What Happened in Palm Beach County?

To: All who are Interested

November 9, 2000

From: Henry E. Brady   Minor Corrections: *7 p.m.*
        Professor of Political Science and Public Policy
        Director, Survey Research Center
        University of California, Berkeley
Re: What Happened in Palm Beach County?

It is now well-known that the design of the ballot in Palm Beach County could have led to numerous mistakes in voting. Moreover, it is easy to show that the Buchanan vote in Palm Beach County is a substantial outlier considering the Bush and Gore votes in that county. Buchanan obtained about .8% of the vote, but his expected vote was probably no more than .2% to .3%. Indeed, statewide, Buchanan received about .3% of the three candidate (Bush, Gore, and Buchanan) vote. Thus, something like 2000 of Buchanan's 3,407 votes in Palm Beach County may have been mistakes.

The question is: Who made the mistakes? Was it voters who wanted to vote for Bush or those who wanted to vote for Gore? Given the construction of the ballot, it is possible that voters from either camp made the mistakes.

This memorandum shows that almost all of the mistakes probably came from voters who intended to vote for Al Gore but who mistakenly voted for Pat Buchanan. Thus, as many of 2000 of the votes for Pat Buchanan in Palm Beach County were probably mistakenly cast by Gore supporters.

Data Used – The data on Florida counties used in this memorandum were obtained from the CBS news web site and the precinct level data were obtained from the Palm Beach County web site. The files were created quickly, but tests were conducted to make sure that they were correct. Nevertheless, it seems likely that further work will refine this analysis.

Summary of the Argument – A prima facie case can be made for the proposition that mostly Gore voters mistakenly voted for Buchanan in Palm Beach County by considering the relationship between the Buchanan vote and Bush vote across Florida's counties compared to the observed relationship between the Buchanan and Bush vote in Palm Beach Counties' precincts. Across the 67 counties of Florida, Buchanan vote increases with Bush vote. Thus, across these counties, the fraction of Buchanan voters in the county <u>increases</u> by about .9% for every ten percent increase in Bush vote. Across the precincts of Palm Beach County, however, Buchanan vote <u>decreases</u> with Bush vote.[1] Thus, in the precincts of Palm Beach County, the fraction of Buchanan voters in the precinct <u>decreases</u> by about 1.4% for every ten percent increase in Bush vote.

---

[1]   All of this analysis could have been done with Gore vote which is negatively correlated with Bush vote at about -.999 in the counties and -.993 across the precincts.

There is no reason, however, to suppose that Palm Beach County Buchanan voters are substantially different from other Buchanan voters in the state of Florida. Thus, it is hard to square the results across all Florida counties with those across Palm Beach County precincts unless voters who intended to vote for Al Gore mistakenly marked the ballot for Pat Buchanan. If this happened, then the observed Buchanan vote would be composed of "true" Buchanan supporters whose percentage would increase with the percentage of the Bush vote in the precincts of Palm Beach County and "mistaken" Buchanan voters whose percentage would increase with the percentage of the Gore vote across the precincts. Because the number of mistaken Buchanan supporters is greater than the number of true supporters, we find that the percent of observed Buchanan voters in Palm Beach county precincts increases with the observed percent of Gore supporters.[2]

A Model of What is Going On – The following develops a simple mathematical model of what is going on, and it presents statistical estimates, using ordinary least squares, of the size of the effects.

Assume that in each precinct in Palm Beach County there was a true George W. Bush vote percentage denoted by (TGB), a true Al Gore vote denoted by (TAG), and a true Pat Buchanan vote denoted by (TPB). We do not observe these votes, however. Instead, we have observed George W. Bush votes (OGB), an observed Al Gore vote (OAG), and an observed Pat Buchanan vote (OPB). Then assume that a fraction (*f*) of people in each precinct who wanted to vote for George W. Bush mistakenly voted for Pat Buchanan and that a fraction (*g*) of people in each precinct who wanted to vote for Al Gore mistakenly voted for Pat Buchanan.

