

Voting Irregularities in Palm Beach County

Jonathan N. Wand Kenneth W. Shotts Jasjeet S. Sekhon
Walter R. Mebane, Jr. Michael C. Herron

November 28, 2000
Version 1.3

(Authors are listed in reverse alphabetic order)

Abstract

It is well known that Reform Party candidate Pat Buchanan received an unusually high share of the presidential vote in Palm Beach County, Florida. It has been alleged that the non-standard ballot used in Palm Beach County was responsible for this insofar as the ballot caused individuals who wanted to vote for Al Gore to instead vote for Buchanan. In light of this alleged irregularity we analyze presidential voting in Palm Beach County and reach the following conclusions.

Compared to all the 4,317 reporting districts (counties or townships) that cover 46 of the 50 United States, Palm Beach County is one the second most irregular in terms of having exceptionally high support for Buchanan that deviates from expected patterns. Among districts with more than 10,000 voters, Palm Beach County is the *most* irregular. Furthermore, based on census and electoral data from all 67 Florida counties we find that Buchanan's true support in Palm Beach County was significantly less than his 0.79 percent vote share.

Our analysis shows that in Palm Beach County Buchanan did better in precincts that strongly supported Gore. In addition, we show that liberal precincts within Palm Beach County tended to have higher proportions of ballots that were not counted for the Presidential election either because no holes were punched or multiple holes were punched. This evidence supports the claim that the ballot format in Palm Beach County led some Gore supporters mistakenly to vote for Buchanan and, in some cases, to vote for multiple presidential candidates.

Overall, we offer several different analyses of presidential voting in Palm Beach County, and each analysis leads to the same result: *The vote totals in Palm Beach County are irregular.* In particular, Buchanan received far more votes in Palm Beach County than we should expect given the county's characteristics and historical voting patterns. Moreover, patterns of voting within the county indicate that excess votes for Buchanan came primarily from Gore supporters.

Availability

A current draft of this paper as well as technical details related to its statistical models are available from <http://elections.fas.harvard.edu/>

Author Affiliations and Contact Information

Jonathan N. Wand

Department of Government, Cornell University
Lamarck, Inc.
jnw4@cornell.edu (phone: 617-905-9332)

Kenneth W. Shotts

Department of Political Science, Northwestern University
k-shotts@nwu.edu (phone: 847-491-2628)

Jasjeet S. Sekhon

Center for Basic Research in the Social Sciences, Harvard University
Lamarck, Inc.
jsekhon@fas.harvard.edu (phone: 617-496-2426)
[HTTP://jsekhon.fas.harvard.edu/](http://jsekhon.fas.harvard.edu/)

Walter R. Mebane, Jr.

Department of Government, Cornell University
wrm1@cornell.edu (phone: 607-255-3868)
[HTTP://macht.arts.cornell.edu/wrm1](http://macht.arts.cornell.edu/wrm1)

Michael C. Herron

Center for Basic Research in the Social Sciences, Harvard University
Department of Political Science, Northwestern University
mherron@latte.harvard.edu (phone: 617-495-4134)

1 Introduction

Many supporters of Al Gore expressed concern on election day that they may have mistakenly voted for Pat Buchanan or mistakenly marked votes for two presidential candidates. Buchanan did receive a surprisingly large number of votes in Palm Beach County, Florida. These facts have ignited a highly-charged controversy over the format of the Palm Beach County ballot.

We contribute to the debate about what happened in Palm Beach County by analyzing the magnitude of Buchanan’s vote share in Palm Beach County compared to other counties in Florida and to counties and townships across the entire country.¹ We find that Buchanan’s Palm Beach County vote total is not merely large but that in statistical terms it is extraordinary. This result supports the serious concern that the Palm Beach County ballot led to voting irregularities.

Furthermore, we examine voting patterns within Palm Beach County and find strong statistical evidence that Buchanan voters are concentrated in the most liberal precincts of Palm Beach County. We also find that invalid, double-punched ballots—presumably double-punched for Gore and Buchanan—tend to come from relatively liberal precincts. These two findings are evidence for the claim that the ballot format in Palm Beach County led some Gore supporters mistakenly to vote for Buchanan and, in some cases, to vote for multiple presidential candidates. These findings also support the assertion that Palm Beach County’s presidential election voting was irregular.²

2 National Analysis

Recent analyses of vote returns in Florida by journalists, academics, and others have focused on allegations of ballot irregularities and surprisingly high levels of support for Buchanan in Palm Beach County. We compare the election returns of Palm Beach County with those of 4,317 reporting units (counties or townships) across the United States.³ Such a comparison is essential to clarify whether the Palm Beach County electoral returns are exceptional.

We start with Florida. For each county in Florida, we compute the number of votes expected for Buchanan given the shares of the votes in the county for George Bush and for Ralph Nader. For each county we then compute the discrepancy between the actual number of votes for Buchanan and the expected number. We adjust the discrepancies to make it possible to compare them across counties.

Figure 1 presents the distribution of the discrepancies across Florida’s 67 counties. Most Florida counties are regular, which means their discrepancy values are very close to the zero

¹Alaska, Delaware and Hawaii were omitted because the number of reporting districts (counties or townships) used in those states is too small to allow our models to be estimated. Michigan was omitted because Buchanan did not appear on the ballot there.

²All data used in this paper are based upon pre-recount vote totals.

³For most states in our dataset, especially the large ones, the reporting units are counties. But for some states such as CT the reporting units are smaller than counties, usually townships. See <http://www.cnn.com/ELECTION/2000/results/president/index.html> for details. As indicated in footnote 1, data from Alaska, Delaware, Hawaii and Michigan were omitted.

point of the horizontal axis. The Buchanan vote in these regular counties does not greatly differ from the expected values. But Figure 1 shows Palm Beach County to have a very large positive discrepancy. It is by far the most irregular county in Florida.

Extending such an analysis to the whole country shows Palm Beach County to have the second largest discrepancy from expectations in the entire United States. We compute discrepancies for the reporting units in all states, using the same method for each state as we used for Florida. The same adjustments that let the discrepancies be compared across counties (or townships) within a state also allow the discrepancies to be compared not only among the reporting units in each state but also across states.

