

Bush. In short, these results are inconsistent with the exurb hypothesis and provides additional support that the 2002 official results for Baldwin deviate from the expected vote significantly, the solid dot representing Baldwin County is well outside the boundaries of the 95% confidence interval, and in the direction consistent with the theft hypothesis.

How Baldwin County Results Could Have Been Manipulated:

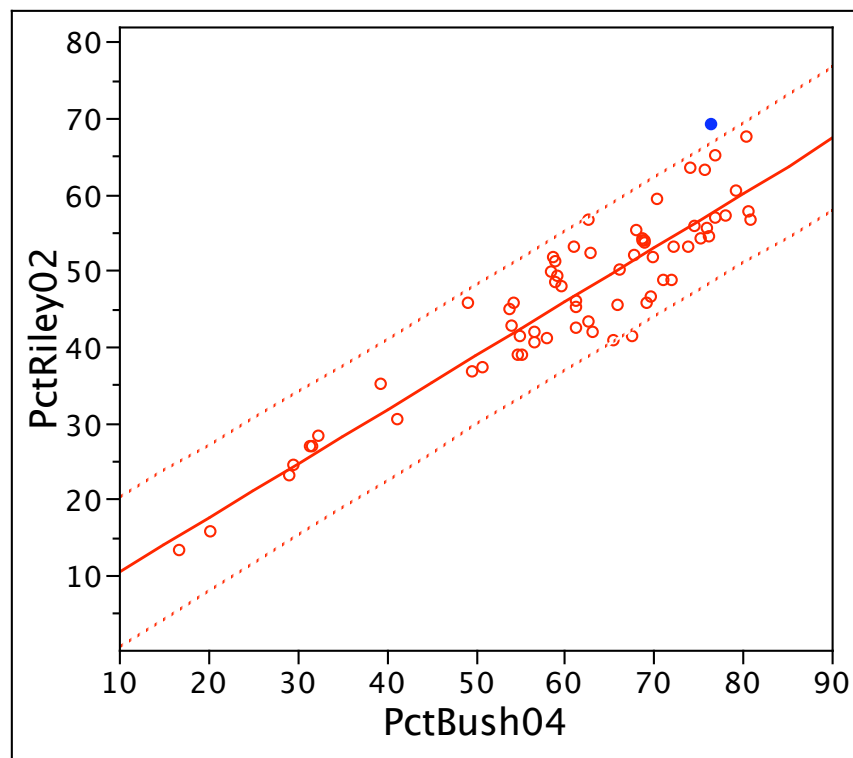
The other change I want to make from the 2003 version of this paper is the description of how the returns could have been manipulated. In 2003 I suggested the errors could have been made through the careless pasting of spreadsheet calculation formulas into spreadsheets holding the voting data. I have since learned that the data are not put into spreadsheet format until after they are tabulated and formatted for reporting. The data are collected and tabulated in Access format and manipulation of the data would require using Access tools, not spreadsheet formulas. However, it is easy to make the same apparent mistake, adding the stolen votes back to Siegelman's total as well as Riley's, as I described in the earlier paper and above.

Conclusion:

In this paper I update an earlier study that showed how some relatively simple statistical analysis techniques can be used to identify probable electronic manipulation of voting results in the 2002 Baldwin County Governor's election results. It had been suggested that the earlier results could have been the results of a dramatic shift in the political orientation of the Baldwin County voters because it is one of the new "exurbs" that are known to be the core of new Republican growth. The exurb hypothesis is rejected, the view that the 2002 Governor's election was stolen through manipulation of the Baldwin County vote is again found to be most consistent with the data.

consistent with the republican exurbs theory as the cheating theory. At the time I had no way to test that exurb hypothesis. However, with the 2004 election, we now have election results to test this theory. Specifically we can use the results for the 2004 election to predict the official 2002 results and see if Baldwin County moves back close to the regression line, results that would support the exurb hypothesis, or remains an outlier, results that support the theft hypothesis. The results are presented in the Figure 5 below:

Figure 5: Plot of Percent of County Vote for President Bush in 2004 and Percent of County Vote for Riley in 2002 with Final Baldwin County Results.



