

# A Statistical Assessment of Buchanan's Vote in Palm Beach County

Richard L. Smith

*Abstract.* This article presents a statistical analysis of the results of the 2000 U.S. presidential election in the 67 counties of Florida, with particular attention to the result in Palm Beach county, where the Reform party candidate Pat Buchanan recorded an unexpectedly large 3,407 votes. It was alleged that the “butterfly ballot” had misled many voters into voting for Buchanan when they in fact intended to vote for Al Gore. We use multiple regression techniques, using votes for the other candidates and demographic variables as covariates, to obtain point and interval predictions for Buchanan's vote in Palm Beach based on the data in the other 66 counties of Florida. A typical result shows a point prediction of 371 and a 95% prediction interval of 219–534. Much of the discussion is concerned with technical aspects of applying multiple regression to this kind of data set, focussing on issues such as heteroskedasticity, overdispersion, data transformations and diagnostics. All the analyses point to Buchanan's actual vote as a clear and massive outlier.

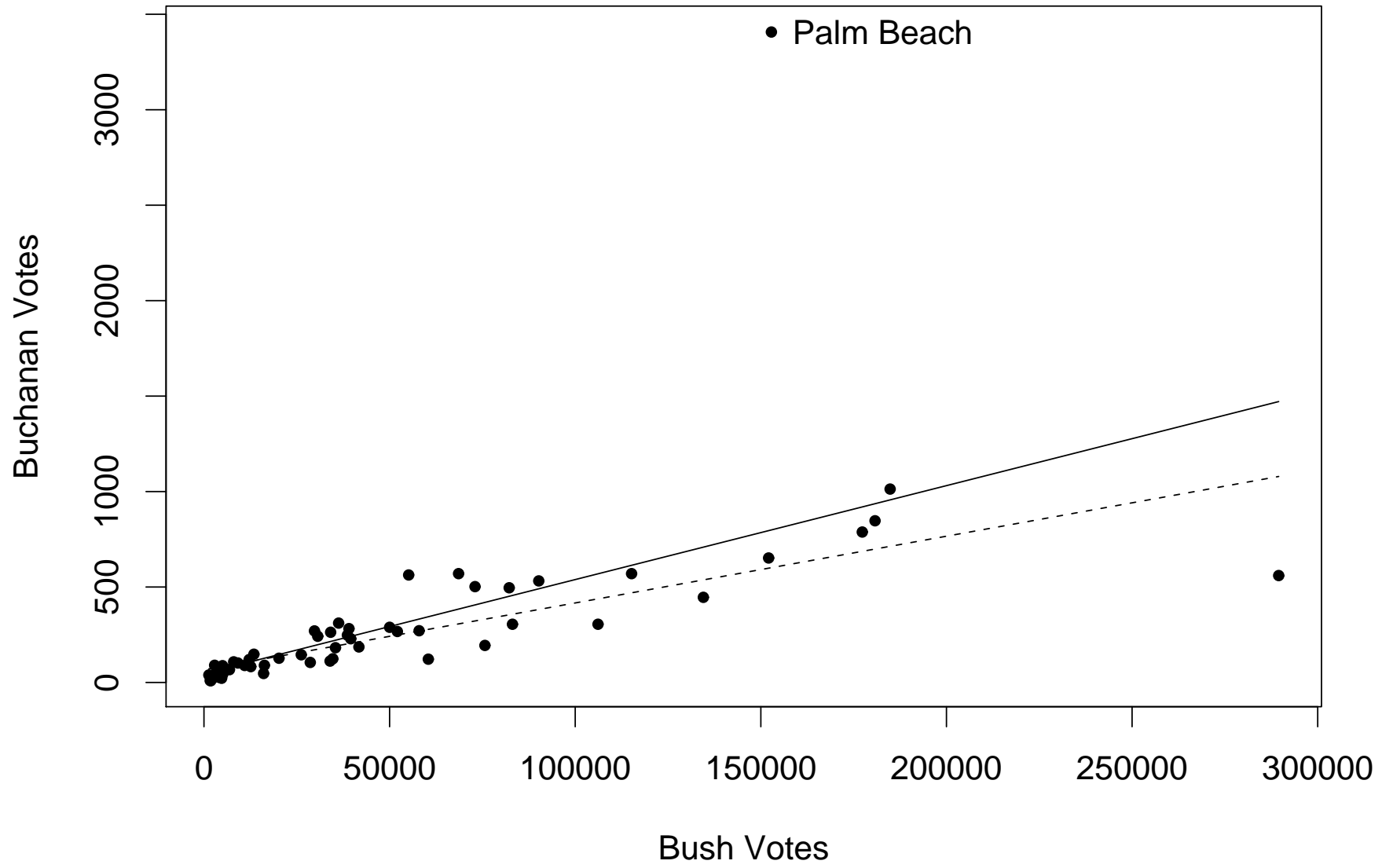
OFFICIAL BALLOT, GENERAL ELECTION  
PALM BEACH COUNTY, FLORIDA  
NOVEMBER 7, 2000

<p><b>ELECTORS FOR PRESIDENT AND VICE PRESIDENT</b></p> <p>(A vote for the candidates will actually be a vote for their electors.)</p> <p>(Vote for Group)</p>	<p>(REPUBLICAN)</p> <p><b>GEORGE W. BUSH</b> - PRESIDENT      3 →</p> <p><b>DICK CHENEY</b> - VICE PRESIDENT</p>
	<p>(DEMOCRATIC)</p> <p><b>AL GORE</b> - PRESIDENT      5 →</p> <p><b>JOE LIEBERMAN</b> - VICE PRESIDENT</p>
	<p>(LIBERTARIAN)</p> <p><b>HARRY BROWNE</b> - PRESIDENT      7 →</p> <p><b>ART OLIVIER</b> - VICE PRESIDENT</p>
	<p>(GREEN)</p> <p><b>RALPH NADER</b> - PRESIDENT      9 →</p> <p><b>WINONA LaDUKE</b> - VICE PRESIDENT</p>
	<p>(SOCIALIST WORKERS)</p> <p><b>JAMES HARRIS</b> - PRESIDENT      11 →</p> <p><b>MARGARET TROWE</b> - VICE PRESIDENT</p>
	<p>(NATURAL LAW)</p> <p><b>JOHN HAGELIN</b> - PRESIDENT      13 →</p> <p><b>NAT GOLDBABER</b> - VICE PRESIDENT</p>

