The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida

JONATHAN N. WAND *Cornell University*
KENNETH W. SHOTTS *Northwestern University*
JASJEET S. SEKHON *Harvard University*
WALTER R. MEBANE, JR. *Cornell University*
MICHAEL C. HERRON *Northwestern University*
HENRY E. BRADY *University of California, Berkeley*

We show that the butterfly ballot used in Palm Beach County, Florida, in the 2000 presidential election caused more than 2,000 Democratic voters to vote by mistake for Reform candidate Pat Buchanan, a number larger than George W. Bush's certified margin of victory in Florida. We use multiple methods and several kinds of data to rule out alternative explanations for the votes Buchanan received in Palm Beach County. Among 3,053 U.S. counties where Buchanan was on the ballot, Palm Beach County has the most anomalous excess of votes for him. In Palm Beach County, Buchanan's proportion of the vote on election-day ballots is four times larger than his proportion on absentee (nonbutterfly) ballots, but Buchanan's proportion does not differ significantly between election-day and absentee ballots in any other Florida county. Unlike other Reform candidates in Palm Beach County, Buchanan tended to receive election-day votes in Democratic precincts and from individuals who voted for the Democratic U.S. Senate candidate. Robust estimation of overdispersed binomial regression models underpins much of the analysis.

Beginning on election day November 7, 2000, Palm Beach County (PBC), Florida, attracted national and eventually international attention because thousands of voters in the county complained that they had difficulty understanding the now infamous butterfly ballot. As a result, they claimed that they had cast invalid or erroneous presidential votes. Lawyers working for the Democratic Party reportedly collected 10,000 affidavits sworn by voters with complaints about some aspect of their election-day experiences in the county (Associated Press 2000b; Firestone 2000a, 2000b; Van Natta 2000; Van Natta and Moss 2000). Shortly after election day, eleven groups of PBC voters filed independent lawsuits seeking relief, claiming they and others had made mistakes in their votes for president because of the confusing format of the ballot.[1] Many of them stated that they had intended to vote for Democratic candidate Al Gore but by mistake chose Reform candidate Pat Buchanan. The number of votes involved was more than enough to have tipped the presidential vote in Florida from Republican candidate George W. Bush to Gore, thus giving him Florida's 25 electoral votes and the presidency.[2]

PBC is a heavily Democratic, politically liberal county that conventional wisdom says should provide few Buchanan votes. Two days after the election, Bay Buchanan, Pat Buchanan's sister and campaign manager, said "she was startled to hear Bush strategist Karl Rove argue Thursday that Buchanan has strong support in a county where his campaign never bought an ad and never paid a visit" (Garvey 2000).[3] Yet, the

Jonathan N. Wand is a doctoral candidate, Department of Government, Cornell University; Lamarck Inc., 171 Newbury Street, 5th Floor, Boston, MA 02116 (jnw4@cornell.edu). Kenneth W. Shotts is Assistant Professor of Political Science, Northwestern University, Evanston, IL 60208-1006 (k-shotts@northwestern.edu). Jasjeet S. Sekhon is Assistant Professor of Government, Harvard University, 34 Kirkland Street, Cambridge, MA 02138 (jsekhon@fas.harvard.edu). Walter R. Mebane, Jr., is Associate Professor of Government, Cornell University, Ithaca, NY 14853-4601 (wrm1@cornell.edu). Michael C. Herron is Assistant Professor of Political Science, Northwestern University, Evanston, IL 60208-1006 (m-herron@northwestern.edu). Henry E. Brady is Professor of Political Science, University of California, Berkeley, CA 94720-5100 (hbrady@bravo.berkeley.edu).

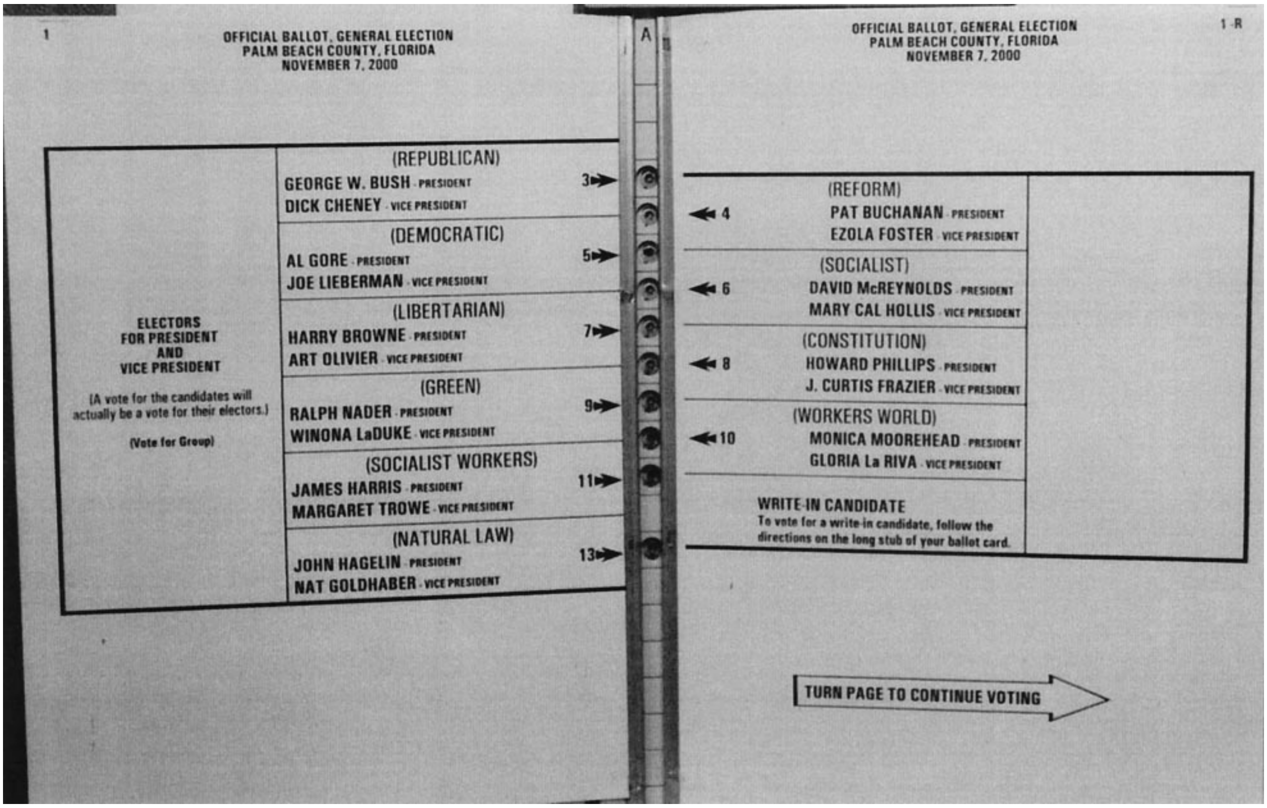
Herron worked on this paper while a Postdoctoral Research Fellow at the Center for Basic Research in the Social Sciences, Harvard University. Previous versions were presented at the 2001 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Atlanta, GA, August 5–9, and at the 2001 annual meeting of the Public Choice Society, San Antonio, TX, March 9–11. The authors, who are listed in reverse alphabetical order, served as unpaid expert witnesses in Case CL 00-10992AF in the Fifteenth Judicial Circuit of Florida, West Palm Beach: *Beverly Rogers and Ray Kaplan* v. *The Palm Beach County Elections Canvassing Commission; et al.* (see Brady et al. 2001). The authors thank Todd Rice and Lamarck, Inc., for their generous support and provision of computing resources, Kevin Matthews of the Department of Geography, George Mason University, for mapmaking services, Benjamin Bishin and Laurel Elms for research assistance, and seminar participants at Columbia University, Cornell University, Harvard University, and Northwestern University for helpful comments.

[1] The cases filed in the Fifteenth Judicial Circuit of Florida, West Palm Beach, were CL 00-10965, CL 00-10970, CL 00-10988AE, CL 00-109922AF, CL 00-11000AH, CL 00-11084AH, CL 00-11098AO, CL 00-1146AB, CL 00-1240AB, CL 00-129OAB, and CL 00-113O2AO. These were consolidated by Administrative Order No. 2.061-11/00. Texts of the filings and of the Fifteenth Circuit Court rulings in the cases are available from http://www.pbcountyclerk.com/.

[2] Bush received 271 electoral votes, one more than needed to win, and Gore received 266. One Elector pledged to Gore from Washington, DC, left her ballot blank, reducing Gore's count from 267 (Mitchell 2001; Stout 2000).

[3] The story further observes that "longtime Reform members in the state described a party in 'disarray' with little organization, much less

FIGURE 1. The Palm Beach County Bufferfly Ballot



Source: AP Worldwide Photos, Gary I. Rothstein. Reprinted with permission.

county supplied 19.6% of Buchanan's votes in Florida. In contrast, only 5.4% of his Florida votes came from PBC in the 1996 Republican presidential primary, which did not use a butterfly ballot.[4]

The butterfly ballot, shown in Figure 1, was an innovation of Theresa LePore, Supervisor of Elections for PBC.[5] The distinctive format was used only in PBC and only for election-day ballots for president. It is a "butterfly" because two columns of names of candidates (the wings of the butterfly), all for the same office, sandwich a single column of punch holes between the names. These punch holes are alternately for the left-hand and right-hand side of the ballot. Thus, the first valid punch hole (identified on the ballot as #3) is for Bush, the first candidate on the left-hand side. The second valid punch hole (identified on the ballot as #4) is for Buchanan, the first candidate on the right-hand side. On the left, however, the second candidate listed is Gore, and someone who scanned down the left-hand column without looking to the right could mistakenly conclude that the first two punch holes corresponded, respectively, to Bush and Gore. Having made such an incorrect reading, a Bush voter would still be likely to punch the first hole, but a Gore voter might mistakenly punch the second and vote for Buchanan.

Sinclair et al. (2000) report experimental evidence that a double-column ballot format like the one used in PBC can be more confusing and cause more voter errors than a single-column ballot. Other published research on the effects of ballot design is scarce and does not provide much guidance regarding the errors the PBC butterfly ballot may have induced (Campbell and Miller 1957; Darcy 1986; Hamilton and Ladd 1996).

Did the butterfly ballot cost Gore the presidency? The lawsuits filed by citizens of PBC were thrown out because the Supreme Court of Florida ruled that the ballot was not illegal,[6] but the ruling neither depended upon nor implied anything about the ballot's effect on

a groundswell of support for Buchanan in a place even he concedes is not his base."

[4] In 2000, Buchanan received 0.787% of the presidential vote in PBC while garnering only 0.3% of the overall Florida presidential vote. In contrast, Ross Perot, the Reform candidate for president in 1996, received only 7.7% of the PBC vote and garnered 9.1% of the Florida vote. These data are from the Florida Department of State.

[5] Reportedly, LePore "split the names over two pages to make the type larger." Two days after the election she was quoted as saying: "Hindsight is 20-20, but I'll never do it again" (Engelhardt 2000). Merzer and *Miami Herald* (2001) describe how LePore went about designing the ballot and many other defects in the administration of the election in Florida.

[6] The court stated: "Even accepting appellants' allegations, we conclude as a matter of law that the Palm Beach County ballot does not constitute substantial noncompliance with the statutory requirements mandating the voiding of the election" (*Supreme Court of Florida, Fladell, et al.* v. *Palm Beach County Canvassing Board, etc. et al.* Case Nos. SC 00-2373 and SC 00-2376). The cases did not progress to hearings regarding the facts.

voter behavior. The court did not rule on whether Gore lost the election because of the ballot. Our goal is to investigate that question by determining whether the butterfly ballot caused several thousand Gore supporters to vote mistakenly for Buchanan.

OVERVIEW

Among scholars who posted an analysis on the Internet shortly after the election, a consensus quickly formed that the vote for Buchanan in PBC was anomalously large. Of the 3,407 votes that Buchanan received in the initial, uncertified count of PBC ballots, a typical estimate was that he received about 2,800 more votes than were to be expected based on voting patterns elsewhere in Florida.[7] The number of apparently accidental votes for Buchanan exceeded Bush's official margin of victory in Florida, which was only 537 votes more than Gore.[8] But the early analysis did not show unambiguously that the butterfly ballot was the cause or that the erroneous Buchanan votes would otherwise have gone to Gore.

