Factors Associated with the Excessive CD-13 Undervote in the 2006 General Election in Sarasota County, Florida*

Walter R. Mebane, Jr.† David L. Dill‡

Initial draft: 18jan2007; this draft: 18jan2007

*We thank Martha Mahoney for helpful discussion and Herbie Ziskend for assistance.

†Professor, Department of Government, Cornell University. 217 White Hall, Ithaca, NY 14853–7901 (Phone: 607-255-2868; Fax: 607-255-4530; E-mail: wrm1@cornell.edu).

‡Professor, Department of Computer Science, Stanford University. Gates Building 3A, Stanford, CA 94305-9030 (Phone: 650-725-3642; Fax: 650-725-6949; E-mail: dill@cs.stanford.edu).

Introduction

Based on statistical analysis of detailed electronic ballot and event log data from the Sarasota general election ending November 7, 2006, we find that none of the many theories advanced so far to explain the extraordinarily high undervote rate in the Florida Congressional District 13 (CD-13) race adequately account for systematic covariations between that undervote and other identifiable factors, such as voting patterns in other contests and unusual events on the voting machines. At this time, we are unable to propose a convincing mechanism based on voter, machine or ballot characteristics that completely explains the phenomenon.

This paper describes the data and statistical analysis we have performed and evaluates explanations that have been advanced by others as well as plausible explanations that we propose in light of our analysis. Our results are suggestive but in important respects puzzling. In a nutshell, **the excessive CD-13 undervote rate in Sarasota County is not yet well-understood, and will not be understood without further investigation. On its own, further statistical analysis of the kind of data we examine here probably cannot explain the undervotes.**

Even though important components of the CD-13 undervote rate are readily explicable, the available explanations do not fully account for the phenomenon. Some factors that correlate with the CD-13 undervote connect to simple and general plausible explanations. Several hundred CD-13 undervotes come from voters who, based on votes they cast for other offices, seem to be disinclined to vote for candidates affiliated with either the Democrat or Republican party: the CD-13 race did not offer a third-party alternative and did not allow write-in votes. Several thousand CD-13 undervotes appear on ballots that also have undervotes for other offices, notably for the five statewide offices that appeared on Florida ballots in 2006. It is plausible that most of these undervotes reflect voters who simply did not care to vote for several offices. These undervotes are not related to any voting machine characteristics we were able to observe.

Other correlates of the CD-13 undervote rate are straightforward to describe, but the general explanations that may connect to them are not so clear. We find differences of hundreds of CD-13 undervotes when we compare voting machines that have different observable characteristics. Hundreds of these undervotes are related to a specific error message in the event log file (the event log file supposedly reports every transaction that occurred on each voting machine). The CD-13 undervote rate varies substantially across voters, differences that correlate significantly with the partisan balance among the votes recorded for the five statewide offices. Ballots that have even one additional statewide office with a vote for a Democrat rather than for a Republican tend to have higher rates of CD-13 undervoting.

The principal question we cannot answer is whether these patterns reflect voluntary behavior or artificial errors or manipulations. The urgency of this question is highlighted by the fact that the relationship between the CD-13 undervote rate and the statewide office voting pattern differs depending on whether a particular error message occurs on the voting machine on which the votes were cast. In the data for votes cast on election day, this interaction effect is statistically significant. The same effect is observable but not significant in the data for votes cast during early voting. The number of CD-13 undervotes that appear to be directly implicated in this interaction effect is relatively small, but it is worrisome to see any sign that an error in the voting machines' operation is a marker for otherwise comparable voters having their votes recorded differently. Moreover it is difficult to conclude that persistent differences in basic propensities to undervote explain the differences in CD-13 undervoting among ballots that have different statewide voting patterns, because we find that the same subsets of ballots have substantially different undervote rates for other offices. For instance, the CD-13 undervote rate is higher on ballots that have votes for Democrats for all five statewide offices than it is on ballots that have all five votes for Republicans, but it is the straight-Republican ballots that have the higher undervote rate in votes for Hospital Board Southern District Seat 1. The relationship between partisan voting for the statewide offices and undervoting is peculiar to the CD-13 race.