The regression using the percent of the county vote for President Bush in 2004 to predict the percent of the county vote for Riley in 2002 shows Baldwin county to be even more of an outlier than when predicted from the results of the 1998 Governor's race or the 2000 vote for President

slope for the other voting districts, .854, and subtracting the results from 1.00 you get an estimate of the proportion of the Siegelman vote in each voting district that that apparently disappeared from the official Baldwin County results. This yields a result of .18, which is about half as much as predicted from the hypothesis based on the first reported results. This raised the question about how could a process of moving X number of votes from one candidate to the other results in the mysterious production of erroneous results that were 2X above the final reported results?

The answer is surprisingly simple. This is a common error pattern to appear in programming spreadsheet calculations. My hypothesis is that someone was moving a little more than 3,000 Baldwin County votes from Siegelman to Riley by calculating a fifth of Siegelman's votes in each voting district, rounding it to a whole number, adding the resulting value to Riley's votes in that district and then subtracting that number from Siegelman's vote. However instead of subtracting the calculated number they added it to the vote for Siegelman. This is a common error created by using copy and paste to produce the invisible formulas for cells of spreadsheets. The result was a first report of county vote totals that had percentage distributions close to what was expected but a total vote that much higher than expected. Once they went back and fixed the procedure so that it performed as they desired, a reasonable total vote and Riley winning the election, the difference between the first and second reporting of Siegelman's vote was twice the number of electronically shifted votes. If what I hypothesized happened, then the total votes for Baldwin County was 27,866 votes for Riley and 15,283 votes for Siegelman. This would have produced state totals of 669,039 votes for Riley and 671,652 votes for Siegelman. The only way we will know for sure is if the paper ballots for Baldwin County are recounted.

An Analysis of the 2004 Election Data

After I presented the 2003 paper I was presented with an argument that both of the above these results were also consistent with a view of Baldwin County becoming one of the new republican exurbs and that my results were just as

Table 1: Montgomery and Shelby County Results

Term	Estimate	Std Error	t Ratio	Prob> t	r
Intercept	7.863645	23.97778	0.33	0.7440	0
Siegelman98	0.853541	0.029685	28.75	<.0001	0.96125

Table 2: Baldwin County Results

Term	Estimate	Std Error	t Ratio	Prob> t	r
Intercept	-9.771309	11.75185	-0.83	0.4110	0
Siegelman98	0.697915	0.021321	32.73	<.0001	0.98316

There are two important points to note in comparing these results. First, the correlations, $r = .96$ and $.98$ are quite strong. That means that the 1998 Siegelman vote is an adequate predictor of the Siegelman 2002 vote. Second, is the difference in the estimates. The estimate, or slope, for the voting districts outside of Baldwin County is $.85$. This means that for each vote that Siegelman got in 1998 he got $.85$ votes in 2002. The fact that this matches the pattern in the state outside Baldwin County suggests that the selected voting districts adequately represents the rest of the state. Remember that in the first part of the paper, I showed that Siegelman got 85% of the vote in 2002 that he got in 1998. The regression for the Baldwin County voting districts shows a slope of $.697$. This means that results for Baldwin County are substantially different from the voting districts outside of Baldwin County. A significance test for the difference between the slopes shows that the two slopes are significantly different from each other, $t = 6.19$, $p < .0001$.

The combined findings of a strong relationship between the 1998 and 2002 votes in Baldwin County as well as outside, and the different slope strongly suggests a systematic manipulation of the Baldwin County voting results. In addition, a comparison of the slopes provides a way to estimate the apparent nature of the manipulation of the results. By dividing the Baldwin County slope, $.697$, by the

