

Working Paper: The Effect of Electronic Voting Machines on Change in Support for Bush in the 2004 Florida Elections

Michael Hout, Laura Mangels, Jennifer Carlson, Rachel Best
With the assistance of the
UC Berkeley Quantitative Methods Research Team

Motivation

Public discussion of changing voting technology raised concern that some forms of electronic voting might produce a discrepancy between voters' intentions and tabulations of the election's outcome. In particular, touch-screen voting machines were criticized for being unverifiable unless they printed out a hard copy that voters could certify as correct and election officials could keep in case a recount was ordered. Without a paper trail, statistical comparisons of jurisdictions that used e-voting are the only tool available to diagnose problems with the new technology.

In our research we used ordinary least squares and more sophisticated linear modeling approaches to assess the statistical properties of e-voting. In particular we develop models that predict both the percentage of the votes registered for the incumbent - President Bush - and the amount that percentage changed between 2000 and 2004. These models can incorporate adjustments for a large number of factors that we or others thought might help explain the patterns. These include socioeconomic and demographic factors like the typical family's income or its ethnic ancestry. We also adjust for ecological factors like the size of the county. Most importantly we adjust for its voting history, reaching back not only to the 2000 election but farther to the 1996 election. To this list of factors we add consideration of whether the county's voting technology was e-touch machines or optical scanning equipment.

Finally we translated percentage differences into vote totals in two ways. The first was to assume that the vote margin was due to the appearance of "ghost votes" - votes registered for in a way that helped one candidate but did not reduce the total for the other. Mechanisms that would produce this outcome include having votes electronically registered in the machine prior to any voters using the machine or after the last voter used it - through software errors or hacking - and other flaws that interfere with counting after some limit is reached - reports indicate that some machines may have been programmed to stop counting or subtract votes after some limit is reached. The second count assumes a misattribution by the machine, i.e., a vote intended for candidate A that gets counted for candidate B. Since every vote miscast for candidate B costs candidate A one too, the difference is doubled, so we double our initial estimate to get our estimate of the miscount under this type of error. A combination of one type of error and the other would yield a vote total in between.

Finding

Electronic voting raised President Bush's advantage from the tiny edge he held in 2000 to a clearer margin of victory in 2004. The impact of e-voting was not uniform, however. Its impact was proportional to the *Democratic* support in the county, i.e., it was especially large in Broward, Palm Beach, and Miami-Dade. The evidence for this is the statistical significance of terms in our model that gauge the average

impact of e-voting across Florida's 67 counties and statistical interaction effects that gauge its larger-than-average effect in counties where Vice President Gore did the best in 2000 and slightly negative effect in the counties where Mr. Bush did the best in 2000. The state-wide impact of these disparities due to electronic voting amount to 130,000 votes if we assume a "ghost vote" mechanism and twice that - 260,000 votes - if we assume that a vote misattributed to one candidate should have been counted for the other.

Data

We used three types of data: election data, demographic data, and voting-machine data.

Election Data

Our 2000 data for Florida elections was taken from US Together (<http://ustogether.org/election04/FL2000.htm>). Our 2000 for Ohio and 2004 election data for Florida and Ohio were taken from CNN.com's online coverage of the 2000 and 2004 elections (2000: <http://www.cnn.com/ELECTION/2000/results/national.html>, 2004: <http://www.cnn.com/ELECTION/2004/pages/results/president>). The data are organized by state and then by county; we used these data for Florida and Ohio at the county level. We re-checked our Florida data as of Nov. 11, 2004. Our Ohio data was entered on Nov. 6 but, not subsequently re-checked. This is due to the fact that, as of preliminary tests, e-voting did not appear to be a significant factor in the change of percentage of Bush support from 2000 to 2004 in Ohio (see Results).

We also used 1996 election data, which were taken from the Atlas of U.S. Presidential Elections (www.uselectionatlas.org).

Demographic Data

We used 2000 Census data on median income and Hispanic population for Florida counties. We did not collect demographic data for Ohio.

Voting Machine Type Data

We collected data on voting machine type by county from the Verified Voting Foundation (<http://www.verifiedvoting.org/verifier/>) for Florida and Ohio. A dummy variable was introduced to designate a county's use of electronic voting. Optical scanning and paper ballots were coded as '0'; electronic voting machines were coded as '1'.

