
Was this election rigged?

55,000 Residents Desperately Need Your Help!

Andrew Gelman

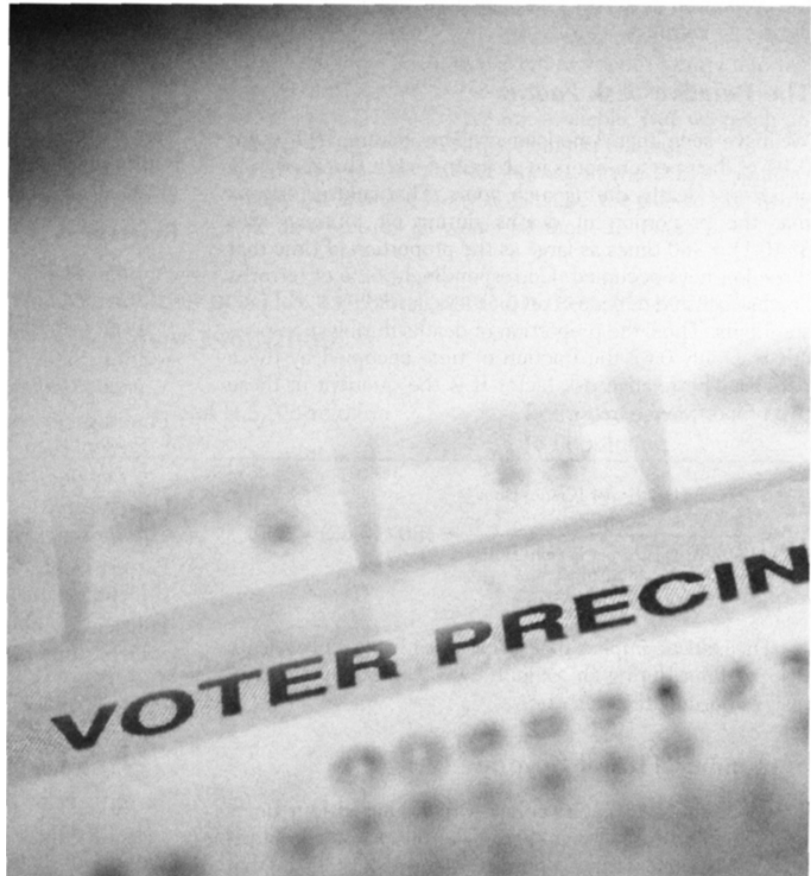
One day last summer, we received a fax, entitled HELP!, from a member of a residential organization:

Last week we had an election for the Board of Directors. Many residents believe, as I do, that the election was rigged and what was supposed to be votes being cast by 5,553 of the 15,372 voting households is instead a fixed vote with fixed percentages being assigned to each and every candidate making it impossible to participate in an honest election.

The unofficial election results I have faxed along with this letter represent the tallies. Tallies were given after 600 were counted. Then again at 1200, 2444, 3444, 4444, and final count at 5553.

After close inspection we believe that there was nothing random about the count and tallies each time and that specific unnatural percentages or rigged percentages were being assigned to each and every candidate.

Are we crazy? In a community this diverse and large, can candidates running on separate and oppo-



Clotelia Smith	208	416	867	1259	1610	2020
Earl Coppin	55	106	215	313	401	505
Clarissa Montes	133	250	505	716	902	1129

Figure 1. Subset of results from the cooperative board election, with cumulative votes for each candidate (names altered for anonymity) tallied after 600, 1200, 2444, 3444, 4444, and 5553 votes. These data were viewed as suspicious because the proportion of votes for each candidate barely changed as the vote counting went on. (There were 27 candidates in total, and each voter was allowed to choose 6 candidates.)

