Evaluating a New Generation of Expansive Claims about Vote Manipulation*

Justin Grimmer[a], Michael C. Herron[b], and Matthew Tyler[c]

[a]Democracy and Polarization Lab, Department of Political Science, and Hoover Institution, Stanford University. 616 Jane Stanford Way, Stanford CA 94305 (jgrimmer@stanford.edu).

[b]Program in Quantitative Social Science, Dartmouth College. 6108 Silsby Hall, Hanover NH 03755 (michael.c.herron@dartmouth.edu).

[c]Postdoctoral Fellow at the Democracy and Polarization Lab and the American National Election Studies, Stanford University. 616 Jane Stanford Way, Stanford CA 94305 (mdtyler@stanford.edu).

First Draft: October 25, 2022
This Draft: July 15, 2023

Abstract

In the wake of Donald Trump's attempt to overturn the 2020 presidential election, a cottage industry of conspiracy theorists has advanced ever more expansive claims of vote manipulation, going so far as to allege that all American elections are subject to manipulation—even in largely Republican states. In the extreme, these conspiracy theorists argue that candidates in U.S. elections are *selected* rather than *elected*. We evaluate two recent sets of claims about vote manipulation that allege algorithms are used to shift votes towards preferred candidates. Even though these claims are distinct, they fail for similar reasons. For example, both sets of claims assert that "unnaturally" accurate predictions of election results are evidence of vote manipulation, an allegation that is a result of predicting a variable with itself. Furthermore, both

*Grimmer and Herron served as expert witnesses in *Gilbert v. Sisolak* (Case No.: 22 OC 000851B), an election contest discussed at length in this paper. Grimmer also served as an expert witness in *Washington County v. Tim Sippel* (Case No.: 22 CV 07782), where the defendants unsuccessfully attempted to include Douglas Frank as an expert on elections.

claims make easily refuted errors in logic and data analysis and in addition misrepresent historical election patterns. While recent claims about vote manipulation are *prima facie* outlandish, their effects on policy and the public are real. Refuting false claims about vote manipulation is essential to ensuring the continued functioning of U.S. elections and American democracy more generally.

Contents

Appendices

1 Introduction

Donald Trump's refusal to concede the 2020 presidential election was a shock to the public's faith in the U.S. electoral system. And almost two years on, Trump continues to claim that he won the 2020 presidential race (Hansen 2022). In response to allegations that the 2020 election was stolen (Hasen 2022), scholars, journalists, and the courts have resoundingly refuted a multitude of claims about illicit vote manipulation (Cassidy 2021; Danforth et al. 2022; Eggers et al. 2021). While other presidential elections have been the subject of allegations of malfeasance (Cottrell et al. 2018), the scale of claims made in the aftermath of 2020 election dwarfs what the United States has previously experienced.

Despite the failure of Trump's objections to the 2020 election, the consequences of his unprecedented allegations continue to unfold, having laid a foundation for even more expansive claims of vote manipulation. Unlike Trump's initial post-2020 election allegations, these new claims of malfeasance indict nearly every election in the United States, including Republican primary elections, elections in deep red states that resulted in Republican legislative super-majorities, and even city council elections in relatively small towns.[1] In the extreme, these new conspiracies allege that winners of U.S. elections are *selected* rather than *elected*.

We document two high-profile examples of expansive claims of vote manipulation and demonstrate how supposed evidence in favor of the claims fails to establish any evidence of malfeasance. Our first example considers allegations made by the second-place candidate in the 2022 Nevada Republican gubernatorial primary, Joey Gilbert, who claimed that vote manipulation led to his June 2022 loss to Joseph Lombardo, the Trump-endorsed sheriff of Clark County, Nevada. Gilbert's claims about vote manipulation were the foundation of an election contest filed on July 15, 2022 in Nevada's first judicial district (*Gilbert v. Sisolak*, Case No.: 22 OC 000851B). Gilbert's contest alleged that an unspecified group of conspirators deployed an algorithm to shift votes in his gubernatorial primary from Gilbert to Lombardo and votes across primary elections to the incumbent Democratic Nevada governor Steve Sisolak. The contest was dismissed on August 11,

[1] See Frank (2022c), a presentation by Douglas Frank in Idaho, and both Thayer (2022) and Corasaniti and Berzon (2022) on allegations of vote manipulation in Grand Junction, Colorado.

2022, with Carson City District Court Judge James Wilson subsequently ordering sanctions against Gilbert and requiring him to pay attorney fees for Lombardo (Seeman 2022).

Claims similar to Gilbert's resurfaced in 2022 during a County Clerk election in Will County, Illinois. In this election, Republican Gretchen Fritz alleged that her loss to Lauren Staley Ferry was manipulated by an unspecified group of conspirators who sought to deliver the election to her opponent. Fritz's contest was also dismissed and, like the Gilbert campaign, Fritz's campaign received sanctions for the lawsuit (Kukulka 2023b).

Our second example considers an expansive election conspiracy theory that is regularly pitched to local election officials and civic groups as part of a campaign to influence the administration of future elections (Mollenkamp et al. 2022). To wit, we examine Douglas Frank's claims that an algorithm determines the results of all U.S. elections. As in the Gilbert and Fritz contests, Frank alleges that an algorithm is used to obtain preferred election results. Unlike in Gilbert/Fritz, though, Frank alleges that this fraud occurs across all U.S. elections and at all levels. Frank alleges that an unnamed group of conspirators use "phantom voters" to stuff the ballot box, causing the support for conservative candidates to be diluted or to enable the election of establishment Republican and Democrats. Frank peddles his claims of election manipulation to local organization across the U.S., to audiences of both election officials and local integrity groups. Frank's claims have been amplified by Mike Lindell, a pillow magnate and Donald Trump supporter on his eponymous television network, with movie-length arguments that elections are stolen (Thayer 2022) and annual symposia like the "Cyber Symposium" (Marks 2021) and the "Moment of Truth Summit." Frank's claims have also surfaced in court proceedings, including an Oregon state-level case involving access to voting machine information (*Washington County v. Tim Sippel* (Case No.: 22 CV 07782)) and a federal lawsuit alleging voter fraud throughout the state of Oregon (*Thielman et al. v. Fagan et al.* (Case No.: 22 CV 01516)).

The reasons we focus on claims made in the Gilbert/Fritz contests and by Frank are twofold. First, these claims are important examples of expansive claims of fraud that are motivating the efforts of grassroots activists to change how American elections are conducted. This is exemplified by the manner in which Douglas Frank tours the country, spreading his message about election integrity, meeting with election officials during the day, and providing public talks at night (e.g., Ervin (2023)). In front of audiences, Frank repeats his argument that American elections are hopelessly manipulated and therefore a wholesale reform of U.S. election policy is necessary. Frank has already affected election administration: based in part on his efforts (Frank 2022b), Shasta County, California, canceled its contract with Dominion voting systems and is now selecting a new vote counting method, which could include hand counting all ballots returned in the county (Arthur 2023).

Also inspired by activists like Frank are grassroots organizations often called "election integrity groups," who form audit teams and go door-to-door in an effort to find evidence that confirms conspiracies made about contemporary elections (Frank 2022b). The groups have the explicit goal of uncovering evidence they can use to persuade local officials to alter how their local elections are conducted. And there is compelling evidence that proponents of election conspiracies increasingly occupy the time of elections officials. They flood local electoral offices with public records requests (Gardner and Marley 2022) and lead voters to ask officials about "rigged" voting machines and whether mail-in ballots are legal (Kang 2022).

Beyond grassroots movements seeking to change how American elections are conducted, the claims we investigate exert real costs on the infrastructure of American elections. Candidates who have sufficient financial support, exemplified by Joey Gilbert and Gretchen Fritz, can file lawsuits requiring hearings, depositions, and expert reports. This exerts a financial cost and can delay election certification.

A second reason for our consideration of Gilbert/Fritz and Douglas Frank is that the two corresponding schools of claims exhibit flaws that are common across many claims about fraud in American elections. The most important flaw is a lack of validation: no one has shown that the methods advocated by the contestants in Gilbert and in Fritz and by Douglas Frank are actually reliable at diagnosing vote manipulation. In particular, in both the voter fraud allegation cases that we examine here, there is no evidence that, in the absence of vote manipulation, we would observe different patterns in election data than the patterns labeled by the proponents of Gilbert and Fritz and by Frank as anomalous.

Not only do the Gilbert/Fritz contestants and Douglas Frank not provide any evidence their proposed methods are accurate at diagnosing fraud, we provide extensive evidence that the tests that these individuals propose cannot possibly identify vote manipulation. Indeed, we show that supposedly anomalous patterns in election data highlighted by the Gilbert/Fritz contestants and by Douglas Frank are the result of poor research designs and confused analyses. Even though these individuals allege that manipulated elections lead to highly predictive relationships, their arguments are flawed. We show (1) that the Gilbert/Fritz contestants and Douglas Frank have duped themselves, by discovering (in their own unique ways) that a variable will be highly correlated with itself. And, (2) we highlight the lack of any analysis implying that predictability in U.S. elections is an indication of fraud. In fact, we provide both historical and synthetic examples where there is an absence of manipulation yet we observe what the Gilbert/Fritz contestants and Douglas Frank might call suspicious predictability.

Our research is useful because it provides tools to members of the public, judges, and election officials to use when scrutinizing claims about vote manipulation. Fair elections, a cornerstone of democratic politics, are at risk when unsupported claims about malfeasance cause logjams in the court system and cause voters to lose trust in democracy (Berlinski et al. 2021). The best response to claims of vote manipulation is a careful examination of the underlying evidence.

Our work is also useful because it provides a set of simple questions anyone can ask to obtain basic evidence of whether a particular claim of vote manipulation is based on a sound methodology. We return to this matter in our conclusion when we offer two questions anyone without any statistics training can pose when confronted with alleged evidence of vote manipulation.

2 Vote manipulation in contemporary American elections

There is an extensive empirical literature studying the prevalence of vote manipulation in the United States. Much of this literature has developed in the shadow of claims about voter impersonation that have been used to publicly motivate voter identification laws (Ansolabehere and Persily 2008; Mazo 2018).

While millions of Americans report believing that voter fraud is a serious problem in the United States, research on the subject of vote manipulation and election malfeasance more generally finds no evidence in support of broad concerns. There is no compelling evidence of systematic and widespread double-voting in the 2012 general election (Goel et al. 2020), and studies of the 2016 and 2020 elections similarly find no evidence in support of claims that these two elections were subject to widespread malfeasance (Cottrell et al. 2018; Eggers et al. 2021). Other analyses of U.S. elections reach similar conclusions, uncovering no evidence of systematic problems with election integrity in the United States (Levitt 2007; Minnite 2007, 2010; Wu et al. 2020). Scholars do not argue that malfeasance literally never occurs (e.g., Herron 2019), but Li et al. (2022) is a good example of how post-Election Day changes in Democratic candidate support rates in California, Colorado, and North Carolina may appear suspicious but actually reflect systemic patterns in vote counting procedures.

2.1 Assessing vague public claims about vote manipulation

We examine two instances of broad claims about vote manipulation in U.S. elections. Even though both claims are either contained in hundreds of pages of expert reports and hundreds of hours of public presentations, it is difficult to infer the specific details of what the claims are alleging. Neither set of claims has a clearly articulated theory of vote manipulation, and neither substantiates a connection between a theory of vote manipulation and a set of statistical methods used to ostensibly diagnose whether manipulation occurred. Instead, the advocates of our two sets of claims rely on implication and suggestion to advance their claims, or they merely assert that some patterns are anomalous without a basis for that assertion. This make engaging with their claims difficult and, in some instances, potentially impossible. To counteract this, throughout this paper we make our best effort to articulate the strongest and most coherent version of the claims we address before refuting them. This conservative approach ensures that our critiques engage with the best possible versions of the theories of vote manipulation that motivate our research.

3 Gilbert/Fritz and “Irrefutable Geometric Proof” of vote manipulation

The 2022 Nevada Republican gubernatorial primary saw Joseph Lombardo, a Trump-endorsed candidate and sheriff of Clark County, face off against Joey Gilbert, a retired boxer and personal injury attorney who participated in the January 6, 2021 riot at the United States Capitol (Gentry 2022), former U.S. Senator Dean Heller, and several other lower-profile candidates. Lombardo prevailed in this primary, winning approximately 38.4 percent of the Republican vote — 26,023 more votes than Gilbert and 55,674 more votes than Heller out of 228,570 total votes cast in the gubernatorial race.[2] Despite Lombardo’s decisive victory, Gilbert refused to concede, declaring that, “I smell a lawsuit because this STINKS!” (Cole 2022).

Subsequent to a recount that confirmed his loss to Lombardo (Robison and Lacanlale 2022), Gilbert followed through on his threat on July 15, 2022, filing a formal election contest apparently funded by cryptocurrency mogul and election conspiracy theorist Robert Beadles.[3] The contest, which alleged that reported results for Nevada’s 2022 Republican gubernatorial primary were “mathematically impossible” and therefore the election stolen (p. 15, ¶ 50), is a 194 page document which includes a set of seven appendices, the most important of which (Appendix A) was a report written by Edward Solomon, a former swing- set installer who has risen to prominence with his analysis of supposed election irregularities (Gordon 2021). Also among the contest’s appendices were reports authored by Oliver A. Hemmers, a physicist who had previously worked as an expert on noise pollution and drunk driving (Appendix B);[4] G. Donald Allen, a former mathematics

---

[2] Official statewide results for the 2022 Republican gubernatorial primary can be found at the website of the Nevada Secretary of State. See https://silverstateelection.nv.gov/NVOther (last accessed September 13, 2022).

[3] The contest is available at https://s3.documentcloud.org/documents/22088788/gilbert-v-sisolak-et- al.pdf (last accessed September 12, 2022). On Beadles’s background, see Robison (2022); and on Beadles’s support of Gilbert, see his statement at “Uncharted Waters. . . Here We Go!,” available at https: //operationsunlight.com/2022/07/15/uncharted-waters-here-we-go (last accessed September 13, 2022).

[4] On Dr. Hemmers’s previous expert work, see lines 15-24 on p. 14 of the transcript of his July 28, 2022 deposition in *Gilbert v. Sisolak*.

professor at Texas A&M University (Appendix D); and, Walter C. Daugherity, a former computer science lecturer at Texas A&M (Appendix F).[5]

The same type of claims surfaced in a post-election dispute for County Clerk in the 2022 general election in Will County, Illinois (Kukulka 2023a). This election saw the Democratic candidate Lauren Staley Ferry win 121,833 votes and the Republican candidate Gretchen Fritz, 108,629 votes. According to a contest filed on December 28, 2022, Fritz initially became suspicious because "Upon examination of the November 8, 2022 election returns for Will County, petitoner noticed that in Will County, Democrat Governor Candidate JB Pritzker received 117,475 votes, or 4,358 fewer votes than respondent Lauren Staley Ferry" (p. 1 ¶ 7). The contest claimed that, "It appeared quite unusual that a candidate for Will County Clerk, listed at least eight offices below the office with the most media coverage and largest political spending in the state, would receive more votes than the gubernatorial candidate of her party" (p. 2, ¶ 8). As in Gilbert's Nevada-centered contest, Fritz's contest contained expert reports authored by Solomon (Exhibit B) and Daugherity (Exhibit D). The contest claims that "Mr. Solomon's analysis reveals that the results of Will County Clerk election are not the result of a free and fair election" (p. 2, ¶ 11). And, that "The algorithm and The Bivariate Cubic Manifold on page 8 Exhibit B solves for the final votes and percentages in each category: petitioner's Election Day Vote, Early Vote and Mail-In Vote and respondent's Election Day Vote, Early Vote and Mail-In Vote with partial knowledge determines the final results in all 310 precincts, a mathematical an geometric impossibility in a fair, not predetermined election" (p. 2, ¶ 12). The Fritz contest summarizes Daugherity's analysis similarly, alleging that he provides conclusive evidence that, "Will County Clerk election are not the results of a free and fair election" (p. 2, ¶ 17).

[5] While the contest's Appendix A was not signed by Edward Solomon, the body of the contest states that Solomon was the appendix's author (pp. 7-8, ¶ 25) Moreover, Allen in his report identifies Appendix A as Solomon's work, writing in the contest's Appendix D that, "I have reviewed, mathematically, the reports by Edward Solomon furnished to me which mathematically analyzes (sic) the June 14, 2022, Republican gubernatorial primacy in Clark County, Nevada, as well as other races" (p. 1, ¶ 4). Daugherity wrote similarly of Appendix A, stating that, "I have reviewed the reports by Edward Solomon furnished to me which mathematically analyze the June 14, 2022, Republican primary in Clark County, Nevada, as well as other races" (Appendix F, p. 2, ¶ 6).