OFFICIAL BALLOT, GENERAL ELECTION  
PALM BEACH COUNTY, FLORIDA  
NOVEMBER 7, 2000

← 4	<p>(REFORM)</p> <p><b>PAT BUCHANAN</b> - PRESIDENT</p> <p><b>EZOLA FOSTER</b> - VICE PRESIDENT</p>
← 6	<p>(SOCIALIST)</p> <p><b>DAVID McREYNOLDS</b> - PRESIDENT</p> <p><b>MARY CAL HOLLIS</b> - VICE PRESIDENT</p>
← 8	<p>(CONSTITUTION)</p> <p><b>HOWARD PHILLIPS</b> - PRESIDENT</p> <p><b>J. CURTIS FRAZIER</b> - VICE PRESIDENT</p>
← 10	<p>(WORKERS WORLD)</p> <p><b>MONICA MOOREHEAD</b> - PRESIDENT</p> <p><b>GLORIA La RIVA</b> - VICE PRESIDENT</p>
<p><b>WRITE-IN CANDIDATE</b></p> <p>To vote for a write-in candidate, follow the directions on the long stub of your ballot card.</p>	

**TURN PAGE TO CONTINUE VOTING** →



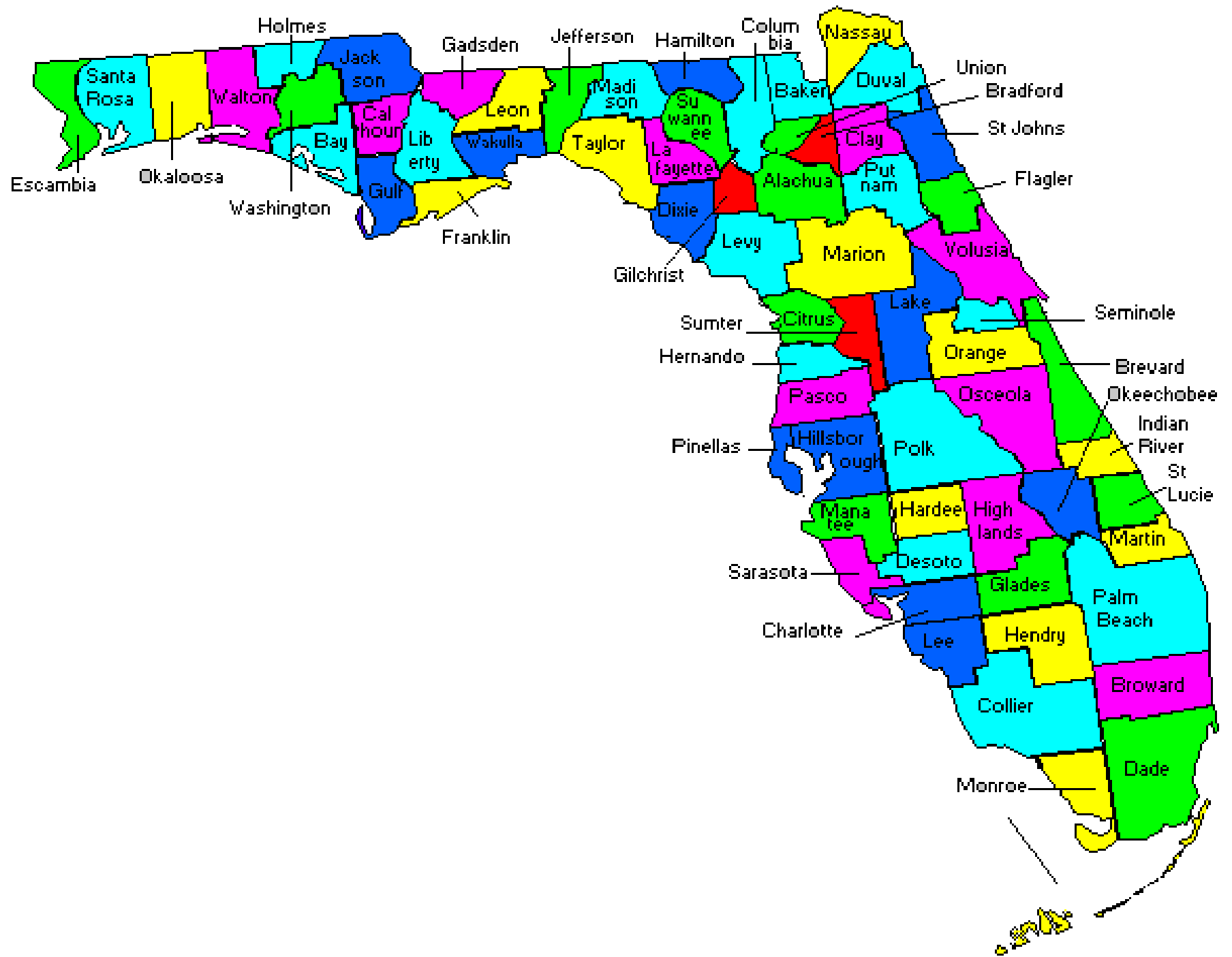
# Objective

Buchanan's vote of 3407 in Palm Beach County appears to be a gross outlier compared with his votes in other Florida counties. Crude analyses suggest his natural vote should have been about 500.

Concept of statistical analysis:

Predict Buchanan's vote from relevant covariates (other candidates' votes plus demographics) in other 66 counties, then apply resulting regression model to Palm Beach