An analysis of Voting District Data

Given that the county level analysis deepens rather than reduces suspicions that the Baldwin County results are manipulated to make Riley the winner, I then extended the analysis to a set of voting districts to see if the suspicious pattern was a significant deviation from expected results. For this analysis I produced a data set for comparable voting districts in Baldwin, Montgomery, and Shelby Counties for the Governor's race in 1998 and 2002. It should be noted that determining which voting districts are comparable over two elections that are four years apart is difficult and tedious because several voting districts were changed either in name or boundaries or both between the two elections. I attempted to add the voting districts for Jefferson County but was unable to verify sufficient compatibility between the 1998 and 2002 results to include them in the analysis. Out of Montgomery and Shelby Counties' 122 voting districts I was able to verify sufficient consistency in names and boundaries to use 70. Of Baldwin County's 65 reported voting districts I was able to produce 39 comparable voting districts for both 1998 and 2002. Most of this reduction in numbers was due to aggregating boxes or beats to insure comparability. For example Fairhope Civic Center has boxes for four voting districts that shifted their boundaries between them from 1998 to 2002. By aggregating the four boxes into one geographic unit for both 1998 and 2002 I was able to create one comparable unit of analysis for both years. I conducted regression analyses similar to the county level analyses above for two sets of voting districts, those outside of Baldwin County and those in Baldwin County. If the first reported results were accurate for Siegelman but inflated for Riley, you would expect the slope for Baldwin County to be about a third lower than the slope for the other voting districts. The results are shown below:

Figure 3: Plot of Percent of County Vote for President Bush in 2000 with Percent of County Vote for Riley in 2002 with First Baldwin County Results.

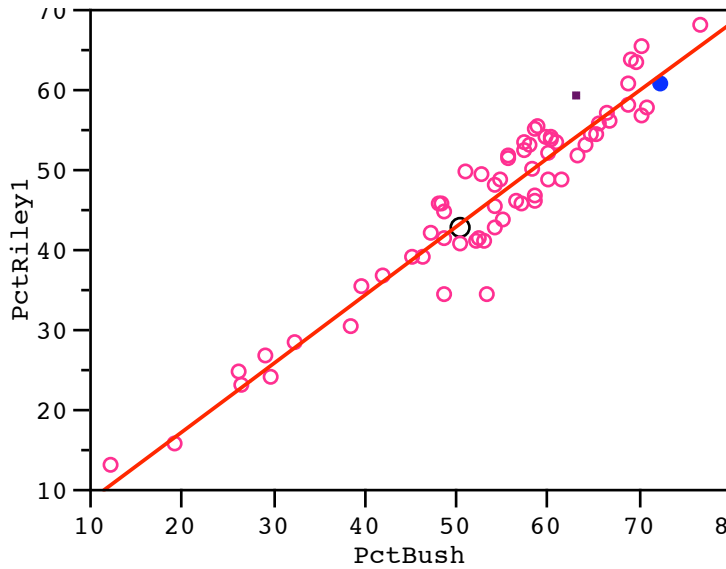
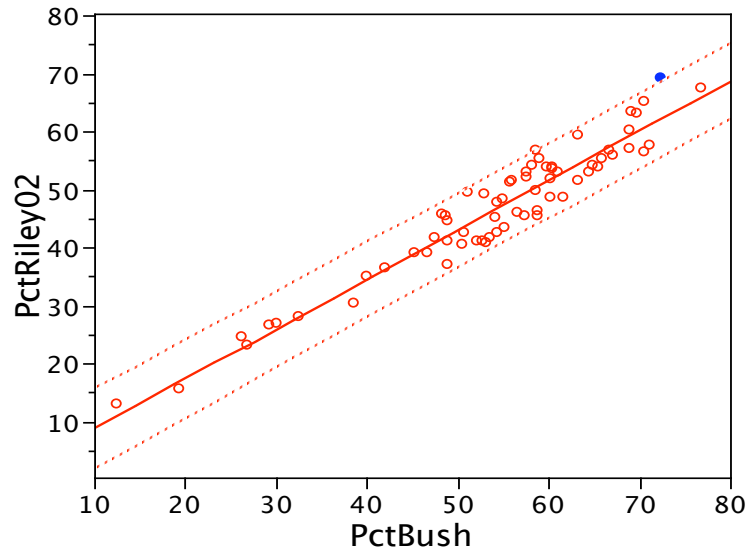


Figure 4: Plot of Percent of County Vote for President Bush in 2000 with Percent of County Vote for Riley in 2002 with Final Baldwin County Results.