Figure 2 presents the distribution of the discrepancies across the 4,317 reporting units in the 46 states we analyze. As in Florida, most reporting unit are regular, with discrepancy values near zero. Only one reporting unit, located in South Carolina, has a discrepancy greater than the discrepancy for Palm Beach County. Palm Beach County is the second most irregular place in terms of having exceptionally high support for Buchanan that deviates from the expected level. Further examination shows that among reporting units with more than 10,000 voters, Palm Beach is the *most* irregular—see Figure 3.

Please see the technical appendix in Section 9 for more details.

**Histogram of Discrepancies
From Expected Vote for Buchanan
Florida, 67 Counties**

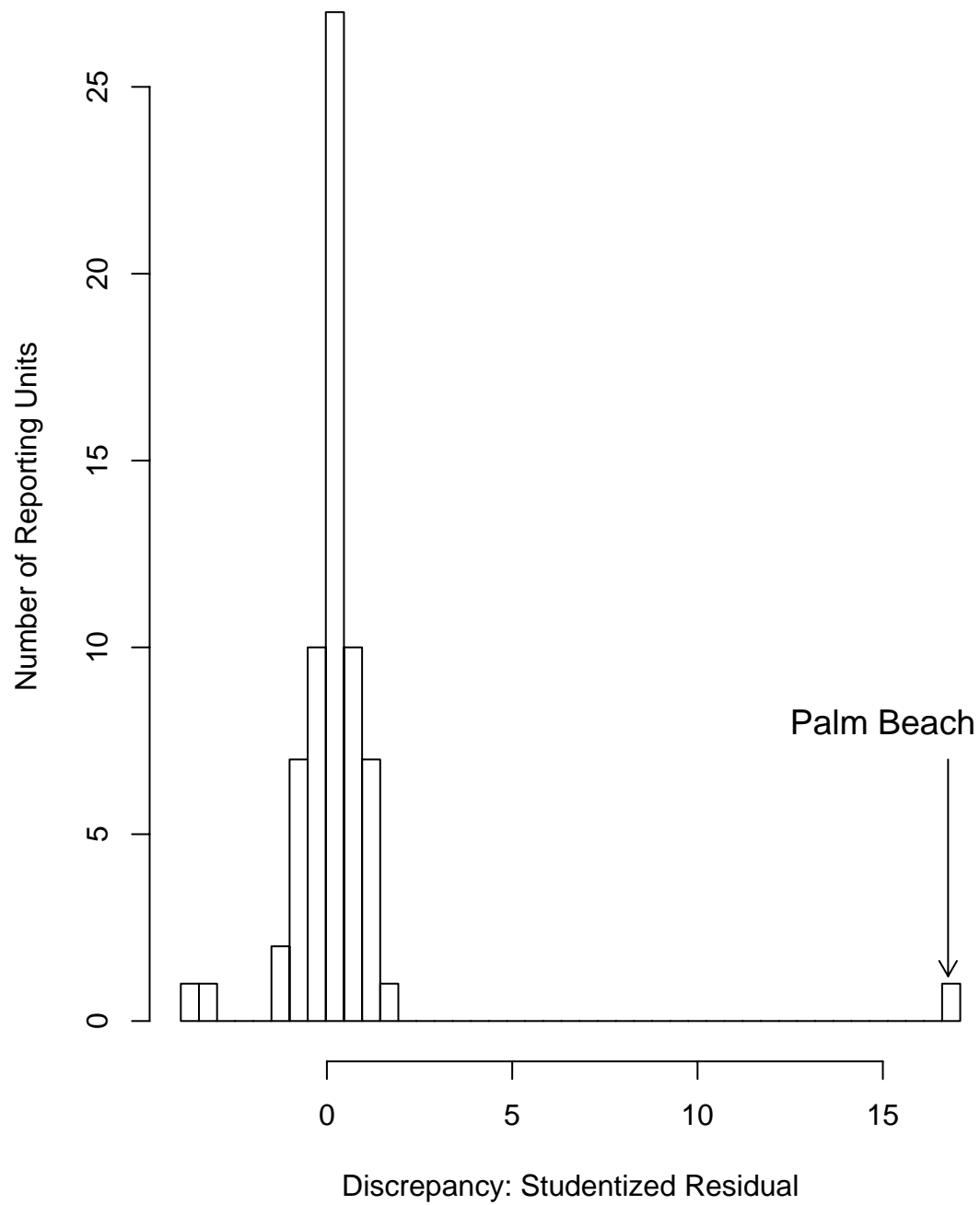


Figure 1: Discrepancies from Expected Vote for Buchanan

**Discrepancies From Expected Vote for Buchanan
By Reporting Units
4317 Reporting Districts in 46 States**

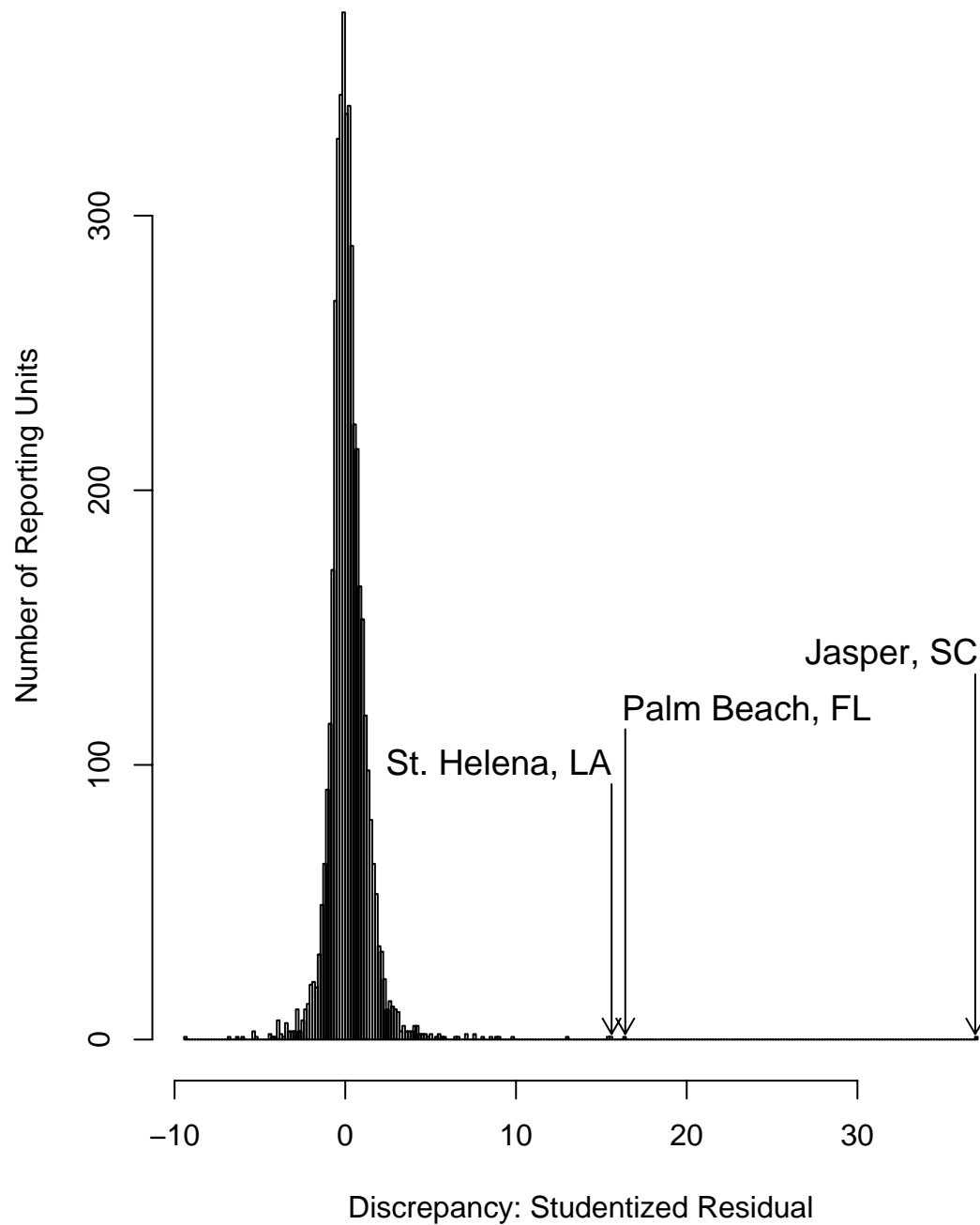


Figure 2: Discrepancies from Expected Vote for Buchanan

**Discrepancies From Expected Vote for Buchanan
Reporting Units with Greater than 10,000 votes
1560 Reporting Districts in 46 States**