In this way, we hope to determine what his "natural vote" would have been



County voting data, part I

County	Bush	Gore	Brow	Nade	Har	Hag	Buc	Mc	Ph	Mo
Alachua	34,124	47,365	658	3226	6	42	263	4	20	21
Baker	5,610	2,392	17	53	0	3	73	0	3	3
Bay	38,637	18,850	171	828	5	18	248	3	18	27
Bradford	5,414	3,075	28	84	0	2	65	0	2	3
Brevard	115,185	97,318	643	4470	11	39	570	11	72	76
Broward	177,323	386,561	1212	7101	50	129	788	34	74	124
Calhoun	2,873	2,155	10	39	0	1	90	1	2	3
Charlotte	35,426	29,645	127	1462	6	15	182	3	18	12
Citrus	29,765	25,525	194	1379	5	16	270	0	18	28
Clay	41,736	14,632	204	562	1	14	186	3	6	9
Collier	60,433	29,918	185	1399	7	34	122	4	10	29
Columbia	10,964	7,047	127	258	1	7	89	2	8	5
Desoto	4,256	3,320	23	157	0	0	36	3	8	2
Dixie	2,697	1,826	32	75	0	2	29	0	3	2
Duval	152,098	107,864	952	2757	37	162	652	15	58	41
Escambia	73,017	40,943	296	1727	6	24	502	3	110	20
Flagler	12,613	13,897	60	435	1	4	83	3	3	12
Franklin	2,454	2,046	17	85	1	3	33	0	3	2
Gadsden	4,767	9,735	24	139	3	4	38	4	7	6
Gilchrist	3,300	1,910	52	97	0	1	29	0	2	4
Glades	1,841	1,442	12	56	0	3	9	1	0	1
Gulf	3,550	2,397	21	86	2	4	71	2	2	9
Hamilton	2,146	1,722	12	37	4	1	23	8	7	4
Hardee	3,765	2,339	17	75	0	2	30	0	2	3
Hendry	4,747	3,240	11	103	3	1	22	2	7	2
Hernando	30,646	32,644	116	1501	8	26	242	4	10	22
Highlands	20,206	14,167	64	545	6	16	127	3	7	8
Hillsborough	180,760	169,557	1138	7490	35	217	847	29	68	154
Holmes	5,011	2,177	18	94	1	7	76	3	6	2
Indian River	28,635	19,768	122	950	4	13	105	2	13	10
Jackson	9,138	6,868	40	138	0	2	102	1	4	7
Jefferson	2,478	3,041	14	76	2	1	29	1	0	0
Lafayette	1,670	789	6	26	2	0	10	1	1	0
Lake	50,010	36,571	204	1460	4	36	289	1	21	15
Lee	106,141	73,560	538	3587	30	81	305	5	34	96
Leon	39,053	61,425	330	1932	9	28	282	7	16	31
Levy	6,858	5,398	92	284	1	1	67	1	10	12
Liberty	1,317	1,017	12	19	0	3	39	0	1	2

TABLE 2  
*County voting data, part II*

County	Bush	Gore	Brow	Nade	Har	Hag	Buc	Mc	Ph	Mo
Madison	3,038	3,014	18	54	0	2	29	1	1	5
Manatee	57,952	49,177	242	2491	5	35	271	3	19	26
Marion	55,141	44,665	662	1809	13	26	563	6	22	49
Martin	33,970	26,620	109	1118	14	29	112	7	20	14
Miami-Dade	289,492	328,764	760	5352	87	119	560	35	69	124
Monroe	16,059	16,483	162	1090	1	26	47	0	3	7
Nassau	16,280	6,879	62	253	0	7	90	4	3	3
Okaloosa	52,093	16,948	313	985	4	15	267	2	33	20
Okeechobee	5,057	4,588	21	131	1	4	43	1	3	4
Orange	134,517	140,220	891	3879	13	65	446	7	41	46
Osceola	26,212	28,181	309	732	10	20	145	5	10	33
Palm Beach	152,846	268,945	743	5564	45	143	3407	302	188	103
Pasco	68,582	69,564	413	3393	19	83	570	14	16	77
Pinellas	184,823	200,629	1230	10022	41	442	1013	27	72	170
Polk	90,180	75,193	365	2062	8	59	532	5	46	36
Putnam	13,447	12,102	114	377	2	7	148	3	10	12
Santa Rosa	36,274	12,802	131	724	1	13	311	1	43	19
Sarasota	83,100	72,853	431	4069	11	94	305	5	15	59
Seminole	75,677	59,174	550	1946	6	38	194	5	18	26
St. Johns	39,546	19,502	210	1217	4	11	229	2	12	13
St. Lucie	34,705	41,559	165	1368	4	12	124	10	13	29
Sumter	12,127	9,637	53	306	2	2	114	0	3	17
Suwannee	8,006	4,075	52	180	2	4	108	0	9	5
Taylor	4,056	2,649	4	59	0	3	27	1	8	1
Union	2,332	1,407	15	33	1	0	37	0	1	0
Volusia	82,214	97,063	442	2903	8	36	496	5	20	69
Wakulla	4,512	3,838	30	149	2	3	46	1	0	6
Walton	12,182	5,642	68	265	3	11	120	2	7	18
Washington	4,994	2,798	32	93	0	2	88	0	9	5

# Demographic Variables

- Pop: county population in 1997,
- Whi: percentage of whites in 1996,
- Bla: percentage of blacks in 1996,
- Hisp: percentage of Hispanics in 1996,
- $\geq 65$ : percentage of the population aged 65 and over
- HS: percentage of the population graduating from high school (1990 census),
- Coll: percentage of the population graduating from college (1990 census),
- Inc: Mean personal income (1994).