Statistical Methodology

Technique

We used an ordinary-least-squares regression model (OLS) with and without robust standard errors. We supplemented these calculations with other estimation techniques designed to test the limits of simple methods. First, we used robust regression methods designed to minimize the leverage of a single influential county. Second, we weighted counties according to the number of votes cast - giving the populous counties more weight in the calculations than the smaller ones. Neither of these more complicated methods led to substantively different conclusions about electronic voting (the robust regression methods did suggest that Hispanic voters were more pro-Kerry than the OLS results

led us to believe). In fact, the standard errors for the robust estimates were smaller than those obtained using OLS.

Dependent Variable

The dependent variable is the change in percent voting for Bush by county from 2000 to 2004. This was calculated by subtracting the percent voting for Bush in 2000 from percent voting for Bush in 2004.

Independent Variables

The independent variables in our model are baseline support for Bush (percent voting for Bush in 2000), percent voting for Dole in 1996, change in voter turnout from 2000 to 2004 (number of votes for Bush and Gore in 2000 subtracted from number of votes for Bush and Kerry in 2004), median income, Hispanic population, size of county (number of votes for Bush and Kerry in 2004), and a dummy variable for electronic voting (1 = electronic voting machine, 0 = optical scanning or paper ballot).

We also included a squared term for baseline support for Bush, an interaction effect between baseline support for Bush and electronic voting, and an interaction effect between baseline support for Bush squared and electronic voting. (Table 1)

Table 1 - Description of variables

	N	Mean	SD	Min	Max
Change in % voting for Bush	67	0.037	0.029	-0.03	0.107
% Bush 2000	67	0.563	0.093	0.315	0.755
% Bush 2000 squared	67	0.325	0.103	0.099	0.569
% Dole 1996	67	0.507	0.084	0.288	0.712
Voter Turnout Change	67	24236	31692	663	116327
Size (Kerry + Bush votes)	67	111141	158958	2997	714362
Median Income	67	35385	6343	26032	52244
Hispanic Population	67	0.085	0.100	0.015	0.573
Electronic Voting	67	0.224	0.420	0	1
% Bush 2000 * Electronic Voting	67	0.118	0.226	0	0.702
% Bush 2000 squared * Electronic Voting	67	0.064	0.130	0	0.493

Results

Table 2 presents results for OLS regressions without (model 1) and with control variables: percent voting for Dole in 1996, voter turnout change between 2000 and 2004, median income, and number of Hispanic residents (model 2). Table 3 presents regressions with robust standard errors and fully robust regressions (robust standard errors and

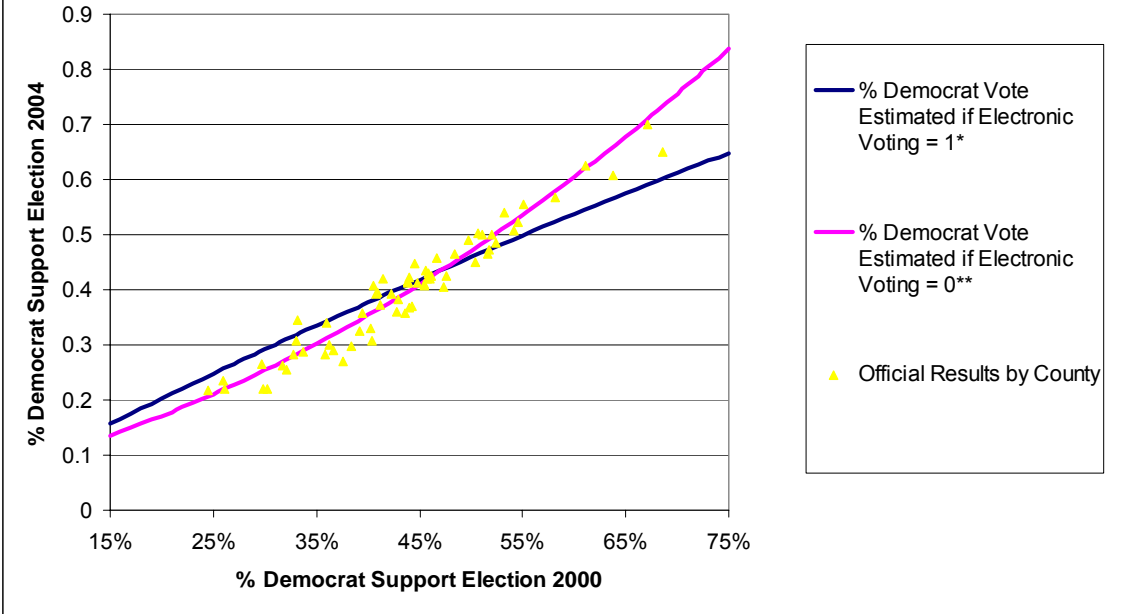
coefficients) for the reduced and complete models. Table 4 presents results for the models with frequency weights for county size, defined as total Bush and Kerry votes in 2004. With the exception of the reduced model with frequency weights, all models show similar and significant effects of electronic voting on change in percent voting for Bush.