To test whether Democratic voters mistakenly voted for Buchanan because of the butterfly ballot, we use multiple methods and diverse data sources to rule out alternative explanations for the Buchanan vote in PBC. First, we show that Buchanan's vote was anomalously high on election day. Specifically, we prove three key facts. (1) Anomalies in the Buchanan vote as large as the one in PBC did not occur in any other county in the country in 2000. (2) The vote for the Reform candidate in the previous presidential election, Ross Perot, was not anomalous in PBC in 1996. (3) PBC voters who used election-day ballots (the butterfly format) recorded unusually high support for Buchanan, while those who used absentee ballots (no butterfly format) evinced the expected level of support for him.

Second, we show that Buchanan's excess of support was almost entirely from Democratic voters. Here we demonstrate two key facts. (1) The unusually high level of support for Buchanan was concentrated in precincts with high levels of support for Democratic candidates for other offices and not in precincts with high levels of support for Reform candidates for other offices. (2) Individual ballots confirm these patterns: Democratic voters (as measured by votes in the U.S. Senate election) who voted on election day were much more likely to support Buchanan than were Democratic voters who used absentee ballots. We conclude that the butterfly ballot caused at least 2,000 Democratic voters to vote mistakenly for Buchanan.

Our conclusions depend upon repeated comparisons: across counties, between election-day ballots and absentee ballots, and across precincts. To determine whether the Buchanan vote is anomalous, we compare the actual number of votes received by Buchanan in a county to the votes predicted by a well-specified statistical model that uses past votes and demographic characteristics of all the counties in the same state to form estimates. When actual votes deviate significantly from what is expected, we have what statisticians call an outlier. Then, to isolate sharply the effect of the butterfly format, we compare Buchanan's share of the votes on election day to his share on absentee ballots across all Florida counties. Next, to determine whether Buchanan's votes in PBC were concentrated in areas with high levels of support for Democratic candidates, we compare how well Democratic strength in a precinct predicts Buchanan votes versus votes for Reform candidates for other offices. Finally, we use a comparison between Buchanan votes on individual election-day (butterfly) and absentee (nonbutterfly) ballots to compute our minimum estimate for the number of Democratic voters the butterfly ballot caused to vote mistakenly for Buchanan instead of Gore.

A simple way to estimate the expected votes for Buchanan is to take the percentage of Buchanan voters in each county and regress it, using ordinary least squares (OLS), on past votes and demographic characteristics. Then an examination of the difference between the actual and expected vote might tell us whether the county was anomalous. This procedure might be misleading, however, because of three major statistical problems. First, the counties differ in size—the smallest have only a few hundred voters, while the largest have millions—leading to the problem of heteroskedasticity. Second, heteroskedasticity is exacerbated because the expected proportion of votes for Buchanan varies over counties. Third, because OLS is notoriously subject to outliers, any OLS-based estimates designed to detect them are suspect. All three problems must be solved in order to get reliable results.

To understand why differing county size is a problem, consider a hypothetical example. Suppose there are two counties, in both of which Buchanan is expected to receive 1% of the vote. One county has 100 voters, so the expected Buchanan vote is one, and the other has 100,000 voters, so the expected vote for Buchanan is 1,000. In the small county a 2% excess of votes over the expected value corresponds to an observed total of three Buchanan votes: $(3 - 1)/100 \times 100\% = 2\%$. In the large county an observed vote of 3,000 for Buchanan would be required to produce the same 2% discrepancy. But two "extra" Buchanan votes by chance in the small county are much more likely than an extra 2,000 in the large county. For instance, using a simple binomial model for the count of votes for Buchanan, the z-score for the discrepancy is $(3 - 1)/\sqrt{100 \times .01 \times .99} = 2$ in the small county but $(3000 - 1000)/\sqrt{100000 \times .01 \times .99} = 64$ in the large county.

To avoid the heteroskedasticity problem, we use a generalization of the standard binomial model to esti-

[7] In Brady et al. (2001) we list the early posters, including ourselves. Brady posted analysis on November 9, and Wand, Shotts, Sekhon, Mebane, and Herron posted on November 11. Lists of empirical work posted on the Internet through the end of November 2000, appear at http://www.bestbookmarks.com/election (created by Jonathan O'Keeffe), http://www.sbgo.com/election.htm (created by Sebago Associates), and http://madison.hss.cmu.edu (created by Greg Adams and Chris Fastnow).

[8] The final, certified results gave Bush 2,912,790 votes and Gore 2,912,253 votes in Florida. A few days after the election, the Associated Press reported a vote margin of 327 based on the initial, automatic recount across Florida (Wakin 2000).

mate expected vote counts, and we use statistics that adjust for the counts' variances when making comparisons across counties. In the model the probability of a vote is a function of past votes and demographic factors. The variance of the vote count can be greater than in a standard binomial model. The resulting model is called an overdispersed binomial model. The statistics we compare across counties are studentized residuals (Carroll and Ruppert 1988, 31–4), which are defined as the simple difference between observed vote count and expected vote count, divided by the estimated standard deviation of the observed count.

The overdispersed binomial regression model treats the count of votes for Buchanan in a particular geographic area as having the mean and variance of a binomial random variable, except that the variance is multiplied by a constant scale factor. As described in detail by McCullagh and Nelder (1989, 125), the scale factor may reflect a process in which each vote count is a sum of vote counts produced at lower levels of aggregation, where each lower-level aggregate is a binomial random variable. According to such a motivation for the model the binomial distribution that describes each lower-level count may have a distinct probability parameter (p. 125). In this way the overdispersed binomial model recognizes the heterogeneity among the clusters of voters whose choices comprise the vote counts we analyze. We describe the overdispersed binomial model in greater detail in Appendix A.

The most important methodological innovation in our analysis is that we develop new methods for robust estimation of the overdispersed binomial regression model. These guard against the possibility, which is very real with OLS estimation, that an outlier may destroy the estimation. Without robust methods, a single large outlier may mask other outliers (Atkinson 1986), which means that the distorted data appear to be the norm rather than the exceptions. Masking also may occur when there are several outliers that are similar in magnitude. For instance, in the simplest case of estimating the average for a set of numbers, it is easy to see that the sample mean can be greatly affected by a single exceptionally large value. Masking occurs when this one very large point so greatly alters the sample mean that other large but smaller points do not appear to be as discrepant as they truly are from the mean value that characterizes most of the sample data. Similarly, masking can occur when a few exceptionally large or small data points, all of about the same size, pull the sample mean toward themselves so that none of them are far from the sample mean value, although in fact they do not represent the same statistical distribution as the rest of the data. Masking makes it difficult to determine which data points really do deviate from what we should expect.

The primary reason to use robust estimators in the PBC situation is that the voter complaints, legal cases, and media reports strongly suggest that the electoral results there were produced by processes substantially different from the standard political factors (partisanship, liberalism-conservatism, policy positions) that cause voters to act predictably from one election to another and that produced the results elsewhere in Florida. We want our models to predict what would have happened without idiosyncratic factors such as a confusing ballot form. Significant departures from those predictions will indicate that idiosyncratic factors must be at work.

Data weakness is another reason for robust estimation. Because our models are based on the results from other elections, anomalies in those elections will give a distorted picture of the standard political factors that predict the vote in the current one. The robust estimators we use protect against the influence of such distortions. A county in which the previous election results are highly distorted will not affect the parameter estimates and, indeed, will itself appear to be an outlier. We describe our robust estimation methods in more detail in Appendix A.

BUCHANAN'S VOTE IN COUNTIES ACROSS THE UNITED STATES

To assess how excessive Buchanan's PBC vote appears to be when compared to outcomes across the country in 2000, we use the overdispersed binomial regression model to estimate the expected number of votes cast for Buchanan out of all votes cast for Buchanan, Gore, Bush, Ralph Nader (Green Party), Harry Browne (Libertarian Party), Howard Phillips (Constitution Party), John Hagelin (Natural Law Party), or any write-in candidates. We predict votes for Buchanan, given the total number of votes cast for all presidential candidates, in each U.S. county in 2000.[9] Two kinds of information, the results of previous elections and the demographic characteristics of the county, are available and highly relevant for making such predictions. The previous election outcome is a proxy not only for the array of interests and party sentiments in each county but also for the strength of local party mobilization. We use two variables to represent the preceding election result: the proportion of votes officially received by the Republican candidate in the 1996 presidential election; and the proportion of votes officially received by the Reform Party candidate in the 1996 presidential election. We supplement them with a set of nine demographic variables. Seven of the variables come from the 2000 Census of Population and Housing: the proportions of county population in each of four Census Bureau race categories (namely, white, black, Asian and Pacific Islander, and American Indian or Alaska Native); 2000 proportion Hispanic; 2000 population density (computed as 2000 population/1990 square miles); and 2000 population.[10] The eighth and ninth are the 1990 proportion of population with a college degree and 1989 median household money

[9] We ignore undervotes (no apparent vote recorded on the ballot), overvotes (votes for more than one presidential candidate on a single ballot), and other spoiled ballots. For discussions of undervotes and overvotes in PBC see Engelhardt and McCabe 2001a, 2001b.

[10] The 2000 Census data were built from Census 2000 Redistricting Data (Public Law 94–171) Summary File, Matrices PL1, PL2, PL3, and PL4 (U.S. Census Bureau, 2001, "FactFinder Tables," accessed April 7, 2001, at http://factfinder.census.gov).

income. We do not use the number of voters registered as Reform Party members because voter registration is nonpartisan in many states.

We cannot use all eleven variables (plus a constant) in our models at once, because we estimate the model separately for each state (except for four states that we pool because each has only a few counties; see below), and with that many variables a state would need to have about 40 to 60 counties to produce reliable estimates. But 21 states (including the District of Columbia) have fewer than 50 counties. Therefore, we include the two vote proportion variables in all the models and use principal components of the demographic data. To maximize the efficiency of the information gained from the demographic variables, we compute principal components of the set of residuals obtained by regressing each demographic variable on the previous election proportion variables and the constant.[11] The regressions on the election variables and the computation of principal components are done separately for each state.

For the results of the county-level analysis that we report in detail here, we use only the first principal component of the Census data. As we explain below, using more principal components does not substantively change the results that bear on PBC. Hence, the expected vote for Buchanan in county i in the given state is based on a linear predictor defined by:

$$x_i'\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i},$$

where x_{1i} is the 1996 proportion of votes received by the Republican candidate, $x_{2\,i}$ is the 1996 proportion of votes received by the Reform candidate, x_{3i} is the principal component, and β_0, β_1, β_2, and β_3 are constant coefficients.

We compare PBC to the 3,053 counties in the United States for which we can robustly estimate the overdispersed binomial model and hence compute studentized residuals. Because they have too few counties to analyze separately, we pool the data for Connecticut, Delaware, Hawaii, and Rhode Island (which have, respectively, eight, three, four, and five counties), using dummy variables to give each state a different intercept but requiring the other coefficients to be the same for all four states.[12] The dependent variable is the count y_i of votes for Buchanan in each county i in the 2000 election. The studentized residuals are comparable across states.[13]

PBC has the largest residual[14] among the 3,053 counties in our analysis. This result can be seen in Figure 2, which presents boxplots that display the distribution of the residuals for each state.[15] The residual for PBC not only has the largest positive value but also is the largest in absolute magnitude. Buchanan received vastly more votes in PBC than predicted by the county's electoral history and demographic profile.

Appendix Table B-1 lists the residuals, expected Buchanan vote proportion, actual Buchanan vote proportion, and number of valid presidential ballots for all counties for which the residual is greater than 4.0 or less than −4.0, namely, all the outliers. There are 68 positive outliers and eight negative outliers. The difference between the number of positive and the number of negative outliers reflects the overall positive skew of the residuals that is visible for most states in Figure 2. Outliers occur in 31 of the 49 states covered by the analysis.