3.1 The Gilbert/Fritz alleged manipulation conspiracy

Vote manipulation is alleged in both the Gilbert and Fritz contests, but neither contest provides clear details about how manipulation ostensibly occurred. Edward Solomon wrote expert reports in both Gilbert and Fritz, and in these reports Solomon alleges that manipulation is evidenced because of the presence of an "illegal formula" (pp. 9-10, ¶ 32), which lead to "geometric interference" (p. 12, ¶ 36). Solomon's Appendix A in Gilbert argues that such predictability was the product of a "Neural Network", a claim he also makes in Fritz (p. 8, Exhibit B).

While Solomon's reports are sparse on details, we are able to assemble from them some sense of the logic Solomon adduces. It appears that both reports focus entirely on manipulation that allegedly occurred after ballots had been cast. In Gilbert, Solomon alleges that a neural network took votes cast for Gilbert and assigned them to Joseph Lombardo (the winner of the Republican primary) and Steve Sisolak, the lone Democratic candidate and incumbent Nevada governor at the time.[6] Solomon is also vague about the specific procedure used to (re)assign votes and how a vote-reassignment formula was calculated, ultimately confessing that the specific manipulation process would "most likely remain as mysterious as the thought processes which incited the Neural Network, Leela Zero, to execute her Immortal Queen Sacrifice against Stockfish." In Fritz, Solomon provides extensive detail about how he thinks a vote-manipulating neural network was trained, the constraints that were used, and the process used for network optimization. But he provides no direct evidence for his claims about a neural network. The judge's opinion that dismissed Fritz's contest states that, "the allegations [of vote manipulation] lack specificity as to who created the algorithm, who loaded it into vote tabulation machines and how" (Anderson 2023, p. 5).

While the contests in Gilbert and Fritz are vague about actual conspiracies, they do nonetheless claim to have uncovered empirical evidence of fraud. In what follows we focus on Gilbert because the claims made in Fritz are essentially repeat those in the former. All of our arguments, therefore, about Gilbert case apply to the Fritz as well.

[6] Neither the Gilbert contest nor Solomon explain why anyone seeking to manipulate a Republican primary would bother to shift votes from a Republican candidate to an unopposed, incumbent Democratic candidate.

3.2 Alleged empirical evidence of vote manipulation in the Gilbert contest

The Gilbert contest alleges that candidates in an election should, within precincts, receive roughly equivalent vote shares across methods of voting. In Nevada, which the locus of the Gilbert contest, this would imply that Gilbert should have received equal shares from in- person voting on Election Day, in-person early voting, and voting via mail ballot. Examining the results from the 2022 Republican gubernatorial primary, the Gilbert contest argued that precinct-level deviations from vote-share equality in each mode of voting provided evidence that vote manipulation tarred this election. On this point, the contest states that, "There is absolutely no correlation between Gilbert's Election Day, Early, and Mail-in Percentages across precincts," labeling this an "irregularity" (pp. 8-9, ¶ 29).

The claims made in the Gilbert contest fail for two reasons, among others. First, they mischaracterize historical election results: in fact, there is no evidence that candidates in Nevada elections receive equal vote shares across different methods of voting. This is unsurprising given that different types of voters may have different preferences for, or varied access to, alternative methods of voting. Second, what the contest argues is surprising predictability is actually the result of using a variable to predict itself — naturally, this cannot be evidence of election malfeasance.

3.2.1 Voting methods and candidate vote shares in Nevada

We now consider the Gilbert contest's claims about candidate vote shares across voting methods and show that they do not hold.

The contest's claims about methods of voting

The body of the Gilbert contest states that,

> *In a fair election, we expect a strong linear correlation between Gilbert's Election Day, Mail-in and Early Vote percentages across the precincts. That is, whatever Gilbert's Election Day percentage is at a particular precinct, we expect both Gilbert's Mail-in percentage and Early Vote percentage to be roughly the same, not exactly, since that would imply causation. . . but roughly, which implies a strong correlation, which would be consistent with Clark County's Historical Election*

Results in all years prior to 2020, both in the Primaries and the General Elections (bold in the original, p. 8).

Moreover, in Solomon's Appendix A,

> *In a fair election, we expect a candidate that [sic] received 10% of the Election Day Vote to get roughly 10% of the Mail-in Vote; likewise if they get 90% of the Election Day Vote, we expect them to get 90% of the Mail-In Vote. Even if Democrats prefer to vote by mail, that should reflect in both percentages across the precincts, not just one of them. In other words, if we plot the election day and mail-in percentages against each other across the precincts, they should array themselves across a 45 degree angle y = x (Appendix A, p. 9).*

Solomon grounds this claim with the following: "[E]ach candidate's proportion of election day, to early, to mail-in ballots, should be roughly the same, as all other candidates, in all races. Again, this is confirmed by historical records of elections prior to 2020 and countless simulations" (Appendix A, p. 7).

In his contribution to the Gilbert contest, Hemmers also reiterates the argument that candidate vote shares should not vary by voting technology, asserting that,

In a fair election, the sum of the Early Day and Election Day votes should produce very similar results to the Mail-in votes when the regressions analysis has a high confidence (usually called R^2), meaning the x-values and the y-values should be similar (when x is 10% then y should be close to 10% as well) and not off by 25% (Appendix B , p. 2, ¶ 5).

We are aware of no academic literature arguing that "fair elections" are those that feature identical — or even similar — candidate vote shares across different voting methods. And, to the best of our knowledge, there are no legal opinions on election fairness that assert this. The Gilbert contest simply asserts that "fair elections" feature similar candidate vote shares across methods of voting without any supporting scientific theory or historical evidence.

Voter preferences for voting methods and state election laws imply that candidate vote shares differ by voting method

Across the United States, voter types sort themselves into different methods of voting and sometimes even into times of voting (Herron and Smith 2012). In the 2020 general election, for example, relatively educated voters cast early and vote-by-mail ballots at greater rates than voters with less education (Scherer 2021). And, in recent presidential elections, the more educated a voter, the more likely the voter was to have supported the Democratic candidate for president (e.g., Sides et al. 2019). If in an election voters with high (or in principle, low) educational attainment systematically vote early and with vote-by-mail ballots, and if this same type of voter is disproportionately Democratic or Republican, the result will be a correlation between voting method and candidate vote share.

The case of Texas provides an example of this phenomenon albeit based on age as opposed to educational attainment. In Texas, voters age 65 and older are permitted to vote absentee without an excuse, but younger voters do not have this privilege. Not surprisingly, older Texas voters cast absentee ballots at greater rates than younger Texans, as noted by Yoder et al. (2021), who study both Texas as well as Indiana, another state which provides special absentee voting dispensation to voters 65 and up. Insofar as age is correlated with political preferences (e.g., Ghitza et al. 2022), Texas law practically guarantees that there is a correlation in the state between voting method and candidate vote share. Louisiana, Mississippi, and South Carolina are three other states in which voters 65 years or older are able to cast absentee ballots without excuses.[7] Even in states without this cutoff, there is likely to be a correlation between preferences and mode of voting. Atkeson et al. (2022) show that in New Mexico older voters and those weary of COVID were more likely to vote by mail in the 2020 election. Thus, the Gilbert contest's claim that fair elections have equal candidate vote shares across voting methods is inconsistent with the interaction between election laws in multiple states in the country and the fact that educational attainment and age are correlated with political preferences.

---

[7] On Louisiana, see "VOTE ABSENTEE," *Louisiana Secretary of State*, available at https://www.sos.la.gov/ElectionsAndVoting/Vote/VoteByMail/Pages/default.aspx (last accessed September 19, 2022); on Mississippi, see "ABSENTEE VOTING FREQUENTLY ASKED QUESTIONS," *Secretary of State*, available https://www.sos.ms.gov/absentee-voting-information (last accessed September 19, 2022); and on South Carolina, see "Absentee Voting," *South Carolina Elections Commission*, available at https://scvotes.gov/voters/absentee-voting/ (last accessed September 19, 2022).

Nevada elections do not feature equal candidate vote shares across voting methods

While the empirical literature on American elections provides no foundation for the Gilbert contest's claim that fair elections feature equal candidate vote shares across methods of voting, it could nonetheless be the case that the contest's claim in this vein was made in a narrow way, simply about Nevada. In light of this possibility, we show that the claim is not supported by historical Nevada election data.

To do this, we examine precinct level election results in historical Nevada elections as well as results from Nevada elections contemporaneous with the 2022 Republican gubernatorial primary. For precincts in Clark County, Nevada, Figure 1 plots winning candidate vote shares (in percentages) by method of voting for the Republican gubernatorial primaries of 2014, 2018, and 2022. Panel (a) of the figure plots Election Day and early voting vote shares and Panel (b), Election Day and mail-in ballot vote shares. Each point in the figure's six plots denotes a Clark County precinct, and points are sized proportional to total votes cast by both relevant voting methods. Nevada election law changed in 2021 with the passage of Assembly Bill 321, which mandates that all registered voters in Nevada receive mail ballots prior to an election.[8] That said, Figure 1 and subsequent figures in this paper treat absentee voting and voting by mail as equivalent.

There are 17 counties in Nevada (treating Carson City as a county), and Figure 1 includes data from Clark County only. This county, unlike the other 16 such jurisdictions in Nevada, regularly publishes machine-readable reports that disaggregate election returns by method of voting. Clark County is by far the most populous county in Nevada; there were 1,821,058 registered voters in the Nevada as of the 2022 primary, and of these 1,279,597 (slightly more than 70 percent) resided in Clark County.[9]

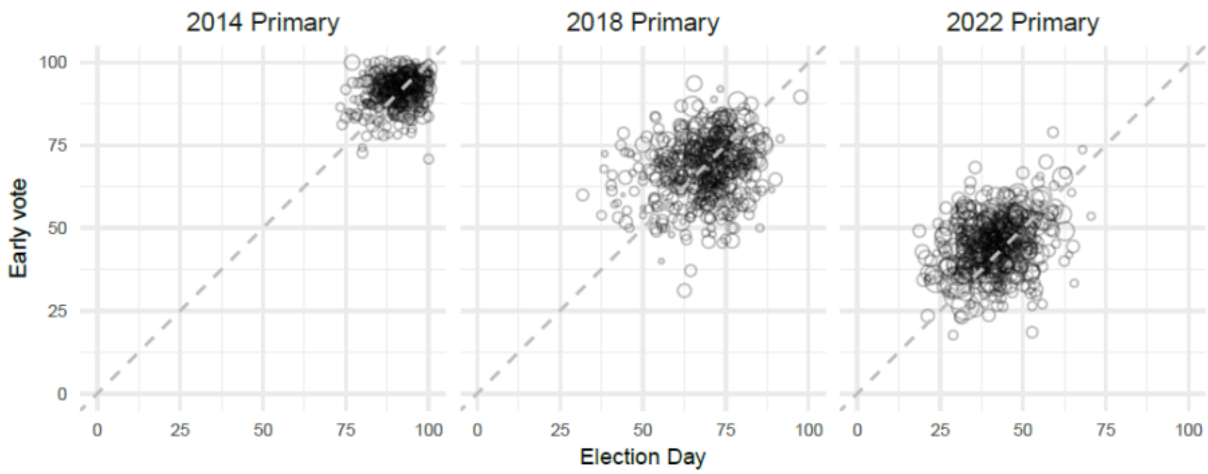
The rightmost plot in Panel (a) of Figure 1 shows that Election Day voters in Clark County precincts supported Joseph Lombardo, the winning 2022 Republican

[8] "Governor Sisolak signs groundbreaking legislation to expand voting access in Nevada, increase education funding," *Nevada Governor Steve Sisolak*, June 2, 2021, available at https://gov.nv.gov/News/Press/2021/ Gov signs groundbreaking legislation expand voting access/ (last accessed September 19, 2022).

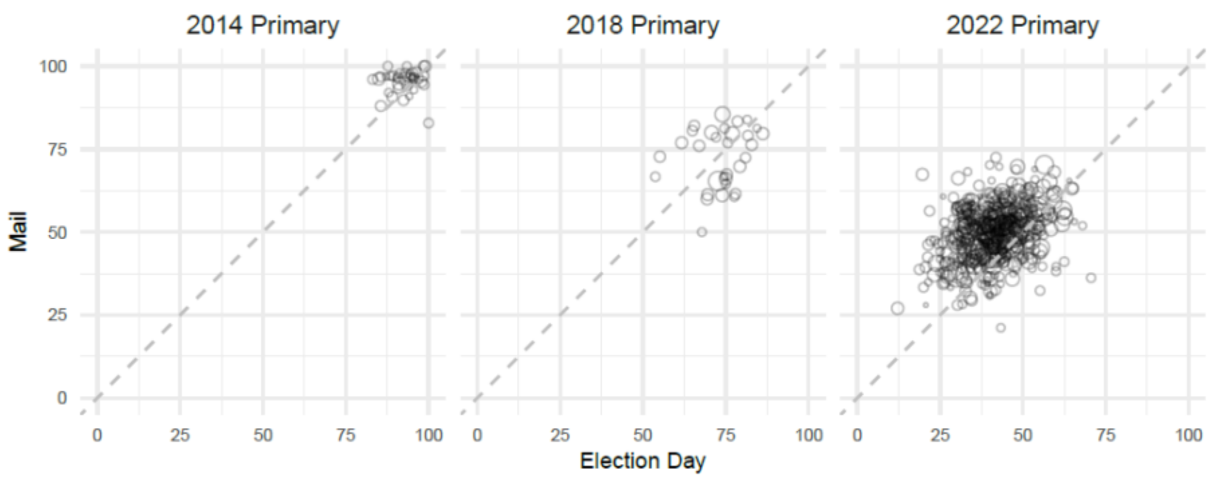
[9] Voter registration statistics are drawn from the website of the Nevada Secretary of State, available at https://silverstateelection.nv.gov/vote-turnout (last accessed September 19, 2022).

gubernatorial candidate, at slightly lower rates than did corresponding early voters. Note that the majority of the points in the rightmost plot in this panel fall *above* the plot's 45-degree line. In the 2014 and 2018 Republican gubernatorial primaries, no such relationship exists for the winners of these two contests, Brian Sandoval and Adam Laxalt, respectively.

Figure 1: Election Day and early voting precinct vote shares of winning candidates in the 2014, 2018, and 2022 Republican gubernatorial primaries in Clark County.



(a) Election Day versus early voting



(b) Election Day versus mail

Note: includes only precincts with at least 25 voters for each voting method.

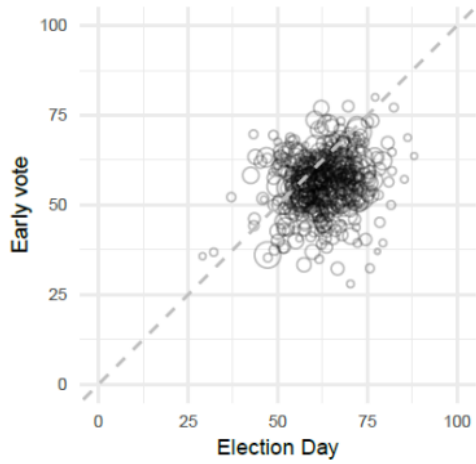
With respect to Panel (b), the cloud of points in the 2022 primary plot shows that mail ballot voters supported Lombardo at greater rates than did corresponding early voters. This occurred in the 2014 Republican gubernatorial primary as well, in which mail ballot voters supported Sandoval at greater rates than did corresponding early voters. There are fewer precincts pictured in the leftmost plot of Panel (b) because, prior to the passage of the aforementioned Assembly Bill 321, mail ballot voting was not nearly as common in

Nevada as it is now. Sandoval won his primary overwhelmingly (the points in the two leftmost plots of Panels (a) and (b) are up and to the right) but a mail versus Election Day difference in candidate support rates is evident in the 2014 Republican gubernatorial primary regardless. There are no credible claims that this primary was subject to vote manipulation.

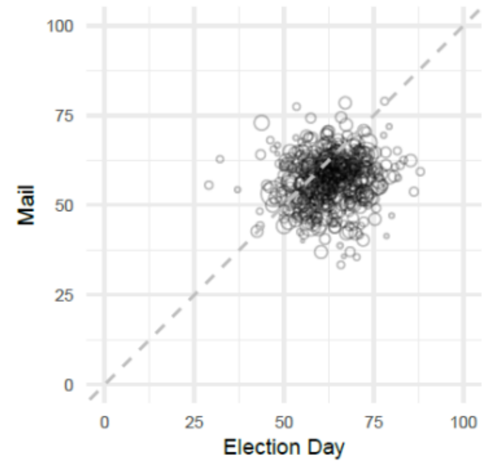
One could in principle argue that some form of vote manipulation affected the 2022 Republican gubernatorial primary by driving down the Election Day vote shares of the primary's winner, who just happened to be Joseph Lombardo. However, a quick look at the 2022 United States Republican and Democratic United States Senate primaries dispels this conjecture. All three of these primaries appeared on 2022 Nevada primary ballots, and vote shares from winning United States Senate candidates are depicted in Figure 2.