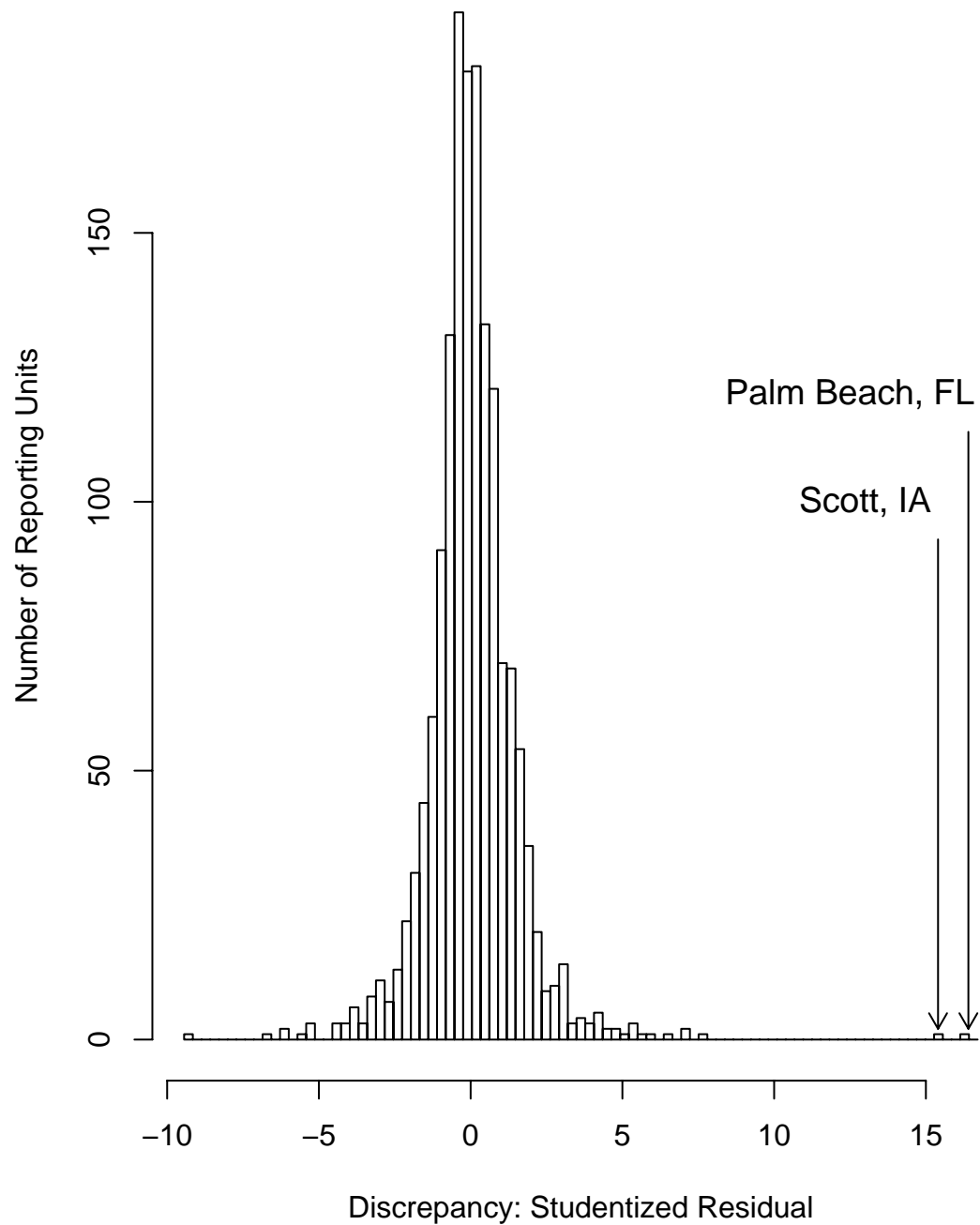


Figure 3: Discrepancies from Expected Vote for Buchanan

3 County-Level Analysis

We extend our analysis of Palm Beach County voting by examining county-level data throughout Florida. In particular, the objective of the county-level analysis concerns the following question: given the characteristics of Palm Beach County, what is the vote share that we would expect Buchanan to have received in the 2000 presidential election? In reality, we know that he received 0.789% of the Palm Beach County presidential vote, or 3407 votes total. One way to assess whether 3407 is a reasonable number is to try to “predict” the Buchanan share of the Palm Beach County vote using all counties in Florida except for Palm Beach County. The word predict is in quotes here because, of course, we need not predict something that is actually known. Nonetheless, given voting patterns in the 66, non-Palm Beach counties, we can determine whether the observed Palm Beach County vote is reasonably close to what we would have predicted had we not observed it.

Let $P_{\text{Bush},i}$ denote the share of votes received by George W. Bush in Florida county i , $i = 1, \dots, 67$. Define $P_{\text{Gore},i}$ and $P_{\text{Buchanan},i}$ similarly. We begin our county-level analysis by defining $P_{\text{Bush-Brogan},i}$ as the share of the gubernatorial vote received by the winning Republican pair Jeb Bush and Frank Brogan in their 1998 race against Democrats Buddy MacKay and Rick Dantzler. Roughly speaking, $P_{\text{Bush-Brogan},i}$ reflects the extent to which Florida county i is politically conservative. Furthermore, the 1998 gubernatorial election took place close to the 2000 presidential election, so it is reasonable to think that Florida counties did not change dramatically between 1998 and 2000.

Thus, we estimate using 66 Florida counties (all but Palm Beach County) a statistical model that links Buchanan vote share $P_{\text{Buchanan},i}$ to Bush-Brogan vote share $P_{\text{Bush-Brogan},i}$.⁴ Our model shows that Buchanan vote share in 2000 was positively related to Bush-Brogan share. This finding is intuitive since the extent to which a Florida county was politically conservative in 1998 should be positively associated with the county’s Buchanan support in 2000. In particular, our statistical model implies that, if the relationship between $P_{\text{Buchanan},i}$ and $P_{\text{Bush-Brogan},i}$ in Palm Beach County were the same as the relationship between these two variables in the other 66 Florida counties, then Buchanan should have received 0.196%, plus or minus 0.447%, of the Palm Beach County vote share in the 2000 election.⁵ Since the observed Buchanan vote share of 0.789% is greater than $0.196 + 0.447 = 0.643\%$, we conclude that the support received by Buchanan in Palm Beach County was not consistent with the county’s level of political conservatism as measured by its support for Bush-Brogan in 1998. In other words, Buchanan received more votes than he should have in 2000, given the nature of Palm Beach County voters.

