

Who Won Florida? Are the Palm Beach Votes Irregular?

Bruce E. Hansen
Stockwell Professor of Economics
Department of Economics
University of Wisconsin
<http://www.ssc.wisc.edu/~bhansen>
bhansen@ssc.wisc.edu

November 11, 2000

1 Summary

A county-level demographic analysis of the November 2000 Presidential vote in Florida shows that the Palm Beach vote is irregular. We can state with 99.9% confidence that 2058 of the 3407 votes cast for Patrick Buchanan are inconsistent with the demographic characteristics of Palm Beach.

A 99.9% confidence interval¹ for the correct Palm Beach vote is [251, 1349]. A vote outside this interval should be viewed as extremely unlikely and irregular. In particular, the reported vote of 3407 should be dismissed as irregular.

An implication is that at least 2058 of the votes attributed to Buchanan are irregular.

Our demographic analysis shows precisely why we can make this claim. The voting patterns of Florida counties show a strong relationship between the votes cast for Buchanan and the demographic characteristics of the county. Factors which contribute to *low* votes for Buchanan are *high* percentages of elderly, Blacks, Hispanics, and college education, and *high* median household income. Palm Beach has a high percentage of elderly and college education, and a high median household income. Palm Beach is therefore expected to have a *low* percentage vote for Buchanan. The recorded high percentage vote is inconsistent with this prediction.

There have been an explosion of papers written on this subject in the past few days. All of the analysis I have seen has looked only at the relationship between the votes for the various candidates. The analysis described here is the first I have seen which explores the Florida vote using cross-sectional demographic variation. Hopefully this sheds additional light on the subject.

The data and programs used in this analysis can be found on my webpage.

¹This means that out of 1000 forecasts, this interval will contain the correct vote 999 times.

2 Data Description

As of the November 10 recount, there were 431,621 votes cast in the Presidential race², with 269,696 (62.5%) for Gore, 152,954 (35.4%) for Bush, 3407 (0.79%) for Buchanan, and 5564 (1.3%) for Nader.

The 3407 votes for Buchanan were the most votes he received in any one Florida county. While this has received great attention in the press, this fact alone should not be viewed as surprising as Palm Beach is one of the largest counties in Florida (with a population just over one million). The relevant measure is the percentage vote. Interestingly, Buchanan's percentage vote, 0.79%, is not out-of-line when compared with other Florida counties. Seven of the other 66 counties had higher percentages for Buchanan; one – Calhoun county – cast 1.7% of its votes for Buchanan.

The raw percentage vote is not very revealing, as the Florida counties are quite heterogeneous. They vary in known³ demographic characteristics such as population and the percentage of the population which are over the age of 65, Black, Hispanic, college-educated, and median household income.

3 Regression Analysis

The analysis is based on a simple regression of the natural logarithm of the percentage of votes cast for Buchanan (P) as a linear function of the percentage of the population over age 65 (A_{65}), the percentage of blacks (B), the percentage of Hispanics (H), the percentage who are college educated (C), and median household income in thousands of dollars (I).

Estimation is by least-squares using the 66 Florida counties *other than* Palm Beach. The estimated equation is

$$\log(P) = 2.146 - 0.0415 A_{65} - 0.0132 B - 0.0349 H - 0.0193 C - 0.0658 I.$$

(0.396) (0.0057) (0.0057) (0.0050) (0.0068) (0.00113)

$$R^2 = 0.73 \quad \hat{\sigma} = 0.334 \quad n = 66$$

Heteroskedasticity-consistent standard errors are reported in parenthesis.

For forecasts of P , we also need an estimate of the conditional variance of the equation error. Since P is a sample average we expect its variance to be roughly inversely proportional to the population size of the county (N). Thus the variance of $\log(P)$ is expected to be roughly inversely proportional to $\log(N)$. We therefore specify the variance of the equation error as a function of $\log(N)$. (Spot checking indicated that no other variable was relevant.) In the Figure we plot a nonparametric estimate⁴ of the equation error variance as a function of the natural logarithm of county population. The relationship is negative for small and moderate populations, leveling out roughly at 0.057 for counties with populations above

²Source: http://www.abcnews.go.com/sections/politics/2000vote/general/FL_county.html

³Source: U.S. Census Bureau. The population variables are for 1996, college attainment is for 1990, and income is for 1993.