From Panels (a) and (b) of this figure, it is evident that, within Clark County precincts, the winner of the Republican Senate primary did better with Election Day voters than with early and mail ballot voters, the *opposite* of what occurred in the 2022 Republican gubernatorial primary. Regarding the Democratic United States Senate primary, for which there were no known credible allegations of vote manipulation, the winning candidate of this contest did worse with Election Day voters than with early and mail ballot voters. Overall, the lack of consistency across the four panels of Figure 2 is not consistent with vote manipulation in 2022 Nevada primaries that, say, systematically diminished vote shares of winning candidates.

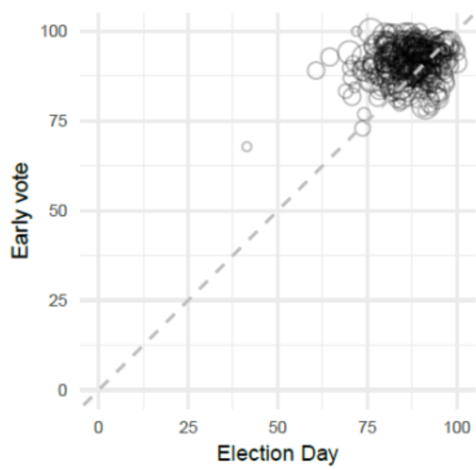
Figure 2: Election Day, early voting, and mail ballot precinct vote shares of winning candidates in 2022 United States Senate primaries in Clark County



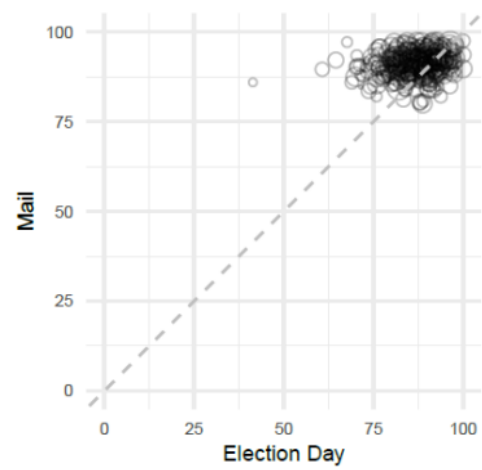
(a) Election Day versus early voting (R)



(b) Election Day versus mail (R)



(c) Election Day versus early voting (D)

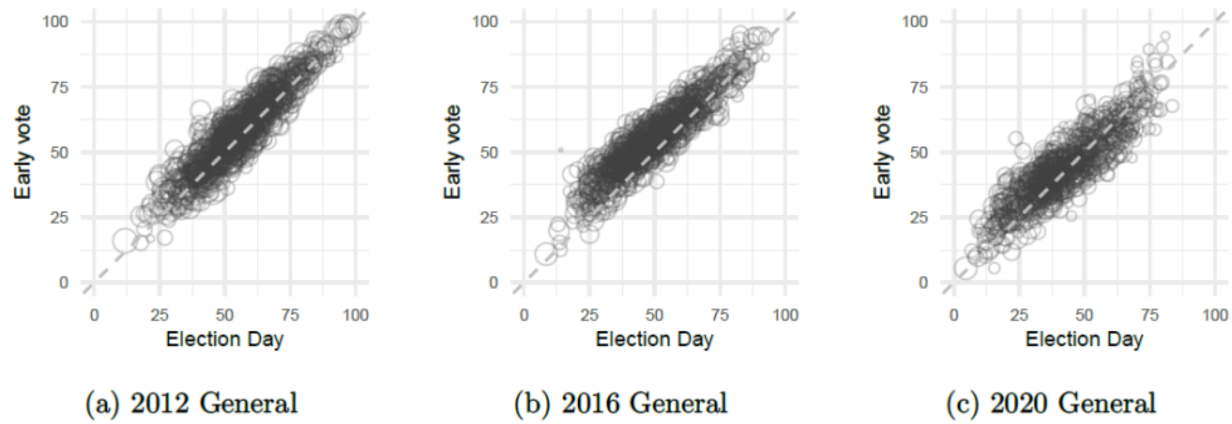


(d) Election Day versus mail (D)

Note: includes only precincts with at least 25 voters for each voting method; "D" denotes the Democratic primary and "R," the Republican primary.

Differing candidate vote shares by voting method are also apparent in the three most recent presidential elections in Nevada, where we find no consistent relationship between performance in early vote totals, Election Day vote totals, and who wins. For Clark County precincts, Figure 3 plots Election Day and early voting vote percentages for Barack Obama, Hillary Clinton, and Joe Biden, the winning (Nevada) presidential candidates in the 2012, 2016, and 2020 presidential elections, respectively. Within Clark County precincts, Obama in 2012 did marginally better among early voters than Election Day voters, Clinton in 2016 clearly had more early voter support than Election Day support, and there is little evidence in Figure 3 that Joe Biden performed better with early or Election Day votes in 2020.

Figure 3: Election Day and early voting precinct vote shares of winning presidential candidates in 2012, 2016, and 2020 general elections in Clark County.

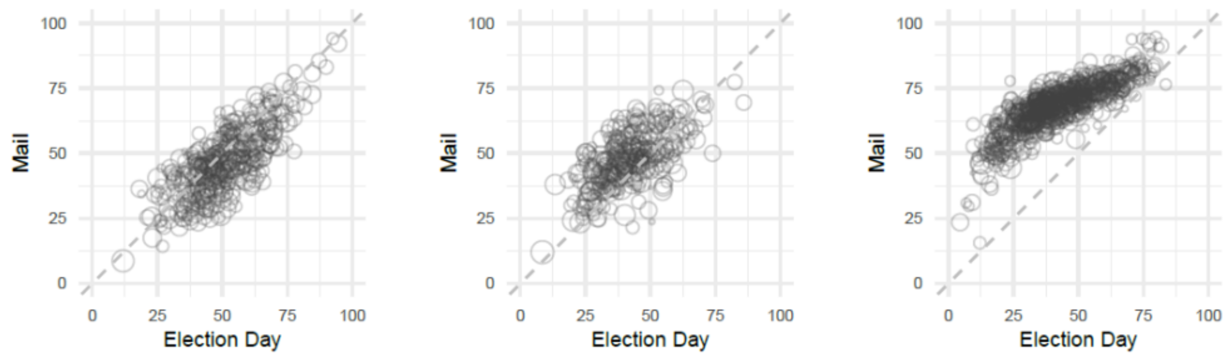


Note: includes only precincts with at least 50 voters for each voting method.

Continuing with our analysis of recent presidential races in Nevada, candidate vote shares from Election Day and mail balloting in 2012, 2016, and 2020 are displayed in Figure 4. In 2012, there is some evidence of a relationship between Obama's mail ballot vote share and his Election Day vote share; namely, Obama performed worse among mail ballot voters than Election Day voters. In both the 2016 and 2020 presidential contests, Clinton and Biden, respectively, clearly performed better among mail ballots than on Election Day ballots, the latter election arguably exemplifying the jump in mail ballot voting that occurred particularly among Democratic voters in 2020 (Yoder et al. 2021).

In summary, a cursory look at only a few elections in Nevada turns up examples of races in which vote shares of winning candidates were not equal across methods of voting. This is not surprising given literature on how voters sort themselves into Election Day, early, and vote by mail voting. Thus, the Gilbert contest's claim — that a "fair election" features similar candidate vote shares across voting methods, something "consistent with Clark County's Historical Election Results in all years prior to 2020, both in the Primaries and the General Elections" (p. 8) — is false.

Figure 4: Election Day and mail precinct vote shares of winning presidential candidates in the 2012, 2016, and 2020 general elections in Clark County.



(a) 2012 General (b) 2016 General (c) 2020 General

Note: includes only precincts with at least 50 voters for each voting method.

Nevada election results are not perfectly predictable

The Gilbert contest alleges that it is possible to predict to a suspicious degree of accuracy Lombardo's mail ballot vote total in any given Nevada precinct using only Gilbert's mail ballot vote total and Lombardo's and Gilbert's early voting totals. The contest contends that this is evidence of vote manipulation, the ostensible logic being the idea that, if a feature of Lombardo's vote share is predictable, then it must have been illicitly fixed prior to an election.

We demonstrate that there is nothing surprising or suspicious about what the Gilbert contest calls the predictability of Lombardo's vote shares in the 2022 Republican gubernatorial primary. In fact, the contest's supposed predictions have little to do with actual election results. Instead, the contest proposes algebraic transformations of observed vote totals that mechanically create the impression of a predictable election result.

Allegations of suspicious predictability appear throughout the Gilbert contest, but the contest's exact claim about this subject is difficult to pin down. For example, Daugherity argues that vote totals for Gilbert and Lombardo disaggregated by mail-in ballots, Election Day ballots, and early ballots are so "tightly dependent" that "the exact number of mail-in votes for Lombardo" can be "exactly predicted. . . [in] every single precinct in [Clark] county"

(Appendix F, p. 4, ¶ 13). The body of the contest makes a related claim, positing that a "formula fits all precincts in [Nevada counties] without any variation to such formula" (p. 9, ¶ 32, 5th bullet point), allowing one to "solve [for a] candidate's aggregate percentage share of the ballots in each [Nevada] precinct with no knowledge of Mail-in to Election Day Votes" (p. 10, ¶ 32). Solomon and Allen make similar claims (Solomon's Appendix A, p. 2, and Allen's Appendix D, p. 2, ¶ 5).

The Gilbert contest's fundamental predictability argument is that information about mail-in votes from one gubernatorial candidate (here, Gilbert) was used across Nevada precincts — ostensibly by an unspecified individual or set of individuals — to determine the number of votes Lombardo would need to win Nevada's 2022 Republican gubernatorial primary. Then, a formula was implemented that manipulated cast ballots to guarantee that Lombardo received the requisite number of votes, this resulting in a functional — and thus highly predictable — relationship between total votes cast for varying candidates. While as evidenced above there are a number of ways across the Gilbert contest's many components in which predictability claims are adduced, here we analyze the prediction argument from Daugherity's Appendix F, this argument being the most coherent of the prediction claims in the contest.

The foundation of Daugherity's prediction argument

Daugherity divides 2022 Nevada Republican gubernatorial primary vote totals into two categories: votes for Lombardo (the primary winner, as we have already noted) and everyone else in the primary, labeled by Daugherity as "Gilbert et al." Daugherity then focuses on four variables for each precinct, defining them as follows (p. 3, ¶ 10):

a = Lombardo's mail-in vote total

b = Gilbert et al.'s mail-in vote total

c = Lombardo's early in-person vote total

d = Gilbert's et al.'s early in-person vote total

Referring to his variables a - d, Daugherity states that, "Clearly these numbers should be independent…[and that] knowing some of the numbers should not allow exactly predicting the other numbers" (p. 3, ¶ 11). This claim is the basis of his predictability argument.

Lombardo's and Gilbert et al.'s mail-in vote totals are linearly dependent

Daugherity's claim about independence of the variables *a* - *d* is vague because it does not specify if Daugherity believes these variables are mutually independent or pairwise independent. The latter is weaker than the former, and thus we adopt this form of independence here. Specifically, it is unclear if Daugherity means that not only are any pair of variables supposed to be independent (satisfying the pairwise independence criterion) or if he also means that all the variables considered simultaneously are independent (satisfying mutual independence).

In any Nevada precinct, consider Daugherity's claim that *a* (Lombardo's mail-in vote total) "should be independent" of *b* (Gilbert et al.'s mail-in vote total). This (pairwise) claim must be false because by definition *a* and *b* sum to the total number of mail-in votes cast in the precinct. Given mail ballot turnout, if in any Nevada precinct *a* increases by one, then *b* must decrease by one. In other words, *a* and *b* are linearly dependent conditional on mail-in voter turnout.[10]

A parallel argument applies to Daugherity's *c* and *d*, which are linearly dependent conditional on early voting turnout. Therefore, Daugherity's claim in the Gilbert contest that his variables *a* - *d* "should be [pairwise] independent" is false.

Daugherity's prediction argument is inconsistent with voter sorting

Linear dependence aside, Daugherity's claim about the independence of *a* - *d* (that there should not be any) fails given known patterns of voter behavior in the United States. Voters across the country evaluate residential locations in a way that is correlated with political party affiliation (Gimpel and Hui 2015), and partisans sort themselves geographically (Brown and Enos 2021). Therefore, one should expect there to be locations in Nevada that had many supporters of Lombardo and other locations with many Lombardo opponents. This is certainly true empirically at the county level, with Lombardo having received a larger vote share in Clark County than in Nevada's more rural Washoe County.

[10] For expositional simplicity, here we ignore mail-in ballots cast without valid votes in Nevada's 2022 Republican gubernatorial primary as well as rejected mail ballots. If we were to include ballots like these, we would instead posit that a and b must sum to the total number of non-rejected mail ballots cast in a precinct with valid gubernatorial votes.

It follows from this logic that, in areas of Nevada where *a* is high, we would expect *c* to be high as well. This does not mean that the correlation between *a* and *c* (and similarly between *b* and *d*) should be expected to be literally one. Rather, if some areas of Nevada were relatively pro-Lombardo (anti-Lombardo) in mid-2022, these areas will have relatively large values of *a* and *c* (*b* and *d*). This logic, gleaned from empirical literature on American voters, renders as false Daugherity's assertion that *a*, *b*, *c*, and *d* "should be independent."

Daugherity's claim that Lombardo's mail-in vote totals can be predicted with certainty is empirically false

Voter sorting notwithstanding, Daugherity's claim that Lombardo's mail in vote total (*a*) "can be exactly predicted from *b*, *c*, and *d*. . . [in] "every single precinct in [Clark county]" (p. 4, ¶ 13) fails because it is simply not true. To assess this claim, we regress Lombardo's mail-in vote total (*a*) on variables *b* (Gilbert et al.'s mail in vote total), *c* (Lombardo's early in ~~person vote total), and *d* (Gilbert et a~~l.'s early in person vote total). If Daugherity's claim were correct, then *a* could be predicted from *b*, *c*, and *d*, with essentially no residual variance.[11]

[11] The regression summarized in Table 1 includes only Clark County precincts that have a > 0, b > 0, c > 0, and d > 0. This is because Clark County's official precinct-level election results suppress reported vote totals (i.e., variables like a - d, among others) when these totals are very small. When, say, a is not reported for a given precinct, we cannot determine its value.

Table 1: Daugherity prediction regression

	Dependent variable:
	a
b	0.886***
	(0.027)
c	0.602***
	(0.053)
d	−0.347***
	(0.045)
Constant	0.011
	(0.834)
Observations	647
R^2	0.863
Residual Std. Error	13.102 (df = 643)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results from our regression are in Table 1, and they show that Daugherity's *a* cannot be predicted by the variables *b*, *c*, and *d*. In fact, our regression has a residual standard error of approximately 13 mail ballot votes per precinct. To put this in context, on average Lombardo received roughly 38 mail-in votes in each Clark County precinct, implying that a residual standard error of 13 is substantial and far from being consistent with an exact prediction. This standard error reflects that there is some inherent variability across districts in how voters cast their ballots and who they support when voting those ballots, even though there may be some correlations.

To ensure that Table 1's lack of predictability in *a* is not due to a functional form limitation in the table's linear regression, we also predicted Lombardo's mail-in vote total *a* based on *b*, *c*, and *d* using (1) a random forest regression and (2) a generalized additive model (GAM).[12] Results from these two models are in Table 2.

[12] We fit our GAM with smooth terms for variables b, c, and d, with cross validation to determine model complexity as implemented in the R package MGCV. For our random forest model we use the **randomForest** package in R. There, we set the number of variables randomly sampled to be split on each iteration as two (following the default in R, which is √3, or the square root of the number of variables in the model) and set the minimum node size to be 5.

Table 2: Root mean squared errors for linear regression, random forest, and generalized additive models (GAM), all of which seek to predict a based on b - d.

Linear regression	Random forest	GAM
13.102	15.8	12.3

The leftmost column of Table 2 repeats the value of 13.102, the residual standard error from our aforementioned linear regression model. In the middle column of the table, we present the root mean squared error from a random forest regression, which is a flexible machine learning tool that does not make explicit functional form assumptions about the relationship between a dependent variable (here *a*) and a set of explanatory variables (*b* - *d*) (Hastie et al. 2009). Using a random forest, we find a root mean square error of approximately

15.8 votes — a non-trivial error given typical Lombardo vote totals per precinct.[13] Finally, in the rightmost column of Table 2, we present the results from our GAM, which is a tool that assumes the relationship between covariates (here, *b* - *d*) and the outcome variable (*a*) is additive but beyond this makes no assumptions about the functional form that describes the relationship between covariates and outcome. Using a GAM, we find a root mean square error of 12.3 votes — again, a substantial error.