It is important to recognize that predictions like the one described above—that we would have expected Buchanan to have received 0.196%, plus or minus 0.447%, of the Palm Beach County vote—need to be treated very carefully. What is most important, in our opinion, is that the top of the range for predicted Buchanan vote share is less than the observed

⁴In particular, we estimate a generalized linear model with a probit link between $P_{\text{Buchanan},i}$ and $P_{\text{Bush-Brogan},i}$. This is an appropriate model as $P_{\text{Buchanan},i}$ must lie within the unit interval. The estimated slope coefficient in the model is 0.455 with an estimated standard error of 0.0274.

⁵The vote share of 0.447% represents 1.96 times the estimated standard error on the Palm Beach County prediction for $P_{\text{Buchanan},i}$.

Buchanan vote share.

We uncover a finding similar to the one above when we employ a statistical model that links Buchanan vote share in Palm Beach County to the vote share of Bill McCollum, the losing Republican Senate candidate in Florida.⁶ Here, we use McCollum vote share per county as a measure of county conservatism; this is a plausible approach since the ballot problems which are alleged to have caused Gore supporters to vote for Buchanan have not been linked with Senate voting problems. Based on our McCollum analysis we find that, in the 66 Florida counties, Buchanan vote share is positively related to McCollum vote share; this seems eminently reasonable since, presumably, a fraction of McCollum supporters are sufficiently conservative to support Buchanan for president. Our statistical model, estimated with data from 66 counties, implies that Buchanan vote share in Palm Beach County should have been 0.210%, plus or minus 0.355%.⁷ Again, we see that, according to the distribution of McCollum votes across Florida, Buchanan received more votes in Palm Beach County than he should have. This is evidence of voting irregularity.⁸

Finally, we compare characteristics of the 67 Florida counties using data from the 1990 census. Unfortunately, data from the 2000 census are not yet available and this means that our measures of Florida county characteristics are rather dated. However, at this point dated county measures appear to be the only measures available, and hence we use them in our analysis with the caveat that measures based on 2000 census data are clearly necessary before a final analysis of Palm Beach County voting is performed.

With this caveat in mind, we consider a statistical model that links Buchanan vote share in 2000 with the following county characteristics: median family income, percentage of African-American residents, percentage of Hispanic residents, percentage of individuals with at least a high school education, percentage of veterans, percentage of residents who are 65 years or older, and the crime rate per 100,000 residents. As before, we apply our model to the 66 Florida counties and then use the results from this model to predict the Palm Beach County vote share for Buchanan. With respect to county characteristics, we find, among other things, that high median income is associated with low Buchanan vote share and that counties with large African-American and Hispanic populations had low Buchanan vote shares. These findings are not surprising, although we emphasize again that they are based on dated county measures.