The outlier status of some counties is readily explained. Jasper County, South Carolina, which has the second largest residual in our analysis, did not receive much national media attention because the outcome was immaterial to the 2000 presidential contest. Bush defeated Gore in South Carolina by 220,376 votes, but only 6,469 presidential ballots were cast in Jasper County. Nonetheless, there were serious problems with a voting machine in the county's Tillman precinct, where Gore and Bush each received one vote, Buchanan 239 votes, and Nader 111 votes. The problems in the precinct affected vote totals for other offices. Indeed, "the State Board of Canvassers unanimously said [...] that problems in the county council election were so numerous that a new election should be held."[16]

PBC is not geographically contiguous to any other outlier, but many of the counties listed in Table B-1 are contiguous to another outlier. Table 1 displays the sets of counties from Table B-1 that are geographically contiguous. Sixteen of the twenty-five largest positive outliers belong to such a cluster. Two clusters include counties from two states (West Virginia and Ohio, Kansas and Missouri), and another includes counties from three states (South Dakota, Iowa, and Nebraska). Because these span state borders, it is highly unlikely that the exceptional support for Buchanan reflects problems of ballot format or electoral administration. Most likely the reason is special success in mobilizing voters for Buchanan in those areas.

Plausible explanations also can be produced for some of the remaining outliers, but we do not emphasize these because the residuals for all of those counties are much smaller than that for PBC, and some outliers are not stable over variations in the model specifications. In contrast, the size and relative position of the residuals are stable for counties such as PBC or Jasper, for which voting irregularities were reported, or Hancock County in West Virginia and Pottawatomie

[11] We standardize the residuals of the demographic variables to have variance equal to 1.0 before computing the principal components.

[12] With only one county, the District of Columbia cannot be analyzed. Our collection of counties (or equivalents) includes all states except Michigan, where Buchanan was not on the ballot and could receive only write-in votes. For Alaska we use the 25 county-equivalents defined by the U.S. Census Bureau for reporting 1990 Census data.

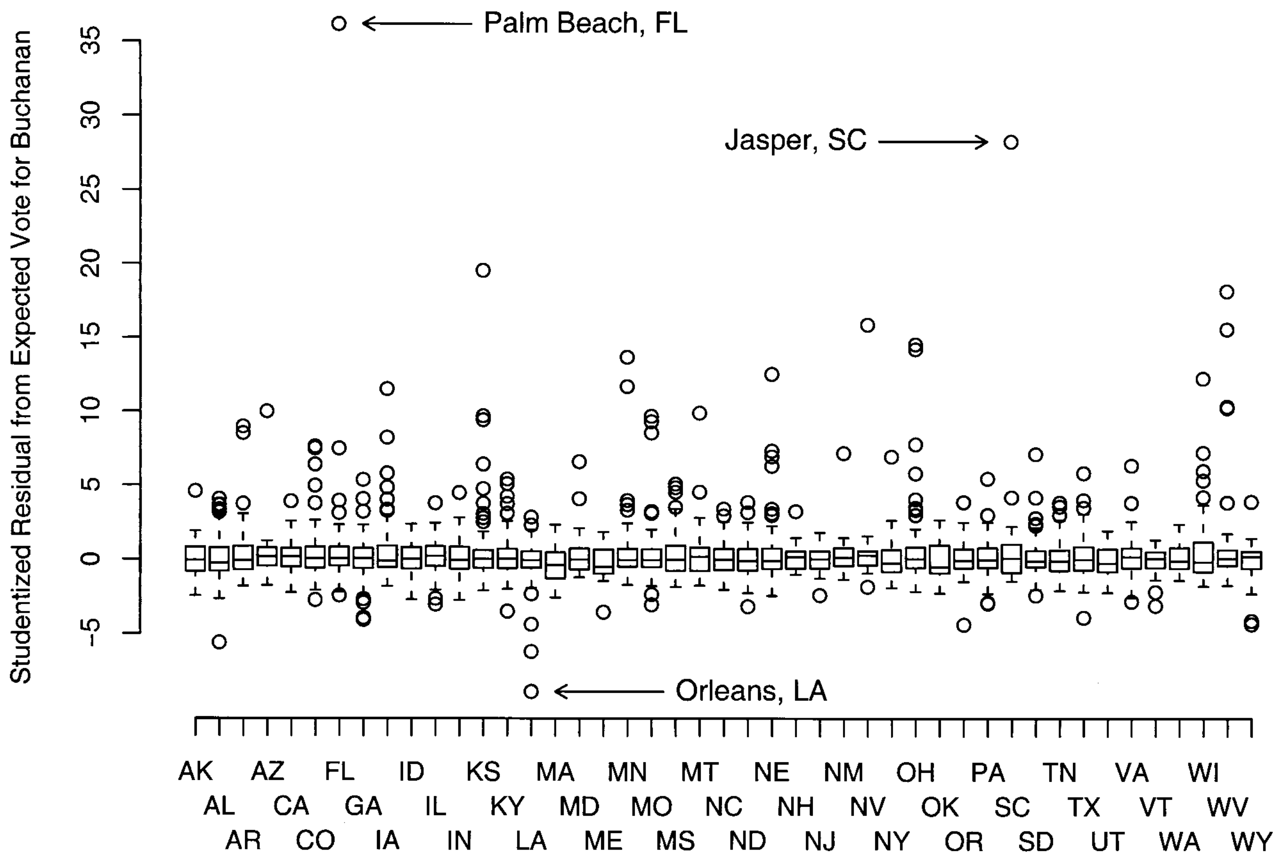
[13] We use the final, certified vote counts for each state. See Appendix B for data sources.

[14] Throughout the rest of this article the word *residual* always refers to the studentized residual, as defined in Appendix A, equation A-1, unless otherwise indicated.

[15] Whiskers extend to the nearest value not beyond 1.5 times the interquartile range. The residuals for the pooled states (CT, DE, HI, and RI) are omitted.

[16] See the December 28, 2000, issue of the *Beaufort Gazette* and Associated Press 2000a for allegations regarding the Tillman precinct.

FIGURE 2. Boxplots of Studentized Residuals in U.S. Counties, by State



County in Kansas, which are in contiguous clusters with unusually large numbers of Buchanan votes. PBC has the largest residual and Jasper County the second largest, whether we use one or two principal components to represent the demographic variables. With three principal components, PBC has the second largest residual (35.5) and Jasper the third largest (20.9). With no principal components (i.e., only past vote proportions), Jasper has the largest residual (25.6) and PBC the second largest (21.5).

No other county in Florida comes close to PBC in terms of excessive votes for Buchanan in the 2000 election. The only other outlier in the state is Pinellas County (see Table B-1). The parameter estimates for Florida give a point estimate of 438 for the number of votes expected for Buchanan in PBC, which implies an excess of 2,973 accidental votes in his certified tally of 3,411 votes.

Although PBC was an outlier in 2000, it is possible that support for the Reform Party is typically unusually high there. Because no butterfly ballot was used in 1996, an unexpectedly high number of Reform votes in the county that year compared to other Florida counties might support an alternative explanation for the 2000 result. Another possibility is that the 1996 Reform vote was exceptionally low in PBC, so the anomality of the 2000 vote could be exaggerated. In that case, the 2000 vote would appear to be excessive even if the county's Reform vote were simply returning to normalcy. The 1996 data support neither of these possibilities.

Using the overdispersed binomial model to analyze the votes received across the counties of Florida by Ross Perot, the Reform candidate in the 1996 presidential election, we find that PBC was not an outlier in 1996. We model the number of votes cast for Perot out of all votes cast either for Perot, Democrat Bill Clinton, or Republican Bob Dole. The regressors are defined to be earlier versions of the county-level variables we used to analyze the 2000 vote data: the proportion of votes officially received by the Republican candidate in the 1992 presidential election; the proportion of votes officially received by the Reform candidate in the 1992 presidential election; and earlier demographic data.[17] We use our robust estimators.

In 1996 no county in Florida has a residual even remotely as large as the one for PBC in 2000. In 1996 only St. Lucie County has a residual of absolute magnitude greater than 4.0 (−4.92). The largest positive residual is for Holmes County (2.30). PBC has the

[17] The race, Hispanic ethnicity, and population variables are taken from the 1990 Census.

TABLE 1. Contiguous Counties among Those with the Largest Studentized Residuals for the Buchanan Vote

State	County Name
OH	Hocking
OH	Athens
MD	Wicomico
MD	Somerset
SD	Union
IA	Sioux
NE	Dixon
NE	Thurston
NE	Dakota
IA	Plymouth
IA	Woodbury
MT	Deer Lodge
MT	Silver Bow
KS	Wabaunsee
KS	Shawnee
KS	Pottawatomie
CO	Arapahoe
CO	Jefferson
CO	Adams
KY	Kenton
KY	Boone
WI	Wood
WI	Lincoln
WI	Marathon
KS	Wyandotte
MO	Jackson
OH	Belmont
WV	Ohio
WV	Marshall
OH	Jefferson
WV	Brooke
WV	Hancock
MO	St. Louis City
MO	St. Louis
MN	Ramsey
MN	Hennepin

Note: Results based on 3,053 counties. This table presents all contiguous counties with studentized residuals of magnitude greater than or equal to 4.0.

seventh most negative residual (−1.86). PBC was not an outlier in 1996.

The 2000 vote for Buchanan in PBC was extremely unusual—clearly, among the most unusual in the entire country. The vast difference between the residual for PBC and the residuals for the other counties confirms that a large anomaly occurred relative to the vote predicted for Buchanan based on PBC's voting history and demographic profile.

A NATURAL EXPERIMENT: FLORIDA'S ELECTION-DAY AND ABSENTEE VOTERS IN 2000

Buchanan's vote total in PBC in the 2000 presidential election was anomalously large, but how can we be sure that the cause was the butterfly ballot? Because that format was not used for absentee ballots, the election gives us a natural experiment: One group of PBC voters (election day) used a butterfly ballot but a second group (absentee) did not. If Buchanan's vote total in the county reflects true support among the voters, then this support should be present in both pools of ballots. But if the butterfly ballot is responsible for Buchanan's vote then his support should come disproportionately from votes on election day.

A limitation of this natural experiment is that the mechanism that allocates voters to either the election-day pool or the absentee pool is not random assignment (Achen 1986). Voters self-select to be in the absentee group.[18] Some, such as military personnel stationed overseas, must cast absentee ballots. Absentee voters may not be representative of voters in general, but there are good reasons to believe that the influences on their voting behavior are similar across the counties. PBC's absentee voters are probably similar to those in at least some of the other counties, and we would expect their level of support for Buchanan to be similar as well. Therefore, if we take the difference between the proportion of those voting for Buchanan on election day and the proportion voting for him on absentee ballots, PBC's difference should cluster with the differences for other counties, unless something like the butterfly ballot has caused it to be distinctly different.

This difference analysis shows disproportionate support for Buchanan among election-day voters in PBC. Figure 3 plots the difference—election-day proportion minus absentee proportion—for all 67 counties of Florida by the number of presidential ballots cast in each county.[19] One can see that PBC has one of the largest differences in the state, although four counties have larger differences, and a few others have differences close in value to that of PBC.[20] But in all those counties the voting population is much smaller than in PBC, where 433,186 ballots were cast for president. In Calhoun, the largest county with a difference greater than PBC's, there were 5,174 ballots cast for president.