Our root mean squared error results in Table 2 reflect in-sample calculations, meaning that we used the same data to estimates our models and then to draw conclusions. This biases our estimates of root mean squared errors towards zero, which is in favor of Daugherity's claim that *a* can be predicted by *b* - *d*.

Thus, according to standard statistical methods of prediction, there is not an exact relationship between Lombardo's mail-in vote total and the other vote totals. To the point, Daugherity's claim that his variable *a* can be predicted by his variables *b* - *d* is false.

[13] While a more flexible method, a random forest regression can perform worse that a linear model if the relationship being modeled is approximately linear and the random forest regression uses information to estimate the functional form assumption from a linear regression.

Daugherity's non-linear transformations mechanically induce high correlations and imply nothing about predictability

The Gilbert contest, and Daugherity in particular in his Appendix F, argues that an exact prediction of a can be obtained based on a set of three non-linear transformations of a, b, c, and d proposed by Solomon. In line with Solomon's argument in his Appendix A (p. 14), Daugherity describes three new variables g, h, and α that are functions of his original a, b, c, and d (Appendix F, p. 4, ¶ 15). These variables are defined as follows:

$$g = \frac{d}{a+d} \tag{1}$$

$$h = \frac{b}{b+c} \tag{2}$$

$$\alpha = \frac{b+d}{a+b+c+d} \tag{3}$$

With these new variables, Daugherity estimates the following precinct-level regression (Appendix F, p. 5, ¶ 18):

$$g = B_0 + B_1 \times \alpha + B_2 \times h + e_i$$

Based on the definitions of g, h, and α, this is equivalent to the following:

$$\frac{d}{a+d} = B_0 + B_1 \times \frac{b+d}{a+b+c+d} + B_2 \times \frac{b}{b+c} + e_i \tag{4}$$

Reinforcing the point made above, the variable a is on both the left- and right-hand sides of the regression in Equation 4, making it obvious that the regression cannot be used to predict a, or any function of a for that matter.

Daugherity's Appendix F provides the following estimates (with the level of precision presented exactly as on ¶ 18 of the appendix):

$$\hat{g} = 0.01818144438 + 1.758536682 \times \alpha - 0.8083882873 \times h \tag{5}$$

Of the values of g, h, and α that produced the regression estimates in Equation 5, Daugherity states that they exemplify "improper dependence [which] confirms that the election results in the June 14, 2022, Republican gubernatorial primary in Clark County, Nevada, were artificially contrived" (Appendix F, ¶ 17). And in a revised appendix to the Gilbert contest, Allen claims to estimate the same regression and reports obtaining an R^2 of 0.96, arguing like Daugherity that this "is extraordinarily high" (¶ 20).

The ostensible intuition on which Daugherity and Allen are drawing here is that a high value of R^2 in Equation 5 implies that the 2022 Republican gubernatorial primary must have been "fixed" prior to the election. In other words, if Nevada's 2022 Republican gubernatorial primary produced values of g that when regressed against h and α return a high R^2 value, then it must be the case, according to Daugherity, that the election suffered from vote manipulation.

Putting aside the fact that a - d are not independent in the first place, the fact that a regression with (1) a function of a on its left-hand side and (2) a different function of a on its right-hand side has a high value of R^2 implies nothing at all.

Moreover, the three non-linear transformations of a - d in Equations 1 - 3 that Daugherity used to generate g, h, and α induce a strong relationship between these three latter variables, *even in instances where there is no malfeasance at all associated with the variables a, b, c, and d.* That is, a high R^2 value in the regression in Equation 4 is a feature of the regression model and thus implies nothing about the underlying data.

We provide details on this claim in Appendix A, which presents three sets of simulations based on Clark County precincts from the 2022 Republican gubernatorial primary. In each such set, we randomly draw the variables a - d in a non-illicit way, i.e., in a way that does not presuppose that there was any vote manipulation. After drawing a - d in what can be called a clean way, we calculate simulated values of g, h, and α and then estimate Equation 4. We find high values of R^2, indicating that what Daugherity and Allen claim is evidence of vote manipulation is in fact evidence of nothing.

Daugherity appears unaware that the underlying regression model of Equation 4, whose estimates appear in Equation 5, will generate a high R^2 regardless of the presence of vote manipulation. Building on his (erroneous) logic that the high R^2 does indeed show malfeasance, Daugherity goes on to claim that, if one knew the values of α, b, c, and d,

then one could predict the value of a (Appendix F, ¶¶ 20-21). Daugherity explains this using a hypothetical example, assuming that $\alpha = 0.463855422$ in a given Clark County precinct. Then, given values of $b = 95$, $c = 45$, and $d = 133$, it follows that $a = 133$.

This does not represent, however, a "prediction" of a. Since α is a function of a, as made clear in Equation 3, it is not possible to know the value of α in a Clark County precinct (or in any precinct for that matter) without first knowing the value of a. To state the obvious and place this issue in terms of the alleged vote manipulation described in the Gilbert contest, if the transformations that underlie g, h, and α are intended to capture how some nefarious actor or actors were to have manipulated Lombardo's mail in vote total (a), the way the manipulation was carried out, as exemplified by the transformations, cannot have depended upon the number of mail-in votes Lombardo received (a).

Despite Daugherity's claim that he can make an exact prediction of Lombardo's mail- in vote total a in each Nevada precinct based on a set of four other variables (α and b - d), nothing in fact in Daugherity's report is actually a prediction. Rather, in his work Daugherity merely demonstrates a series of algebraic manipulations that move from one set of variables (a - d) to another (g, h, and α). As these manipulations can only be carried out when the variables a, b, c, and d are all known in the first place, all that Daugherity shows in his contribution to the Gilbert contest is that, if one assumes a given variable has a particular value in an algebraic equation, then one will find that variable has the value assumed. By way of analogy, Daugherity's analysis is equivalent to someone claiming an ability to predict the price of a house based on knowing (1) the house's square footage and (2) its price per square foot.

3.3 Summary of Gilbert and Fritz

The Gilbert contest was a legal challenge to a 2022 Republican primary election in Nevada. Its claims — that fair elections have equal candidate vote shares across voting methods, that mail-in and early voting vote shares of candidates are independent, and that a form of predictability indicates the presence of malfeasance in the primary, among others — range from unfounded to simply false. Taken together, the claims in the Gilbert contest fail due to manifold errors of statistics, logic, and misrepresentation of election history. They also fail because their supposedly surprising predictive results are merely the consequence of using transformations of variables to predict themselves.

The Fritz contest advanced similar claims that fail for the same reasons. For example, Daugherity claims in his expert report in Fritz that, "knowing s [Fritz's Election Day vote total], t [Staley-Ferry's Election Day vote total], and u [Fritz's vote by mail plus early voting total] would enable an almost exact prediction for v [Staley-Ferry's vote by mail plus early voting total], showing that it was not a fair election, since fair elections are not predictable, and predictable elections are not fair" (Exhibit D, p 6-7, ¶ 17). Yet, a simple examination shows this is not true. A precinct-level linear regression of Staley-Ferry's vote by mail plus early voting total on Staley-Ferry's Election Day vote total, Fritz's Election Day vote total, and Fritz's vote by mail plus early voting total yields an R^2 of only 0.66–hardly an exact prediction. Claims of exact prediction in Fritz thus fail for the same reasons they fail in Gilbert.

While the claims made in both complaints are outrageous, they still portended serious consequences for the American election infrastructure. The claims made in Gilbert's and Fritz's expert reports were the basis of challenges to election winners. They exemplify how an expansive theory of vote manipulation, one not grounded in evidence, can lead to real legal disputes.

4 Douglas Frank's many theories of election malfeasance

After the 2020 election, Douglas Frank, a Ph.D. chemist and high school math teacher, rose to prominence asserting he had clear evidence of malfeasance in the 2020 election (Seidman 2021). "Just about every county in the country was hacked," he said of this election (Murray and Simon 2022). Subsequently, Frank has toured the country, meeting with elections officials during the day and citizen groups during the night, to spread his vote manipulation theory (Gardner et al. 2021; Halpern 2022). Frank alleges an expansive vote manipulation conspiracy that in principle encompasses every election in the United States, in every state and in every county. If true, Frank's theory would call into question the position of essentially all elected officials in the country.

4.1 Douglas Frank's Theory of vote manipulation

While Frank's story sometimes varies, his core assertions about vote manipulation rest on three steps, which we have deduced from videos of his public presentations. First, Frank argues that voter rolls are inflated with "phantom voters" (Frank 2021c). Per Frank, phantom voters are registered voters who are no longer eligible to vote in a particular district or potentially were never eligible to begin with. According to Frank, these individuals could have lost eligibility because they moved, because they died, or because they never existed at all. Phantom voters are important, Frank contends, because they constitute the voters who can be inserted into election results to obtain a desired outcome.

Second, Frank alleges that there exists in U.S. elections a "key" that is determined at the state level and then deployed across a state's counties by an unspecified group of conspirators (Frank 2021d). Specifically, Frank contends that such a state-level key is based on state- level turnout rates for registered voters grouped by age, 18 to 100. According to Frank, for a given election a state's key is used to determine, prior to the election, the reported number of voters from each age group in each county who turn out to vote. In a national election, Frank alleges that each state in the country uses a different key.

Third and finally, Frank argues that, before final election results are reported, an algorithm at the county-level administered by unnamed conspirators compares the number of ballots cast in a county by voters of age a with the prescribed turnout rate for age a based on a state-level key that was determined earlier. The difference between prescribed turnout and actual turnout is then made up by inserting votes from a reservoir of "phantom voters." Specifically, Frank claims that state-level turnout rates for different age groups are multiplied by the number of registered voters in a particular age group in a county. Then, phantom votes are added to the actual number of ballots cast in the county to achieve the required age group turnout.

While Frank alleges a coordinated nationwide effort to determine statewide voter turnout rates, to the best of our knowledge he has never identified the specific perpetrators of the vote manipulation scheme that he claims to have identified. Nor has he identified the organization (or in principle organizations) that have the capacity to conduct this scheme. Frank has made vague allusions about funding from Meta Chief Executive Officer Mark

Zuckerberg (Frank 2022a), he has never specifically explained Zuckerberg's role in an alleged vote manipulation conspiracy.

Before proceeding, it is useful to note what Frank's conspiracy theory does not explain. First, in many states county elections are run independently, so it is unclear how his alleged manipulation could occur. Second, Frank provides no guidance on how supposedly fraudulent ballots are actually inserted into the election counts. Third, there is a large literature that shows that there are exceedingly few instances of illegal ballots being cast from ineligible voters who Frank would call "phantom voters" (Hood III and Gillespie 2012; Wu et al. 2020; Goel et al. 2020). For the moment we set aside these flaws and examine the evidence Doug Frank does present for his claims.

4.2 Douglas Frank's evidence of voter fraud

Frank has never presented direct evidence of the conspiracy that is summarized in the three steps above. Instead, he argues that the existence of the conspiracy can be inferred from indirect evidence obtained from official election returns and statewide voter registration list. A statewide voter registration list is a list of registered voters in a given state that contains voter characteristics like name, address, age, and the historical elections in which voters have turned out. Statewide voter registration lists are often called statewide voter files.[14]

For instance, Frank argues in favor of the existence of "phantom voters" based on well-documented issues associated with voter file maintenance. In particular, he asserts that differences between the total number of reported voters in a given election based on a statewide voter file and actual turnout in the election reported by state officials is evidence of malfeasance. It is not, however. In Appendix B, we explain that what Frank considers troubling reflects the fact that records of registered voters are regularly deleted from statewide voter files when voters move, die, or cancel their voter registrations.

Voter file maintenance issues aside, we focus here on addressing Frank's statistical claim that he has developed a method to diagnose whether an election suffers from vote manipulation. Contrary to Frank's claims, we show that his proposed test for the presence

[14] States vary in the voter characteristics that are recorded in their voter files. Moreover, some states make their voter files public while others do not. See National Conference of State Legislatures (2022).

of manipulated votes will regularly diagnose fair elections as suffering from manipulation. This is because his test effectively correlates a variable with itself, a phenomenon we also saw in our earlier discussion of the Gilbert and Fritz contests.

4.3 Frank's test for illicit vote manipulation

In a given election, Frank's test for vote manipulation is based on calculating the correlation between a predicted number of voters in an age group in a county (where predictions are based on a state-level key that Frank asserts is part of a vote manipulation conspiracy) and the observed number of voters in that age group within a county. Operating across age groups, Frank calculates one correlation per county and reports that his correlations are often close to one for practically all counties in a state, which he asserts is evidence of malfeasance. The ostensible logic here is that a high correlation between observed voter turnout and turnout as predicted by a vote manipulation conspiracy implies that the conspiracy was at work. Here we show, however, that Frank's test fails as a tool to detect vote manipulation and in fact will return a correlation between predicted and observed turnout numbers close to one even if the county being scrutinized suffered from no vote manipulation at all.

Turning now to details, Frank's test operates as follows. In a state s and for voters of age $a \in A = \{18, \ldots, 100\}$, Frank calculates the turnout rate t_{sa}, usually basing such a calculation on observed turnout n_{sa} among voters of age a in state s and dividing this number by the number of registered voters v_{sa} of age a in state s. As a simple matter of algebra, it is useful to note that $n_{sa} = t_{sa} \times v_{sa}$, i.e., that the total number of voters of age a in state s is the product of a turnout rate and a count of registered voters. To calculate age-based turnout rates within each state, Frank uses statewide voter files, which as we have noted often include voter age.

For the age groups $a \in A$ in state s, Frank regresses age-based turnout t_{sa} rates on a sixth order polynomial of underlying age and a constant[15]:

$$t_{sa} = \beta_0 + \beta_1 a + \beta_2 a^2 + \beta_3 a^3 + \beta_4 a^4 + \beta_5 a^5 + \beta_6 a^6 + \epsilon_{sa} \quad (6)$$

[15] Interestingly, Frank claims that he settled on a sixth-order polynomial because that was as "high as excel would go" (Frank 2021a).

For a state s, the estimated coefficients from the regression in Equation 6 yield fitted values $\hat{t}_{sa}$ for $a \in A$. These fitted values are predicted turnout rates, and Frank alleges that each county i in state s will have the same predicted turnout rate for registered voters of age a (we discuss this allegation shortly and for the moment take it at face value). Based on this, Frank calculates a predicted number of voters from each age a in each county i that is part of state s. To do this, for age group a Frank multiplies the statewide predicted turnout rate $\hat{t}_{sa}$ with the number v_{sia} of voters in age group a, county i, and state s. That is, the predicted number of voters from age group a in county i in state s is $\hat{n}_{sia} = v_{sia} \times \hat{t}_{sa}$.[16]

Finally, Frank calculates the correlation *within* each county i between the predicted number $\hat{n}_{sia}$ of voters in each age group a and the observed number of voters n_{sia} in that group:

$$\mathrm{Cor}(\hat{n}_{sia}, n_{sia}) = \mathrm{Cor}(\hat{t}_{sa} \times v_{sia}, t_{sia} \times v_{sia}) \tag{7}$$

where t_{sia} is the observed turnout rate for voters in county i of state s who are a years old. Frank claims that, if this correlation is large, then the state-level "key" determined actual election turnout and, accordingly, that vote manipulation occurred (Frank 2021b). The underlying premise behind this ostensible logic is that county-level election results should not be so regular as to have high correlations between what Frank calls predicted turnout and observed turnout. Frank claims — by way of analogies and simple assertions rather than via any theoretically driven rationale — that high predictability as highlighted by high correlations in Equation 7 is "unnatural" and suspicious. For example, Frank regularly claims that correlations above 0.7 are rare when analyzing human behavior and therefore a correlation close to one is unnatural (for example, Frank 2021b). This is, of course, untrue — there is no general cutoff for a surprising correlation within any field of study.[17] Rather than reasoning by analogy and simple assertion, a rigorous

[16] Note that all of our predictions are in-sample, which means the same data is used to estimate $\hat{t}_{sa}$and then form the predictions. We make out of sample predictions below when evaluating the predictive performance of key's across state-lines.