TABLE 3  
County demographic data, part I

County	Pop	Whi	Bla	Hisp	≥ 65	HS	Coll	Inc
Alachua	198,326	74.4	21.8	4.7	9.4	82.7	34.6	19,412
Baker	20,761	82.4	16.8	1.5	7.7	64.1	5.7	14,859
Bay	146,223	84.2	12.4	2.4	11.9	74.7	15.7	17,838
Bradford	24,646	76.1	22.9	2.6	11.8	65.0	8.1	13,681
Brevard	460,977	88.3	9.2	4.1	16.5	82.3	20.4	19,567
Broward	1,470,758	80.3	17.5	10.9	20.3	76.8	18.8	24,706
Calhoun	12,337	81.6	16.9	1.6	14.3	55.9	8.2	12,570
Charlotte	133,681	94.3	4.4	3.4	33.4	75.7	13.4	18,977
Citrus	112,454	96.2	2.8	2.5	30.7	68.6	10.4	16,060
Clay	135,179	91.0	6.0	3.5	7.9	81.2	17.9	18,598
Collier	195,731	93.3	5.7	17.1	21.5	79.0	22.3	30,906
Columbia	52,856	78.3	20.5	1.9	12.3	69.0	11.0	15,349
Desoto	26,259	80.6	18.1	12.1	18.0	54.5	7.6	16,544
Dixie	12,563	89.8	9.5	1.2	14.4	57.7	6.2	12,035
Duval	732,622	69.4	27.5	3.4	10.7	76.9	18.4	20,686
Escambia	282,604	73.3	22.7	2.6	11.7	76.2	18.2	17,661
Flagler	46,128	88.5	9.8	5.9	23.0	78.7	17.3	15,613
Franklin	10,133	84.5	14.5	1.0	17.8	59.5	12.4	15,735
Gadsden	45,441	37.6	61.8	2.9	11.6	59.9	11.2	14,416
Gilchrist	13,367	90.0	9.3	2.1	13.0	63.0	7.4	12,865
Glades	9,698	79.6	13.7	10.1	15.3	57.4	7.1	14,789
Gulf	13,926	73.9	25.2	1.1	13.6	66.4	9.2	15,482
Hamilton	12,521	56.3	43.0	3.6	10.9	58.4	7.0	12,357
Hardee	22,113	93.1	5.9	28.4	13.3	54.8	8.6	16,812
Hendry	31,634	78.2	18.8	26.6	9.9	56.6	10.0	17,823
Hernando	125,537	94.4	4.6	4.0	29.6	70.5	9.7	16,062
Highlands	76,854	87.1	11.6	6.7	32.4	68.2	10.9	17,655
Hillsborough	909,444	82.8	14.9	16.0	12.3	75.6	20.2	20,167
Holmes	18,382	91.7	6.5	1.7	15.5	57.1	7.4	12,790
Indian River	99,215	89.2	9.9	3.9	26.6	76.5	19.1	28,977
Jackson	45,706	69.5	29.6	3.5	14.4	61.6	10.9	15,519
Jefferson	13,232	49.4	50.1	1.3	13.4	64.1	14.7	15,574
Lafayette	6,289	83.0	16.4	5.1	10.7	58.2	5.2	13,663
Lake	196,214	88.2	10.9	3.8	26.3	70.6	12.7	18,269
Lee	387,091	91.1	7.8	5.9	24.4	76.9	16.4	22,053
Leon	215,170	70.4	27.3	3.1	8.4	84.9	37.1	16,705
Levy	32,254	84.4	14.2	2.6	17.6	62.8	8.3	13,745
Liberty	6,703	78.1	20.9	3.1	10.7	56.7	7.3	14,896

TABLE 4  
*County demographic data, part II*

County	Pop	Whi	Bla	Hisp	≥ 65	HS	Coll	Inc
Madison	17,558	53.9	45.6	1.9	13.8	56.5	9.7	13,002
Manatee	237,159	89.8	9.0	5.8	27.8	75.6	15.5	23,031
Marion	237,308	84.3	14.6	4.0	21.4	69.6	11.5	14,502
Martin	116,087	91.8	6.9	6.2	26.6	79.7	20.3	31,996
Miami–Dade	2,044,600	77.0	21.2	54.4	14.4	65.0	18.8	20,014
Monroe	81,919	92.3	6.2	15.8	15.9	79.7	20.3	25,160
Nassau	54,096	87.3	11.9	1.5	9.8	71.2	21.5	20,874
Okaloosa	167,580	85.3	10.3	4.2	9.1	83.8	21.0	18,959
Okeechobee	33,102	91.1	7.5	14.8	14.6	59.1	9.8	15,162
Orange	783,974	79.1	17.5	12.3	10.4	78.8	21.2	20,469
Osceola	142,128	90.7	6.6	15.3	13.2	73.7	11.2	16,256
Palm Beach	1,018,524	83.9	14.4	9.8	23.7	78.8	22.1	33,518
Pasco	320,253	96.5	2.3	4.4	32.0	66.9	9.1	16,924
Pinellas	871,766	89.1	9.0	3.1	26.6	78.1	18.5	24,796
Polk	448,646	83.3	15.4	5.3	18.2	68.0	12.9	17,824
Putnam	70,430	78.1	20.9	3.4	17.6	64.3	8.3	14,250
Santa Rosa	114,481	92.6	4.6	2.0	8.9	78.5	18.6	17,127
Sarasota	301,644	94.0	5.1	2.8	32.3	81.3	21.9	30,205
Seminole	344,729	87.4	9.8	8.4	10.1	84.6	26.3	21,815
St. Johns	112,707	88.7	10.1	3.0	15.6	79.9	23.6	25,637
St. Lucie	179,559	79.6	19.0	5.2	20.1	71.7	13.1	16,483
Sumter	39,428	81.0	18.1	3.1	20.3	64.3	7.8	14,606
Suwannee	33,077	82.2	16.9	2.0	15.8	63.8	8.2	14,773
Taylor	18,718	77.1	21.5	1.3	12.7	62.1	9.8	15,459
Union	12,359	71.0	27.8	4.8	7.0	67.7	7.9	10,783
Volusia	419,797	88.0	10.5	5.0	22.7	75.4	14.8	17,778
Wakulla	19,172	83.9	14.9	0.9	10.9	71.6	10.9	15,570
Walton	37,914	88.9	8.6	1.2	14.9	66.5	11.9	14,866
Washington	20,221	79.7	17.6	1.5	16.4	60.9	7.4	13,732