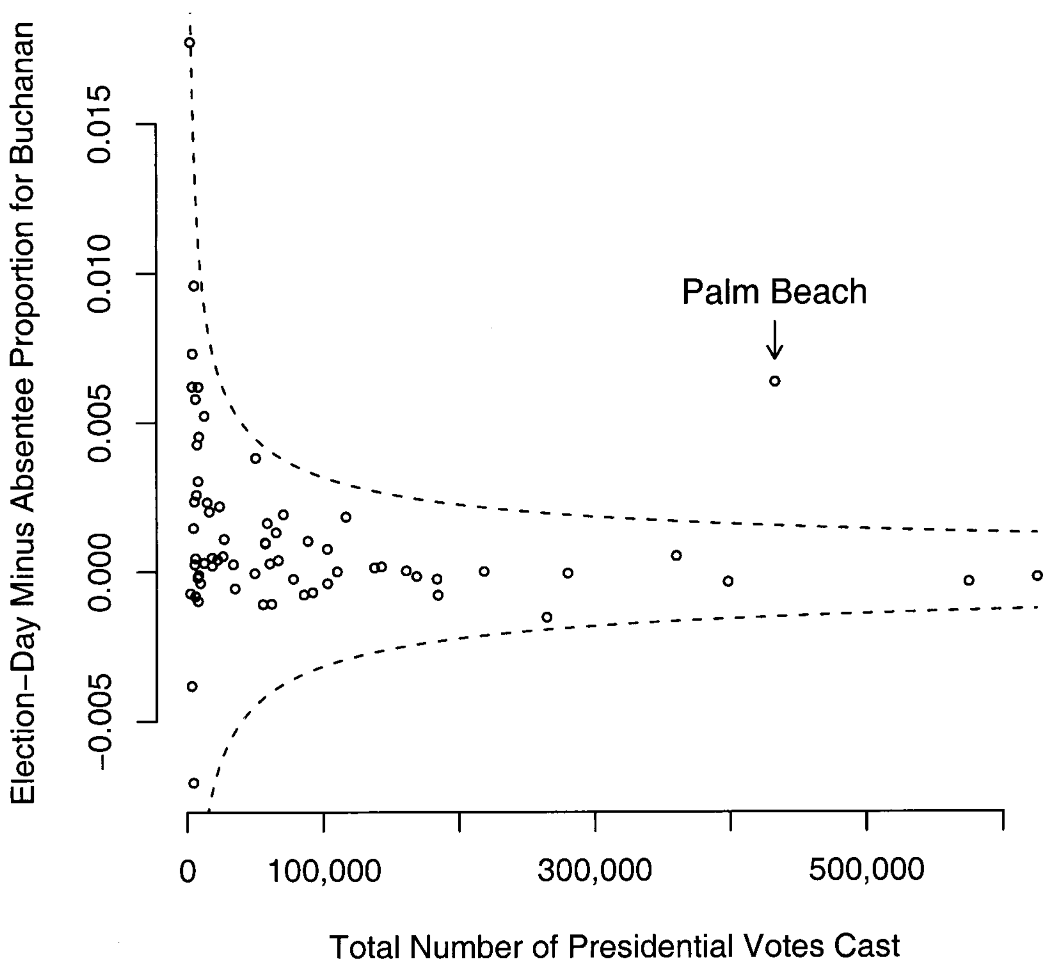
The significance of the population size disparity is that even if voting processes are identical in all the counties, there will be greater variability in the differ-

[18] According to 2000 Florida Statutes, title IX, chapter 101.62, a registered voter need not give any reason for requesting an absentee ballot.

[19] The data are based on certified numbers from the Florida Department of State and precinct-level returns provided by the 67 Florida counties. We used the precinct data to calculate the absentee returns.

[20] The difference in PBC is 0.00634 (433,186 presidential ballots), exceeded by Liberty (0.0177, 2,410 ballots), Calhoun (0.00959, 5,174 ballots), Hamilton (0.00731, 3,964 ballots), and Dixie (0.00706, 4,666 ballots). Union and Baker counties have the next largest differences after PBC, in both cases 0.00620. The numbers of presidential ballots cast in those counties were, respectively, 3,826 and 8,154.

FIGURE 3. Election-Day 2000 Minus Absentee Ballot Support for Buchanan in Florida Counties



ence in proportions in counties where smaller numbers of ballots are cast. With identical processes in all counties, the standard deviation of the difference in proportions will vary approximately in proportion to the reciprocal of the square root of the total number of ballots. If in all counties the proportions of election-day and absentee ballots cast for Buchanan do not systematically differ, then the observed differences between the proportions should be contained within bounds defined by the reciprocal of the square root of the number of ballots.[21] In Figure 3 the dashed lines are drawn at the values $\pm 1/\sqrt{m}$, where m denotes the number of ballots. One can see that only PBC falls outside those bounds. The difference between Buchanan's election-day and absentee proportions in PBC appears to be much greater than one would expect by chance. The differences for the other counties appear to fall within the range one would expect given only random deviations from equality in the processes that generate election-day and absentee votes.

The distorting effects of vastly different population sizes across counties may be fully corrected by explicitly setting up the comparison between election-day and absentee ballots as a test of the hypothesis that the proportion of votes for Buchanan among all the votes cast for president is equal for the two ballot formats. The simplest test is to determine whether the difference between Buchanan's proportion of the election-day ballots differs significantly from his proportion of the absentee ballots. We can do this using a straightforward modification of a difference of proportions test that appears in standard textbooks (Wonnacott and Wonnacott 1990, 275, eq. 8-29). To define the proportions, let A_i denote the total number of absentee votes cast for president in county i, let a_i denote the number of absentee votes cast for Buchanan, let B_i denote the total number of election-day votes cast for president, and let b_i denote the number cast for Buchanan. We again use an overdispersed binomial model for the

[21] If m is the (large) total number of ballots, q is the proportion of election-day ballots, and π is the probability of voting for Buchanan for both types of ballots, then the overdispersed binomial model with dispersion parameter σ^2 implies that the standard deviation of the difference between the observed proportions is approximately $\sigma[(1/m)(1/q + 1/(1 - q))\ \pi\ (1 - \pi)]^{1/2}$. Because $\pi \approx .003$ for Buchanan throughout Florida, containment within the bounds will fail only if σ is large or q is near 1.0—or if the probabilities for the two ballot types are not equal.

TABLE 2. Votes for Reform Candidates by Proportions Voting for U.S. Senate Candidates, for Palm Beach County Precincts

Scope	Reform Candidate	Intercept	Senate: Nelson (D)	Senate: Deckard (Ref)	$\hat{\sigma}$
All precincts	Buchanan	−6.17 (0.15)	2.06 (0.21)	−12.74 (14.79)	1.22
District 35	Buchanan	−7.48 (0.51)	3.85 (0.71)	13.13 (23.89)	1.26
District 35	Lowe	−1.98 (0.34)	−1.86 (0.51)	18.06 (14.46)	1.54
District 16	Buchanan	−7.00 (0.29)	3.32 (0.46)	3.95 (15.09)	1.15
District 16	McGuire	−3.37 (0.30)	−1.00 (0.50)	25.96 (9.10)	1.52

Note: Entries are tanh estimates of coefficient parameters of the overdispersed binomial regression model using precinct-level data from the 2000 election (standard errors are in parentheses). The last column reports the LQD dispersion estimate $\hat{\sigma}$. Number of precincts: all precincts, 515; District 35, 105; District 16, 149.

totals a_i and b_i, with statewide dispersion parameter σ^2. Because A_i and B_i are large, the equality hypothesis implies that the following z-score is normal with mean zero and unit variance:

$$z_i = \sigma^{-1}(b_i/B_i - a_i/A_i)$$

$$\left[\frac{(b_i/B_i)(1 - b_i/B_i)}{B_i} + \frac{(a_i/A_i)(1 - a_i/A_i)}{A_i}\right]^{-1/2}$$

If z_i is significantly greater than zero, then in proportional terms election-day voters cast ballots for Buchanan much more often than did absentee voters. If z_i is significantly less than zero, then support for Buchanan was disproportionately great among absentee voters. To compute z_i we need an estimate of σ. We use the estimate of the scale parameter that we obtained for the 2000 Florida data when we estimated the county-level overdispersed binomial model. That estimate is $\hat{\sigma} = 3.81$.

PBC has $z_i = 6.3$, a value more than six standard deviations away from the value of zero that is expected under the equality hypothesis. The next largest positive value among the remaining counties of Florida is 1.65, and 58 counties have a z-score of magnitude less than 1.0. Only one negative value, a z-score of -1.1 for Duval County, is more than one standard deviation away from zero.

In PBC, election-day votes went to Buchanan vastly more often, in proportional terms, than did absentee votes. The results of the election-day versus absentee ballot natural experiment strongly support the conclusion that Buchanan's anomalous support was caused by the butterfly ballot.

Underlying the dramatic test statistic value for PBC is the fact that county voters supported Buchanan on election day at a rate (.0085) approximately four times that of absentee voters (.0022). If election-day voters had cast ballots for Buchanan at the rate that absentee voters did, he would have received $387{,}356 \times 0.0022 = 854$ election-day votes. In fact he received 3,310 election-day votes. By this method one might gauge the number of accidental votes for Buchanan because of the butterfly ballot as follows: $3{,}310 - 854 = 2{,}456$ votes.

PRECINCT-LEVEL ANALYSIS OF PALM BEACH COUNTY RETURNS

Who made the mistakes on the butterfly ballot? Was it voters who favored Bush or those who wanted to vote for Gore? At an intuitive level, mistakes seem less likely for Bush voters, who had to match the first candidate with the first punch hole, than for Gore voters, who had to match the second candidate with the third punch hole. Furthermore, it was a large number of Democratic voters who complained that the ballot caused them to vote for Buchanan by mistake. Nonetheless, it is important to examine the possibility that Buchanan received votes intended for both candidates.

To assess the asymmetry of voting error—whether Democratic voters mistakenly chose Buchanan at a substantially greater rate than Republican voters—we estimate an overdispersed binomial model for precinct-level election returns across PBC. We use two regressors (plus the constant): the proportions of the vote in each election-day precinct for two U.S. Senate candidates, namely, Democrat Bill Nelson and Reform candidate Joel Deckard. We use our robust estimators. If the butterfly ballot did not cause systematically biased voting errors, then support for Nelson (D) in a precinct should be negatively associated with the Buchanan vote, and support for Deckard (Ref) would be positively associated with that vote. If the butterfly ballot did cause asymmetric voting errors, then support for Nelson (D) should be positively associated with the Buchanan vote.

The first line in Table 2 presents the coefficient estimates for all PBC election-day precincts. Support for Nelson (D) is positively associated with support for Buchanan. This pattern is expected if the butterfly ballot caused many Democrats to vote for Buchanan and supports the claim that his votes tended to come from mistaken Gore supporters.

The pattern does not occur in votes for other offices that included a Reform candidate but did not use the butterfly ballot. In 2000 in PBC, Sherree Lowe (Ref) ran for the State Senate in District 35 and John McGuire (Ref) ran for the U.S. House in District 16. The issue positions stated on their and Deckard's web pages suggest that all three of the Florida Reform candidates were in sympathy with the Buchanan faction. For instance, all were pro-life and looked askance at free trade. Deckard went so far as to express concerns about *legal* immigration.[22]

We robustly estimate two models for election-day precincts in PBC in each district. In both districts we analyze the votes received by Buchanan. For District 35 we also analyze the votes received by Lowe (Ref), and for District 16 we analyze the votes for McGuire (Ref). Because the butterfly ballot was used only for the presidential vote, we expect support for Nelson (D) to be negatively associated with support both for Lowe (Ref) and for McGuire (Ref).

The coefficient estimates in Table 2 match those expectations, while in both districts support for Nelson (D) remains positively associated with the Buchanan vote. These results support the claim that the butterfly ballot caused systematic, biased voter errors that cost Gore more votes than Bush. Democratically inclined precincts (as measured by the Nelson [D] vote proportion) have fewer votes for Reform candidates in general (i.e., Lowe and McGuire) but have more votes for Buchanan. The difference is the butterfly ballot.

These findings refute one possible explanation for the positive association between Nelson's vote share and Buchanan votes, which is that Reform Party members in PBC tend to live among Democrats. Such an explanation is contradicted by the negative association between the proportion of votes for Nelson (D) and the votes for Lowe (Ref) and for McGuire (Ref).

We also can refute another class of alternative explanations for Buchanan's vote total in PBC: It may have been caused by a group of anomalous precincts within the county. In other words, anomalous results concentrated within a few precincts would suggest that excess votes were the result of localized phenomena rather than the butterfly ballot which was used uniformly throughout the county. For example, malfunctioning vote machines—such as the one in Tillman precinct in Jasper County, South Carolina—could have recorded extra votes for Buchanan in a few precincts. Merzer and the *Miami Herald* (2001, 78–80) document the existence of malfunctioning machines in numerous PBC precincts: 96 of 462 ballots from tests conducted immediately before polls opened showed failures. Or there may have been intentional fraud in a few precincts. Finally, pockets of intense election-day Reform mobilization could have delivered the extra votes.