The first plot shows the relationship for the initial returns and the second shows the results for the second set of Baldwin county returns. The solid dot is Baldwin County. Note that in the initial returns, Baldwin County fits closer to the line than most of the other counties. But, in the second Baldwin County is further from the line along the vertical dimension than any other county. These results suggest that the changes made between releasing the first and second set of results made Baldwin County an outlier. This is exactly the opposite of what you would expect if the changes corrected an error in the data. That is, errors usually make the data point deviate from expected patterns and fixing the errors typically moves the data point into the pattern. This kind of statistical irregularity deepens one's suspicions about the final Baldwin County election results.

However, there could be factors operating on the republican side that may account for this. Baldwin County voters, like most in Alabama, voted with strong majorities for President Bush in the last election. And Bush campaigned for Riley in Alabama. It can be argued that this change in the vote simply reflects the effect of Bush on the Baldwin County voters. Analyses using the percent of vote for Bush in 2000 to predict the percent of vote for Riley in 2002 are presented below. The first scatter plot shows the Baldwin County percentage based on the first reported results. Again the first results, those giving the election to Don Siegelman, fit the pattern shown by the other counties. The second results has Baldwin County an extreme outlier showing that Riley is receiving about fifteen percent more of the vote than predicted by the vote for Bush in 2000. It is worth noting that Baldwin County is the only data point that is outside the boundaries of the 95% confidence interval. The small square dot is Clay County, Riley's home. It is worth noting that home county advantage is only about a third as large as the advantage the final count gave Riley in Baldwin County. Thus, the analysis of the republican side of the election also increases the suspicion that Baldwin County results are manipulated.

percent of the county vote for the democratic candidate in 2002, using both sets of Baldwin county's returns, on the percent of the county vote for the same candidate for Governor in 1998. The results of these analyses are presented in Figures 1 and 2 below:

Figure 1: Plot of Percent of County Vote for Siegelman in 1998 and 2002 with First Baldwin County Results.

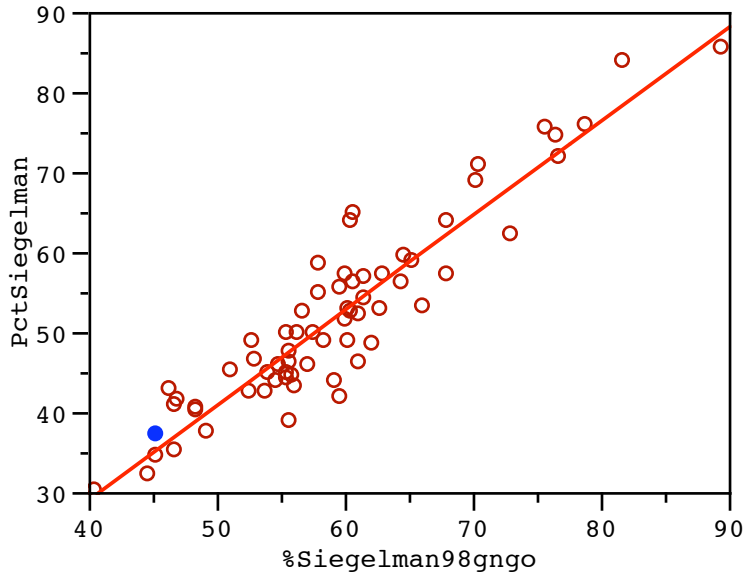
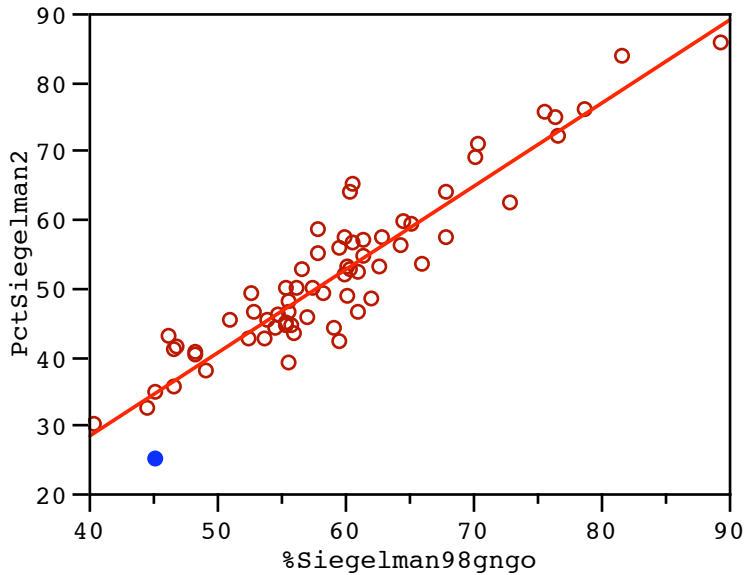