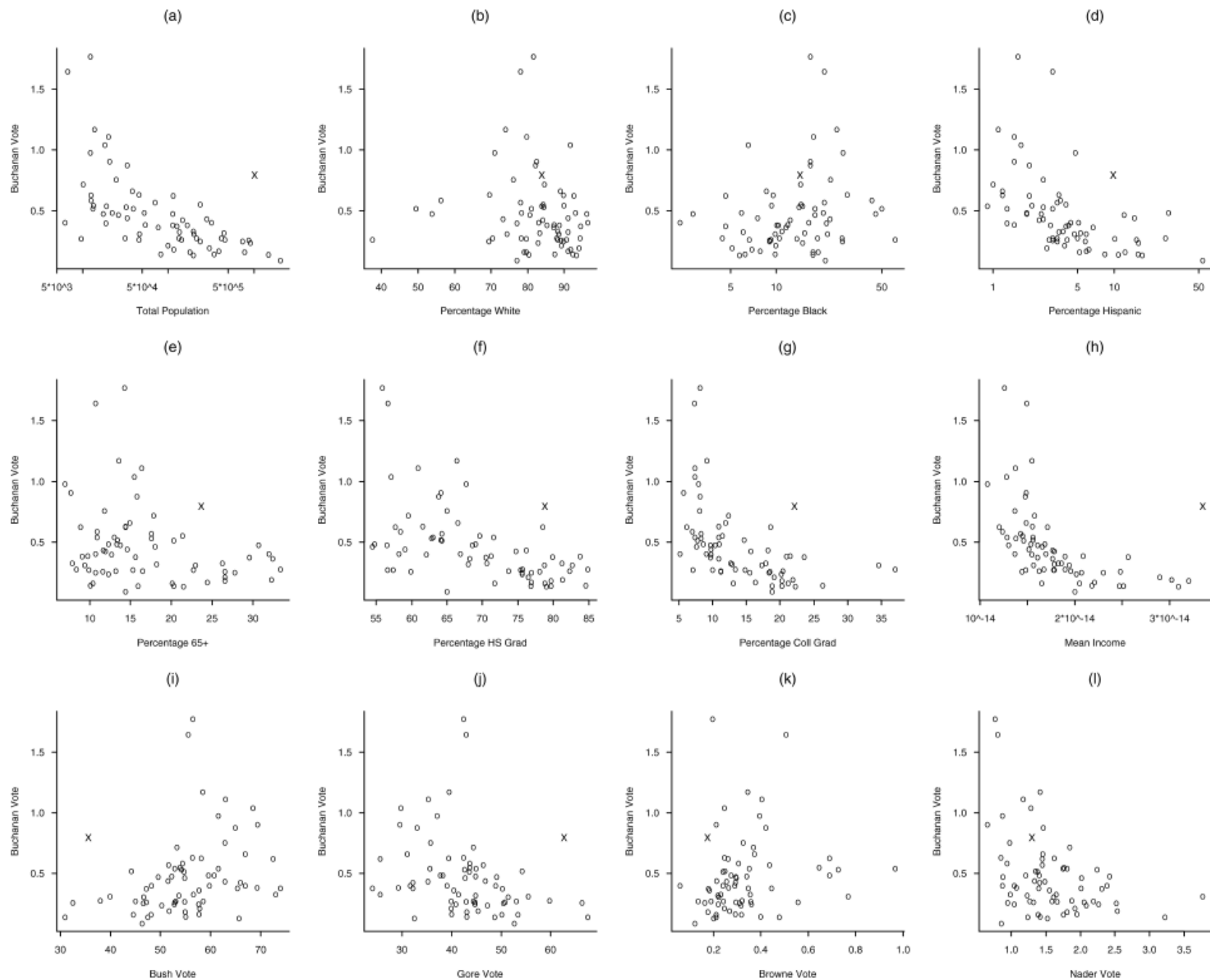


FIG. 1. Percentage of Buchanan vote against 12 covariates. Palm Beach county is marked with an  $\times$ .

TABLE 5  
*List of covariates used in the analysis*

<b>Covariate</b>	<b>Definition</b>
lpop	Log total population size
whit	Proportion of whites
lblac	Log proportion of blacks
lhispanic	Log proportion of Hispanics
o65	Proportion of population aged 65 and over
hsed	Proportion graduated high school
coll	Proportion graduated college
inco	Mean personal income
pbush	Proportion voting for Bush
pbrow	Proportion voting for Browne
pnader	Proportion voting for Nader

## Approach

For 66 counties excluding PBC, build a regression model to express Buchanan's vote as a function of covariates. Then use the model to predict Buchanan's vote in PBC. This can then be compared with the actual vote (3407).

## Difficulties

What to use as dependent variable?

If  $y_i$  is Buchanan vote and  $N_i$  is total votes casts in county  $i$ , we could use  $y_i$  or  $Y_i/N_i$  as dependent variable, but both create difficulties with heteroskedasticity.

Transformations: consider  $h(y_i) = \sum_j x_{ij}\beta_j + \epsilon_i$ . Use  $h_\lambda(y) = C \frac{y^\lambda - 1}{\lambda}$  where  $C = y^{1-\lambda}$ .

Also consider whether to *transform both sides* (Carroll and Rupert): if  $y_i^\lambda$  is dependent variable, also scale independent variables by  $N_i^\lambda$

## Other issues

Variable selection — use Mallows  $C_p$  or backward selection with each of three transformations,  $y_i$  or  $\sqrt{y_i}$  or  $\log(Y_i/N_i)$ .

RSS is lowest for  $\sqrt{y_i}$  as dependent variable

However there is still a problem of *overdispersion* — residual variance based on  $\sqrt{y_i}$  is 2.42, compared with 0.25 for Poisson.

TABLE 6

*Covariates selected by either Mallows's  $C_p$  or backward selection; all models include the rescaled intercept term*

<b>Response variable</b>	<b>Selection method</b>	<b>Variables selected</b>
$y_i$	$C_p$	lpop, whit, lhis, o65, hsed, coll, pbush, pnade
$y_i$	Backward	lpop, whit, lhis, o65, hsed, pbush, pnade
$\sqrt{y_i}$	$C_p$	whit, lhis, o65, hsed, inco, pbrow
$\sqrt{y_i}$	Backward	whit, lhis, o65, hsed, inco, pbrow
$\log(y_i/N_i)$	$C_p$	lpop, lhis, hsed, inco, pbush
$\log(y_i/N_i)$	Backward	lhis, hsed, inco, pbush