Such explanations, however, are quite difficult to reconcile with our precinct-level findings. Given our use of a robust estimator, the coefficient estimates in Table 2 would not be affected by a few precinct-level anomalies produced by irregular voting processes.[23] Moreover, a localized mobilization effort should affect outcomes in multiple races, but the peculiar relationship between Nelson (D) vote share and Reform vote is present only in the presidential race, which used the butterfly ballot.[24]

TABLE 3. Vote for Buchanan by U.S. Senate Vote in Palm Beach County, for Individual Ballots by Ballot Type

Ballot Type	Intercept	Nelson (D)	Deckard (Ref)
Election-day	−5.18 (0.034)	0.61 (0.040)	2.41 (0.138)
Absentee	−6.11 (0.156)	−0.21 (0.236)	3.68 (0.400)

Note: Entries are maximum likelihood estimates of coefficient parameters of the binary logistic regression model using ballot data from the 2000 election (standard errors are in parentheses). Ballots with spoiled presidential votes (undervotes or overvotes) are omitted. Including them does not materially change the results. Number of unspoiled ballots for each type: election-day, 381,449; absentee, 36,412.

COMPARISON USING INDIVIDUAL PALM BEACH COUNTY BALLOTS

We extend our analysis to ballot-level data from PBC,[25] which enable us to compare an individual's presidential vote choice with the individual's choices for other offices. The data include both election-day and absentee ballots but exclude returns for 25 precincts that were overwritten by test data when PBC tested its vote tabulating machines. The number of Buchanan votes lost from those precincts (2.1%) is proportional to the number of precincts lost. We use the ballot-level data to validate the precinct-level regression results. We also produce another estimate of the size of the butterfly ballot effect.

With ballot-level data the binomial regression model reduces to the familiar binary logistic regression model.[26] The dependent variable is whether the voter cast a vote for Buchanan or voted for another candidate. In addition to the constant, the regressors are two dummy variables that respectively indicate whether the ballot records a vote for Nelson (D) or a vote for Deckard (Ref).

The estimates reported in Table 3 for election-day ballots show a positive and significant coefficient on voting for Nelson (D), which matches the positive coefficient on the precinct-level proportion of votes for Nelson (D) in Table 2. Among election-day ballots, voting for Nelson (D) is positively associated with voting for Buchanan. For absentee ballots, however, the Nelson coefficient is negative and insignificant. The confidence intervals of the Nelson coefficients for the two ballot formats do not overlap. Therefore, we reject the hypothesis that the effect of voting for Nelson (D) is the same regardless of the ballot format.

[22] Candidate issue positions were posted at http://www.joeldeckard.com/issues.htm (Deckard), http://www.ronhoward.org/mcissues.htm (McGuire), http://www.sherreelowe.com/ (no longer available) and http://election.dos.state.fl.us/cgi-bin/CandHtml.exe?account=31167&elecid=20001107-GEN (Lowe) (all accessed April 7, 2001).

[23] In the analysis of all 515 precincts, reported in Table 2, ten were outliers.

[24] In addition, Elms and Brady (2001) show that the excessive Buchanan vote is spread throughout the county, and there is a precipitous drop in the Buchanan vote proportion in precincts in adjoining counties.

[25] The ballot data were acquired from the PBC Supervisors of Elections.

[26] In the absence of grouping there cannot be overdispersion (McCullagh and Nelder 1989, 125) and high breakdown estimators do not exist (Christmann 1994).

TABLE 4. Proportion Voting for Buchanan by U.S. Senate Vote Choice and Ballot Type in Palm Beach County

Senate Candidate	Election-Day Ballots		Absentee Ballots	
	Proportion	N	Proportion	N
Bill Nelson (D)	0.0102	228,455	0.0017	17,779
Joel Deckard (Ref)	0.0590	1,000	0.0808	99

Note: Entries are the proportion of ballots with a vote for Buchanan out of the N ballots of each type voted for each Senate candidate, using ballot data from the 2000 election. Ballots with spoiled presidential votes (undervotes or overvotes) are omitted.

Table 4 shows the proportion of votes in PBC going to Buchanan among all ballots that record U.S. Senate votes for either Nelson (D) or Deckard (Ref). The proportions show that PBC voters who support the Democratic Senate candidate are significantly more likely to vote for Buchanan on the butterfly ballot than are their counterparts who use the absentee ballot—indeed, six times more likely. Fewer than two in one thousand absentee voters in PBC who vote for Nelson (D) also vote for Buchanan, but among election-day Nelson voters the figure is ten in one thousand. If we treat the absentee proportion as the proportion of votes truly intended to go to Buchanan, then about 8.5 of every 1,000 Nelson voters in PBC—about 2,300 voters—appear to have voted mistakenly for Buchanan.[27] Because most (89.6%) absentee Nelson voters voted for Gore, we can further conclude that at least 2,000 of the 2,300 would have been Gore votes.

In contrast, Table 4 shows that individuals who vote for Deckard (Ref) are more likely to vote for Buchanan on the absentee ballot. Deckard voters who support Buchanan should not be affected by the butterfly ballot, and the difference between election-day and absentee Buchanan vote proportions is small.

The ballot data add to the evidence that the butterfly ballot caused systematic voting errors in PBC that cost Gore votes. In particular, these data help rule out the possibility that Buchanan's exceptional support in the county was a result of populist appeals he made or policy positions he took that Democrats found attractive. The ballot data show that the appeals would somehow have to have been effective for Democrats who voted on election day but not Democrats who used an absentee ballot.

We gain analytical precision with the ballot data, but the ballot-level analysis complements rather than replaces the county and precinct analyses presented above. Ballot-level data are rarely retained or made available after an election, so it is not generally possible to compare these results across states or counties. Without such comparisons, the ballot-level results must be considered with some caution. Moreover, ballot data from 2% of the precincts in PBC are not available, but no data are missing for our precinct-level analysis. Also, unlike the county and precinct analysis, the analysis of individual ballots cannot use robust estimation techniques. In the absence of robust—high breakdown point—results, it would have been possible to claim that the aberrations we found may be limited to a few idiosyncratic precincts and not characteristic of PBC as a whole. But the aberrations prevail throughout the county.

CONCLUSION

We have examined the source of the anomalous support for Buchanan in PBC by focusing on allegations that the county's use of a butterfly ballot caused systematic voting errors that boosted the number of votes for Buchanan.[28] Robust estimation of overdispersed binomial regression models showed that, with respect to the Reform vote in 2000, PBC is the largest outlier among all counties in the United States we were able to examine. We also showed that PBC was not a Reform vote outlier in 1996, a presidential year in which the county did not use a butterfly ballot. In some counties around the country we found clear auxiliary evidence of problems with ballots, voting machines, or election administration. In still others there are strong indications that Buchanan received an exceptional number of votes because he had exceptional political support in those places. There is no reason to believe that he had such mobilized support in PBC.

Having confirmed that in 2000 PBC was an outlier, we sought to verify whether the butterfly ballot was the cause. A comparison of election-day versus absentee ballot results across all Florida counties shows that Buchanan's success in PBC did not extend to absentee voters, who did not use the butterfly ballot. We examined the claim that Democratic presidential candidate Al Gore in particular was harmed by the butterfly ballot. We found that Buchanan's support in PBC tended to come from more Democratic precincts and from those who voted for the Democratic candidate for the U.S. Senate, which supports the claim that mistaken votes for Buchanan tended to come from Gore supporters.

Was the butterfly ballot pivotal in the 2000 presidential race? The evidence is very strong that it was. Had PBC used a ballot format in the presidential race that did not lead to systematic biased voting errors, our findings suggest that, other things equal, Al Gore would have won a majority of the officially certified votes in Florida.

Our analysis complements the efforts of media groups to inspect ballots throughout Florida in order to assess what result would have been produced by completing the recount that the U.S. Supreme Court terminated, or by conducting a count using uniform standards throughout the state. As of this writing only the results of a statewide inspection conducted by the

[27] This number is calculated by 269,835 × 0.0085 = 2,294. The number 269,835 is the total Nelson (D) received in the entire county on election day, including the precincts missing from the ballot data.

[28] A related allegation is that the PBC ballot led to excessive overvoting in the presidential race (Merzer and *Miami Herald* 2001). The subject of overvoting is beyond the scope of this article (recall footnote 9; also see Bridges 2001).

Miami Herald have been reported.[29] Our analysis answers a counterfactual question about voter intentions that such investigations cannot resolve. The inspections may clarify the number of voters who marked their ballot in support of the various candidates, but the inspections cannot tell us how many voters marked their ballot for a candidate they did not intend to choose.

Citing the results from various scenarios in which votes were counted using one of several reasonable uniform standards, the *Herald* concludes "After study and analysis of 111,261 overvotes and 64,826 undervotes, [. . .] the outcome still depends on the standard used to gauge undervotes. Gore wins narrowly under two undervote standards, by margins of 332 and 242 votes; Bush wins narrowly under two other undervote standards, by 407 and 152 votes (Merzer 2001a; see also Merzer 2001b). Evidently, the number of votes that were intended for Gore but that went to Buchanan because of the butterfly ballot is large enough to have changed the election outcome given any of several reasonable standards that might have been used to count the votes in Florida.

Although we focus here on the butterfly ballot in PBC, our methods could be used on a regular basis as part of an ongoing effort to identify election anomalies and improve the administration of elections by eliminating such anomalies. Our robust estimation and outlier detection methods offer an accurate and powerful technology for detecting irregular vote outcomes. But determining why a particular irregularity occurs requires a strategy of triangulation, such as the one we pursue here. Different models and different types of data need to be marshaled to eliminate plausible alternative explanations. In the case of PBC and the 2000 presidential election, such a strategy leads to the conclusion that "the butterfly did it."

APPENDIX A: STATISTICAL MODEL AND ESTIMATION METHODS

The Overdispersed Binomial Regression Model

The overdispersed binomial regression model is defined as follows. Let i indicate one of n geographic areas, $i = 1, \ldots, n$. Let y_i denote the count of votes for Buchanan in area i, and let m_i denote the total number of ballots cast for all presidential candidates in area i. Given m_i and a probability value π_i, the expected value of y_i is

$$E(y_i \mid m_i, \pi_i) = m_i \pi_i,$$

and the variance of y_i is

$$E[(y_i - m_i\pi_i)^2 \mid m_i, \pi_i, \sigma^2] = \sigma^2 m_i \pi_i (1 - \pi_i),$$

with $\sigma^2 > 0$ (McCullagh and Nelder 1989, 125, eq. 4.20). If $\sigma^2 = 1$, then the variance is the same as the variance of a standard binomial random variable. If $\sigma^2 > 1$, then there is overdispersion relative to a purely binomial model. The probability value π_i is a logistic function of a linear predictor denoted $x_i'\beta$, where x_i is a vector of k regressors, and β is an unknown constant vector of coefficient parameters. The probability π_i is defined by:

$$\pi_i = \frac{1}{1 + \exp(-x_i'\beta)}.$$

We do not treat the overdispersed binomial model as fully characterizing a likelihood function for the data. We treat the model as what McCullagh and Nelder (1989) describe as a "quasi-likelihood." As they explain (pp. 323–48), under fairly mild conditions estimates of parameters of the mean that are based on a quasi-likelihood have desirable properties, such as asymptotic unbiasedness, asymptotic optimality, and asymptotic normality. In the current analysis, the quasi-likelihood approach means that we assume only that the mean and variance formulas we define are good descriptions of the data. In fact, by using the robust estimation methods defined below, we assume somewhat less than that and still obtain estimates with good statistical properties. We require only that the mean and variance formulas are good descriptions for most of the observed vote counts in each collection of data for which we estimate parameters. In the analysis of county-level data, this means that the model is good for most of the counties in each state, and in the analysis of data for PBC precincts, for most of the precincts in PBC.

The idea that the model is good for most of the data does not mean that the model fully characterizes the process that generated most of the vote counts. Obviously, the process is vastly more complicated than our spare model specifications could possibly represent in full. Rather, the idea is that our estimator of the model's parameters converges asymptotically to a unique value for each parameter. In the theory of robust estimation this idea is made precise by the assumption that a probability model exists relative to which the estimator is Fisher consistent (Hampel et al. 1986, 82). If applied to all the data, that property would be for all practical purposes the same as the unique identifiability and convergence properties that White (1994) demonstrates are necessary for what he calls a "quasi maximum likelihood estimator" (QMLE) to have good statistical properties.[30] The robust estimation methods we use have good statistical properties even when the property holds only for a majority of the observations.