Nonetheless, our analysis of county characteristics implies that Buchanan should have received a vote share of 0.0659%, plus or minus 0.178%. Like all of our county-level findings described above, this result again shows that Buchanan's vote share in Palm Beach County was extremely anomalous or, more to the point, irregular.

⁶Our McCollum analysis based on a generalized linear model, probit link, which models $P_{\text{Buchanan},i}$ as a function of McCollum vote share by county.

⁷Furthermore, a generalized linear model that links Buchanan vote share to McCollum share *and* Bush-Brogan share in 1998 predicts that Buchanan share in Palm Beach County should have been 0.194%, plus or minus 0.451%.

⁸Similarly, if we estimate a generalized linear model connecting Buchanan vote share to Gore vote share, we find that counties with many Gore votes tended to have few Buchanan voters. Nonetheless, such a model predicts that Buchanan vote share in Palm Beach County should have been 0.155%, plus or minus 0.433%. Again, according to this model Buchanan received in 2000 a greater percentage of the Palm Beach County vote than he should have.

In summary, our county-level findings are as follows. First, compared to other Florida counties as measured in a number of ways, the Palm Beach County vote share for Buchanan is extremely large. In fact, what we know about other counties in Florida implies that this vote share is so large as to be practically unbelievable. It is virtually certain that there is something unique about Palm Beach County, and the only obvious factor that is unique to Palm Beach County is its ballot format. Namely, arguments regarding the ballot format imply that it should have led to an excess number of Buchanan votes at the expense of Gore votes. We have uncovered evidence of this alleged consequence.

Second, and closely related to the first point, we have found no evidence to support the claim that Pat Buchanan received large numbers of votes from large numbers of Buchanan supporters living in Palm Beach County. Rather, using census data we estimated that Palm Beach County actually contains relatively few Buchanan supporters. Indeed, we believe that a sizable fraction of Buchanan voters did not intend to vote for Buchanan.

4 Precinct Analysis of Palm Beach and Leon Counties

Thus far we have argued that Buchanan’s vote total in Palm Beach County vastly exceeded any reasonable expectations. We now consider two possible counter-arguments to our claim that Buchanan received a substantial number of votes from people who were trying to vote for Gore.

Counter-Argument 1: The ballot in Palm Beach County was confusing to all voters, not just those trying to vote for Gore. Thus, Buchanan received erroneous votes from Bush supporters as well as from Gore supporters, and there is no reason to think the ballot structure influenced the electoral outcome.

Counter-Argument 2 The Reform Party is quite popular in Palm Beach County. The fact that Buchanan received many votes there simply reflects this popularity.

To address these claims we use data on electoral outcomes for each precinct in Palm Beach County.⁹ If the above counter-arguments are correct and if Buchanan did not disproportionately receive votes intended for Gore, then the number of votes Buchanan received in a given precinct should be unrelated to the number of votes that Gore received in that precinct. (Actually we would probably expect Buchanan to receive fewer votes in liberal, pro-Gore precincts). In contrast, if Buchanan received numerous votes intended for Gore, then the number of votes he received in a precinct should be positively correlated with the number of Gore supporters in that precinct.

Our analysis shows that in Palm Beach County Buchanan did better in precincts that strongly supported Gore. A regression-based statistical model which compares Buchanan’s vote share to Gore’s vote share across precincts in Palm Beach County suggests that between 0.8% and 1.6% of the voters who intended to vote for Gore wound up voting for Buchanan.¹⁰

⁹Precinct-level data on Palm County are available at <http://www.pbcelections.org>.

¹⁰Of our 614 precinct-level data points for Palm Beach County, 99 represent absentee ballot tabulations for various precincts. Although we assume that absentee ballots did not use the controversial “Butterfly

Given that Gore received approximately 269,000 votes in the county, this would mean that a substantial majority of Buchanan’s 3,407 votes came from mistaken Gore supporters.¹¹ Thus we conclude that Buchanan did better in Palm Beach County precincts that contained large numbers of Gore supporters, and this pattern is consistent with the claim that Buchanan received a large number of votes intended for Gore. And, notably, the pattern is quite inconsistent with either of the above counter-arguments which assert that Buchanan did not receive votes intended for Gore.

As another way of checking whether the ballot structure in Palm Beach County affected the number of votes received by Buchanan, we conducted a similar statistical analysis for Florida’s Leon County.¹² In Leon County, where the controversial “Butterfly Ballot” was *not* used, we found that there was almost no meaningful relationship between the fraction of votes in a precinct that went to Gore and the fraction that went to Buchanan. Indeed, the relationship was slightly negative, i.e., precincts with large numbers of Gore votes tended to have slightly lower than average numbers of Buchanan votes, and this is intuitive.¹³ This finding further supports the conclusion that the correlation between Gore votes and Buchanan votes in Palm Beach County was caused by the ballot layout.

5 Non-Counted Ballots In Palm Beach and Leon Counties

We also used precinct-level data to analyze the number of ballots that were not counted in the presidential election because they were either double-punched (and hence invalid for the presidential race) or because no presidential candidate was selected. The possibility of large numbers of invalid ballots is important in light of concerns that the confusing Palm Beach County ballot caused Gore supporters unintentionally to invalidate their presidential votes by voting for both Gore and Buchanan.

To analyze non-counted ballots in Florida precincts, we need a measure of the total number of ballots cast in each precinct in Palm Beach and Leon Counties. Since we do not yet have access to such data, we instead used the total number of votes cast in each precinct

Ballot” we have included them in our analysis. When the analysis is re-run using only the non-absentee data points, our results remain basically unchanged.

¹¹These findings follow from ordinary least squares estimation of a precinct-level statistical model which regresses Buchanan’s fraction of the vote on Gore’s fraction. The resulting slope coefficient estimate is approximately 0.012, and it has an estimated, heteroskedastic-consistent standard error of approximately 0.0019. A 95% confidence interval for the coefficient is (0.008, 0.016). Qualitatively identical findings are produced using a generalized linear model which recognizes that Buchanan vote share must lie within the unit interval. For this latter model, the estimated coefficient on Gore’s vote share was 0.712 with an estimated standard error of 0.039. The coefficients for the general linear model cannot be straightforwardly interpreted like linear regression coefficients, but it is nonetheless true that the generalized linear model strongly corroborates the finding that Gore vote share was positively correlated with Buchanan vote share.