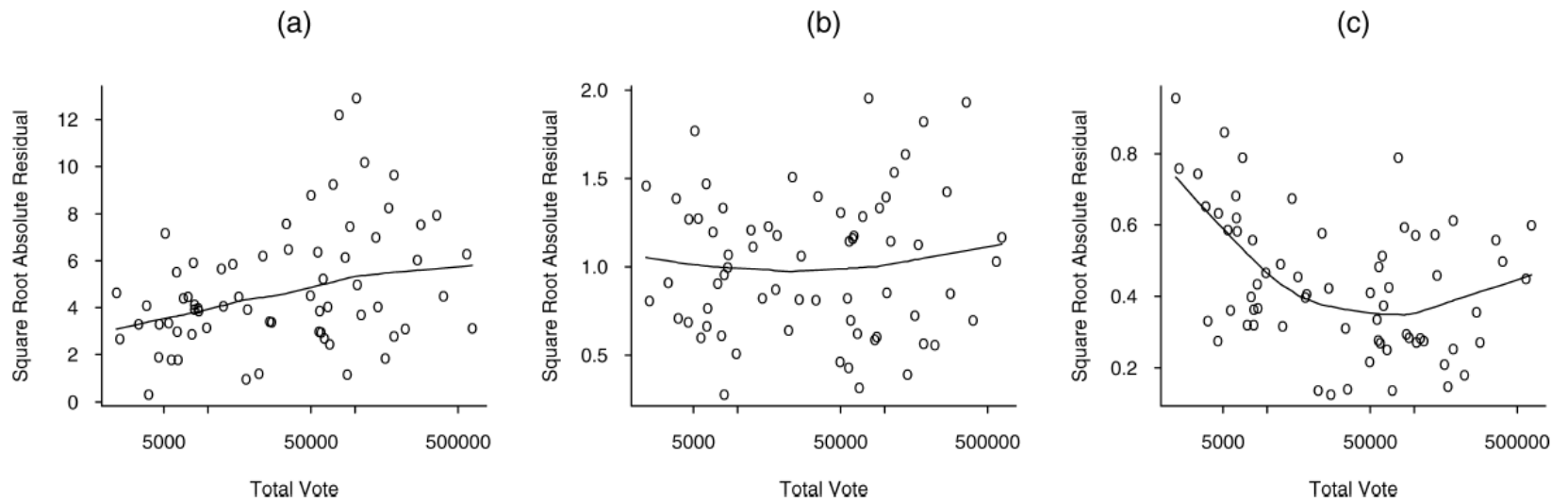


FIG. 2. (a) Plot of the square root of absolute values of studentized residuals for the model with untransformed Buchanan votes as the response variable against total votes in the county, together with a smoothed curve from the lowess function in S-PLUS. Palm Beach county has been omitted from the plot. (b) The same thing based on regression using square root of Buchanan votes as the response. (c) The same thing based on regression using log proportion of Buchanan votes as the response.

## Testing for Homoskedasticity

First approach: White's test (SPEC option in PROC REG)

Regress  $e_i^2$  on all squares and cross-products of covariates, use  $R^2$  as test statistic.

Result: Not significant, even if  $y_i$  or  $\log y_i$  are used as dependent variables!

TABLE 7

*Results of White's heteroskedasticity test applied to the five distinct models of Table 6*

<b>Response variable</b>	<b>Selection method</b>	<b><math>nR^2</math></b>	<b>DF</b>	<b><math>p</math> value</b>
$y_i$	$C_p$	43.52	45	0.53
$y_i$	Backward	47.81	36	0.09
$\sqrt{y_i}$	$C_p$ or Backward	23.02	28	0.73
$\log(y_i/N_i)$	$C_p$	18.20	20	0.57
$\log(y_i/N_i)$	Backward	13.12	14	0.52

## Alternative Tests for Homoskedasticity

Wetherill's statistic:

$$\phi = \frac{n \left\{ \sum_i (\hat{y}_i - \bar{y}) e_i^2 \right\}^2}{\sum_i (\hat{y}_i - \bar{y})^2 \sum_i (e_i^2 - \hat{\sigma}^2)^2},$$

where  $\hat{y}_i$  is  $i$ th fitted value,  $\bar{y}$  is average of  $y_i$  or equivalently  $\hat{y}_i$ ,  $e_i$  is the  $i$ th residual and  $\hat{\sigma}^2 = \sum e_i^2 / n$ . Under the null hypothesis of homoskedasticity,  $\phi$  has an approximate  $\chi_1^2$  distribution.

Godfrey's method:

Regress  $e_i^2$  on selected covariates  $z_{ij}$ . If  $Z$  is some  $n \times p$  covariate matrix and  $r_i = e_i^2 / \hat{\sigma}^2 - 1$ , define

$$G = \frac{1}{2} r^T Z (Z^T Z)^{-1} Z^T r.$$

Asymptotics:  $G \sim \chi_{p-1}^2$  under  $H_0$ .

My approach: define  $z_i = \log N_i$  ( $N_i =$  population in county  $i$ ),  $G = (\sum r_i z_i)^2 / (2 \sum z_i^2)$ . p-value by simulation.

TABLE 8

*Test statistics  $\phi$  and  $G$ , with corresponding  $p$  values, applied to the five distinct models of Table 6*

<b>Response variable</b>	<b>Selection method</b>	<b><math>\phi</math></b>	<b><math>p</math> value</b>	<b><math>G</math></b>	<b><math>p</math> value</b>
$y_i$	$C_p$	2.95	0.04	0.269	0.00
$y_i$	Backward	4.49	0.01	0.328	0.00
$\sqrt{y_i}$	$C_p$ or Backward	1.70	0.18	0.031	0.26
$\log(y_i/N_i)$	$C_p$	5.42	0.02	0.277	0.00
$\log(y_i/N_i)$	Backward	3.18	0.08	0.233	0.00

## Influence Diagnostics

Calculate studentized residuals and DFFITS (based on  $\sqrt{y_i}$  regression, including PBC). Also compute simulated confidence bands by Atkinson method.

PBC is extreme outlier. Studentized residual 17.5. Based on  $t_{58}$  distribution, the p-value associated with that is about  $10^{-72}$ .

If PBC is omitted, plots of standardized residuals and DFFITS look OK.

We also considered range of  $\lambda$  values in Box-Cox transformation (omitting PBC) — optimal value about  $\lambda = 0.4$ , but  $\lambda = 0.5$  not significantly different.