[17] Frank justifies this claim in at least one video presentation (Frank 2021e), which references Schober et al. (2018). But even this paper argues, after providing cutoff points that "These cutoff points are arbitrary and inconsistent and should be used judiciously" (Schober et al. 2018). As an example for why we can't simply observe a correlation and assert manipulation, we would expect strong correlations if we are measuring the relationship between the price of a stock at the open of the market and the price of a stock one minute after open. Finding a high correlation in these prices would not be indicative of fraud or other market manipulation.

evaluation of Frank's test would examine whether correlations close to one in Equation 7 occur only in elections where votes have been manipulated in some fashion. We now provide such an evaluation and show that a correlation close to one should be expected in almost all instances based on the calculation in Equation 7, whether there is vote manipulation or not.

4.4 Intuition about Frank's test and why it fails

A fundamental feature of Frank's test for vote manipulation is that it calls for calculating the correlation in Equation 7 between the predicted and actual *number* of voters in an age group at the county level. This will mechanically result in correlations that are close to one. This occurs because Frank's test essentially "discovers" the regularities that (1) age groups with more registered voters tend to have more individuals turning out to vote while (2) age groups with fewer registered voters conversely tend to have fewer people turning out to vote. Because in any county there is considerable variation in the *number* of voters in any age group, and relatively less variation in the *rate* at which individuals in an age group turn out to vote (these rates being constrained between zero and one, by definition), observed numbers of voters (n_{sia}) and "predicted" numbers of voters ($\hat{n}_{sia}$) move together relatively closely, which results in a correlation close to one in Equation 7. In fact, one way to view Frank's test encapsulated in this equation is that it calls for effectively correlating a variable with itself: the number of registered voters v_{sia} of age a in county i of state s dominates both the left and the right input of the correlation.

4.5 Why Frank's test for vote manipulation is meaningless

To see this self-correlation, let us revisit Equation 7, paying particular attention to the fact that Frank's test involves a correlation of $\hat{n}_{sia}$ and n_{sia}. Accordingly, and fixing state s and county i, notice that the *number* v_{sia} of registered voters in age group a appears in both the left and the right input of the desired correlation. The result of this is that variation in the *number* of registered voters will cause both predicted (left side of the correlation) and observed numbers of voters (right hand side) to move together. Ultimately, this results in correlations close to one.

A simple simulation suffices to demonstrate the mechanical correlation inflation that occurs when correlating numbers of predicted voters and numbers of observed voters in age group a rather than predicted turnout rates and observed turnout rates for these individuals. The simulations we construct are free of vote manipulation. Therefore, to the extent that they have high correlations based on Equation 7 — particularly as we increase the variance in the number of individuals in each age group — it follows that Frank's test fails as a diagnostic of vote manipulation.

To construct our simulations we suppose, like Frank that there are 83 distinct ages in a hypothetical county within a single state, individuals ranging from 18 to 100 years old. To show that variation across age group size drives the correlation at the root of Frank's test, we further suppose that 82 of these age groups have 100 individuals — so, 100 18 year olds, 100 19 year olds, and so forth. In the simulations, however, we vary the number of people in the final age group, which allows us to observe how the correlation calculated in Equation 7 changes as the variation in the number of people in each age group changes. Specifically, across simulations we suppose that the number of registered voters in the target age group varies from 100 — the same size as other age groups — to 1000, i.e., ten times the number of individuals in the largest age group relative to other age groups.

Conditional on numbers of voters, we simulate turnout in an election in a way that is independent of age. In each simulated election, we suppose that the turnout rate of each age group a is 40 percent plus random error that is independent and identically distributed Gaussian noise. Formally, our simulations posit that voter the turnout rate for age group a is $0.4 + y_a$ where $y_a \sim \text{Normal}(0,0.025)$. To assess Frank's test, within each simulation we fit a sixth order polynomial with Equation 6 to obtain a "key" and then, within a simulated county, use this to calculate a predicted number of voters in each age group.

Table 3 presents the results of our simulations. The left-most column in the table ("Largest age group size") lists the size of the largest age group in each of the simulated elections. The middle column ("Number of voters") reports average correlations between Frank-predicted turnout *numbers* and observed turnout *numbers*. And, the right-most column of the table ("Turnout rates") reports average correlations between predicted turnout *rates* and observed turnout *rates*.

Table 3: Simulated correlations used in Frank's test for vote manipulation

	Average correlation	
Largest age group size	Number of voters	Turnout Rates

100	0.26	0.26
200	0.68	0.26
300	0.87	0.26
400	0.94	0.26
500	0.96	0.26
600	0.98	0.26
700	0.98	0.26
800	0.99	0.26
900	0.99	0.26
1000	0.99	0.26

The key implication of Table 3 is straightforward: as the size of the largest voter age group increases relative to other age groups, the correlation between numbers of predicted voters based on Frank's "key" model and numbers of actual voters approaches one. This is because, when there is more variation in the number of individuals in a set of age groups, predicted and observed numbers of voters move closely together, which is merely the result of the correlation in Equation 7 uncovering the unsurprising regularity that larger age groups tend to have more voters. In contrast, and as made clear in the rightmost column in Table 3, if one focuses on turnout *rates*, which control for age group size, the correlation between predicted and observed turnout holds steady at 0.26. In short, Table 3 demonstrates that the supposedly unnatural correlations from Frank's test for vote manipulation can occur merely because of variation in the number of individuals across different age groups.

We also see a mechanical increase in correlation when we analyze actual election data. Because Frank has not responded by July 14, 2023 to our request for data and replication code for his nationwide analysis (see Appendix C), we independently evaluated his claims about the utility of his proposed test for vote manipulation. To do this, we used voter file data from L2 and calculated turnout rates for registered voters for each age from 18-100 for all states where we observed a birth date for more than 90% of registrants.[18] Our L2 voter file data derive from snapshots of states' voter files; since states do not synchronize updates to their respective voter files, pull dates vary across states between December 2020 and May 2021, as described in Appendix E. With turnout rates calculated at the state level

[18] Excluded states were Alaska, the District of Columbia, Hawaii, Maryland, Mississippi, New Hampshire, North Dakota, Wisconsin, and Wyoming.

for each age, we followed Frank's procedure and Equation 6 and regressed these rates on a six-degree polynomial to create predicted turnout rate for each age. State by state, we then projected predicted turnout rates to counties. Continuing to follow Frank's procedures, we calculated the correlation within each county between observed numbers of voters and predicted numbers of voters by age. For a given age, predicted numbers of voters were obtained by multiplying the number of registrants of that age by the predicted turnout rate for that age where prediction is based on Frank's key procedure.

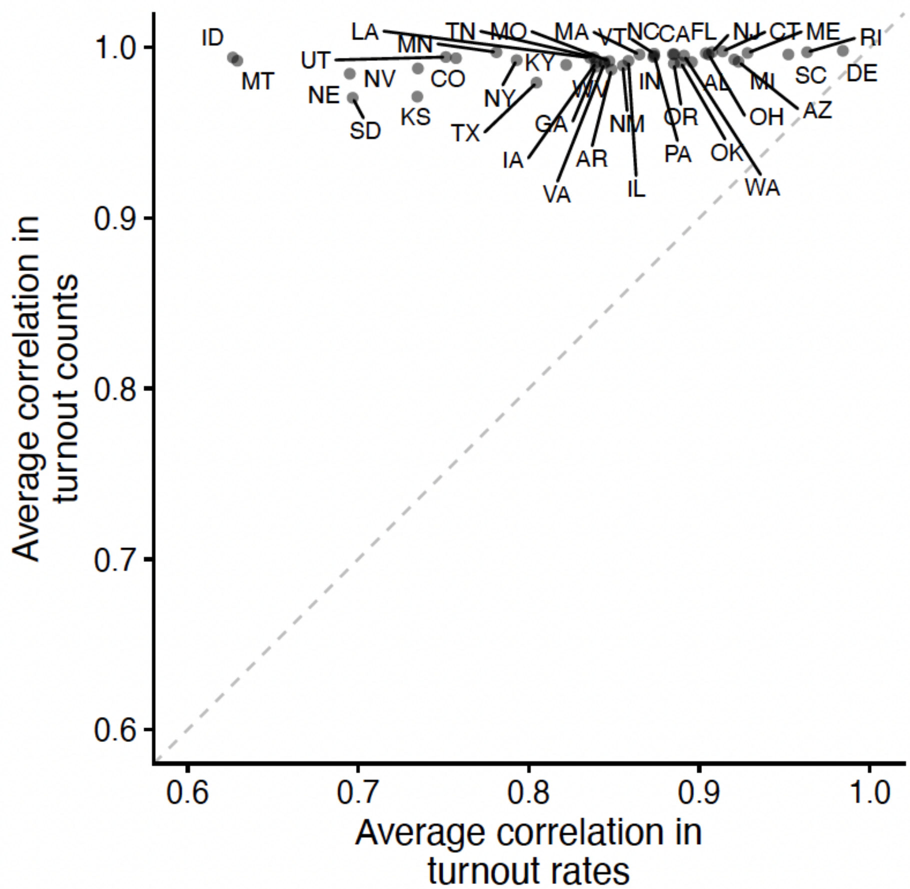
In Figure 5, we present the results of applying Frank's analysis to the 42 states where we have adequate age data. In this figure, the vertical axis describes statewide average correlation coefficients between predicted and observed turnout numbers (i.e., average values of $\mathrm{Cor}\,(\hat{n}_{sia}, n_{sia})$). To generate these numbers, in each county contained in our set of 42 states we calculate a correlation across age groups between predicted and actual turnout numbers. Then, for each state we average these county-level correlations to create a single statewide correlation. Continuing, Figure 5's horizontal axis describes statewide average correlation coefficients between predicted and observed turnout rates (i.e., average values of $\mathrm{Cor}\,(\hat{t}_{sa}, t_{sia})$ where $\hat{t}_{sa}$ is the predicted turnout rate for voters of age a in county i of state s, where $\hat{t}_{sa}$ is not indexed by i because Frank makes a single turnout rate prediction for age group a in the entire state). Parallel to the procedure described above, in each county in each state we calculate the correlation between the predicted and actual turnout rates. Then, in each state we take the average of the county-level correlations to obtain a single statewide average correlation between predicted and actual turnout rates. Figure 5 includes a dashed 45-degree line, and all 42 points in the figure lie above it, indicating that, in all 42 states we study, the average correlation coefficients between observed and predicted turnout numbers is greater than the average correlation coefficients between observed and predicted turnout rates.

Figure 5 shows that in every state we find a higher average correlation between predicted and actual turnout numbers than we do between the predicted and actual turnout rates. In fact, turnout number correlations average 0.991 across the U.S. counties in our data set, with an interquartile range from 0.9905 to 0.9978. In contrast, rate correlations average 0.825 across counties in our data set, with an interquartile range of 0.758 to 0.922. Taken together, this shows correlating turnout numbers artificially inflates the reported correlations.

4.6 Frank's key is not unique to a state

As a supposed demonstration of his method's validity, Frank often asserts that the state-level turnout rate predictions he uses only perform well within a particular state. To the best of our knowledge, he fails to provide evidence for this claim. To assess this claim, we used Frank's methodology, but for each county we calculated the predicted counts using the predicted turnout rates from every state, not just the county's home state. We then calculated the correlations between these out of state predictions and the actual turnout rates.

Figure 5: High correlations between actual and predicted ballots in counties are found in each state.



The results of this analysis are presented in the left-hand half of Figure 6. The top left-hand facet presents the distribution of correlations between the predicted and actual turnout counts using out of state turnout rates to make the prediction and the bottom left-hand facet shows the distribution of correlations using the turnout rates from the county's same state to make the predictions. Both correlations are close to 1. In fact, the average correlation using the turnout rate from the county's same state is 0.99, while the average correlation using out of state predicted turnout rates is 0.984. In 16.9% of instances the predictions based on out of state turnout rates yields a *higher* correlation than the predictions based on the within state turnout rates. The right-hand plot correlates the county-level turnout rates with the state-level predictions about turnout rates. The top right-hand facet performs this correlation across state lines and the bottom right-hand facet performs this correlation in the same state. The correlations between predicted and actual turnout rates is much lower than the correlation between predicted and actual turnout counts, both within the same state and across state lines. This is because focusing on counts, rather than rates, artificially inflates Frank's correlations.

4.7 Frank's claims about constant voter turnout rates across counties are false

In our discussion of Frank's theory of vote manipulation, we noted that the theory implies that, within states, voter turnout rates by age would be effectively constant across counties. This is most decidedly not the case, which undermines Frank's argument that turnout rates are fixed by a "key."

Figure 7 displays by state voter turnout rates (vertical axis) by age (horizontal axis) for the 2020 general election. Each point in this figure is a turnout rate for a particular age group in a county, and what Frank would call the "key" for each state is approximated by the red line.[19]

Consider Texas, which had the second largest registered voter pool in the 2020 general election. Turnout rates of the youngest Texans are shown on the left side of Figure 7's Texas panel, and turnout rates of the oldest Texans are shown on right side of the panel. The red curve in the Texas panel is roughly concave down, indicating that, in the 2020 general election, younger and older Texans had lower turnout rates than middle age Texans.

Within each state and conditional on age, Figure 7 highlights considerable county-to-county variation in turnout rates across the 42 states included in our analysis. Continuing to examine Texas, there are 254 counties in this state. Focusing on the turnout rates of the youngest voters in the Texas (age of 18), county points vary and appear well below and above the red curve. This is true as well of 19 year old voters and in fact of voters of all ages. That is, voters of age $a \in [18, 100]$ did *not* have identical turnout rates across Texas counties. This is inconsistent with Frank's vote manipulation theory.

Figure 6: Average correlations between county-level predicted and actual turnout counts are very similar (left-hand plots), regardless of whether the turnout rates from the same state or different states is used to make predictions. This is the case even though the correlation between predicted

[19] The red lines in the figure are loess smoothers, with bandwidth determined by cross-validation.

and actual turnout rates are much lower, particularly when correlating predicted turnout rates across state lines.