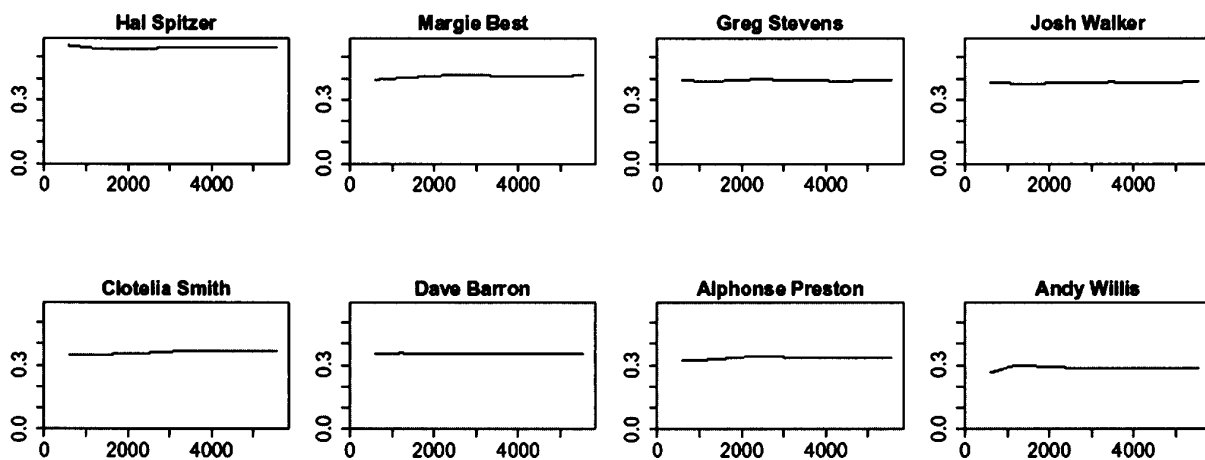


Figure 2. Proportion of votes received by each candidate in the cooperative board election, after each stage of counting: 600, 1200, 2444, . . . , 5553 votes. There were 27 candidates in total; for brevity we display just the leading 8 vote-getters here. The vote proportions appear to be extremely stable over time; this might be misleading, however, since the vote at any time point includes all the previous vote tallies. See Figure 3.

site slates as well as independents receive similar vote percentage increases tally after tally, plus or minus three or four percent? Does this appear random to you? What do you think? HELP!

Figure 1 shows a subset of the data. These vote tallies were deemed suspicious because the proportion of the vote received by each candidate barely changed throughout the tallying. For example, Clotelia Smith’s vote share never went below 34.6% or above 36.6%. How can we HELP these people and test their hypothesis?

We start by plotting the data: for each candidate, the proportion of vote received after 600, 1200, . . . votes; see Figure 2. These graphs are difficult to interpret, however, since the data points are not in any sense independent: the vote at any time point includes all the votes that came before. We handle this problem by subtraction to obtain the number of votes for each candidate in the intervals between the vote tallies: the first 600 votes, the next 600, the next 1244, then next 1000, then next 1000, and the final 1109, with the total representing all 5553 votes.

Figure 3 displays the results. Even after taking differences, these graphs are fairly stable—but how does this variation compare to what would be expected if votes were actually coming in at random? We formulate this as a hypothesis test and carry it out in five steps:

1. The *null hypothesis* is that the voters are coming to the polls at random. The fax writer believed the data contradicted the null hypothesis; this is what we want to check.
2. The *test statistic* is some summary of the data used to check the hypothesis. Since the concern was that the