Robust Estimation

Breakdown Points. The key property of the robust estimators we use is that they have a high breakdown point. In a finite sample, the breakdown point of an estimator is the smallest proportion of the observations that must be replaced by arbitrary values in order to force the estimator to produce values arbitrarily far from the parameter values that generated the original data (Donoho and Huber 1983). The general concept of breakdown point (e.g., the asymptotic breakdown point) has the same connotation, although extensive technical apparatus is required to achieve full generality (Hampel 1971). To illustrate the breakdown point idea, we consider again the case of estimating the average for a set of numbers. For concreteness, suppose that the numbers originate from an unbiased sample of size n from a process that has mean zero and finite variance. As an estimator of the true mean, the sample mean has a breakdown point of $1/n$, because only one of the n data points needs to be replaced to

[29] See *Miami Herald* 2001 and Merzer and *Miami Herald* 2001.

[30] The McCullagh and Nelder (1989) concept of quasi-likelihood, which has to do with using a model that one believes is correct only for the first two moments of the data (the mean and covariance matrix), is not the same as the QMLE of White (1994), which has to do with the asymptotic properties of a model that is misspecified.

force the sample mean to take a value arbitrarily far from the true mean. If one data point is replaced with another value far from zero but all the other data remain unchanged, then the value of the sample mean moves toward the distorted data value; if the distorted data value is moved indefinitely far from zero, the sample mean moves indefinitely far from zero. Asymptotically, as the sample size n increases to infinity, the sample mean has a breakdown point of zero, because $1/n \downarrow 0$. The same is true for any least-squares estimator or, indeed, for any estimator that always puts positive weight on all the observed data. In contrast, as an estimator of the true mean in our example, the sample median has a breakdown point of 1/2. In order to move the sample median arbitrarily far from the true mean of zero, at least half the data points have to be replaced with values arbitrarily far from zero.

The highest possible asymptotic breakdown point for an estimator of a regression model's mean parameters is 1/2. We use estimators that achieve that maximum for all possible ways of distorting the data. We do not use the well-known minimum absolute deviations estimator (also known as the L_1 estimator) because it has a breakdown point of only $1/n$ relative to distortions of the regressors (Rousseeuw and Leroy 1987, 10–2). A robust estimator called least median of squares (LMS) (Rousseeuw 1984) does achieve the maximum breakdown point and is widely used. Western (1995) discusses high breakdown estimation of linear models using LMS and suggests an approach to robust estimation of generalized linear models. Christmann (1994) discusses application of LMS to a grouped binomial model (albeit without overdispersion). But LMS converges more slowly and is less efficient than the estimators we use.

The high breakdown point of the estimators we use means that the estimates of model parameters remain stable even when unusual voting processes occur in several of the n geographic areas for which we are estimating the model. When we use the estimates to compute studentized residuals for counties, a large anomaly in one county will not mask comparable or perhaps somewhat smaller anomalies that occur in other counties. Hampel et al. (1986, 67) discuss the relationship between the breakdown point and masking. Without a high breakdown point estimator we would underestimate the frequency of highly anomalous election results. When we directly interpret the parameters of precinct-level models, large departures from the model in several precincts will not make the interpretations untrustworthy.

The robust methods we use also perform well in the absence of anomalies. If there are no anomalies, the robust estimator is consistent and is almost as efficient as an estimator, such as simple iteratively reweighted least squares, that ignores the possibility of anomalous observations.

Two Robust Estimators. To obtain robust estimates of the parameters of the overdispersed binomial model, we combine two different estimators. We use one to estimate the scale, $\sigma = \sqrt{\sigma^2}$, and the vector of coefficients, β, and then we use a second to obtain a much more efficient estimate of β. The second estimator depends on the first one's estimate of the scale, $\hat{\sigma}$. The first estimator is called the least quartile difference (LQD) estimator (Croux, Rousseeuw, and Hossjer 1994; Rousseeuw and Croux 1993), and the second is called the hyperbolic tangent (tanh) estimator (Hampel, Rousseeuw, and Ronchetti 1981).

Both estimators minimize functions of particular forms of residuals. Given a vector of estimates $\hat{\beta}$, let $\hat{\pi}_i = [1 + \exp(-x_i'/\hat{\beta})]^{-1}$ denote the estimated probability of voting for Buchanan in geographic area i. The residual that we use in the LQD estimator is:

$$r_i^* = \frac{y_i - m_i\hat{\pi}_i}{\sqrt{m_i\hat{\pi}_i(1-\hat{\pi}_i)}}.$$

Ordinarily, with an overdispersed binomial model in the absence of outliers, a good moment estimator for σ^2 may be defined in terms of r_i^* (McCullagh and Nelder 1989, 126–7, eq. 4.23). The distributional assumption we make is that for most of the observations i, computing r_i^* with $\hat{\beta} = \beta$ would produce a set of independent normal random variables each having variance σ^2. If the distributional assumption held for all the observations, including in the asymptotic limit as $n \uparrow \infty$, then the LQD estimate, $\hat{\sigma}$, would be consistent for the scale σ. If some observations (up to half the data) do not satisfy the distributional assumption, then asymptotically the difference between $\hat{\sigma}$ and σ remains bounded. The LQD estimator is defined by choosing $\hat{\beta}$ to minimize approximately the first quartile of the set of absolute differences $\{|r_i^* - r_j^*| : i < j\}$ Further details are given below.

Given a scale estimate $\hat{\sigma}$, the tanh estimator for β is based on the residual:

$$r_i = \frac{y_i - m_i\hat{\pi}_i}{\hat{\sigma}\sqrt{m_i\hat{\pi}_i(1-\hat{\pi}_i)}}.$$

The tanh estimator uses a function $\psi(r_i)$ to downweight observations that have large residuals. The weight applied to each observation is defined by:

$$w_i = \begin{cases} \psi(r_i)/r_i, & \text{for } r_i \neq 0 \\ 1, & \text{for } r_i = 0 \end{cases}.$$

The definition of ψ has a complicated functional form that we state in equation A-2. Here we characterize the weight values that ψ implies. The weights change qualitatively at thresholds defined by two constants, $p = 1.8$ and $c = 4.0$) (we further explain the constants below). Observations that have residuals in the range $-p \leq r_i \leq p$ have $w_i = 1$: They are not downweighted. Observations that have residuals in the range $p < |r_i| \leq c$ have weights that gradually diminish to zero as $|r_i|$ approaches c. Observations that have residuals of magnitude greater than c, meaning $c < |r_i|$, have $w_i = 0$. The tanh estimator completely rejects such observations so that they have no effect on the tanh estimate $\hat{\beta}$.

We define an outlier to be any observation that has $w_i = 0$.

The usual method used to estimate an overdispersed binomial regression model is an iteratively reweighted least-squares algorithm, which is equivalent to maximum likelihood estimation of the coefficients of a standard binomial regression model with the dispersion σ^2 estimated subsequently (McCullagh and Nelder 1989, 114–28). To implement the tanh estimator we modify that algorithm by weighting observation i by w_i, with w_i computed using the coefficient estimates from the previous iteration and the LQD estimate $\hat{\sigma}$. The estimate of σ remains unchanged throughout the estimation. Numerical convergence is required for both the $\hat{\beta}$ values and the weights. To start the iterations we use the LQD estimates for the initial coefficients and use the LQD values $(r_i^* - \text{med}_i r_i^*)/\hat{\sigma}$ for an initial set of residuals r_i (med$_i$ r_i^* denotes the median of the r_i^* values, $i=1, \ldots, n$).

To estimate the asymptotic variance of the coefficient estimates and hence standard errors, we use the sandwich estimator avar$(\hat{\beta}) = \hat{\sigma}^2\hat{J}^{-1}\hat{I}\hat{J}^{-1}$, where $\hat{I}$ is the outer product of the score, and $\hat{J}$ is the Hessian for a standard binomial likelihood evaluated at $\hat{\beta}$, weighting each observation by w_i. Justification for this estimate, based on White (1994), appears below.

Studentized Residuals Definition. To define the studentized residuals that we use to compare counties, we adjust the residuals r_i for variation associated with each county's regressors. We use the usual adjustment for the influence (or leverage) each observation y_i has on the estimated expected vote counts $m_i\hat{\pi}_i$, modified to take into account the weighting that occurs with the tanh estimator via w_i. The adjustment is a function of quantities h_i, which are defined as follows. Using W to denote a diagonal matrix that has diagonal entries $W_{ii} = w_i$, V to denote a diagonal matrix with $V_{ii} = [m_i\hat{\pi}_i(1 - \hat{\pi}_i)]^{-1/2}$, and X to denote the $n \times k$ matrix of the regressors (row i of X is x_i'), define the matrix $H = VX(X'VWVX)^{-1}X'V$. If no observations are downweighted, so that $w_i = 1$ for all $i = 1, \ldots, n$, then H is the matrix defined in equation 12.3 of McCullagh and Nelder (1989, 397). In that case, each diagonal element of H (i.e., H_{ii}) measures the influence of y_i on $m_i\hat{\pi}_i$. When observation i is downweighted ($w_i < 1$), the influence interpretation still applies when $w_i > 0$. If $w_i = 0$, then y_i has no effect on $m_i\hat{\pi}_i$, and $-H_{ii}$ measures error variation associated with forecasting observation i. Hence we define $h_i = H_{ii}$ if $w_i > 0$ and $h_i = -H_{ii}$ if $w_i = 0$, and the studentized residual is:

$$\tilde{r}_i = r_i / \sqrt{1 - h_i}. \qquad \textbf{(A-1)}$$

The adjustment by $(1 - h_i)^{-1/2}$ makes the variance of the residuals $\tilde{r}_i$ constant for all observations that have $w_i > 0$, $i = 1, \ldots, n$. For the outliers ($w_i = 0$) we do not make any claim about what distribution the residuals may have, but the adjustment should at least reduce one source of variation among them.

Because in the county-level analysis we estimate the parameters of the model separately for each state (except for four states that have few counties), the studentized residuals for counties in a state that are not outliers all have the same variance and so are directly comparable. Across states the variances differ slightly because of the following technical variations in the sampling distributions. The variance for each state is approximately the variance of a t-distribution with degrees of freedom equal to the number of counties in the state that are not outliers. We do not attempt to adjust for that source of variation across states because it is negligible compared to the variation caused by the seriously anomalous processes that occurred in some counties, turning them into outliers. In no case would one expect to observe a residual of magnitude greater than 4.0—which would make the county an outlier—for a nonanomalous county.

Robust Estimation Method Details

The LQD Estimator and Its Properties. One may understand the LQD estimator intuitively as an extension of the idea of using the interquartile range to estimate the dispersion of a set of data. The LQD estimator focuses on the $\binom{h_k}{2}$ order statistic of the set $\{|r_i^* - r_j^*| : i < j\}$ of absolute differences, where $h_k = [(n + k)/2]$ and $\{r_i^* - r_j^*| : i < j\}$ has $\binom{n}{2}$ elements. Following Croux, Rousseeuw, and Hossjer (1994) we use the notation

$$Q_n^* = \{|r_i^* - r_j^*| : i < j\}_{\binom{h_k}{2}:\binom{n}{2}}$$

to denote that order statistic. For large n and k small relative to n, $\binom{h_k}{2}/\binom{n}{2} \approx 1/4$, so that Q_n^* is approximately the first quartile of the absolute differences. To implement LQD we choose estimates $\hat{\beta}$ to minimize Q_n^*. Let $\hat{\beta}_{\mathrm{LQD}^*}$ designate the estimated coefficient vector, and let $\hat{Q}_n^*$ designate the corresponding minimized value of Q_n^*. The LQD scale estimate is

$$\hat{\sigma} = \frac{\hat{Q}_n^*}{\sqrt{2}\Phi^{-1}(5/8)},$$

where Φ^{-1} is the quantile function for the standard normal distribution. The normality assumption we make about r_i^* justifies the factor $[\sqrt{2}\Phi^{-1}(5/8)]^{-1}$ (Rousseeuw and Croux 1993, 1277).