Figure 2: Plot of Percent of County Vote for Siegelman in 1998 and 2002 with Final Baldwin County Results.



the second set of returns Siegelman was reported as receiving 12,736 votes while Riley's total did not change.

In addition to the issues reported in the press, the above review suggests three points about Baldwin County's election results that should make one suspicious. One is the unusually large increase in the votes for the republican candidate from the 21,004 for James, the local resident in 1998 to the 31,052 for Riley in 2002. While Riley ran better across the state than James did, the only other county that showed a similar increase for Riley over James from '98 to '02 was Riley's home, Clay County, which went from 2,122 to 3,176.

The second factor that raises suspicions is the size of the decline in Siegelman's reported vote. The difference between the two reported votes for Siegelman is a decline of almost exactly one third of the total votes finally reported for Siegelman. A one third reduction is commonly found in data that is intentionally changed but rarely the result of random errors.

The final point that raises suspicions is that there should be no way to produce two different results with the computerized vote tabulation. That is, the system should not allow access to computer code or procedures that can produce different results. Computers do not accidentally produce different totals. Someone is controlling the computer to produce the different results. Once a computer produces different election results, any results produced by the same equipment operated by the same people should be considered too suspect to certify without an independently supervised recount.

A County Level Analysis: Baldwin County as an Outlier

The primary method of analysis in the earlier paper is an examination of changes in the level of votes for the democratic candidate between 1998 and 2002. The data were obtained from the elections page of the web site of the Secretary of State of Alabama. The Alabama Secretary of State site provides a substantial amount of data for Alabama election results in Excel format free for the downloading. The first set of analyses regresses the

Election Theft, A Second Look: The Case of Baldwin County
in Electing Alabama's Governor in 2002

By

James H. Gundlach
Auburn University

Paper Presented at the Annual Meetings of
The Alabama Political Science Association
Jacksonville, Alabama, April 2, 2005

About two years ago, I presented a paper at the 2003 Annual Meetings of The Alabama Political Science Association in Troy Alabama. In that paper I presented data and argument that made three points. The first point was based on an analysis of election return data that suggests that Baldwin County results in the 2002 Governor's race were sufficiently anomalous to provide statistically significant evidence that the election had been stolen. The second point of the paper noted patterns in the returns that were consistent with a theory that one in five of the votes for Don Siegelman were electronically taken from Siegelman and given to Bob Riley. The third point I made discussed possible mechanisms by which the probable electronic theft of the election could have taken place. In this paper, I briefly restate those original points, address a major criticism of the earlier paper.

Some election background

In the earlier paper I first looked at the election results by dividing the state into two components, Baldwin County and the rest of the state. I then compared at the results for the 1998 and 2002 Governor's races for these two components. In the 1998 election Don Siegelman received 742,766 votes in the rest of Alabama compared to 533,772 votes for his republican opponent, Fob James. In Baldwin County, which is where Fob James lived at the time, Siegelman received 17,389 votes compared to 21,004 votes for James. In 2002, Siegelman received 635,545 votes compared to 623,145 for Bob Riley in the rest of Alabama. In the first set of returns for Baldwin County Siegelman received 19,070 votes to 31,052 for Riley. In