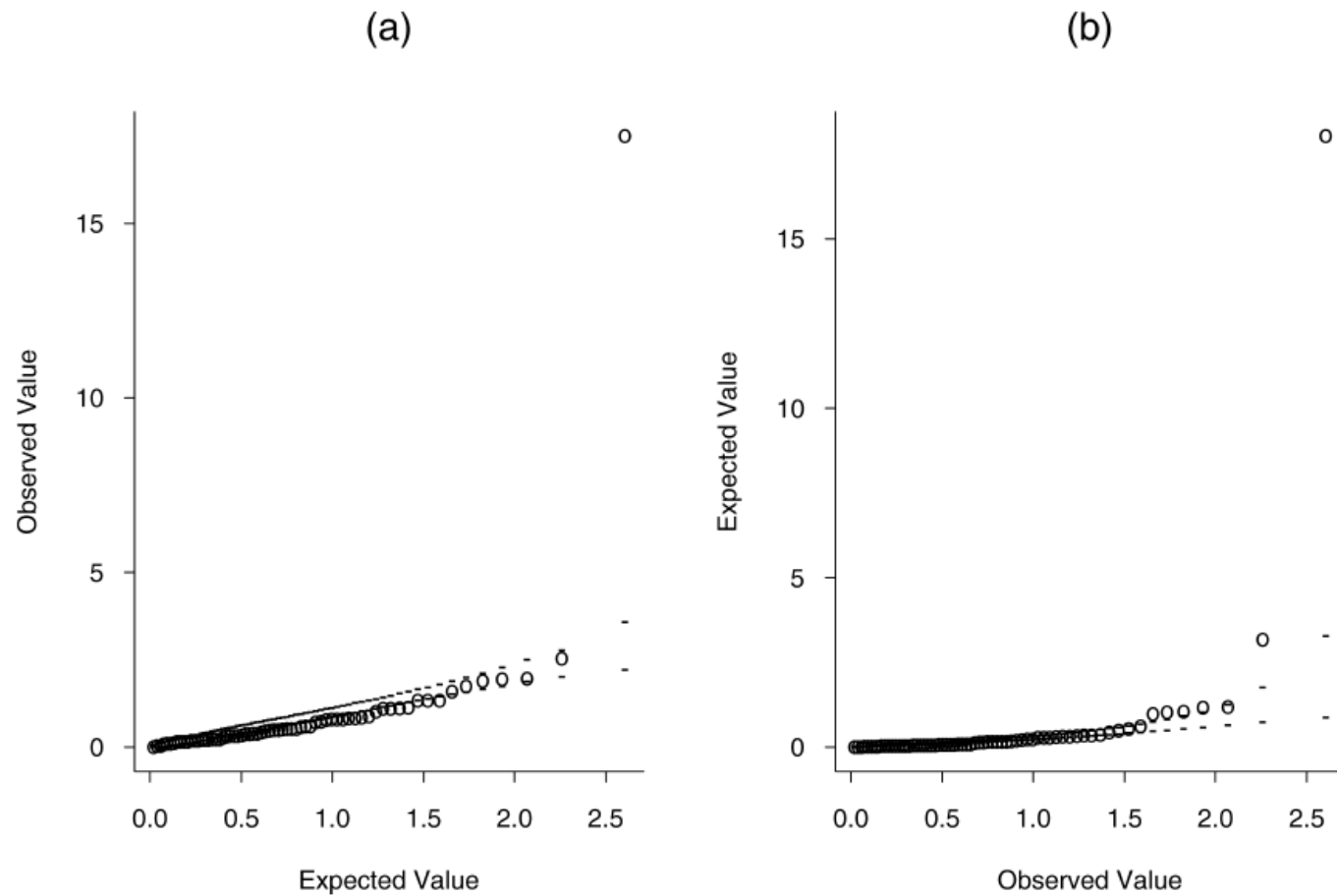


FIG. 3. (a) *Half-normal plot of ordered studentized residuals for the model based on square root Buchanan votes, with pointwise 90% simulation bounds.* (b) *Half-normal plot of ordered DFFITS for the model based on square root Buchanan votes, with pointwise 90% simulation bounds. Palm Beach county is the large outlier on both plots.*



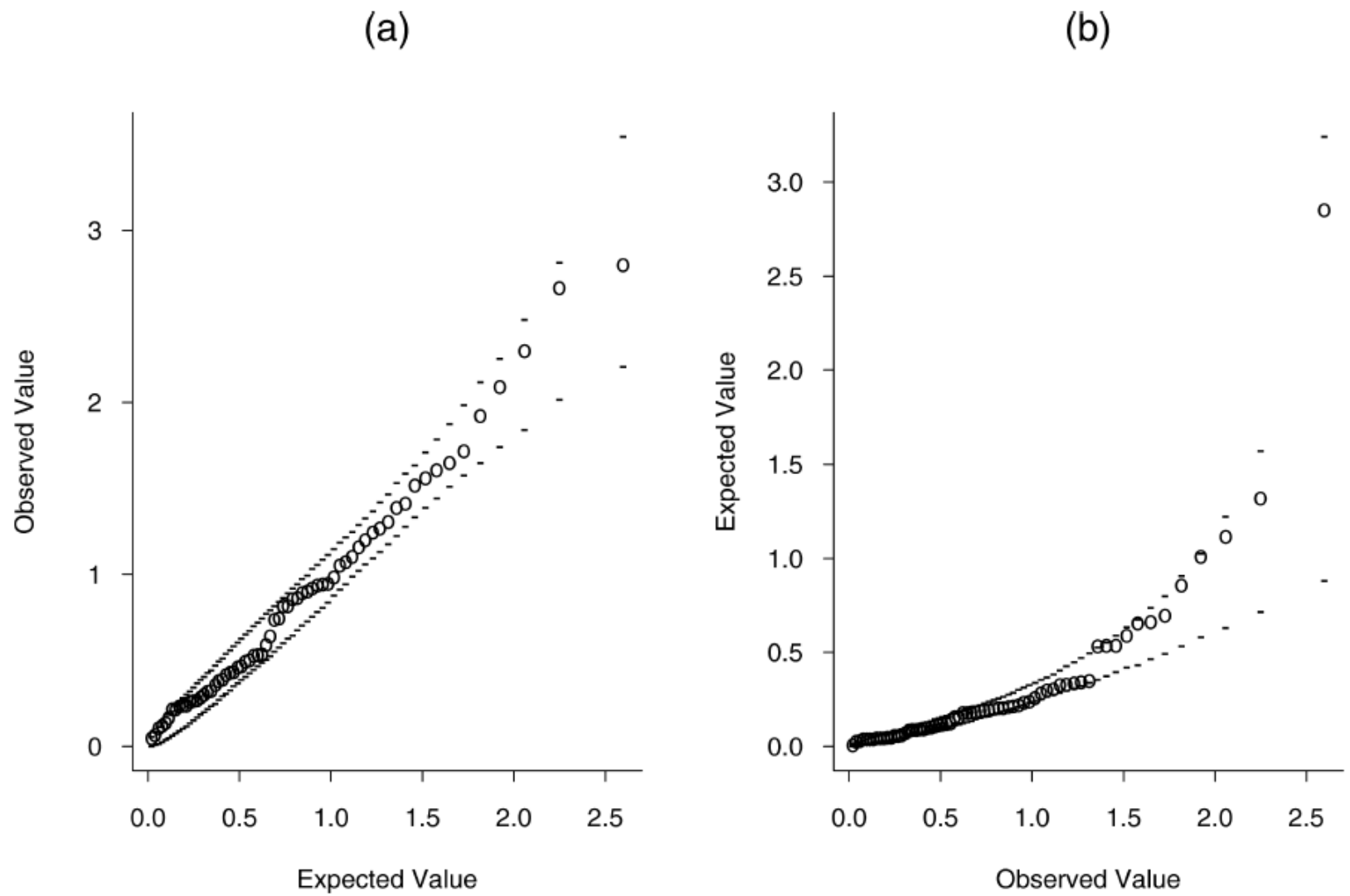


FIG. 4. Same as Figure 3, but omitting Palm Beach altogether.