As baseline support for Bush increases in Florida counties, the change in percent voting for Bush from 2000 to 2004 increases, but at a decreasing rate. Electronic voting has a main, positive effect on the dependent variable. Furthermore, there is an interaction effect between baseline support for Bush and electronic voting, and between baseline support for Bush squared and electronic voting. Support for Dole in 1996, county size, median income, and Hispanic population had no significant effect net of the other effects. Essentially, net of other effects, electronic voting had the greatest positive effect on change in percent voting for Bush from 2000 to 2004 in democratic counties. (Tables 2 - 4) (Figure 1)

We also examined the effect of electronic voting machines and baseline support for Bush on change in percent voting for Bush in Ohio. The OLS regression model used percent voting for Bush in 2004 by county as the dependent variable and baseline support for Bush and electronic voting as independent variables, as well as an interaction effect between baseline support for Bush and electronic voting. Without controlling for change in voter turnout, size, median income, Hispanic population, or percent voting for Dole in 1996, we found no effect of electronic voting on change in percent voting for Bush from 2000 to 2004 in Ohio.

In order to understand the effects of electronic voting in terms of the number of votes for Bush, we translated our dependent variable (percent voting for Bush in Florida in 2004) into raw votes. By setting Electronic Voting equal to zero, we created a predicted percentage change in support for Bush without the effect of electronic voting. We added the predicted percentage change in support for Bush to the percentage of votes he received in 2000. This gave us a predicted percentage of votes for Bush in 2004, which we multiplied by the number of votes in each county to get a predicted number of votes without the effect of electronic voting. We then subtracted this number from the number of votes Bush received, as estimated by the full regression model, including the Electronic Voting effect. Summing these effects for the fifteen counties with electronic voting yields the total estimated excess votes in favor of Bush associated with Electronic Voting; this figure is 130,733.

Figure 1: Democratic Support 2000 Election versus Democratic Support 2004 Election by Voting Type



*(dy/dx | e-voting=1) = $-.2994439 + (1.102387 * [\% \text{Bush } 2000]) - (.8492126 * [\% \text{Bush } 2000_sq]) - (1.477718 * [\% \text{Bush} * \text{E-voting}]) + (.4940859 * [\text{E-voting}]) + (1.02589 * [\% \text{Bush } 2000_sq * \text{E-voting}]) - ((9.13e-08) * 111140.8)$

** (dy/dx | e-voting=0) = $-.2994439 + (1.102387 * [\% \text{Bush } 2000]) - (.8492126 * [\% \text{Bush } 2000_sq]) - ((9.13e-08) * 111140.8)$

Table 2 - Change in % Voting for Bush from 2000 to 2004 in Florida Counties: OLS Regression

	Model 1 β t	Model 2 β t
% Bush 2000	1.102 3.5***	1.028 3.19**
% Bush 2000 sq	-0.849 -3.06**	-0.664 -2.36*
Size (Votes for Kerry + Bush 2004)	-9.13E-08 -3.55***	-3.93E-08 -.59
Electronic Voting	0.494 3.26**	0.417 2.79**
%Bush 2000 * Electronic Voting	-1.478 -2.6*	-1.284 -2.31*
%Bush 2000 sq * Electronic Voting	1.026 1.93	0.938 1.81
% Dole 1996		-0.152 -1.3
Voter Turnout Change		-2.67E-11 0
Median Income		-8.17E-07 -1.08
Hispanic Population		-0.053 -1.71
constant	-0.299 -3.38***	-0.213 -2.26*
<i>R-squared</i>	0.449	0.537

*p <.05

**p<.01

***p<.001

Table 3 - Change in % Voting for Bush from 2000 to 2004 in Florida Counties: Robust Regression