The LQD objective function is difficult to optimize. Because high breakdown point estimators are not smooth functions of the data, optimization typically does not rely on derivative information but instead depends on combinatorial algorithms (Stromberg 1993). We use the global optimizer GENetic Optimization Using Derivatives (GENOUD) (Sekhon and Mebane 1998), which combines global evolutionary algorithm methods with a local, quasi-Newton method to solve difficult unconstrained optimization problems.[31] Sekhon and Mebane discuss the theoretical basis for expecting GENOUD to find the global optimum when the objective function, as in the LQD case, is not a smooth function of the data. They also present Monte Carlo experiments and examples that demonstrate the algorithm's effectiveness. The evolutionary algorithm component of GENOUD excels at finding the neighborhood of the optimum even when the objective function is highly irregular. Local derivative information, when it exists, is useful for efficiently going from a neighborhood of the optimum to the optimum. Even when the derivatives provide no useful information, the evolutionary algorithm component of GENOUD is able to find the global optimum. Optimization is more efficient, however, when the derivatives are informative. In the LQD case, derivatives provide useful local information. Therefore, the quasi-Newton component of GENOUD significantly improves the efficiency of the overall optimization algorithm. But the global optimization properties of GENOUD come from its evolutionary algorithm component.

For a linear regression model, Croux, Rousseeuw, and Hossjer (1994) show that the LQD estimator converges at a rate of $n^{-1/2}$ and has a Gaussian efficiency of 67.1% for all the coefficients except the intercept, which in a linear regression model LQD does not estimate.[32] Gaussian efficiency refers to the efficiency of the estimator when the disturbance is an identically and independently distributed Gaussian random variable with conditional mean zero (which implies that the model is correctly specified for all observations). LMS, in contrast to LQD, converges at the slower rate of $n^{-1/3}$ and has a Gaussian efficiency of 37.0% (Rousseeuw and Croux 1993, 1279). LQD also provides a good estimate of the dispersion when the disturbance has an asymmetric distribution, because LQD does not estimate the scale by measuring a symmetric spread of the residuals around a central location value (Rousseeuw and Croux 1993).

To use LQD to estimate σ^2 we need to define residuals that are reasonably well described by a reference model of normality with zero mean and variance σ^2. If the overdispersed binomial model correctly describes the mean and variance for all the data, then given any consistent estimate for β, the residuals r_i^* are approximately normal with the desired mean and variance.

The approximate normality of r_i^* depends on conditions such as independence across i and sufficiently large values for $m_i\pi_i(1 - \pi_i)$. In view of the small proportion of the votes that Buchanan received, it is important to consider for how small

[31] See http://jsekhon.fas.harvard.edu/rgenoud/ for an **R** version of the GENOUD software and http://jsekhon.fas.harvard.edu/genoud/ for more information.

[32] In the linear model the constant cancels in the differences $(r_i^* - r_j^*)$. In the overdispersed binomial regression model such cancellation does not occur because of the nonlinear factor $[\hat{\pi}_i(1 - \hat{\pi}_i)]^{-1/2}$ in r_i^*. Hence, the LQD estimator of the overdispersed binomial model has information about the constant.

a value of $m_i\pi_i$ the normality model is plausible. Notice that $X^2 = \Sigma_{i=1}^{n} (r_i^*)^2$ is the Pearson chi-squared statistic. Larntz (1978, 255–6) enumerates the exact distribution of X^2 for binomial data and concludes that, even for sample size n as small as 10 or 15, inferences reached by using the asymptotic chi-squared distribution are reasonable when all expected values $m_i\pi_i$ and $m_i(1 - \pi_i)$ are greater than 1.0. We take such results as supporting our use of the reference normal model for the county data. No state has more than a few counties with m_i so small that the expected vote for Buchanan is less than 1.0.

The case is less clear for the precinct data. Because the expected proportion of the vote for Buchanan in PBC is about 0.001 (see Table B-1), a typical precinct would have to have $m_i = 1{,}000$ voters to have one expected vote for Buchanan. Among the 515 election-day precincts in PBC, 172 have $m_i > 1{,}000$, but 56 have $m_i < 100$. Koehler and Larntz (1980, 337) show (for Poisson variables) that the Poisson information kernel declines rapidly to zero as the expected value goes from one to zero. Koehler and Larntz (p. 338) observe that, consequently, the asymptotic (in n) mean of the likelihood-ratio chi-squared statistic "can be much smaller than the chi-squared mean when many expected frequencies are smaller than one-half"; McCullagh (1986) demonstrates one important respect in which the chi-squared approximation survives better for the Pearson statistic than for the likelihood-ratio statistic as m_i decreases to one. That may suggest that our estimates of σ are biased somewhat downward for the precinct data.

The Tanh Estimator and Its Properties. To estimate β, a tanh estimator that uses the LQD scale estimate is more efficient than the LQD estimator alone. In addition to achieving the maximum breakdown point, tanh estimators minimize the asymptotic variance of the estimates for a given upper bound d on how sensitive the variance is to a change in the distribution of the data. Hampel et al. (1986, 125–36) explain the concept of the change-of-variance function that makes the foregoing concept of sensitivity precise. Given a good estimate of the scale, tanh estimators are by definition the most efficient possible estimators of location that have a finite rejection point, which means that some observations may receive zero weight, and the key robustness property of bounded response to arbitrary changes in parts of the data (Hampel et al. 1986, 166). Hampel, Rousseeuw, and Ronchetti (1981) prove existence and optimality properties of tanh estimators (see also Hampel et al. 1986, 160–8). Hampel et al. (1986, 328) conjecture that the tanh estimator of location extends directly to a tanh estimator for linear regression. We are applying the estimator to a nonlinear regression model.

The tanh estimator is a redescending M-estimator of location that is based on the function

$$\psi(u) = \begin{cases} u, & \text{for } 0 \le |u| \le p \\ (A(d-1))^{1/2}\tanh[\tfrac{1}{2}((d-1) \\ \quad \times B^2/A)^{1/2}(c-|u|)]\operatorname{sign}(u), & \text{for } p \le |u| \le c \\ 0, & \text{for } c \le |u| \end{cases} \tag{A-2}$$

where p, c, d, A, and B are constants satisfying $0 < p < c$ and other conditions (Hampel et al. 1981, 645). For $c > |u|$, the sign of $\psi(u)$ equals the sign of u. For u increasing from p to c, $\psi(u)$ descends from its maximum value $\psi(p) = p$ to $\psi(c) = 0$, and for u increasing from $-c$ to $-p$, $\psi(u)$ descends from $\psi(-c) = 0$ to its minimum value $\psi(-p) = -p$. Hampel, Rousseeuw, and Ronchetti (1981, 645) and Hampel et al. (1986, 162) visually depict the shape of ψ. Optimality requires that along the curves between $\psi(p)$ and $\psi(c)$ and between $\psi(-c)$ and $\psi(-p)$, the value of ψ is such that there is a constant ratio between the sensitivity of the variance to a change in the data and the asymptotic variance. Indeed, the value of the ratio is d. Given a choice of d and of the truncation point c, minimizing the asymptotic variance of the estimator while satisfying such conditions implies unique values for p, A and B of (2) (Hampel et al. 1986, 162–4). We use $c = 4.0$ and $d = 5.0$, which imply values $p = 1.8$, $A = 0.86$ and $B = 0.91$ (Hampel, Rousseeuw, and Ronchetti 1981, 645, Table 2).[33]

Given a scale estimate $\hat{\sigma}$ and a vector of trial estimates $\hat{\beta}$, we compute the residuals r_i and then weights w_i. Observation i is weighted by w_i in what is otherwise the usual iteratively reweighted least-squares algorithm to estimate β. Full iteration makes this weighted estimator equivalent to the tanh M-estimator (Hampel et al. 1986, 116). Because redescending M-estimators such as the tanh estimator have multiple solutions, starting values affect the results.

Asymptotic Covariance Matrix of Tanh Coefficient Estimates. The sandwich estimator avar($\hat{\beta}$) is valid for the robust estimator insofar as the conditions for Theorem 6.4 of White (1994, 92) hold. The necessary assumptions are 2.1 (complete probability space), 2.3 (measurability, compact parameter space, and continuity), 3.1 (uniform convergence), 3.2′ (interior identifiably unique maximizers), 3.6 (continuous differentiability), 3.7(a) (uniform convergence of score vector), 3.8 (uniform convergence of Hessian), 3.9 (negative definite Hessian), and 6.1 (score obeys central limit theorem with positive definite covariance matrix). Hampel et al. (1986, 82) directly assume 2.1 and 2.3. The assumption of Fisher consistency (p. 83) implies 3.1. Given a suitable model parameterization, their Theorem 5 (pp. 160–2) implies 3.6 and 3.7(a) for all observations that are not outliers, and optimization by Newton's method from suitable starting values (compare p. 152) implies 3.2′, 3.8, 3.9, and 6.1.

APPENDIX B: COUNTY-LEVEL ELECTION RETURNS DATA SOURCES AND TABLE OF OUTLIERS

For AL, AR, AZ, CO, CT, GA, IA, ID, IL, IN, KS, KY, LA, MA, MD, MN, MO, MS, MT, ND, NE, NH, NJ, NM, NV, NY, OH, OR, PA, RI, SC, SD, TN, TX, UT, VA, WA, and WI, the data are certified (or "official") election results from each state's Secretary of State (or comparable office), or results reported on the Secretary of State's website, originally collected by David Leip (with updates as of December 16, 2000) and posted at http://www.uselectionatlas.org (accessed January 10 or 21, 2001). For NY we obtained data separately for the City of New York.[34] For ME (December 18, 2000) and WV (December 20, 2000) we accessed data from http://www.uselectionatlas.org on the indicated dates.

For the remaining states except AK we obtained data from official websites, as follows: CA, California Secretary of State,

[33] Hampel, Rousseeuw, and Ronchetti (1981) use k for the tuning parameter of ψ that we have denoted by d. The same information about the tuning parameters appears in Hampel et al. (1986, 163, Table 2) with notation r and k used for the parameters we have denoted by c and d.

[34] http://www.vote.nyc.ny.us/BoePages/Results/2000General/allg2000.pdf (accessed February 21, 2001).

TABLE B-1. U.S. Counties with the Largest Positive and Negative Studentized Residuals from 2000 Vote for Buchanan

State	County Name	Studentized Residual	Vote Proportion Expected	Vote Proportion Actual	Number of Ballots	Order
FL	Palm Beach	36.14	0.0010	0.0079	433,186	1
SC	Jasper	28.26	0.0013	0.0379	6,469	2
KS	Pottawatomie	19.53	0.0096	0.0528	7,731	3
WV	Hancock	18.16	0.0048	0.0289	13,472	4
NV	Clark	15.88	0.0037	0.0080	381,845	5
WV	Brooke	15.57	0.0046	0.0286	9,405	6
OH	Jefferson	14.56	0.0066	0.0303	34,636	7
OH	Athens	14.23	0.0049	0.0292	25,447	8
MN	Hennepin	13.67	0.0010	0.0040	573,846	9
NE	Douglas	12.52	0.0018	0.0038	183,156	10
WI	Milwaukee	12.25	0.0010	0.0037	433,537	11
MN	Ramsey	11.67	0.0009	0.0057	244,278	12
IA	Woodbury	11.49	0.0038	0.0107	37,896	13
WV	Marshall	10.35	0.0054	0.0196	13,498	14
WV	Ohio	10.25	0.0031	0.0126	17,964	15
AZ	Maricopa	9.99	0.0057	0.0080	899,808	16
MT	Silver Bow	9.89	0.0100	0.0265	16,703	17
KS	Shawnee	9.67	0.0048	0.0097	74,373	18
MO	St. Louis	9.66	0.0012	0.0024	486,884	19
KS	Sedgwick	9.41	0.0034	0.0062	163,417	20
MO	St. Louis City	9.30	0.0005	0.0025	124,752	21
AR	Union	8.95	0.0052	0.0186	15,609	22
MO	Jackson	8.55	0.0014	0.0030	272,062	23
AR	Faulkner	8.50	0.0052	0.0148	29,216	24
IA	Plymouth	8.20	0.0043	0.0143	10,118	25
OH	Belmont	7.79	0.0074	0.0218	30,141	26
CO	Adams	7.60	0.0049	0.0093	107,852	27
CO	Jefferson	7.48	0.0034	0.0058	235,491	28
FL	Pinellas	7.47	0.0010	0.0025	398,472	29
NE	Dakota	7.33	0.0046	0.0144	6,061	30
WI	Marathon	7.22	0.0041	0.0116	58,374	31
NM	Bernalillo	7.19	0.0011	0.0019	204,319	32
SD	Minnehaha	7.13	0.0034	0.0062	61,369	33
NY	Erie	6.95	0.0050	0.0079	424,654	34
NE	Thurston	6.95	0.0063	0.0245	2,082	35
MD	Somerset	6.59	0.0025	0.0064	7,604	36
KS	Wyandotte	6.42	0.0025	0.0057	48,272	37
CO	El Paso	6.39	0.0047	0.0073	200,757	38
VA	Warren	6.35	0.0026	0.0088	11,166	39
NE	Dixon	6.31	0.0080	0.0233	2,792	40
WI	Lincoln	5.99	0.0046	0.0178	14,239	41
TX	Harris	5.84	0.0014	0.0017	974,822	42
OH	Ashland	5.81	0.0061	0.0176	21,258	43
IA	Dubuque	5.79	0.0040	0.0074	40,323	44
PA	Allegheny	5.46	0.0024	0.0036	582,478	45
KY	Boone	5.41	0.0021	0.0043	31,984	46
GA	Whitfield	5.35	0.0045	0.0081	23,169	47
WI	Wood	5.34	0.0048	0.0123	35,761	48
KY	Kenton	5.10	0.0021	0.0036	55,987	49
MS	Leflore	5.09	0.0012	0.0039	11,276	50
CO	Arapahoe	4.96	0.0031	0.0048	189,942	51
MS	Washington	4.85	0.0012	0.0033	18,328	52
IA	Sioux	4.84	0.0018	0.0056	14,692	53
KS	Wabaunsee	4.76	0.0098	0.0260	3,420	54
AK	Matanuska-Susitna	4.60	0.0191	0.0311	21,254	55
MT	Deer Lodge	4.55	0.0101	0.0254	4,534	56
MS	Wilkinson	4.54	0.0022	0.0078	4,098	57
IN	Allen	4.48	0.0059	0.0083	114,320	58
KY	Graves	4.25	0.0032	0.0063	14,212	59
WI	Brown	4.22	0.0032	0.0060	107,769	60
SC	Abbeville	4.19	0.0022	0.0084	8,374	61
SD	Union	4.19	0.0100	0.0189	5,772	62