Such peculiarity raises a substantial question about the popular idea that the large CD-13 undervote was caused by the format of the ballot (Mahlburg and Tamman 2006; Tamman and Doig 2006; Frisina, Herron, Honaker, and Lewis 2006). Clearly the ballot format caused many of the CD-13 undervotes, but it is not at all clear how many. The ballot format cannot explain the differences we observe between voting machines and between subsets of ballots. The ballot format cannot explain why the distribution of undervotes differs significantly between the race for Hospital Board Southern District Seat 1 and the CD-13 race. The alternatives for the Hospital Board Southern District Seat 1 race appear on the ballot with notable features that are very similar to those that are often credited with causing the problems with the CD-13 race.

A brief history of the CD-13 undervote problem

Almost two months after the 2006 election, the election results for one Federal office are in dispute: the 13th Congressional District. After a recount mandated by state law because of the close race, Vern Buchanan, the Republican candidate, has a margin of victory of 369 votes over Christine Jennings, the Democrat, out of 238,249 total votes.

Unfortunately, this margin of victory is dwarfed by 18,412 undervotes recorded in the CD-13 race in Sarasota County, a proportion of undervotes that is five-to-ten times greater than the undervote in the same race in previous elections. There are also strong suspicions that many of the undervotes were not intentional, since there were numerous complaints from voters about two problems specific to the CD-13 race that could each lead to lost votes. Many voters reported not being able to find the race on the ballot; even more reported selecting a candidate for the CD-13 race, then finding their vote missing when the machine displayed a summary of votes before the ballot was cast.

Because of questions about whether the election results represented the intent of the voters, the election result is still unresolved. A “manual recount” of paper printouts of the electronic ballots stored stored by the voting machines has been performed, and two rounds of “parallel testing” have been performed, where simulated votes were cast on several machines, including some with high undervote rates from the election. No major problems were discovered as a result of this testing. A team of experts is now reviewing the source code (the human-readable representation of the voting machine application software), but no results have yet been reported.

In addition to the ongoing investigation by the Florida Department of Elections, lawyers for Jennings and non-partisan groups are calling for a re-vote, as well as an independent investigation into the voting machine software. Although Buchanan was recently seated in the House of Representatives, the House leadership was careful to state that the investigation into the election should run its course. The ultimate outcome of the election is still very much in doubt, and may hinge on the findings of investigations that are not yet complete (or not even started).

Voting in Sarasota County

There were three major kinds of voting in the election: absentee voting, early voting, and polling-place voting. Absentee and early voting occur for an interval of several weeks before election day, while voting in polling places occurs on election day. Absentee voting is conducted on paper ballots, marked by the voters, which are counted by ballot scanners at a central location. Early voting and polling-place voting are conducted on the ES&S iVotronic touch-screen voting machine, which stores several copies of each vote record on several non-volatile memory devices in the machine.

The ballot on the iVotronic is presented in a series of display screens.[1] The voter selects a candidates by touching an area where the candidate's name is displayed on the screen. When a candidate is selected, an "X" appears in a rectangle next to the candidate's name and the candidate's name is highlighted and changes color. Once a candidate has been selected, a different candidate may be selected by touching the new choice.

A single iVotronic may serve several precincts and allow several different ballot styles to be used. Multiple precincts are especially important because early voting is conducted in relatively few locations, so early voting machines have ballots from many precincts. Election day machines may also have ballots from several precincts, because one polling place may serve several precincts. The same precinct may have several different ballot styles.

The iVotronic has several features to reduce voter errors. Before a ballot is cast, it displays a summary of the voter's selections (the "summary screen"). If the voter sees an error (such as an accidental undervote), he or she can back up or return to the previous screens to add or change a vote. When satisfied, the voter can cast the vote by touching an indicated area on the screen.

The iVotronic uses a small device called a "Personalized Electronic Ballot" (PEB) for several purposes.[2] The PEB is about the size of a carpenter's tape measure, and rests in an indentation in the top of the machine. When so placed, it can communicate electronically with the machine.

PEBs are used for administrative functions, including "opening" the machine (configuring it to accept votes) at the beginning of the election, "closing" the machine at the end of the election, cancelling ballots that have not been cast, or casting ballots when a voter forgets to cast a ballot before leaving the polling place.

