rep. Frank Ryan	AP News
More votes than voters nationally	Trump tweet	PolitiFact
Biden won record low number of counties	Charlie Kirk tweet	USA Today
Unexplained vote bumps in MI, WI, GA	Nick Adams tweet	Reuters
Felons, minors, deceased voted	Trump tweet	Sterling (GA)
Residents who moved out voted	Navarro Report	FactCheck.org
Pro-Biden split ticket in swing states	Epoch Times	538
$1/10^{15}$ chance of Biden victory	Supreme Court case	PolitiFact
Trump won more bellwether counties	Tweet	USA Today
Trump won more bellwether states	Tweet	USA Today
Lower rejection rate of absentee ballots	Trump tweet	Sterling (GA)
Tweet		
Missing absentee ballots	Expert Witness: Miller	Expert Report: King
Dominion machine manipulation	Trump tweet	Sterling (GA)

A Overview of Trump Election Fraud Claims

B Additional County Figures and Tables

C Dominion Voting, Auxiliary Analyses

D Lott's (2021) Analysis of Absentee Voting in GA and PA

Table 6: 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.

	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

Table 7: 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				✓

Table 8: 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				✓

Table 9: 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 10: 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 11: 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		✓		✓

Table 12: 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 13: 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 14: 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 12: Distribution of Estimates for Alternative Precinct Differencing Orders, Pennsylvania Provisional Ballots

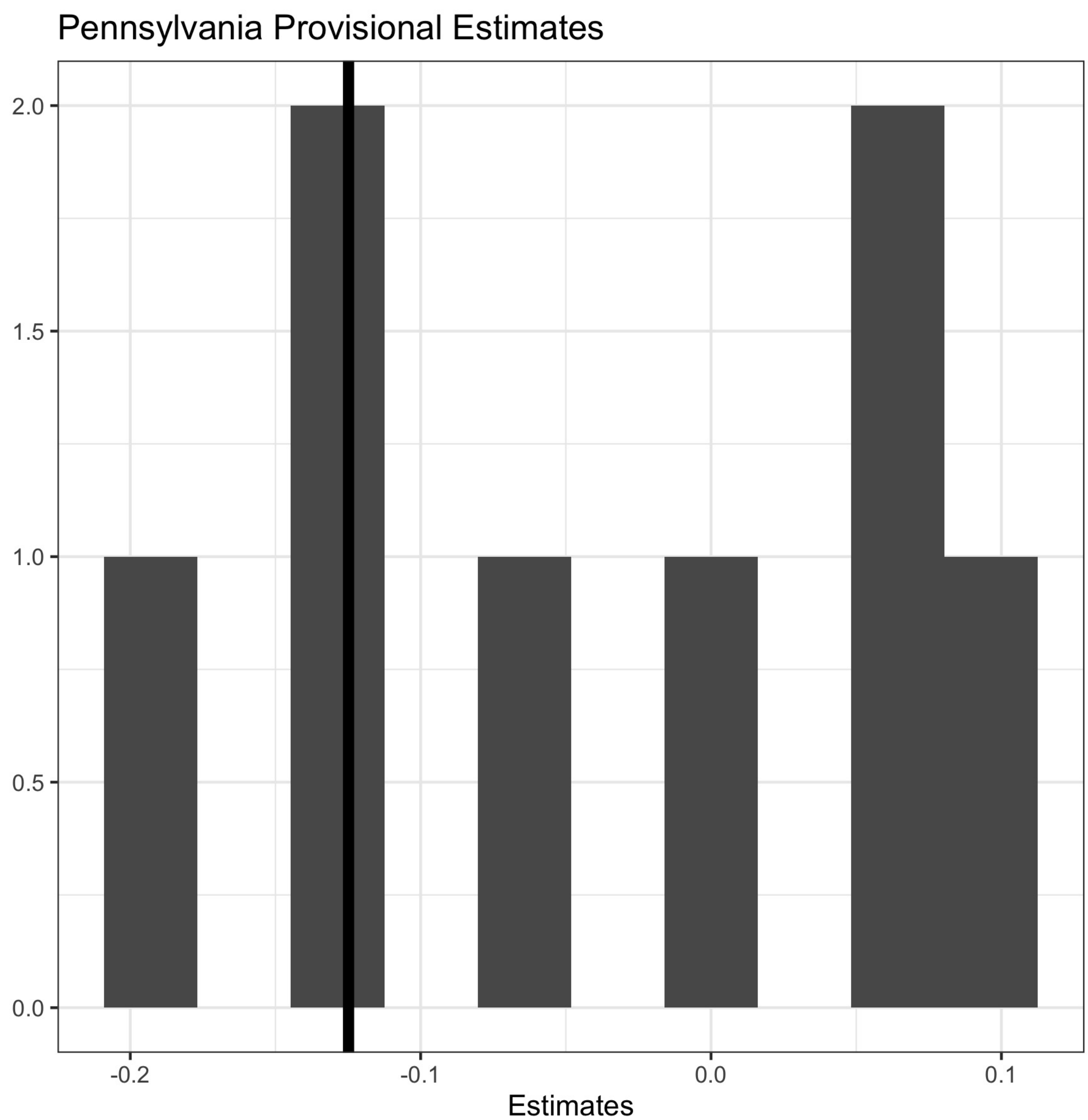


Figure 13: Distribution of Estimates for Alternative Precinct Differencing Orders, Share of Biden Ballots from Pennsylvania Provisional Ballots