¹²Leon County data are available from <http://www.co.leon.fl.us/elect/homepage.htm>. We include Leon County in our analysis because it was the only county besides Palm Beach County for which we could obtain precinct-level data. As data from other counties become available, we will incorporate them into our analysis.

¹³Estimation of generalized linear model, probit link, where Buchanan vote share is regressed on Gore vote share yields a slope coefficient estimate of -0.435 with an estimated standard error of 0.150.

in the Florida Senate election. Then, for each precinct we define the fraction of non-counted ballots as the total number of Senate election votes minus the total number of Presidential election votes, all divided by the total number of Senate election votes. We plan to update our analysis and use actual ballot figures once we know the total number of valid and invalid ballots cast in each precinct.¹⁴

There are two competing explanations for the causes and effects of non-counted ballots in Palm Beach County.

Claim 1: No Effect Non-counted ballots occur in many elections and generally inflict equal damage on both candidates. Thus, they did not influence the election outcome in Palm Beach County.

Claim 2: Harm to Gore Non-counted ballots in Palm Beach County were primarily submitted by Democrats who first punched the hole for Buchanan and then punched the hole for Gore. Thus, non-counted ballots did influence the electoral outcome.

Because we do not have access to individual ballots, we cannot directly determine how individual voters behaved in the voting booth. However, the above claims can be assessed by looking at precinct-level data. Under the *No Effect* claim, the number of non-counted ballots should be unrelated to the number of Gore voters in a precinct. Under the *Harm to Gore* claim, in contrast, precincts with large numbers of Gore voters should have large numbers of non-counted ballots.

The evidence strongly supports the *Harm to Gore* claim. Specifically, our precinct-level statistical analysis of Palm Beach County found a strong positive correlation between the number of non-counted ballots and the number of Gore voters.¹⁵ We also found a qualitatively similar relationship between the fraction of non-counted ballots and support for the Democratic Florida Senate candidate Bill Nelson. This implies that Palm Beach County precincts that were pro-Democratic, as measured by support for Nelson, were also disproportionately likely to cast ballots that did not include presidential votes.

To assess whether this pattern is due to the ballot structure in Palm Beach County, we conducted a similar analysis for the precincts in Florida’s Leon County. In Leon County we found no meaningful relationship between the number of non-counted ballots and the number of Gore voters in a precinct.¹⁶ Similarly, there was no meaningful relationship between the extent to which a Leon County precinct supported Bill Nelson and the extent to which its ballots lacked valid presidential votes. This is an extremely important point, it is captured in Figure 4, and it implies that there was something unique to the Palm Beach County ballot and that this unique feature had deleterious effects on Gore votes.

¹⁴To conduct our analysis with the Senate vote as a proxy for total votes, we need to make the assumption that the fraction of all voters who cast a vote in the Senate election did not systematically vary across Palm Beach and Leon precincts.

¹⁵For a precinct-level regression of the fraction of non-counted ballots on the percentage of Gore votes, the estimated slope coefficient is 0.137 with a heteroskedastic-consistent estimated standard error of 0.031. As in the previous section, the results of our analysis remain unchanged if we analyze only non-absentee data points. Also, we find that among absentee precincts considered separately there was no correlation between non-counted presidential votes and support for Gore.

¹⁶In Leon County the estimated slope was -0.00467 with an heteroskedastic-consistent estimated standard error of 0.00744.