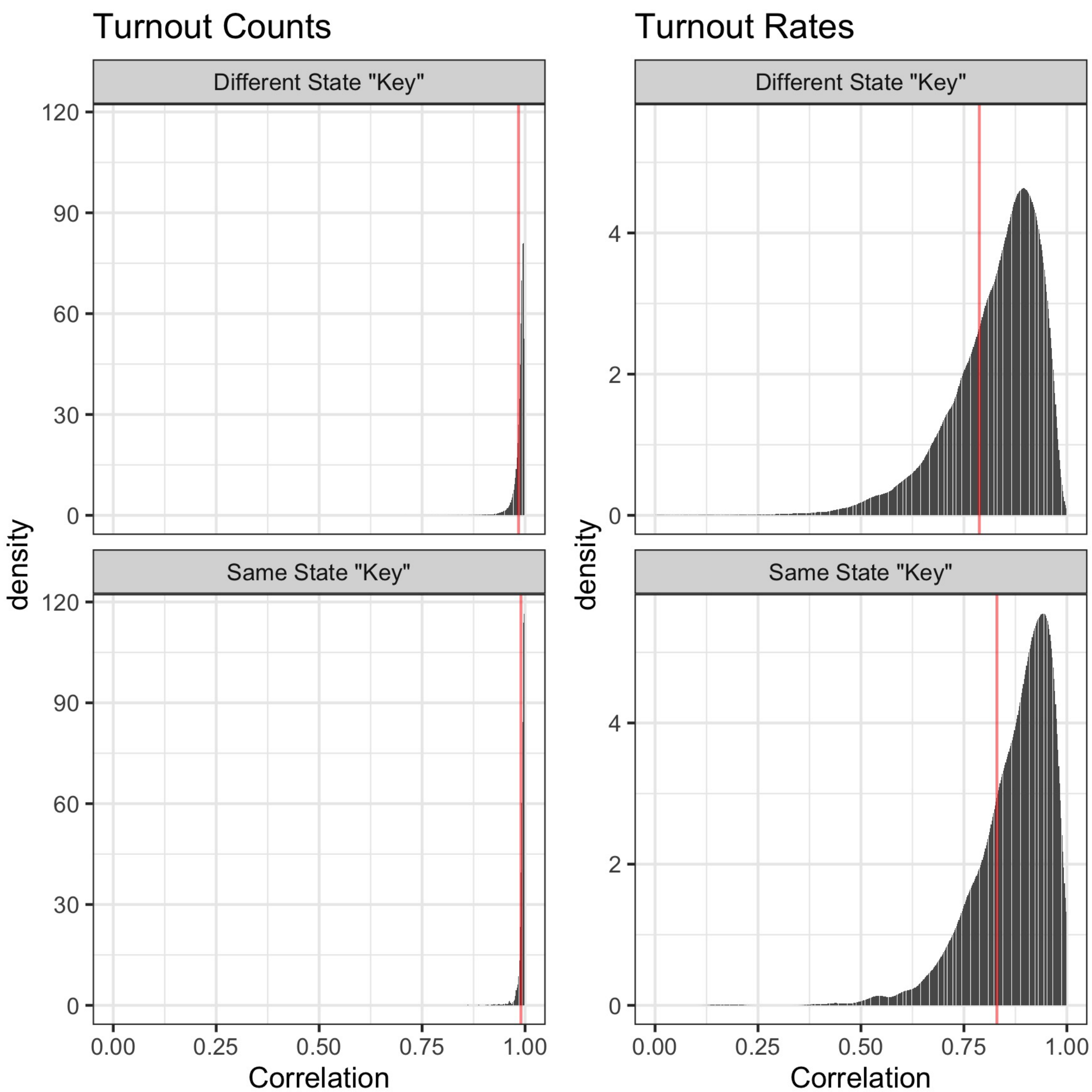
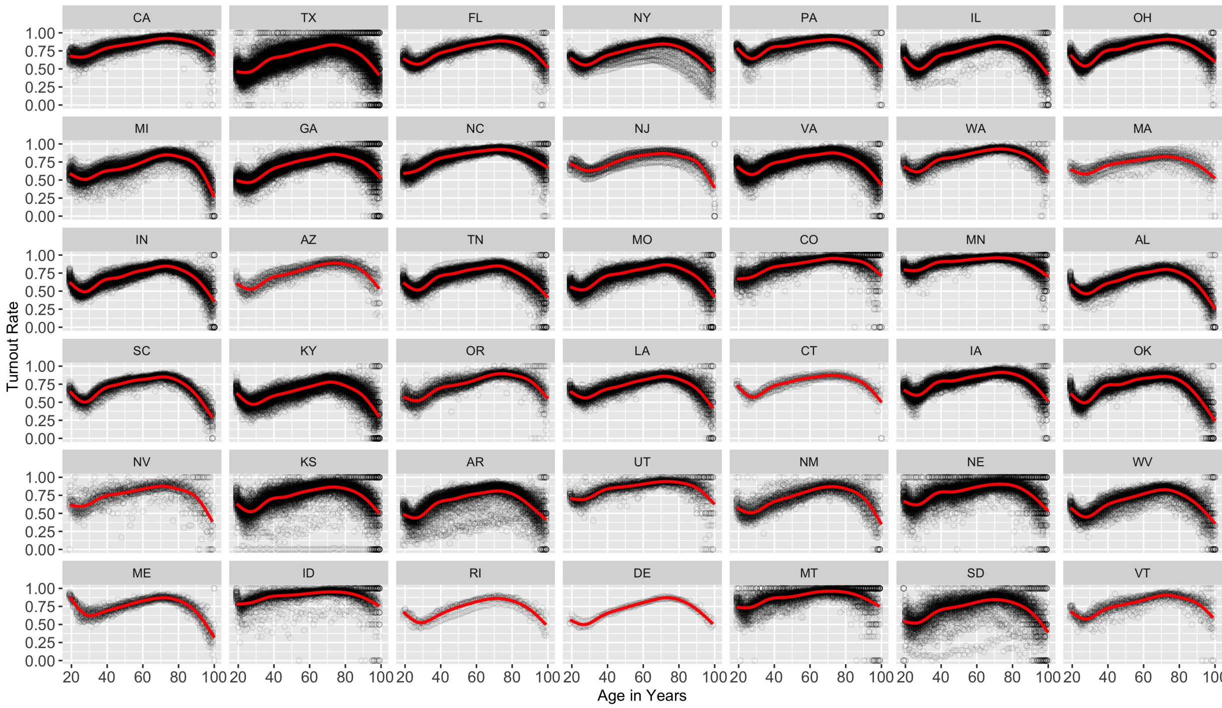


Figure 7: Voter turnout rates and age across states in the 2020 general election



Note: states are ordered by number of registered voters as of the 2020 general election.

4.8 Summary of Frank's claims about American elections

In a variety of videos and speeches made since 2020, Douglas Frank has articulated a theory of vote manipulation that covers essentially the entire United States. If Frank's theory were correct, practically every contemporary election in the country has been marred by malfeasance.

In his work, Frank proposes a test for vote manipulation. While Frank asserts that the test is truly diagnostic, in reality it shows nothing. Rather, Frank's test for voter manipulation fails because it always classifies elections as suffering from vote manipulation due to its construction. As we have explained, this is because Frank's test confuses variation in the number of individuals across age groups in counties and states with evidence of an unnatural manipulation of ballots.

Not only is Frank's test for vote manipulation worthless, but a key principle on which it is based — that various precinct-level turnout rates are constant within and across countries

— is wrong. In short, Frank's findings have no implications at all for the presence of vote manipulation in the United States.

5 Refuting conspiracy theories about election malfeasance

We have considered two prominent exemplars of new, expansive theories of vote manipulation. While the claims made in Gilbert/Fritz and by Douglas Frank involve distinct forms of "evidence," they share features that are common across many claims of election fraud. Both the Gilbert/Fritz and Frank theories of vote manipulation suppose that machines are manipulating U.S. elections, and both suppose that mathematical analysis of aggregate election results can reveal patterns indicative of manipulation.

The evidence provided in both exemplars is baseless. While both the theories of vote manipulation in Gilbert/Fritz and those adduced by Frank assert that the ability to predict elections is indicative of fraud, we show that neither Gilbert/Fritz nor Frank explains why some aspects of elections should not be predictable. And, more critically, we show that the proponents of the theories in Gilbert/Fritz and Frank fail to provide any evidence of prediction. Instead, they have developed distinct ways to demonstrate the unsurprising fact that variables are strongly correlated with themselves. Gilbert's contest about vote manipulation misrepresented prior election results, asserted conclusions that fail to follow from their premises, and the arguments in both Gilbert and Fritz merely show that it is possible to "predict" a variable when that variable is known. Douglas Frank's claims about vote manipulation similarly vacuous. His statistical tests merely show the unsurprising fact that groups with more registered voters have more people who turn out to vote. The refutations of the theories of vote manipulation we analyze require statistical knowledge. But it is still possible for individuals without statistical training—such as lawyers, judges, and politicians—to spot dubious election claims.

The simplest and most critical question to ask of an advocate of a claim about voter fraud might be, how accurate is your statistical method at identifying fraud? Do you know the proportion of elections that actually were manipulated when your method identifies an election result as "surprising" or "impossible?" A related question to ask in the face of an allegation that a given pattern in election results is evidence fraud is, what evidence is there that the pattern in data would not be present in the absence of fraud?

A persistent challenge in refuting claims about vote manipulation is that those making such claims remain remarkably uninterested in recognizing contradictory evidence. For example, in a hearing to dismiss the Gilbert contest, Gilbert's lawyer asserted that, "Nobody from his side. . . has said that the formulas are wrong or that they generate errors in the mathematical computation of the purported mail in ballots" despite reading lengthy reports from two of the authors of this paper where we demonstrated considerable errors (Rumble 2022, 18:30). Both Solomon and Daugherity filed expert briefs in Fritz's contest that repeated errors present in the Gilbert contest. While Frank now acknowledges that the critiques presented in this paper, he fails to engage with the evidence we provide and instead implies that one of us argued that "math is racist" (Frank 2023).

The path to defeat baseless election claims is (1) demonstrating that they are false and (2) publicly explaining why. This is the motivation for this paper, whose results should be valuable to all those who care about the infrastructure of American elections. These people include election officials, who have to engage baseless concerns from concerned citizens, judges who have to weigh the evidence of malfeasance in court cases, and citizens who remain concerned about American elections but are still open to the value of scientific evidence.

6 References

Anderson, Judge John C. (2023). Anderson opinion. Case No. 22 MR 421.

Ansolabehere, Stephen and Nathaniel Persily (2008). Vote fraud in the eye of the beholder: The role of public opinion in the challenge to voter identification requirements. *Harvard Law Review 121*(7), 1737–1774.

Arthur, Damon (2023). With more costs to come, shasta county will spend 950, 000 on new voting system. (last accessed on July 6th, 2023).

Atkeson, Lonna Rae, Wendy L Hansen, Maggie Toulouse Oliver, Cherie D Maestas, and Eric C Wiemer (2022). Should I vote-by-mail or in person? The impact of covid-19 risk factors and partisanship on vote mode decisions in the 2020 presidential election. *Plos one 17*(9), e0274357.

Berlinski, Nicolas , Margaret Doyle, Andrew M. Guess, Gabrielle Levy, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, and Jason Reifler (2021). The effects of unsubstantiated claims of voter fraud on confidence in elections. *Journal of Experimental Political Science*, 1–16.

Brown, Jacob R. and Ryan D. Enos (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour 5*(8), 2397–3374.

Cassidy, Christina A. (2021). Far too little vote fraud to tip election to Trump, AP finds. *AP News*. https://apnews.com/article/voter-fraud-election-2020-joe-biden-donald-trump-7fcb6f134e528fee8237c7601db3328f (last accessed October 12, 2022).

Cole, Brendan (2022, June 15). Jan. 6 Attendant Loses Nevada Primary by 11 Points, Blames Election Fraud. *Newsweek*. https://www.newsweek.com/joey-gilbert-nevada-governor-fraud-trump-republican-1716051 (last accessed October 12, 2022).

Corasaniti, Nick and Alexandra Berzon (2022, June 26). The strange tale of tina peters. *The New York Times*. https://www.nytimes.com/2022/06/26/us/politics/tina-peters-election-conspiracy-theories.html (last accessed October 21, 2022).

Cottrell, David, Michael C. Herron, and Sean J. Westwood (2018). An exploration of Donald Trump's allegations of massive voter fraud in the 2016 General Election. *Electoral Studies 51*, 123–142.

Danforth, John , Benjamin Ginsberg, Thomas B. Griffith, David Hoppe, J. Michael Luttig, Michael W. McConnell, Theodore B. Olson, and Gordon H. Smith (2022). LOST, NOT STOLEN: The Conservative Case that Trump Lost and Biden Won the 2020 Presidential Election. https://lostnotstolen.org//wp-content/uploads/2022/07/Lost-Not-Stolen-The- Conservative-Case-that-Trump-Lost-and-Biden-Won-the-2020-Presidential-Election-July- 2022.pdf (last accessed October 12, 2022).

Eggers, Andrew C. , Haritz Garro, and Justin Grimmer (2021). No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election. *Proceedings of the National Academy of Sciences 118*(45), e2103619118.

Ervin, Malcolm (2023). Letter to Chuck Gray from county clerk's association of Wyoming. https://wyofile.com/wp-content/uploads/2023/04/2023-Letter-regarding-Dr.-Frank.pdf.

Frank, Douglas (2021a). Dr Frank and Dr Draza at Skagit County WA give behind the scenes view of cyber symposium and detail strategy to expose election fraud in all 50 states. https://patriotbites.com/dr-frank-and-dr-draza-at-skagit-county-wa/, Last Accessed 7/14/2023.

Frank, Douglas (2021b). Dr. frank's presentation at washington election integrity public hearing. https://rumble.com/vmjuoo-dr-franks-presentation-at-washington-election-integrity-public-hearing.html.

Frank, Douglas (2021c). What is a phantom voter? https://rumble.com/vk36ty-what-is-a-phantom-voter.html.

Frank, Douglas (2021d). What is a phantom voter? https://rumble.com/vqj701-dr-frank-speaking-at-election-committee-12-8-21.html.

Frank, Douglas (2022a). Dr. Douglas Frank presents election data. https://rumble.com/v1ckbar-july-17-2022.html.

Frank, Douglas (2022b, September 13,). Dr Frank's presentation to Shasta board of supervisors. https://rumble.com/v1k4zlp-election-fraud-in-northern-california.html (last accessed July 6th, 2023).

Frank, Douglas (2022c). Election fraud in Idaho (Dr Frank presentation). https://rumble.com/v1639mg-election-fraud-in-idaho-dr.-frank-presentation.html.

Frank, Douglas G. (2021e). Primer 1: Correlation coefficient. https://rumble.com/vk3720-primer-1-correlation-coefficient.html, Last Accessed on 7/14/2023.

Frank, Douglas G (2022d, August 22). Dr. Douglas G. Frank Speaks at The Moment of Truth Summit. https://frankspeech.com/video/dr-douglas-g-frank-speaks-moment-truth- summit. (last accessed October 12, 2022).

Frank, Douglas G. (2022e). Placer county update. https://rumble.com/v1i8wuv-placer-county-update.html.

Frank, Douglas G. (2023). January 24th, 2022 truth social post. https://truthsocial.com/\spacefactor\@m{}DrDouglasGFrank/109745732255192943 Last Accessed on 7/14/2023.

Gardner, Amy , Emma Brown, and Josh Dawsey (2021). Inside the nonstop pressure campaign by Trump allies to get election officials to revisit the 2020 vote. The Washington Post. https://www.washingtonpost.com/politics/trump-election-officials-pressure-campaign/2021/12/22/8a0b0788-5d26-11ec-ae5b-5002292337c7_story.html (last accessed October 18, 2022).

Gardner, Amy and Patrick Marley (2022, September 11). Trump backers flood election offices with requests as 2022 vote nears. *The Washington Post.* https://www.washingtonpost.com/ nation/2022/09/11/trump-election-deniers-voting (last accessed October 12, 2022).

Gentry, Dana (2022). Gubernatorial hopeful Joey Gilbert: No regrets about Jan. 6. *Nevada Curent.* https://www.nevadacurrent.com/2022/01/06/gubernatorial-hopeful-joey-gilbert- no-regrets-about-jan-6 (last accessed October 12, 2022).

Ghitza, Yair , Andrew Gelman, and Jonathan Auerbach (2022). The Great Society, Reagan's Revolution, and Generations of Presidential Voting. Forthcoming, *American Journal of Political Science*. https://doi.org/10.1111/ajps.12713 (last accessed October 24, 2022).

Gimpel, James G. and Iris S. Hui (2015). Seeking politically compatible neighbors? the role of neighborhood partisan composition in residential sorting. *Political Geography 48*, 130–142.

Goel, Sharad , Marc Meredith, Michael Morse, and David Rothschild (2020, May). One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections. *American Political Science Review 114*(2), 456–469.

Goel, Sharad , Marc Meredith, Michael Morse, David Rothschild, and Houshmand Shirani- Mehr (2020). One person, one vote: Estimating the prevalence of double voting in us presidential elections. *American Political Science Review 114*(2), 456–469.

Gordon, Aaron (2021). 'Expert Mathematician' on Election Fraud Actually a Swing Set Installer, Lawsuit Claims. *Vice*. https://www.vice.com/en/article/88nbk4/expert-mathematician-on-election-fraud-actually-a-swing-set-installer-lawsuit-claims (last accessed October 12, 2022).

Halpern, Sue (2022). The Election Official Who Tried to Prove 'Stop the Steal'. *The New Yorker* . https://www.newyorker.com/news/american-chronicles/the-election-official-who- tried-to-prove-stop-the-steal (last accessed October 12, 2022).

Hansen, Claire (2022, October 14). As Jan. 6 committee discloses new details, Trump responds with recycled lines. *U.S. News & World Report.* https://www.usnews.com/news/politics/articles/2022-10-14/as-jan-6-committee-discloses- new-details-trump-responds-with-recycled-lines (last accessed October 24, 2022).

Hasen, Richard L. (2022). Record Election Litigation Rates in the 2020 Election: An Aberration or a Sign of Things to Come? *Election Law Journal 21*(2), 150–154.

Hastie, Trevor , Robert Tibshirani, and Jerome H. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2 ed.). New York: Springer.

Herron, Michael C. (2019). Mail-in absentee ballot anomalies in North Carolina's 9th congressional district. *Election Law Journal 18*(3), 191–213.

Herron, Michael C. and Daniel A. Smith (2012). Souls to the polls: Early voting in Florida in the shadow of house bill 1355. *Election Law Journal 11*(3), 331–347.

Hood III, MV and William Gillespie (2012). They just do not vote like they used to: A methodology to empirically assess election fraud. *Social Science Quarterly 93*(1), 76–94.

Kang, Cecilia (2022).The Most Dominant Toxic Election Narratives Online. *The New York Times*. https://www.nytimes.com/2022/09/23/technology/midterm-elections-misinformation.html (last accessed October 12, 2022).

Kukulka, Alexandra (2023a). Failed candidate alleges election fraud in will county clerk's race based on mathematical algorithms. https://www.chicagotribune.com/suburbs/daily-southtown/ct-sta-will-county-clerk-election-fraud-lawsuit-st-0201-20230131-m4zksjuiozcxfb4dnnf4gbd4ya-story.html.

Kukulka, Alexandra (2023b). Will county judge orders republican clerk candidate, attorney to pay 35,000 in sanctions for election fraud lawsuit., Last accessed on 7/14/2023.

Levitt, Justin (2007). The Truth About Voter Fraud. Technical report, Brennan Center for Justice at New York University School of Law. https://www.brennancenter.org/our- work/research-reports/truth-about-voter-fraud (last accessed October 12, 2022).