Then we have:

(1) $OPB = TPB + f*TGB + g*TAG$

(2) $OGB = (1\text{-}f)*TGB$

(3) $OAG = (1\text{-}g)*TAG$

The first equation says that the observed Buchanan vote consists of the true Buchanan vote plus a fraction of Bush supporters who mistakenly voted for Buchanan and a fraction of Gore supporters who mistakenly voted for Buchanan. The second equation says that the observed Bush vote is equal to the true Bush vote reduced by the fraction of Bush supporters who mistakenly voted for Buchanan. The third equation says that the observed Gore vote is equal to the true Gore vote reduced by those who mistakenly voted for Buchanan. Note that this model ignores the possibility that Buchanan voters may have mistakenly voted for Bush or Gore. Given the small numbers of Buchanan voters, this must be considered a second order and essentially negligible possibility.

With the Palm Beach County precinct level data, we can observe the means for the three

---

[2] There is also the possibility of some other kind of ecological inference problem, but it is hard to develop a convincing alternative hypothesis.

observed variables, their variances, and their covariances. Consider the means which we denote by "expectation" symbols "E". That is, E(OPB) = Mean of OPB:

(4)   E(OPB) = E(TPB) + f * E(TGB) + g*E(TAG)

(5)   E(OGB) = $(1-f)$*E(TGB)

(6)   E(OAG) = $(1-g)$*E(TAG)

Now, in this analysis, I have just considered proportions of the three candidate vote, Buchanan, Bush, and Gore. Analytically, this probably has only a small impact on the total result. It does mean, however, that the three equations above have to sum to one by the assumption that we are only considering shares of the three candidate vote. (Adding other candidates might help a bit, but probably not very much unless we were sure that their vote was unaffected by the ballot form in Palm Beach County. Even if, say, Nader's share was not much affected, I would be nervous about predicting other candidate's vote from Nader's share.)

From these three equations, we find that the expected observed Pat Buchanan share is:

(7)   E(OPB) = E(TPB) + [$f/(1-f)$]*E(OGB) + [$g/(1-g)$]*E(OAG).

Note that if f and g are zero, then as we would expect, E(OPB) = E(TPB). But if they are not, then the observed Pat Buchanan will be bigger than it should be.

If we take state of Florida county by county data and we omit Palm Beach County, we can estimate the relationship between true Buchanan vote and true Bush vote. The results are, using OLS (without weighting for size of county):

(8)   E(TPB) = -.001607   + .01123 * E(TGB)

This suggests that as the true Bush vote increases, the true Pat Buchanan vote increases. I assume that this same relationship holds in Palm Beach County.

Assume that Bush Voters Don't Make Mistakes – Now, <u>if f is zero</u> and "true" Bush voters don't make mistakes on the ballot in Palm Beach County, then observed Bush vote is the true Bush vote and we can estimate the expected true Pat Buchanan vote in each precinct using (8). In fact, we can just substitute (8) into (7) and assume that $f=0$ so that E(OGB) = E(TGB):

(9) E(OPB) = -.001607   + .01123 * E(OGB) + [$g/(1-g)$]*E(OAG).
But we can't just regress OPB on OGB and OAG because we know that OPB = 1 - OGB - OAG. Hence, we must substitute out for OGB and get the following:

(10)   E(OPB) = [(-.001607 + .01123)/1.01123] + E(OAG)* {[$g/(1-g)$] - .01123}/1.01123

With some algebra, we get:

(11) $E(OPB) = .00964 + E(OAG)* \{[g/(1-g)] - .01123\}/1.01123$

This equation predicts that in a regression of the observed Pat Buchanan vote on the observed Al Gore vote in Palm Beach County, the sign of the Gore term will be positive if g is large enough to overcome the fact that in most places in the state, the vote for Pat Buchanan is negatively correlated with that for Gore. In fact, a regression of OPB on OGB (with no constraints on the intercept term), leads to a statistically significant coefficient for OGB of about .0116. This suggests a value of g of about 2.2%.[3] It is hard to exact error bounds for these numbers given the complexities of this model and my quick estimation methods, but something like one-half percent seems reasonable based upon my output.

Now, the fraction of the three candidate (Bush, Gore, Buchanan) vote that went to Buchanan in Palm Beach County is about .8%. This figure is about .5% to .6% higher than we would expect given the observed Bush Vote and equation (8) above and assuming that the observed Bush vote in Palm Beach County is equal to the true Bush vote.