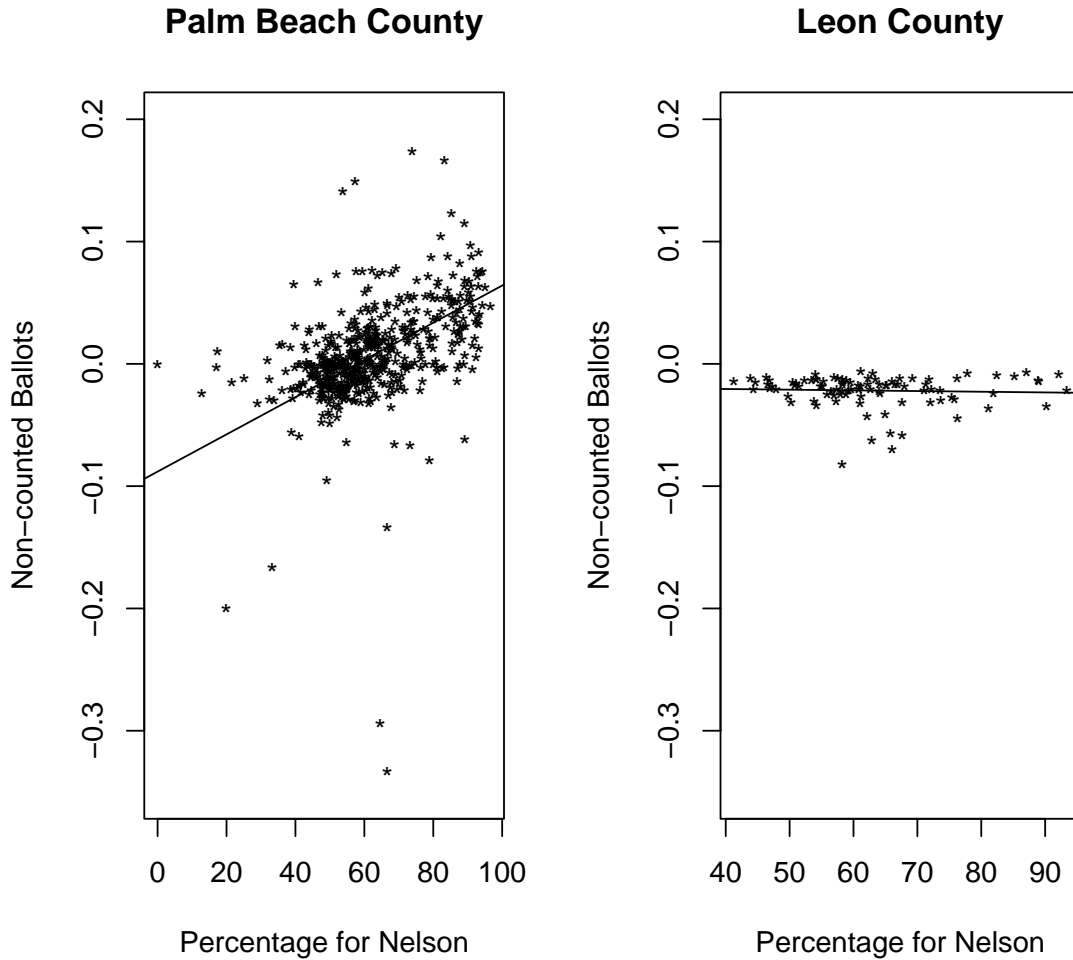


Figure 4: Trends in Non-counted Ballots and Support for Bill Nelson

One other feature of the precinct level data is worth noting. In Palm Beach County, more than half of the precincts had *more* votes counted for Senate than for President. This is very unusual since people normally vote for President and then “roll off” by failing to vote in elections that require reading farther down the ballot.¹⁷ Leon County was much more typical in terms of its roll off behavior. In all 95 precincts there were more votes for President than for Senate. This point supports our claim that the ballot structure in Palm Beach County was uniquely problematic.¹⁸

¹⁷For a discussion of roll off, see Walter Dean Burham, “The Changing Shape of the American Political Universe,” *American Political Science Review* 59(1): 7–28, 1965

¹⁸All data analysis in this paragraph ignores absentee ballots.

6 Buchanan in 1996 and 2000

The final component of our analysis is a historical comparison of Buchanan’s performance in Palm Beach County. In the 2000 presidential election, Palm Beach County supplied 19.6% of Buchanan’s votes in the state of Florida. If this total reflects strong Buchanan support in Palm Beach, then this effect should have been evident when Buchanan competed there in the 1996 Republican presidential primary. However, in the 1996 primary (which we assume did not use the controversial “Butterfly Ballot”) only 5.4% of Buchanan’s Florida votes came from Palm Beach County. This strongly suggests that the ballot Palm Beach County used in the 2000 presidential election was a major factor contributing to the number of votes Buchanan received.

Some commentators have argued that the fact that Buchanan received 8,788 votes in Palm Beach County in 1996 indicates that it is reasonable to believe that he could receive 3,407 votes there in 2000. However, this argument ignores the fact that in 2000 Buchanan received only 17,356 votes in Florida whereas in 1996 he received 162,713 votes in Florida. Given the dramatic differences in these statewide totals, it is necessary to compare Buchanan’s percentages rather than his raw numbers in the two elections.

7 A Methodological Caveat

Having presented evidence in support of irregularity in Palm Beach County voting, we now offer a methodological caveat. Statistical analysis of voting data is a very complicated subject. We, like all other researchers who have studied voting behavior in Palm Beach County, only possess a very limited amount of information.¹⁹ In particular, we have data only at the level of counties, townships or precincts—what is known as aggregate data. Our conclusions are based on the aggregate totals of votes in reporting units and not on information about each individual’s choices and each person’s characteristics. Using aggregate data to draw conclusions about the behavior of voters requires facing what is called an “ecological inference problem,” one of the classic statistical problems in political research (Achen and Shively, 1995).

The best way to analyze voting in Palm Beach County would be to obtain individual ballots from Palm Beach County along with ballots from other counties and townships in the United States. One could then use voting patterns from non-Palm Beach County ballots in conjunction with well-founded statistical methods to estimate how individuals would have cast presidential votes in Palm Beach County. If these estimated votes did not match actual votes in Palm Beach County, then one would have direct, individual-level evidence of voting irregularities.

Unfortunately, at the time of this writing we do not have access to individual Palm Beach County ballots and must rely on data that are currently available, i.e., we must rely solely on aggregate data. This limitation restricts us, and other Palm Beach County researchers as well, to simple and arguably problematic statistical models. Nonetheless, we believe that

¹⁹See <http://socrates.berkeley.edu/~ucdtpums>, <http://madison.hss.cmu.edu/>, and <http://www.econ.jhu.edu/people/ccarroll/carroll.html> for three analyses of Palm Beach County voting.

our results are of great relevance given the issues and debates at hand, and we have tried to present the most thorough analysis possible given available data. The fact that we obtain similar results using several different statistical methodologies significantly strengthens our confidence in our findings. We and other researchers must continue to reanalyze Palm Beach County voting as better data become available.

8 Conclusion

This paper considers whether and why Palm Beach County, Florida, produced an abnormally large vote share for Pat Buchanan in the 2000 Presidential Election. The evidence we have suggests that, indeed, Palm Beach County was exceptional in its support for Buchanan and exceptional in the sense that the residents of Palm Beach County appear to have voted in a way that is very hard to reconcile with the nature of the county. Furthermore, evidence indicates that Democratic areas of Palm Beach County were disproportionately likely to have generated invalid presidential ballots and that this decreased the vote share of Al Gore.

A key aspect of this paper's analysis is that it draws on many different levels of data: county and district data across the United States, county data within Florida, and precinct data from two Florida counties. And, importantly, all of our analyses testify to the same general result. Given the weakness of the data currently available on Palm Beach County, the fact that all of our analyses testify to the exceptional nature of Palm Beach County suggests that our findings are reasonably robust.