Li, Yimeng , Michelle Hyun, and R. Michael Alvarez (2022). Why do election results change after election day? the "blue shift" in California elections. *Political Research Quarterly 75* (3), 860–874.

Marks, Joseph (2021, August 11). The Cybersecurity 202: My Pillow cyber symposium is yet another font of election fraud lies. *The Washington Post*. https://www.washingtonpost.com/politics/2021/08/11/cybersecurity-202-my-pillow-cyber- symposium-is-yet-another-font-election-fraud-lies/ (last accessed October 12, 2022).

Mazo, Eugene D. (2018). Finding Common Ground on Voter ID Laws. *Univeristy of Memphis Law Review 49*, 1233–1273.

Minnite, Lorraine (2007). Election Day Registration: A Study of Voter Fraud Allegations and Findings on Voter Roll Security. Technical report, Dēmos. http://www.demos.org/sites/default/files/publications/edrfraud.pdf (last accessed October 12, 2022).

Minnite, Lorraine C. (2010). *The Myth of Voter Fraud*. Ithaca, NY: Cornell University.

Mollenkamp, Allison, Miles Parks, and Nick McMillan (2022). Election deniers are spreading misinformation nationwide. Here are 4 things to know. *National Public Radio*. https://www.npr.org/2022/07/05/1109538056/election-deniers-are-spreading-misinformation-nationwide-here-are-4-things-to-kn (last accessed October 12, 2022).

Murray, Sara and Jeff Simon (2022, January 29). The 2020 election wasn't stolen. but douglas frank and his bogus equation claiming otherwise are still winning over audiences. *CNN.com*. https://www.cnn.com/2022/01/18/politics/douglas-frank-big-lie/index.html (last accessed October 19, 2022.

National Conference of State Legislatures (2022). Access to and use of voter registration lists. https://www.ncsl.org/research/elections-and-campaigns/access-to-and-use-of-voter- registration-lists.aspx (last accessed October 19, 2022).

Placer County Elections Office (2020). Placer county 2020 general election results. https://www.placercountyelections.gov/uploads/documents/11032020/11032020 _ Final Results.pdf.

Robison, Mark (2022, March 14). Robert Beadles is wealthy, new to town and wants to upend Reno's politics. Who is he? *Reno Gazette-Journal*. https://www.rgj.com/story/news/2022/03/14/robert-beadles-california-washoe-county-politics/9419737002/ last accessed October 12, 2022).

Robison, Mark and Rio Lacanlale (2022, July 7). Lombardo retains win over Gilbert as Washoe, Clark certify recount of GOP governor primary. *Reno Gazette-Journal*. https://www.rgj.com/story/news/2022/06/16/joey-gilbert-pro-trump-ally-refuses-concede- voter-fraud-claims-nevada-primary-election-results/7653105001/ (last accessed October 12, 2022).

Rumble (2022, August 10). Gilbert vs Lombardo, Hearing for Summary Judgement. https://rumble.com/v1fjglj-unbelievable-the-contest-was-denied.html. (last accessed October 12, 2022).

Scherer, Zachary (2021, April 29). What Methods Did People Use to Vote in the 2020 Election? Technical report, United States Census Bureau. https://www.census.gov/library/stories/2021/04/what-methods-did-people-use-to-vote-in- 2020-election.html (last accessed October 12, 2022).

Schober, Patrick , Christa Boer, and Lothar A Schwarte (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia 126*(5), 1763–1768.

Seeman, Matthew (2022, September 22). Judge orders sanctions against Joey Gilbert in failed lawsuit over GOP primary. *KRNV-DT* . https://mynews4.com/newsletter-daily/judge-orders-sanctions-against-joey-gilbert-in- failed-lawsuit-over-gop-primary-nevada-governor-election-gubernatorial-las-vegas-reno- carson-city-district-court-joe-lombardo-politics (last accessed October 12, 2022).

Seidman, Andrew (2021, July 26). Kathy Barnette's futile hunt for voter fraud outside Philadelphia. *Philadelphia Inquirer* . https://www.inquirer.com/news/a/pennsylvania-voter-fraud-audit-kathy-barnette-20210726.html (last accessed October 12, 2022.

Sides, John , Michael Tesler, and Lynn Vavreck (2019). *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America* (2nd ed.). Princeton, NJ: Princeton University Press.

Thayer, Matthew (2022). (S)Election Code. Producer: Mike Lindell. https://frankspeech.com/video/selection-code-full-movie (last accessed October 12, 2022).

Wu, Jennifer, Chenoa Yorgason, Hanna Folsz, Sandy Handan-Nader, Andrew Myers, Toby Nowacki, Daniel M. Thompson, Jesse Yoder, and Andrew B. Hall (2020). Are Dead

People Voting By Mail? Evidence From Washington State Administrative Records. Unpublished working paper. https://stanforddpl.org/papers/wu et al 2020 dead voting (last accessed October 12, 2022).

Yoder, Jesse , Cassandra Handan-Nader, Andrew Myers, Tobias Nowacki, Daniel M. Thompson, Jennifer A. Wu, Chenoa Yorgason, and Andrew B. Hall (2021). How did absentee voting affect the 2020 U.S. election? *Science Advances 7*(52), eabk1755.

A Solomon's transformations mechanically induce high R^2 values

The "extraordinarily high" relationship between the variables g, h, and α — which the Gilbert contest interprets as evidence of vote manipulation — is induced mechanically as a result of Solomon's transformations of the variables a - d, summarized in Equations 1 - 3. As evidence of this, in this appendix we describe a set of simulation showing that, even when underlying vote counts a - d are generated without any illicit manipulation, the regression in Equation 4 returns a high value of R^2. This should not be considered surprising insofar as the same variables (a - d) appear in both the regression's left-hand side and its right-hand side.

We offer three versions of our simulation, which have qualitatively identical results. Each simulation is based on a set of 814 precincts, which mimics Clark County as of November 2022.

In our first simulation ("Poisson"), for each precinct we draw mail ballot turnout from a (rounded) uniform distribution on the interval [100, 300] and we draw early voting turnout similarly and independently. For each precinct, we then draw a from a Poisson distribution with rate parameter equal to one-half of mail turnout and c from a Poisson distribution with rate parameter equal to one-half of early voting turnout. We then set b equal to the difference between mail ballot turnout and a and likewise set d equal to the difference between early voting turnout and c. Finally, we transform our our variables a - d using Equations 1 - 3 and then estimate Equation 4, recording the value of R^2.

In our second simulation ("Uniform"), for each of our 814 precincts we draw mail ballot turnout and early voting turnout as described in the first simulation immediately above. We then draw a from a (rounded) uniform distribution bounded to the left by one and to the right by mail ballot turnout. And, we similarly draw c from a (rounded) uniform distribution bounded to the left by one and to the right by early voting turnout. As in the first simulation, we then set b equal to the difference between mail ballot turnout and a and likewise set d equal to the difference between early voting turnout and c. Finally, we transform our randomly drawn a - d using Equations 1 - 3 and then estimate Equation 4, recording the value of R^2.

In our third simulation ("beta"), for each of our 814 precincts we draw mail ballot turnout and early voting turnout as described in the first simulation. We then specify a for each precinct by multiplying mail ballot turnout with a randomly drawn value from a beta distribution whose shape parameters are the maximum likelihood estimates produced by fitting a beta density to Lombardo's mail-in ballot support rate among precincts in Clark County in the 2022 Republican gubernatorial primary. We similarly specify c for each simulated precinct by multiplying early voting turnout with a randomly drawn value from a beta distribution whose shape parameters are the maximum likelihood estimates produced by fitting a beta density to Lombardo's observed early voting support rate from precincts in Clark County.[20] As in the first simulation, we then set b equal to the difference between mail ballot turnout and a and likewise set d equal to the difference between early voting turnout and c. Finally, we transform our randomly drawn a - d using Equations 1 - 3 and then estimate Equation 4, recording the value of R^2.

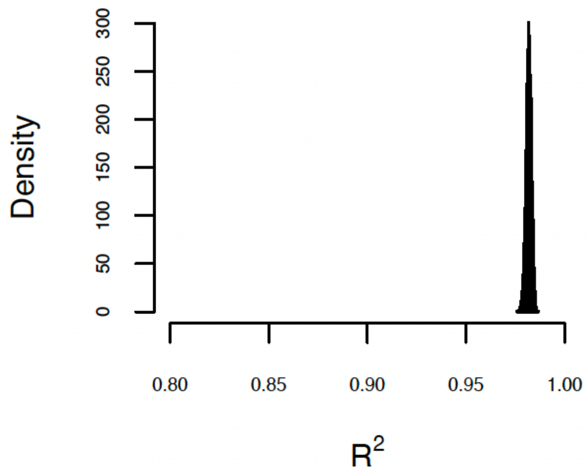
Figure A.1 displays histograms of R^2 values from our three simulations, each of which

was run 5,000 times. In all cases, these values are close to one. In the "Poisson" simulation, the average R^2 was 0.98, and the simulated R^2 values were between 0.95 and one. In the "Uniform" simulation, the average R^2 value was 0.85 and in the "Beta" simulation, 0.94.

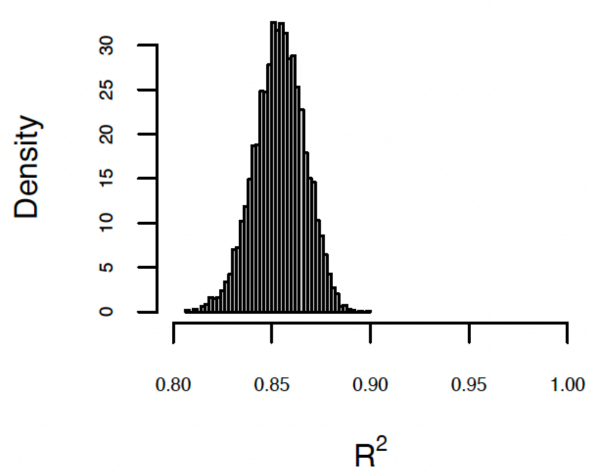
None of our simulations incorporated any form of vote manipulation. Nonetheless, what we see from this histograms in Figure A.1 is that the transformation of Daugherity's variables a - d, specified in Equations 1 - 3, induces a high R^2 value in a regression of g on h and α. Daugherity interprets the high R^2 value he observes as evidence of vote manipulation, but our simulations make clear that in fact the R^2 value noted by Daugherity is in fact evidence of nothing.

[20] When we fit beta distributions to Lombardo's mail ballot and early voting support rates by Clark County precinct, we disregard rates that are zero, one, or undefined on account of zero mail or early voting ballots having been cast.

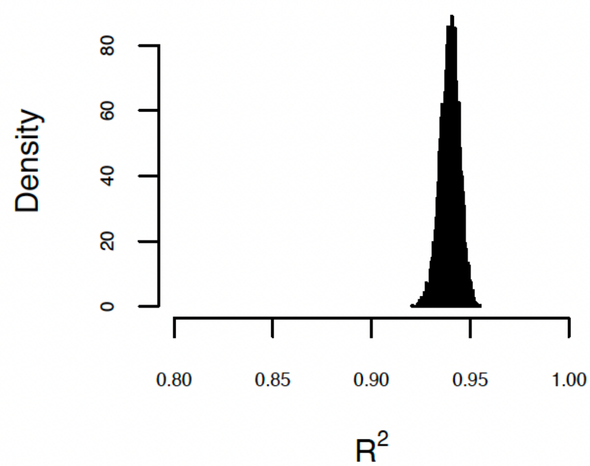
Figure A.1: Histograms showing R2 values across three simulations based on Daugherity's regression



(a) "Poisson" simulation



(b) "Uniform" simulation



(c) "Beta" simulation

B Frank's "phantom voters" are nothing more than voter file maintenance

A key feature of Frank's claims of vote manipulation is that voter rolls are contaminated with "phantom voters" who can be deployed to overturn election results. Frank asserts that these voters serve as turnout credits that can be used to determine election outcomes.

The basis for Frank's claims about phantom voters rests upon discrepancies between statewide voter files — which we previously noted are lists of registered voters in a state — and official election results. Drawing on voter files, proponents of the idea that phantom voters exist point out that the total number of voters in any particular voter file who are listed as having turned out to vote in a given election tends to be less than official voter turnout in said election. One obvious explanation for this is that voters are removed from voter rolls because they have moved, died, or otherwise canceled their registrations. When a voter is removed from a voter file, this voter will no longer be counted when using the voter file to calculate overall turnout in elections in which the voter participated. Over time, for any given election in a state this will lead to the state's voter file having fewer voters who participated in said election than actually voted in it.

For example, in public testimony in Wisconsin, Frank Frank argued on December 8, 2021 that drops in the number of registered voters are a surprising feature of the Wisconsin voter registration system.[21] In fact, Frank appears unaware that such drops coincide with efforts to identify voters who have moved. Rather than reasonably attributing differences between counts of voter file voters and official turnout statistics to voter file maintenance, Frank argues that these these differences reflect fictitious voters added to the rolls so as to determine election outcomes.

To assess his phantom voter theory, Frank runs non-random surveys, which he calls "canvassing," to identify what he deems to be suspicious activity (e.g., Frank 2021b). In

[21] The Wisconsin legislature has records of Frank's testimony at https://legis.wisconsin.gov/assembly/59/ramthun/media/1348/dr-frank-12-8-2021-wisconsin.pdf (last accessed October 24, 2022).

particular, it is our understanding that Frank coordinates with local groups to go door-to-door in neighborhoods to assess discrepancies between official vote records and the recollections of residents. While, to the best of our knowledge Frank has not presented official statistics on individuals he believes did not vote in a given election but are recorded as having voted in it, he has claimed in presentations that he uncovers individuals who fit this description as well as residents who claim that individuals recorded as voting in the election do not live at their supposed addresses. Frank has not addressed alternative explanations for the patterns he thinks are evidence of malfeasance — including the possibility that voters may fail to recall the elections in which they have, that residents of multi-unit buildings may not know all residents of the building, and that poorly trained canvassers may make errors when identifying the current residents of a building.

To better understand who is removed from the voter rolls and the extent to which they are actually real individuals, we take advantage of a policy established in Georgia that calls for maintaining lists of all individuals who vote by mail, i.e., absentee voter lists. These lists are not updated when voters are removed from the voter rolls. Using unique Georgia voter identification numbers, we can link individuals from, say, a current Georgia voter file to absentee voting list from prior elections. After identifying individuals who were removed from a voter file (these individuals are on an absentee voter list but not in said voter file), we can engage in a simple search to confirm that they are real individuals.

In the 2020 presidential election in Georgia, the total number of voters who cast a vote for president was 4,998,482, and total turnout in the election was 5,023,159.[22] The number of voters in the December 2020 voter file, according to L2 was 4,905,383 a difference of 117,776 (about 2.3 percent of the total votes cast). We then used the absentee voter file to determine if we could identify who was removed from the file. Of those who are removed, we were able to find 66,863 who had cast absentee ballots or approximately 56.8 percent. Simple internet searches reveal it is clear that these are real individuals. In fact, we can obtain records of individuals who moved after the election and therefore changed their voter registrations.

[22] For total presidential votes cast in the 2020 general election in Georgia, see data produced by the Georgia Secretary of State, available at https://results.enr.clarityelections.com/GA/105369/web.264614/#/ summary (last accessed October 24, 2022). For total turnout, we downloaded official Georgia turnout by demographic data at https://sos.ga.gov/sites/default/files/bulk/Voter Turn Out By Demographics.zip (last accessed October 24, 2022), extracted the file for the 2020 general election, and summed column BN.

In short, there is little reason to suspect that individuals who voted in the 2020 election but do not appear in the December 2020 Georgia voter file are phantom voters.

The absentee ballot records maintained by the Georgia Secretary of State also show the Georgia system canceling the ballots of individuals who are no longer eligible to vote. For example, in our searches we were able to identify an individual who had requested an absentee ballot for the 2020 general election before passing away before it. This individual's ballot was canceled through an administrative process.[23]

In short, a simple analysis of a state, Georgia, that provides information on who is removed from its voter reveals no evidence of nefarious "phantom voters." Instead, what it shows is a state where election officials update registration records and seek to block deceased individuals from receiving ballots.

[23] To avoid revealing personal information, we are not publicizing names of such individuals. We are able to discuss these names with interested researchers.

C Emails to Douglas Frank for replication code and providing memo

Replication Code + Data for Analysis of Election Results
3 messages

Justin Grimmer <justin.grimmer@gmail.com> Mon, Jun 14, 2021 at 7:35 AM
To: Doug@toolsforanalysis.com

Dear Professor Frank,

I hope you are doing well.