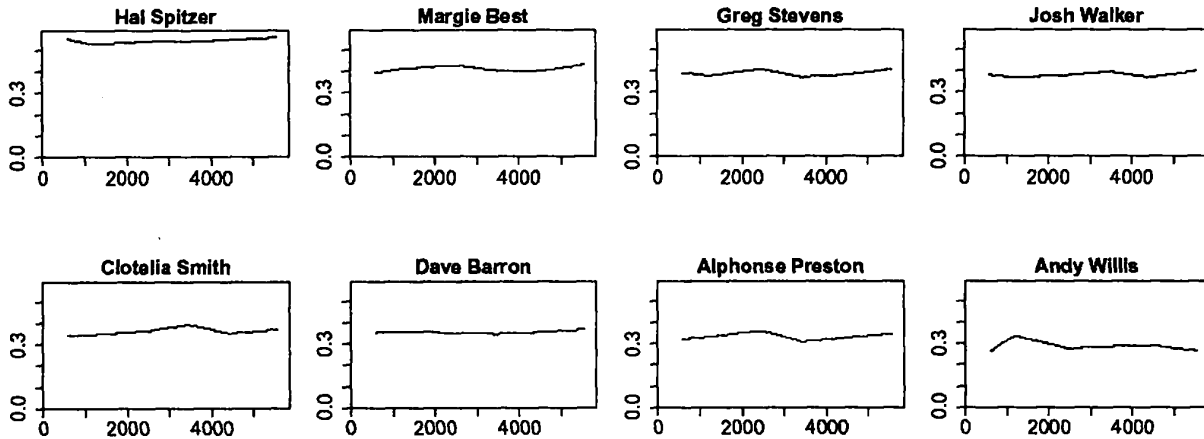


Figure 3. Proportion of votes received by each of the 8 leading candidates in the cooperative board election, at each disjoint stage of voting: the first 600 votes, the next 600, the next 1244, the next 1000, the next 1000, and the final 1109, with the total representing all 5553 votes.

votes were unexpectedly stable as the count proceeded, we define a test statistic to summarize that variability. For each candidate i , we label y_{i1}, \dots, y_{i6} to be the numbers of votes received by the candidates in each of the six recorded stages of the count. (For example, from Figure 1, the values of $y_{i1}, y_{i2}, \dots, y_{i6}$ for Earl Coppin are 55, 51, \dots , 104.) We then compute $p_{it} = y_{it}/n_i$ for $t = 1, \dots, 6$, the proportion of the votes received by candidate i at each stage. The test statistic for candidate i is then the sample standard deviation of these six values p_{i1}, \dots, p_{i6} , a measure of the variation in his or her votes:

$$T_i = \text{sd}_{t=1}^6 p_{it}.$$

3. Determine the theoretical distribution of the test statistic if the null hypothesis were true. Under the null hypothesis, the six subsets of the election are simply six different random samples of the voters, with a proportion π_i who would vote for candidate i . From the binomial distribution, the proportion p_{it} then has a mean of π_i and a variance of $\pi_i(1 - \pi_i)/n_i$. On aver-

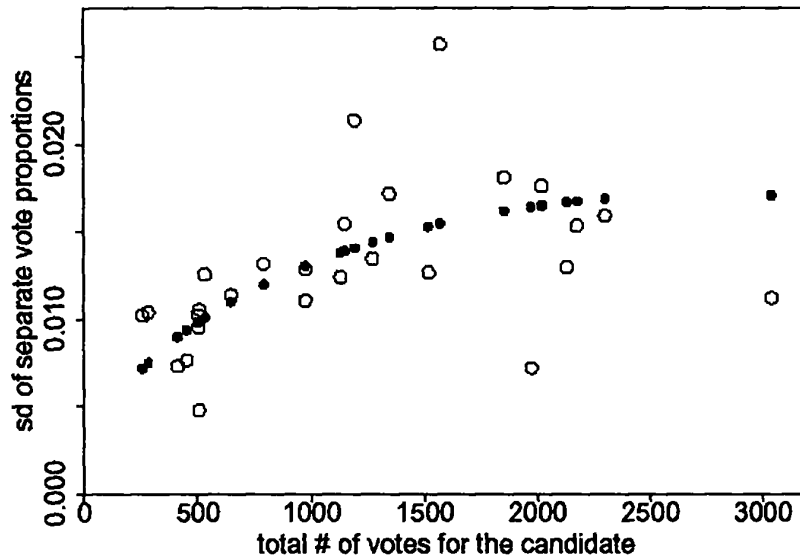


Figure 4. The open circles show, for each of the 27 candidates in the cooperative board election, the standard deviation of the proportions of the vote received by the candidate in the first 600, next 600, next 1244, \dots , and the final 1109 votes, plotted vs. the total number of votes received by the candidate. The solid dots show the expected standard deviation of the separate vote proportions for each candidate, based on the binomial model that would be appropriate if voters were coming to the polls at random. The actual standard deviations appear consistent with the theoretical model.