9 Appendix: National Analysis Methodology

Our model is a generalized linear model (GLM) (McCullagh and Nelder, 1989) of the binomial family with a logistic link, allowing for overdispersion. The dependent variable is the proportion of the presidential vote in reporting unit i that was cast for Buchanan, denoted $P_{\text{Buchanan},i}$, out of the total number of votes cast for Browne, Buchanan, Bush, Gore, Hagelin, Nader, Phillips (when the candidate appears on the ballot). Let $N_{\text{Buchanan},i}$ denote the number of votes for Buchanan and let N_i denote the total number of votes cast for either Buchanan, Bush, Gore or Nader in reporting unit i . The proportion we study is $P_{\text{Buchanan},i} = N_{\text{Buchanan},i}/N_i$. We base the GLM's linear predictor, denoted μ_i , on the proportions of the vote in reporting unit i that were cast for Bush and for Nader, denoted respectively $P_{\text{Bush},i}$ and $P_{\text{Nader},i}$. The linear predictor is defined as

$$\mu_i = \hat{\beta}_0 + \hat{\beta}_1 P_{\text{Bush},i} + \hat{\beta}_2 P_{\text{Nader},i} , \quad (1)$$

where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated coefficient values. The estimate for the proportion of the vote for Buchanan in reporting unit i , based on the model, is

$$\hat{P}_{\text{Buchanan},i} = \frac{\exp(\mu_i)}{1 + \exp(\mu_i)} . \quad (2)$$

We are interested in the discrepancy between the actual number of votes for Buchanan in reporting unit i ($N_{\text{Buchanan},i}$) and the predicted number of votes, denoted $\hat{N}_{\text{Buchanan},i} =$

$N_i \hat{P}_{\text{Buchanan},i}$. The simplest measure of that discrepancy is the simple residual defined by

$$r_i = N_{\text{Buchanan},i} - \hat{N}_{\text{Buchanan},i} . \quad (3)$$

A value of r_i that is much larger for reporting unit i than for other reporting units would indicate that the excess of the actual vote for Buchanan over the expected vote is much larger in unit i than it is in other areas.

A problem with the simple residuals is that, in a sense, the size of residual that we should expect to occur depends on the size of the support for Buchanan that the model predicts. As the size of the expected proportion $\hat{P}_{\text{Buchanan},i}$ increases from zero toward 0.5, the chances of observing a larger residual increases. This may be a real problem where the main question is whether support for Buchanan in a particular reporting unit is excessively large. The residual for a reporting unit may be large relative to the residuals for other reporting units merely because the expected support for Buchanan is truly larger among the voters in that reporting unit. If one determines whether Buchanan vote in an area is excessively large by using a test based on simple residuals, the resulting test results will be biased in the sense of tending to find such excesses when they do not really exist.²⁰

It is important to understand how this phenomenon occurs. The reason one expects to see larger residuals when the baseline support for Buchanan is truly bigger is that as the baseline proportion of votes for Buchanan increases from zero up to 0.5, the variance of the actual proportion of votes around the baseline expected value increases. This means that for any particular “large” size for a possible residual that one might specify (within the range zero to $N_i/2$), the chances of seeing a residual as large as that size increase as the baseline proportion increases. If $\hat{P}_{\text{Buchanan},i}$ is the baseline expected value and one analyzes the vote for Buchanan while treating the total number of votes N_i as a fixed quantity (known as conditioning on the total), then the variance of $N_{\text{Buchanan},i}$ is

$$\text{var}(N_{\text{Buchanan},i}) = \hat{\sigma}^2 N_i \hat{P}_{\text{Buchanan},i} (1 - \hat{P}_{\text{Buchanan},i}) . \quad (4)$$

So the variation of $N_{\text{Buchanan},i}$ around the expected value $\hat{N}_{\text{Buchanan},i}$ increases as $\hat{P}_{\text{Buchanan},i}$ increases, as long as $\hat{P}_{\text{Buchanan},i}$ is less than 0.5. This result follows from assuming that the number of votes for Buchanan in reporting unit i , given N_i , is a binomial random variable with probability $\hat{P}_{\text{Buchanan},i}$, with overdispersion that is approximated in the GLM by the estimated value $\hat{\sigma}^2$.

To make the discrepancies from different reporting units comparable to one another it is necessary to eliminate the variations that stem from the heteroscedasticity (differing variances) among the observed votes for Buchanan. The way to do that is to divide each simple residual by the square root of the variance $\text{var}(N_{\text{Buchanan},i})$. In this way we compute what’s known as the studentized residual, s_i :

$$\begin{aligned} s_i &= r_i / \sqrt{\text{var}(N_{\text{Buchanan},i})} \\ &= \frac{N_{\text{Buchanan},i} - \hat{N}_{\text{Buchanan},i}}{[\hat{\sigma}^2 N_i \hat{P}_{\text{Buchanan},i} (1 - \hat{P}_{\text{Buchanan},i})]^{1/2}} . \end{aligned} \quad (5)$$

²⁰Greg Adams’s analysis (<http://madison.hss.cmu.edu>) amounts to a demonstration that the simple residual (from a linear regression model) for Palm Beach County is large relative to the residuals for other counties in Florida.

If the model we use to compute $\hat{P}_{\text{Buchanan},i}$ correctly approximates the process that generates the vote for Buchanan in each and every reporting unit, then the chances of observing a studentized residual of any particular size are the same for all reporting units. There is no longer a built-in bias which makes the observed discrepancies tend to have larger magnitudes whenever the baseline expected support for Buchanan is larger. If the studentized residual is much larger for one reporting unit than it is for other reporting units, then we can have confidence that the votes for Buchanan in the unusual reporting unit were generated by a process substantially different from what went on in the other units.

For each state we estimate a separate set of parameter values $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ of equation (1) and overdispersion value $\hat{\sigma}$. The studentized residuals are comparable across the reporting units from each state and also across states.

To implement a more powerful assessment of the discrepancy for each reporting unit, we use a jackknife method: the parameter values used to compute the residual for reporting unit i are estimated using the data from all the reporting units in the same state as i but omitting the data for i . The histogram in Figure 1 shows the jackknife studentized residuals from counties in Florida. The histogram in Figure 2 pools such residuals from all 46 states for which the model of equation (1) could be estimated.

References

- Achen, Christopher H. and W. Phillips Shively. 1995. *Cross-Level Inference*. Chicago: University of Chicago Press.
- McCullagh, Peter and John A. Nelder. 1989. *Generalized Linear Models*. London: Chapman Hall, second edition.