However, the most frequent use of PEBs is the normal casting of votes. Before each voter votes, the PEB is initialized by another device with the appropriate precinct and ballot style for that voter. During an election, the iVotronic is unresponsive until a PEB is inserted. After that happens, the voter can select candidates and cast a ballot. At that point, the machine causes the PEB to be inactivated, so that no more votes can be cast until it is initialized.

The Help America Vote Act requires at least one voting machine that is accessible by voters with disabilities at each polling place. Such machines have audio interfaces and a device with raised buttons for visually-impaired voters. In general, there is a special iVotronic that is designated as the accessible machine at each polling place.

---

[1] Our description of how the voting machines work in Sarasota is based on the video `touchscreen384.wmv` posted at `http://www.srqelections.com/ivotronic/touchscreen_video.htm` (accessed January 17, 2007).

[2] An iVotronic machine is displayed with a description of its parts at `http://www.srqelections.com/ivotronic/ivotronic.htm`.

Possible explanations for the undervote rate

Many theories have been advanced for the high undervote rate by journalists, academic researchers, advocacy groups, the state and local elections officials, the voting machine manufacturer, and the candidates and their experts. We summarize them below and review how the results of our statistical analysis bear on them.

Partisan voting patterns A striking phenomenon is the link between variations in the CD-13 undervote and patterns of votes cast for other offices. There are five races for statewide offices on the Florida ballot: U.S. Senator; Governor; Attorney General (AG); Chief Financial Officer (CFO); and Commissioner of Agriculture (CA). Early inspections of the vote image log showed that ballots that have CD-13 undervotes "overwhelmingly supported Democratic candidates in the five statewide races" (Stratton 2006). Partly this means that ballots with votes for Democrats for all five of the statewide offices have proportionally more CD-13 undervotes than do straight-Republican ballots. But our examination of the data shows that even ballots that have only one additional office voted for a Democrat instead of for a Republican tend to have higher rates of CD-13 undervoting. **An especially strong pattern is that ballots that have a vote for the Democrat candidate for Commissioner of Agriculture are more likely to have a CD-13 undervote than are ballots that have a vote for the Republican running for that office. Ballots that have a Democrat vote for the U.S. Senate are also more likely to have a CD-13 undervote than are ballots that have vote for the Republican.**

Two other patterns involving the votes recorded for the statewide offices need to be accounted for before the pattern that is tied to the balance of Democrat versus Republican votes for those offices can be seen clearly. First, **the CD-13 undervote rate is significantly higher for ballots that do not have a vote recorded for at least one of the statewide offices**.[3] It is not possible to say whether this contagion of undervoting reflects intentional actions by the affected voters or a defect in the voting machines that corrupted the votes for multiple offices on those ballots. We do not find any evidence that the relationship between undervoting for the statewide offices and undervoting in the CD-13 race is related to voting machine characteristics we are able to observe.

Second, **the CD-13 undervote rate tends to be substantially higher for ballots that have a third-party, "No Party Affiliation" (NPA) or write-in vote for one of the statewide offices**. Such third-party and write-in votes were recorded only for Senator (four NPA candidates) and Governor (three NPA candidates and one Reform Party candidate). Because the CD-13 race presented a choice between only a Democrat and a Republican, it is perhaps not surprising that voters who avoided the major-party alternatives when given a chance to do so decided to abstain when only major-party alternatives were present.

To study the relationship between the CD-13 undervote rate and the Democrat-Republican split among votes cast for the statewide offices, we focus on ballots that have Democrat or Republican votes for all five statewide offices.

For ballots that do not have a CD-13 undervote, the votes cast for the five statewide offices allow one to predict the vote likely to be cast for the CD-13 race with great accuracy. Someone who votes for five Republican candidates for the statewide offices is extremely likely to vote for

[3] In an odd way this result echoes the result observed by Herron and Sekhon (2003) for punchcard ballots in Broward and Miami-Dade counties during the 2000 election: ballots that had an overvote for president also tended to have overvotes for other offices.

the Republican U.S. House candidate, and someone who votes for five Democrats is extremely likely to vote for the Democrat U.S. House candidate. If the greater prevalence of CD-13 undervotes among the straight-Democrat ballots than among the straight-Republican ballots is not a result of voluntary nonvoting for that office, then the number of thwarted votes that were intended to go to Jennings is much larger than the number that were intended to go to Buchanan, and the difference greatly exceeds the official margin of victory.