	Robust SE		Robust SE & Coeffs	
	Model 1	Model 2	Model 1	Model 2
	β t	β t	β t	β t
% Bush 2000	1.102387 5.34***	1.028138 4.99***	1.142 3.46***	1.203 4.49***
% Bush 2000 sq	-0.8492126 -4.52***	-0.6641244 -4.05***	-.889 -3.06**	-.730 -3.12**
Size (Votes for Kerry + Bush 2004)	-9.13E-08 -3.96***	-3.93E-08 -0.65	-9.24E-8 -3.43***	-1.92E-8 -0.35
Electronic Voting	0.4940859 5.43***	0.4165562 4.76***	.510 3.21**	.366 2.94**
%Bush 2000 * Electronic Voting	-1.477718 -4.05***	-1.283671 -3.89***	-1.542 -2.60*	-1.185 -2.56*
%Bush 2000 sq * Electronic Voting	1.02589 2.80**	0.9380135 3.00**	1.086 1.95	.915 2.12*
% Dole 1996		-0.1520883 -1.28		-.274 -2.81**
Voter Turnout Change		-2.67E-11 -0.00		2.33E-7 0.93
Median Income		-8.17E-07 -1.11		-1.05E-6 -1.67
Hispanic Population		-0.0525564 -1.09		-.130 -5.09***
constant	-0.2994439 -5.51***	-0.2130006 -3.86***	-.308 -3.32**	-.220 -2.81**
<i>R-squared</i>	.450	.538		

*p <.05
 **p<.01
 ***p<.001

**Table 4 - Change in % Voting for Bush from 2000 to 2004
in Florida Counties: OLS regressions with frequency
weights for county size**

	Model 1 β t	Model 2 β t
% Bush 2000	0.213 .60	0.535 1.94*
% Bush 2000 sq	-0.151 -.50	-0.431 -1.83
Electronic Voting	0.152 1.29	0.236 2.46*
%Bush 2000 * Electronic Voting	-0.512 -1.17	-0.707 -1.98
%Bush 2000 sq * Electronic Voting	0.389 .94	0.478 1.42
% Dole 1996		-0.021 -0.21 -2.1E- 07
Voter Turnout Change		-3.22 -4.9E- 07
Median Income		-0.86
Hispanic Population		-0.054 -4.13***
constant	-0.047 -.046	-0.091 -1.15
<i>R-squared</i>	<i>0.154</i>	<i>0.570</i>

*p < .10
**p < .05
***p < .01

Appendix

Reviewers raised concerns about our use of the total votes for Kerry and Bush as a proxy for county size. Table 5 presents the results for the reduced and complete models using the natural logarithm of county population (obtained from Census 2000 data) instead of our previous size variable. The results are substantively equivalent to those reported above.

Table 5 - Change in % Voting for Bush from 2000 to 2004 in Florida Counties: OLS Regression with Robust Standard Errors

	Model 1	Model 2
	B	B
	t	t
% Bush 2000	.936 4.75***	1.026 4.9***
% Bush 2000 sq	-.707 -4.05***	-0.672 -4.25***
ln(population)	-.009 -3.59***	-0.001 -.22
Electronic Voting	.342 4.13***	.391 4.56***
%Bush 2000 * Electronic Voting	-.997 -2.89**	-1.212 -3.71***
%Bush 2000 sq * Electronic Voting	.661 1.91	.889 2.87**
% Dole 1996		-0.142 -1.05
Voter Turnout Change		-1.25E-07 -.59
Median Income		-7.56E-07 -1.04
Hispanic Population		-.059 -1.38
constant		-0.205 -2.68**
R-squared	.464	.535

*p <.05 **p<.01 ***p<.001

Other reviewers asked if the results would be different if we used the simpler percent for Bush instead of percent change in Bush support as the dependent variable. Since we used percentage of votes for Bush in 2000 as an independent variable, it essentially appears twice in our regression equations. Problems would arise primarily with percentages that approach zero or one hundred; our data range from 31% to 75%. However, to address these concerns, table 6 presents results for the reduced and complete models using percent voting for Bush in 2004 as the dependent variable; the results are substantively equivalent to those reported above.

**Table 6 - % Voting for Bush in 2004 in Florida Counties:
OLS Regression with Robust Standard Errors**

	Model 1	Model 2
	β	β
	t	t
% Bush 2000	2.102 10.19***	2.028 9.84***
% Bush 2000 sq	-.849 -4.52***	-.664 -4.05***
Size (Votes for Kerry + Bush 2004)	-9.13E-8 -3.96***	-3.93E-8 -0.65
Electronic Voting	.494 5.43***	.417 4.76***
%Bush 2000 * Electronic Voting	-1.478 -4.05***	-1.284 -3.89***
%Bush 2000 sq * Electronic Voting	1.026 2.80**	.938 3.00**
% Dole 1996		-.152 -1.28 -2.67E-11
Voter Turnout Change		11 0.00
Median Income		-8.17E-7 -1.11
Hispanic Population		-.053 -1.09
constant	-.299 -5.51***	-.213 -3.86***
<i>R-squared</i>	.960	.967

*p <.05 **p<.01 ***p<.001