⁴A kernel regression of the squared OLS residual on the log of population, using an Epanechnikov kernel and a bandwidth of 2.3.

60,000. The estimated standard deviation for Palm Beach is 0.24. In contrast, if the variance were treated as constant, the regression standard deviation is about 39% higher.

From these estimates and the demographic characteristics of Palm Beach, we can calculate point and interval forecasts for the Palm Beach vote using equation (3) of the Appendix. Briefly, the point forecast for $\log(P)$ is -2.004 with a standard error of 0.2556 (the latter properly taking into account *both* estimation error and the equation error variance). This allows us to calculate a point forecast for P of 0.1387% and hence for the votes cast for Buchanan of 599. The 99.9% confidence interval for $\log(P)$ is $[-2.845, -1.163]$, for P is $[0.0581, 0.313]$, and for the number of votes cast for Buchanan is $[251, 1349]$.

Another measure of the degree of irregularity is to observe that based on the fitted regression for $\log(P)$, Palm Beach has a “prediction error” of 1.77. This is the difference between the predicted and actual values. In sample, the regression residuals vary from -0.72 to 0.89. The prediction error is twice as large as the largest regression residual!

4 Interpretation of Regression Equation

The regression equation takes a semi-log form. So the regression coefficients, multiplied by 100, are the estimated percentage change in P (the percentage of votes cast for Buchanan) due to a one-unit change in the regression variable. This is also the estimated percentage change in the number of votes for Buchanan. Hence:

- A 1% increase in the population over age 65 decreases Buchanan’s vote by 4.1%.
- A 1% increase in the Black population decreases Buchanan’s vote by 1.3%.
- A 1% increase in the Hispanic population decreases Buchanan’s vote by 3.5%.
- A 1% increase in the College-education population decreases Buchanan’s vote by 1.9%.
- A \$1000 increase in median household income decreases Buchanan’s vote by 6.6%.

Among counties in Florida, the population in Palm Beach is largely elderly (23.7%), college-educated (22.1%), and high-income (\$33,518). This is why the regression analysis predicts a low vote for Buchanan, implying that the recorded high vote for Buchanan is inaccurate.

5 Appendix: Forecast Intervals

Let the regression be written as $\log(P_i) = x'_i\beta + e_i$ with point estimate $\hat{\beta}$ and estimated covariance matrix \hat{V} . We assume that e_i is approximately normally distributed. Let x denote the values of the regressors for Palm Beach. Let $\hat{\sigma}^2(x)$ denote the kernel estimate of $\text{Var}(e_i | x_i = x)$ described in the text.

The point forecast for $\log(P)$ is $x'\hat{\beta}$ with standard error

$$\hat{s} = \sqrt{x'\hat{V}x + \hat{\sigma}^2(x)} \quad (1)$$

and approximate 99.9% confidence region

$$\left[x'\hat{\beta} - 3.291 \cdot \hat{s}, \quad x'\hat{\beta} + 3.291 \cdot \hat{s} \right]. \quad (2)$$

The point forecast for P is $\exp(x'\hat{\beta} + \hat{\sigma}^2(x)/2)$, which takes into account the nonlinearity in expectations. The approximate 99.9% confidence region is found by exponentiating the endpoints of (2), namely,

$$\left[\exp(x'\hat{\beta} - 3.291 \cdot \hat{s}), \quad \exp(x'\hat{\beta} + 3.291 \cdot \hat{s}) \right].$$

Let N denote the number of votes cast in Palm Beach. The point forecast for the number of votes, $V = N \cdot P/100$, is $N \cdot \exp(x'\hat{\beta} + \hat{\sigma}^2(x)/2) / 100$. The 99.9% confidence interval is

$$\left[N \cdot \exp(x'\hat{\beta} - 3.291 \cdot \hat{s}) / 100, \quad N \cdot \exp(x'\hat{\beta} + 3.291 \cdot \hat{s}) / 100 \right]. \quad (3)$$

This is the formula for the confidence intervals reported in the text.