The observed Gore vote was about 63% and the true vote can't be much higher than that. Hence, if about 2.2% to 2.7% of the true Gore voters mistakenly voted for Buchanan, we would expect to find about 1.4% to 1.7% more Buchanan voters than we observe. But we do not find quite this many, suggesting that this estimate is somewhat high. If we consider the error bounds suggested above, then these estimates might have a lower bound of something like .6% which is similar to what we observe. In addition, the estimate might be high because of the use of ordinary least squares in an ecological regression which sometimes provides estimates that are higher than more refined methods.

Assume that Gore Voters Don't Make Mistakes – The analysis above assumed that Bush voters don't make mistakes. But that may be false. Maybe it was Bush voters who made the mistakes. If that is true, then we can do an analysis like the one above.

Now we assume that _g is zero_ and "true" Gore voters don't make mistakes on the ballot in Palm Beach County. Then the observed Gore vote is the true Gore vote and we can estimate the expected true Pat Buchanan vote in each precinct using an equation like (8) based upon the county level data. The equation for Gore from the county data (excluding Palm Beach County) is the following:

(12) $E(TPB) = .01 - .01225 * E(TAG)$

We substitute (12) into (7) (remembering that TAG = OAG) and assume that _g=0_:

(13) $E(OPB) = .01 - .01225*E(OAG) + [f/(1-f)]*E(OGB)$.

---

[3] In an earlier version of this paper, the intercept term was mistakenly reported as -.001589. In this version of the paper, I also forced the regression line to go through the intercept term which is asking a lot of the data. With the new value of the intercept term, it appears to be asking too much, and I think that it is best to go with the unconstrained estimation reported in the text.

But we can't just regress OPB on OAG and OGB because we know that OPB = 1 - OGB - OAG. Hence, we must substitute out for OAG and get the following:

(10) $E(OPB) = -.00228 + E(OGB)* \{[f/(1-f)] + .01225\}/.98775$

This equation predicts that in a regression of the observed Pat Buchanan vote on the observed George Bush vote in Palm Beach County, the sign of the Bush term will be positive because the term in braces is always positive.

In fact, we find that the regression coefficient is -.0136 with a standard error of .02. Hence the coefficient has the wrong sign and it is statistically insignificant. *The data decisively reject the model in which all the "extra" Buchanan vote comes from Bush supporters who mistakenly marked Buchanan on the ballot.*

A Mixed Model – This still leaves open the possibility that the correct model involves both *f* and *g*. In this case, we get the following equation:

(11) $E(OPB) = \{.01 + [f/(1-f)]\}/\{1 + [f/(1-f)]\} + E(OAG)*\{[g/(1-g)] - [f/(1-f)] - [.01225/(1-g)]\}/\{1+[f/(1-f)]\}$

In order to estimate the parameters of this equation, we have to use both the estimate of the intercept and the regression coefficient. The intercept "a" of a regression of the observed Pat Buchanan vote on the observed Al Gore vote equals the following:

$a = \{.01 + [f/(1-f)]\}/\{1 + [f/(1-f)]\}$

and the regression coefficient, "b," equals:

$b = \{[g/(1-g)] - [f/(1-f)] - [.01225/(1-g)]\}/\{1+[f/(1-f)]\}.$

Note that the intercept *a* must be greater than zero but the slope *b* can have any sign depending upon the values of *g* and *f*.

Equation (11) presents a messy estimation problem if we try to impose the constraints implied by the non-linear way in which *f* and *g* enter the equation. But a simple and statistically acceptable way to proceed is to just estimate the regression of the observed Buchanan vote across the precincts on the observed Gore vote. If we estimate this regression we get the following values for the intercept and the slope coefficients:

a = .001411, with a standard error of .001

b = .0116, with a standard error of .002.

The only plausible value of *f* is zero from this estimation because the estimated intercept is so near zero, and any value of *f* greater than zero will imply that the intercept should be even larger. The most plausible value of g, once *f* is set to zero, is what we obtained before, namely about 2.2%.

Conclusion -- Using data from the 67 Florida counties along with data from precincts in Palm Beach County, I find that there is a strong likelihood that over 2000 of the Buchanan votes in Palm Beach County were cast by Gore supporters who mistakenly punched Buchanan's name. There is no evidence for the proposition that the extra Buchanan voters were Bush supporters.