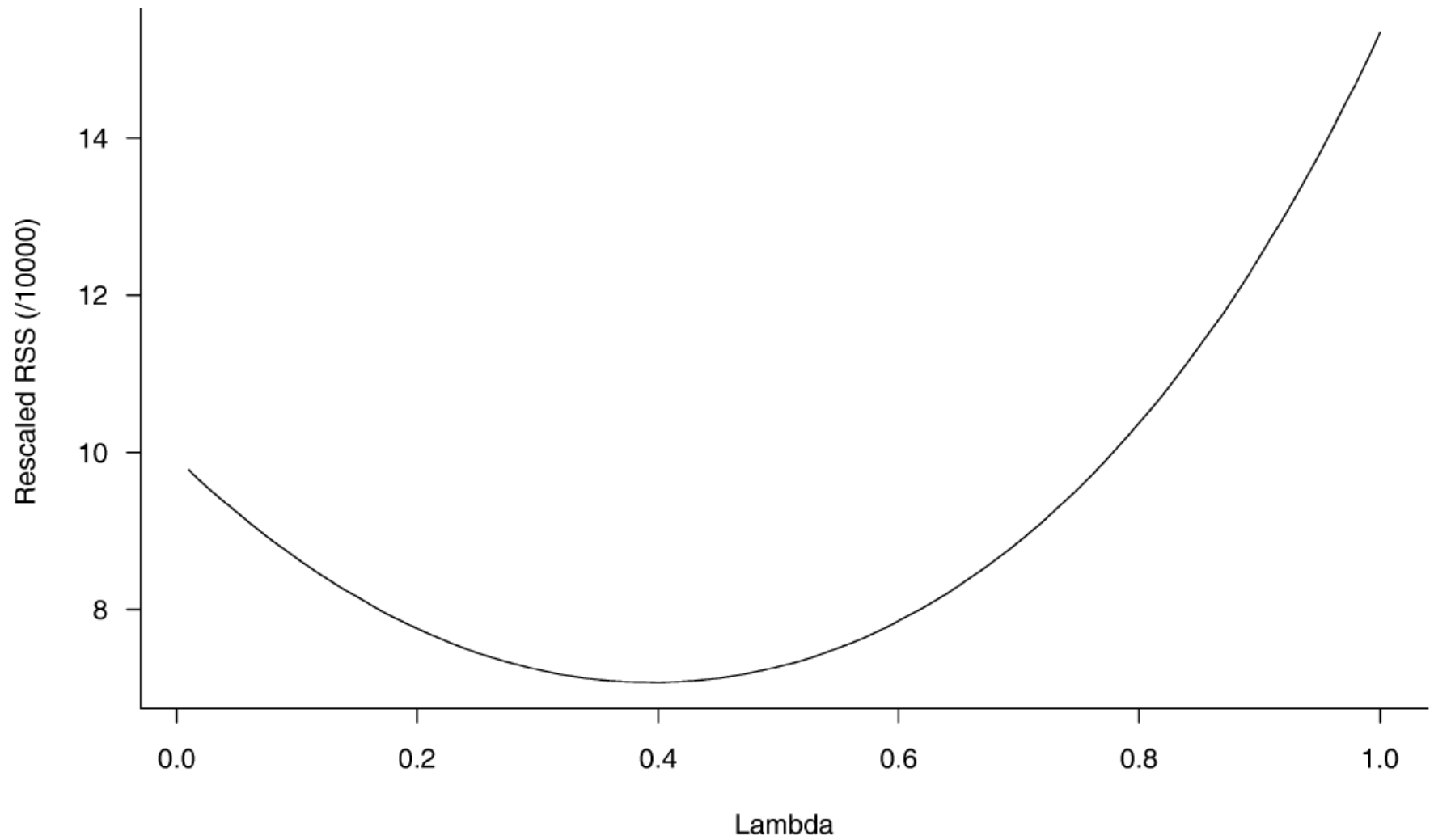


FIG. 6. *Selecting the transformation: Plotting the residual sum of squares against transformation parameter  $\lambda$ , for the normalized transformation.*

## Prediction Intervals

Based on the model fitted to the other 66 counties, we computed a point prediction and 95% prediction interval for Buchanan's vote in PBC, using several of the regression models previously fitted.

We also developed some equivalent "binary data" models based on logistic regression (suggested by Alan Agresti)

All cases shows a point predictor  $< 400$ , and an upper bound of the 95% prediction interval that is  $< 800$ .

TABLE 9

*Point predictions and prediction intervals under three versions of linear model*

<b>Response variable</b>	<b>Variable selection</b>	<b>Point predictor</b>	<b>Prediction interval</b>
$\sqrt{y_i}$	$C_p$ or Backward	371	(219, 534)
$\log(y_i/N_i)$	$C_p$	363	(180, 735)
$\log(y_i/N_i)$	Backward	371	(182, 758)

TABLE 10

*Point predictions, confidence and prediction intervals under four versions of logistic regression*

<b>Method</b>	<b>Point estimate</b>	<b>Confidence interval</b>	<b>Prediction interval</b>
No overdispersion	379	(349, 412)	(330, 447)
Deviance	379	(293, 491)	(237, 606)
Pearson	379	(293, 491)	(237, 606)
Williams	345	(211, 562)	NA

## Conclusions

In all analyses, PBC is an enormous outlier.

The point predictors of Buchanan's vote in PBC are all under 400, and the upper bounds of the 95% prediction intervals are under 800.

Buchanan's actual vote in PBC was 3407.

Therefore, it appears that Buchanan gained at least 2500 excess votes in Palm Beach County.

Had these votes been accredited to Gore, he would have been President!