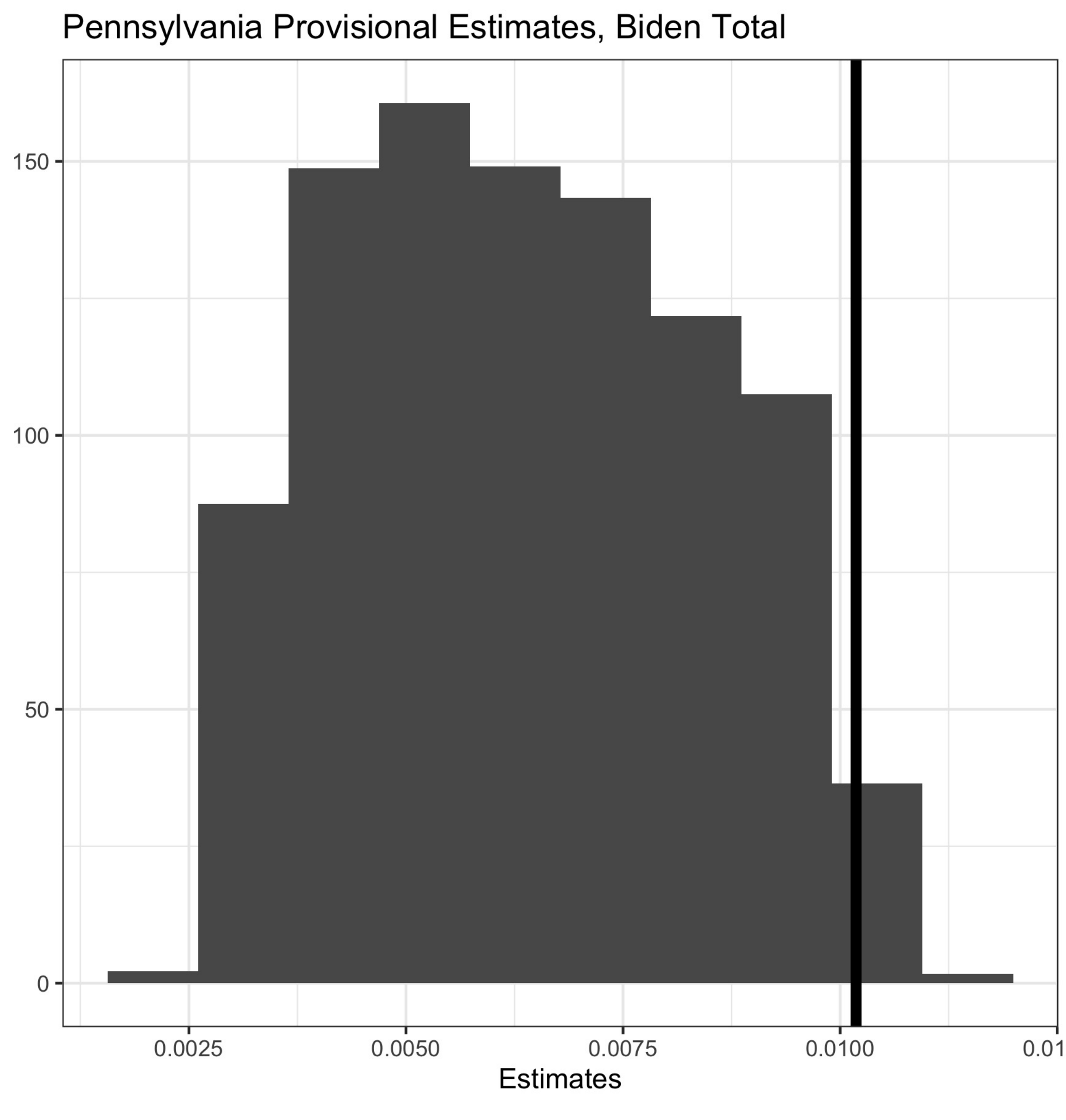


Table 15: Pennsylvania Provisional Ballot Results, Total Ballots

	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	0.385	0.342
		(0.078)	(0.103)	(0.082)
Allegheny County	0.010	0.007	0.007	0.007
	(0.004)	(0.002)	(0.002)	(0.002)
Observations	87	174	174	174
Precinct-Pair Fixed Effects			✓	
County-Pair Fixed Effects				✓