My research group has been investigating statistical claims about voter fraud in the 2020 election and we have come across your analyses.

I'm curious if you'd be willing to share the code + data that generated your results? Our research group has expertise in modeling, statistics, and programming so we're happy to take the data in whatever form is most convenient for you.

If that isn't possible, it would be helpful to know how you fit your sixth degree polynomial.

I'm happy to talk on a call if that is easier.

All the best
Justin Grimmer
Professor, Department of Political Science
Senior Fellow, Hoover Institution
Stanford University

Justin Grimmer <justin.grimmer@gmail.com> Wed, Jun 30, 2021 at 7:29 AM
To: Doug@toolsforanalysis.com, Matthew Tyler <mdtyler@stanford.edu>

Hi Professor Frank,

Matt Tyler (post-doc at the Democracy and Polarization Lab) and I recently authored a report about your election analysis, which is available here. https://www.dropbox.com/s/jibv67zh9lwdlwq/FrankMemo.pdf?dl=0

We show that the high correlations you identify are an artifact of correlating a variable with itself and therefore not evidence of fraud. While we do not include the mathematical proof here, we're happy to discuss the conditions where this will occur generally.

I'm happy to discuss this with you. I do think it is important that you update your analysis given the clear and obvious mistake that is driving the results (and how you might be deceiving people based on that mistake)

All the best
Justin Grimmer
Professor Department of Political Science
Senior Fellow, Hoover Institution
Stanford University

Justin Grimmer <justin.grimmer@gmail.com> Tue, Aug 10, 2021 at 11:27 AM
To: Doug@toolsforanalysis.com, Matthew Tyler <mdtyler@stanford.edu>

HI Professor Frank,

I just watched your appearance at the Cyber Symposium with great interest. I wanted to be sure that you saw our analysis that shows your conclusions about the election are inaccurate.

https://www.dropbox.com/s/jibv67zh9lwdlwq/FrankMemo.pdf?dl=0

You mention a desire to discuss the results with experts. I'm very happy to discuss this with you and explain why your analysis is flawed. The core reason is that you're essentially correlating a variable with itself.

All the best
Justin Grimmer
Professor Department of Political Science
Senior Fellow, Hoover Institution
Stanford University

D Frank's Claims about variation in precinct-level partisan turnout across California counties

At Mike Lindell's "Moment of Truth" summit, held August 20-21, 2022, Frank claimed to have evidence that vote counts in California had been manipulated in the 2020 general election. Specifically, Frank asserted that precinct-level partisan turnout rates were constant across counties in California, the sort of regularity that Frank takes to be evidence of vote manipulation.

To make his case, Frank first focused on the relatively conservative Placer County, located on the California-Nevada border near Lake Tahoe. Frank asserted that, "88% of the registered Republicans in EVERY Placer County precinct voted…Exactly 88% of all Republicans voted in that Precinct" (Frank 2022d). After asserting that this is implausible, Frank presented a slide where he opined that, "we might suspect that someone was stuffing ballots up to a target limit of 88%" (Frank 2022d). He went on to assert that the basic pattern of across-precinct consistency in partisan turnout rates held for Democratic and overall turnout in Placer County and for precinct-level partisan turnout in California counties statewide claiming that, "this is all over the state, this is happening everywhere" (Frank 2022d).

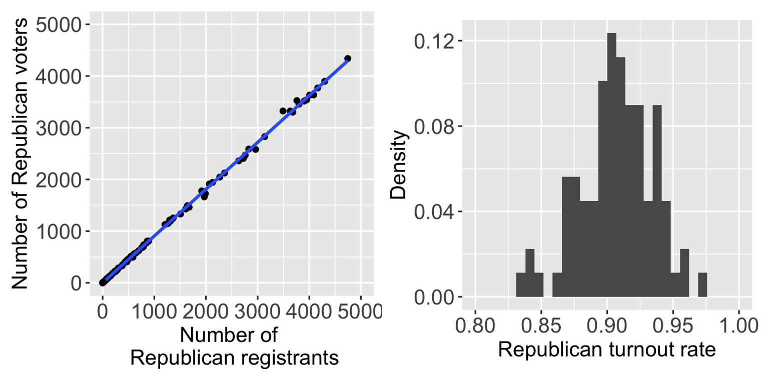
Frank's analysis of California, however, is simply inaccurate.[24] Once (1) Frank's statistical misunderstandings addressed and (2) we correctly disentangle turnout rates and turnout numbers, there is no evidence of any surprising patterns in the 2020 general election results in California.

[24] Frank claims that there were 35 precincts in Placer County in the 2020 general election. This is incorrect: there were 127 voting precincts in the county. Further, the sizes of the precincts Frank described at the "Moment of Truth Summit" are not at all reflected in official election results (Placer County Elections Office 2020). Finally, at the summit Frank displayed a map of Placer County divided into rectangular grids. We cannot determine such a network of grids corresponds to precincts in Placer County, if it does at all. Frank has subsequently explained in a *Rumble* video that he grouped together precincts based on their location (Frank 2022e). Our conversations with Placer County election officials confirm this is erroneous and does not correspond to any division of election districts actually used in the county.

D.0.1 Placer County

The left-hand plot in Figure D.2 replicates the plot Frank presented at the aforementioned "Moment of Truth" summit in August 2022, but using the correct precinct divisions in Placer County. To replicate Frank's plot, we used precinct-level data from the 2020 general election in California.[25] The horizontal axis in the left-hand plot of Figure D.2 shows the number of Republican registered voters and the vertical axis, the number of Republicans who turned out to vote. Pictured as well in the plot is a blue least squares regression line, which summarizes the displayed points.

Figure D.2: Republican voter turnout in Placer County in the 2020 general election



Note: shows only Placer County precincts with more than 100 registered voters.

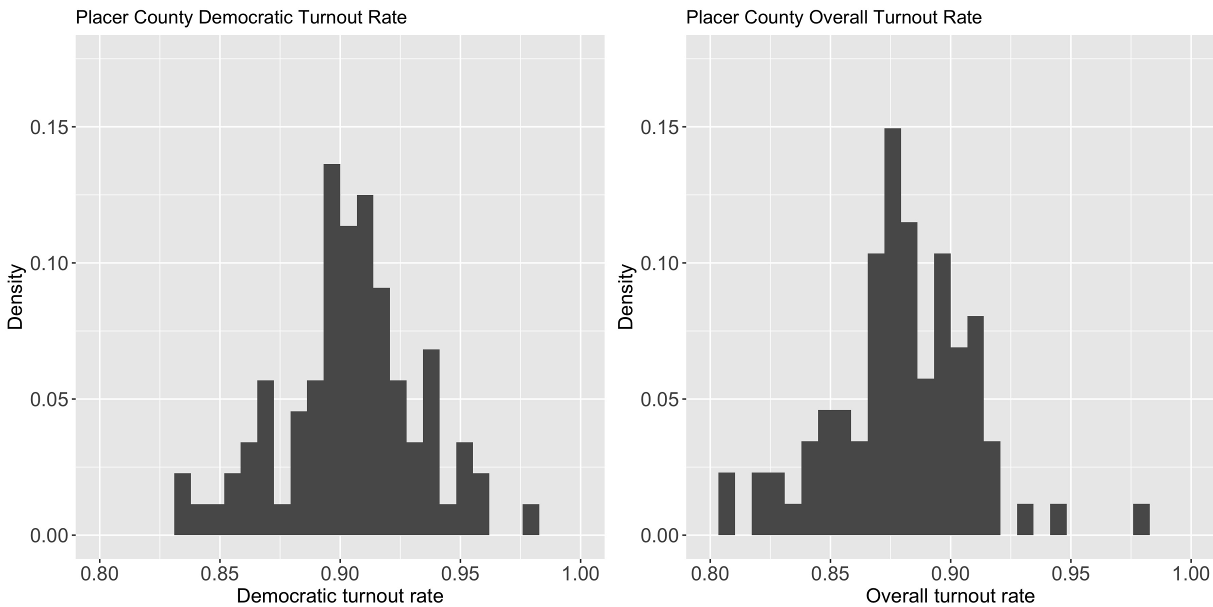
The left-hand plot in Figure D.2 and its blue regression line, which is sloped up, reveals the fact that, in the 2020 general election, larger precincts in Placer County had both more registered Republicans and more Republican voters. This is not surprising.

The figure's regression line, however, cannot speak to Frank's claim about equal precinct-level turnout rates in Placer County. That said, in his comparable figure, Frank reports a regression slope of 0.88 where the slope of the blue line in the left-hand plot of Figure D.2 is 0.9. Armed with his slope, Frank asserts that *every* precinct in Placer County saw 88% of Republicans voting in the 2020 election. This is demonstrably false, as made clear by the

[25] Our data on Placer County and other counties in California was gathered from the California Statewide Database, available at https://statewidedatabase.org, with the specific data obtained at https://statewidedatabase.org/d10/g20.html (last accessed October 21, 2022).

right-hand plot in Figure D.2, which presents precinct-level Republican turnout rates across Placer county. This plot shows that there is considerable variation in these turnout rates — with a precinct-level high of 98 percent of registered Republicans and a low of 78 percent. We observe the same pattern for Democrats and overall turnout in Placer County, which we show in Figure D.3.

Figure D.3: Democratic and overall turnout rates in Placer County in the 2020 general election.



D.0.2 Precincts in California

Frank's claim of a single, county-specific, precinct-level turnout rate across counties in California is also false. Figure D.4 shows the distribution of precinct-level Republican turnout rates in the 24 largest counties in California. Unsurprisingly, across precincts we see considerable variation in precinct-level Republican turnout rates. In short, there is no basis for Frank's claim that there is a single precinct-level turnout rate for Republicans and Democrats in each California county.

Frank's allegations about constant partisan turnout rates across counties fails because he confuses the number of voters who turnout to vote with observed turnout rates. Frank's aforementioned presentation plots the number of Republican voters in each precinct against the number of registered Republicans. Not surprisingly, there is a strong relationship between these two variables: precincts with more registered voters also tend to have more voters. Indeed, across counties in California the R^2 from a regression of precinct-level Republican turnout against precinct-level Republican registration within a county has an interquartile range from 0.9956 to 0.9975, with a similar interquartile range of R^2's for Democrats (0.9843 to 0.9940), and overall turnout (0.9808 to 0.9922). This is not, however, evidence of fraud or other illicit activity.

In fact, even if turnout rates by partisan groups in a precinct were randomly determined, we would expect to observe a large R^2 when regressing the number of partisans voting on the number of partisans who are registered to vote. To show this, we constructed a simple simulation using the actual number of registered voters in actual California counties. Specifically, we supposed that, in each simulation and for each precinct i, in each county j, the turnout rate for partisans p, $t_{ijp} \sim \text{Uniform}(0.825, 0.975)$ and we then calculate the number of voters from the iteration of the simulation as $\tilde{n}_{ijp} = t_{ijp} \times v_{ijp}$. Finally, to replicate Frank's analysis, in each county we then regressed $\tilde{n}_{ijp}$ on the number of registered voters v_{ijp} and recorded the R^2 for county j in that iteration of the simulation R^2. We repeated this simulation 1,000 times for each county.

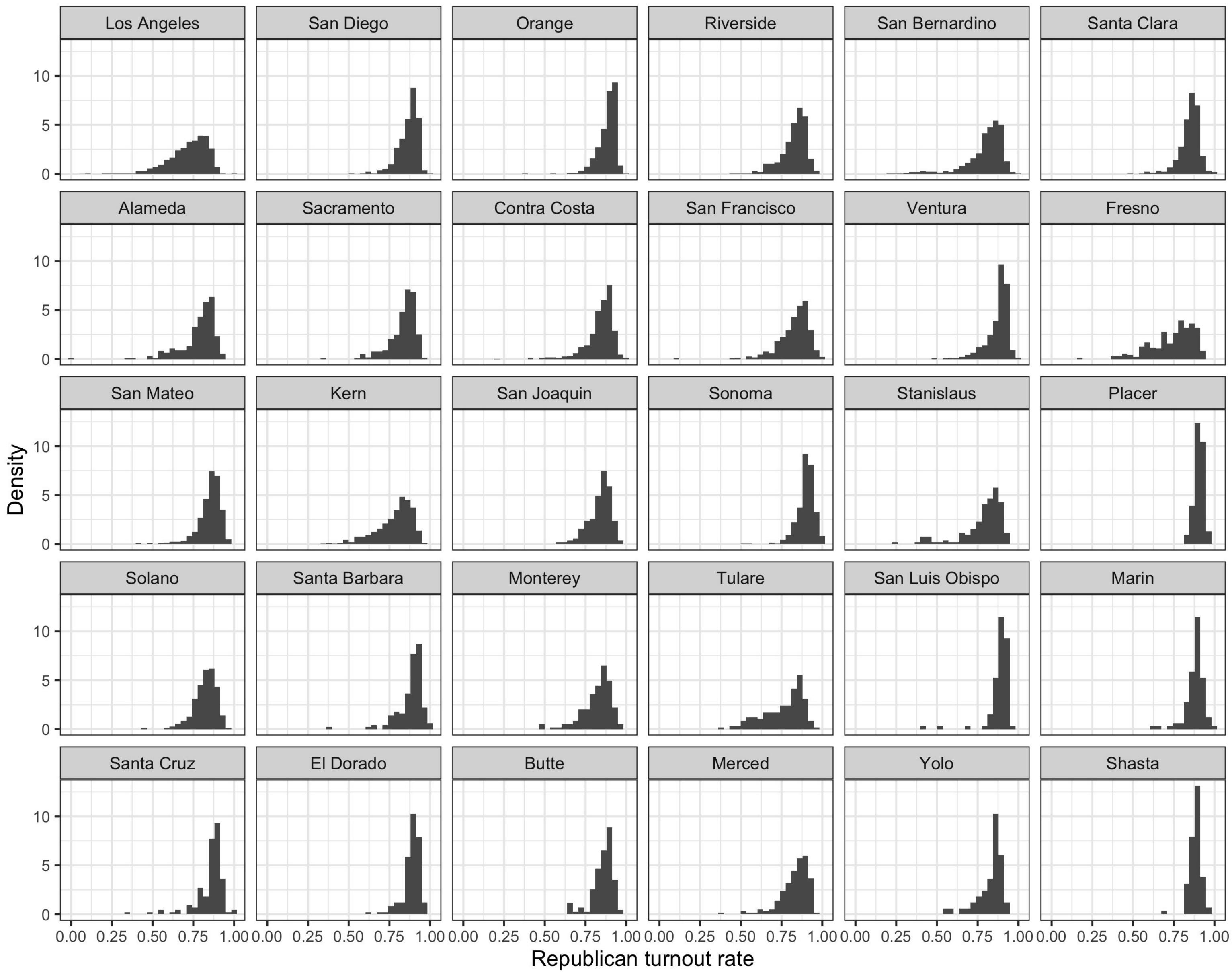


Figure D.4: Republican Turnout Rates for the 24 California Counties with the most registered voters. Across the counties there is considerable variation in the turnout rate of precincts

Note: counties order by numbers of registered voters in the 2020 general election.

From our simulation, we find a very high average R^2. The average R^2 for Democrats, Republicans, and for overall turnout is 0.992, 0.994, and 0.992, respectively. This high average R^2 occurs even though the turnout rate in each precinct is independently drawn at random. In short, Frank's supposed evidence of vote manipulation in California precincts is not evidence of manipulation at all. Rather, it demonstrates that precincts with more registered voters tend to have more voters turnout to vote.

E L2 statewide voter files

In the body of the paper, we described how we used statewide voter files from L2 to characterize age-based voter turnout by county and state. Table 4 in this appendix lists the dates on which the 42 voter files were pulled.

Table 4: Pull dates for L2 statewide voter files

State	Pull date
AL	2021-02-04
AR	2021-03-16
AZ	2021-05-20
CA	2021-05-02
CO	2021-05-18
CT	2021-03-30
DE	2021-03-24
FL	2021-05-19
GA	2021-04-16
IA	2021-03-04
ID	2021-03-16
IL	2021-03-05
IN	2021-01-15
KS	2021-03-16
KY	2021-05-11
LA	2021-01-22
MA	2021-01-19
ME	2021-05-11
MN	2021-02-14
MO	2021-02-11
MS	2021-03-23
MT	2020-12-14
NC	2021-05-18
NE	2021-01-20
NJ	2021-03-11
NM	2021-02-25
NV	2020-12-17
NY	2021-03-15
OH	2021-04-09
OK	2021-02-08
OR	2021-02-05
PA	2021-05-20
RI	2021-03-16
SC	2021-04-16
SD	2021-01-22
TN	2021-03-29
TX	2021-03-25
UT	2021-03-26
VA	2021-02-18
VT	2021-03-04
WA	2020-12-09
WV	2021-03-11