TABLE B-1 (Continued)

State	County Name	Studentized Residual	Vote Proportion Expected	Vote Proportion Actual	Number of Ballots	Order
MD	Wicomico	4.09	0.0023	0.0042	31,795	63
AL	Lauderdale	4.07	0.0056	0.0092	32,137	64
GA	Richmond	4.06	0.0030	0.0044	57,359	65
OH	Hocking	4.06	0.0078	0.0202	10,756	66
IA	Benton	4.02	0.0038	0.0081	11,766	67
TX	Cottle	4.01	0.0023	0.0132	757	68
GA	Fulton	−4.08	0.0028	0.0021	261,945	3046
WY	Albany	−4.12	0.0131	0.0054	13,163	3047
WY	Teton	−4.35	0.0132	0.0033	9,667	3048
LA	East Baton Rouge	−4.37	0.0072	0.0041	168,989	3049
OR	Deschutes	−4.37	0.0074	0.0042	57,885	3050
AL	Madison	−5.61	0.0056	0.0028	113,318	3051
LA	Caddo	−6.21	0.0105	0.0035	95,639	3052
LA	Orleans	−8.89	0.0112	0.0035	181,221	3053

Note: Results based on 3,053 counties. This table presents all counties with studentized residuals greater than or equal to 4.0 or less than or equal to −4.0.

Election 2000[35]; DE, Secretary of State, Department of Elections[36]; FL, Florida Department of State, Division of Election, "Data Download Utility"[37]; HI, State of Hawaii Office of Elections[38]; NC, North Carolina State Board of Elections, Official Results[39]; OK, Election Results and Statistics, 2000, Oklahoma State Election Board[40]; VT, 2000 Vermont Election Results, State of Vermont, Office of the Secretary of State, Elections and Campaign Finance Division[41]; WY, 2000 Official Election Results, Wyoming Secretary of State, Election Administration.[42]

For AK, we use the 25 county equivalents that were in effect during the 1996 election, leaving Denali Borough within Yukon-Koyukuk Census Area, and Yakutat City and Borough within Skagway-Hoonah-Angoon Census Area. We aggregate precinct vote data to create the county-equivalent units. Precinct-level results for 2000 are in pdf files dated December 5, 2000,[43] and for 1996 they are in text files dated November 27, 1996.[44] To map the data into county equivalents we used voting district and state legislative district data provided by the Census Bureau.[45]

[35] http://vote2000.ss.ca.gov/Returns/pres/59.htm (accessed January 21, 2001).
[36] http: //www.state.de.us/election/index.htm (accessed June 26, 2001).
[37] http://enight.dos.state.fl.us/report.asp?Date=001107 (accessed January 21, 2001).
[38] http://www.state.hi.us/elections/ (accessed July 24, 2001).
[39] http://www.sboe.state.nc.us/y2000elect/stateresults.htm (accessed January 22, 2001).
[40] http://www.state.ok.us/~elections/00result.html (accessed April 4, 2001).
[41] http://vermont-elections.org/2000geresults.htm (accessed January 22, 2001).
[42] http://soswy.state.wy.us/election/2000/results/g-usp.htm (accessed January 22, 2001).
[43] http: //www.gov.state.ak.us/ltgov/elections/elect00/00genr/index.shtml (accessed June 19, 2001).
[44] http://www.gov.state.ak.us/ltgov/elections/results/sov.zip (accessed June 19, 2001).
[45] http://factfinder.census.gov/servlet/DTGeoSearchByListServlet?ds_name=DEC_2000_PL_U&state=dt&_lang=en&_ts=11623041369 (accessed June 19, 2001).

REFERENCES

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Associated Press. 2000a. "One Election Result Confirmed: Second Election Ruled a Do-Over." Associated Press State and Local Wire, December 13.

Associated Press. 2000b. "Voting Scrutinized All Over Florida." Associated Press, November 9.

Atkinson, A. C. 1986. "Masking Unmasked." *Biometrika* 73 (3): 533–41.

Brady, Henry E., Michael Herron, Walter R. Mebane, Jr., Jasjeet S. Sekhon, Kenneth Shotts, and Jonathan Wand. 2001. "Law and Data: The Butterfly Ballot Episode." *PS: Political Science & Politics* 34 (March): 59–69.

Bridges, Tyler. 2001. "Ballots Offer Clues on Intent: Experts See Patterns in the Overvotes." *Miami Herald*, May 11, Internet edition.

Campbell, Angus, and Warren E. Miller. 1957. "The Motivational Basis of Straight and Split Ticket Voting." *American Political Science Review* 51 (June): 293–312.

Carroll, Raymond J., and David Ruppert. 1988. *Transformation and Weighting in Regression*. New York: Chapman & Hall.

Christmann, Andreas. 1994. "Least Median of Weighted Squares in Logistic-regression with Large Strata." *Biometrika* 81 (2): 413–17.

Croux, Christophe, Peter J. Rousseeuw, and Ola Hossjer. 1994. "Generalized S-Estimators." *Journal of the American Statistical Association* 89 (December): 1271–81.

Darcy, Robert. 1986. "Position Effects with Party Column Ballots." *Western Political Quarterly* 39 (December): 648–62.

Donoho, D. L., and P. J. Huber. 1983. "The Notion of Breakdown Point." In *A Festschrift for Erich L. Lehmann*, ed. Peter J. Bickel, Kjell A. Doksum, and J. L. Hodges. Belmont, CA: Wadsworth. Pp. 157–84.

Elms, Laurel, and Henry E. Brady. 2001. "Mapping the Buchanan Vote Escarpment in Palm Beach County, Florida." Paper presented at the annual meeting of the Public Choice Society, March 9–11, San Antonio, Texas.

Engelhardt, Joel. 2000. "Elections Chief on Firing Line." *Palm Beach Post*, November 9, Internet edition.

Engelhardt, Joel, and Scott McCabe. 2001a. "Election 2000: Over-Votes Cost Gore the Election in FL." *Palm Beach Post*, March 11, Internet edition.

Engelhardt, Joel, and Scott McCabe. 2001b. "Election 2000: Under-Votes Could Have Meant Victory for Gore." *Palm Beach Post*, March 11, Internet edition.

Firestone, David. 2000a. “Democrats Still Prefer to Keep Counting.” *New York Times*, November 11, Internet edition.

Firestone, David. 2000b. “Party Mobilized on Ballot Even as Votes Were Cast.” *New York Times*, November 11, Internet edition.

Garvey, Megan. 2000. “Bay Buchanan Sees Something Peculiar in Palm Beach Voting.” *Los Angeles Times*, November 10, Internet edition.

Hamilton, James T., and Helen F. Ladd. 1996. “Biased Ballots? The Impact of Ballot Structure on North Carolina Elections in 1992.” *Public Choice* 87 (3/4): 259–80.

Hampel, F. R. 1971. “A General Qualitative Definition of Robustness.” *Annals of Mathematical Statistics* 42 (December): 1887–96.

Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.

Hampel, Frank R., Peter J. Rousseeuw, and Elvezio Ronchetti. 1981. “The Change-of-Variance Curve and Optimal Redescending M-Estimators.” *Journal of the American Statistical Association* 76 (September): 643–8.

Koehler, Kenneth J., and Kinley Larntz. 1980. “An Empirical Investigation of Goodness-of-Fit Statistics for Sparse Multinomials.” *Journal of the American Statistical Association* 75 (June): 336–44.

Larntz, Kinley. 1978. “Small-Sample Comparisons of Exact Levels for Chi-Squared Goodness-of-Fit Statistics.” *Journal of the American Statistical Association* 73 (January): 253–63.

McCullagh, Peter. 1986. “The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data.” *Journal of the American Statistical Association* 81 (March): 104–7.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*, 2d ed. New York: Chapman & Hall.

Merzer, Martin. 2001a. “‘Overvotes’ Leaned to Gore, But to Win, He Needed Help of Dimpled Ballots.” *Miami Herald*, May 11, Internet edition.

Merzer, Martin. 2001b. “Review Shows Ballots Say Bush, But Gore Backers Have Some Points to Argue.” *Miami Herald*, April 4, Internet edition.

Merzer, Martin, and the staff of the *Miami Herald*. 2001. *The Miami Herald Report: Democracy Held Hostage*. New York: St. Martin’s.

Miami Herald. 2001. “About This Project.” *Miami Herald*. May 11, Internet edition.

Mitchell, Alison. 2001. “Results Official, at Last: Bush Defeats Gore.” *New York Times*, January 7, Internet edition.

Rousseeuw, P. J. 1984. “Least Median of Squares Regression.” *Journal of the American Statistical Association* 79 (December): 871–80.

Rousseeuw, Peter J., and Christophe Croux. 1993. “Alternatives to the Median Absolute Deviation.” *Journal of the American Statistical Association* 88 (December): 1273–83.

Rousseeuw, Peter J., and Annick M. Leroy. 1987. *Robust Regression and Outlier Detection*. New York: Wiley.

Sekhon, Jasjeet Singh, and Walter R. Mebane, Jr. 1998. “Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models.” *Political Analysis* 7: 189–203.

Sinclair, Robert C., Melvin M. Mark, Sean E. Moore, Carrie A. Lavis, and Alexander S. Soldat. 2000. “An Electoral Butterfly Effect.” *Nature* 408 (December 7): 665–6.

Stout, David. 2000. “As Electors Cast Their Votes, Bush Confers in Washington.” *New York Times*, December 18, Internet edition.

Stromberg, Arnold J. 1993. “Computation of High Breakdown Nonlinear Regression Parameters.” *Journal of the American Statistical Association* 88 (March): 237–44.

Van Natta, Don, Jr. 2000. “Gore to Contest Recount Result in Palm Beach.” *New York Times*, November 25, Internet edition.

Van Natta, Don, Jr., and Michael Moss. 2000. “Democratic ‘War Room’ Tries to Oversee the Battle for Florida, to Mixed Results.” *New York Times*, November 11, Internet edition.

Wakin, Daniel J. 2000. “Bush Campaign Pushes for a Gore Concession.” *New York Times*, November 10, Internet edition.

Western, Bruce. 1995. “Concepts and Suggestions for Robust Regression Analysis.” *American Journal of Political Science* 39 (August): 786–817.

White, Halbert. 1994. *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1990. *Introductory Statistics*, 5th ed. New York: Wiley.