We can say with great confidence that **the greater prevalence of CD-13 undervotes among the straight-Democrat ballots does not reflect a general tendency for those ballots to have more undervotes for competitive partisan races**. For instance, in the race for Hospital Board Southern District Seat 1, the undervote rate is higher for the statewide straight-Republican ballots than it is for the straight-Democrat ballots.

Voting machine properties Even before the end of early voting, concerns were being raised about voting machine malfunctions that could lead to a high undervote rate (Ruger 2006). Many voter complaints were reported by the Herald-Tribune and by Jennings's attorneys (Gulliver, Allen, and Gluck 2006; Wallace 2006b; Coffey and Herron 2006; Coffey, Herron, Verrilli, Hirsch, Hauck, and Amunson 2006). One report notes, "The problem that has consistently arisen is that when a voter makes a selection for Ms. Jennings on the electronic voting equipment, her opponent's name is either immediately highlighted, or appears incorrectly as the choice in the machine summary" (Wallace 2006a).

It is an unfortunate fact that no feasible amount of testing or examination of modern computer systems can *rule out* machine malfunctions, which are often subtle and unreproducible. This is a classic case where one "cannot prove a negative." Hence, machine malfunctions can never be ruled out as a cause of another problem, unless the problem has been traced to other causes with certainty.

Testing or examination might be able to *demonstrate* a malfunction. As noted above, testing of the machines has so far not shown there to be a machine problem causing the high undervote rate. Such a problem may still be revealed, since the State's investigation continues. It is still possible that the Jennings team or nonpartisan organizations will be allowed to test the machines and examine their design, and it is possible that this would yield interesting results.

Statistical analysis may also point to machine malfunctions as a cause of the undervote. For example, a very strong statistical link between a specific error message in the machine's event log, and a high undervote rate on that machine may be an indication of a mechanical cause of high undervotes. We examine this possibility and find a statistically significant relationship between certain machine events and variations in undervoting. **One particular error message ("Invalid vote PEB") is both directly associated with variations in the CD-13 undervote rate and related to differences in the relationship between the CD-13 undervote rate and the pattern of votes cast for the five statewide offices**. Hundreds of CD-13 undervotes are related to variations in the occurrence of this error message. Both the prevalence of the error message and its direct relationship with the CD-13 undervote differ between the early voting and election day data. In the early voting data the error message occurs on about 54 percent of the voting machines, most of the events occur during two hours in the morning, and the CD-13 undervote rate is lower on machines that have the error. In the election day data the error message occurs on about 12 percent of the voting machines, the events are scattered relatively evenly throughout the day, and the CD-13 undervote rate is higher on machines that have the error.

Of course, other machine properties besides malfunctions might lead to variations in undervote rates. We find **the CD-13 undervote rate is related in a complicated way to two features of how the voting machines are configured or deployed**. A few machines—one per precinct on election day—are configured for accessibility by voters with disabilities. We identify such machines based on a distinctive code associated with the PEBs recorded as being used with them. Among all the machines, most are used with multiple PEBs and these PEBs are shared among multiple machines. **In the election day data the CD-13 undervote rate tends to be higher on the accessibility configured machines than on the other machines, and the CD-13 undervote rate is significantly lower on the machines that were not accessibility configured but were used with a unique combination of PEBs. In the early voting data the CD-13 undervote rate is higher on machines that were accessibility configured or that shared PEBs with machines that were so configured than it is on the other machines.** Coffey et al. (2006) identify a tendency for the CD-13 undervote rate to be higher on machines that have a later date for the "Terminal clear and test" event. The machines with the later "Terminal clear and test" events are predominantly machines configured for accessibility.[4]

Ballot design/display The currently prevailing explanation of the high CD-13 undervote focuses on the format of the ballot as displayed on the iVotronic screen (Tamman and Doig 2006). In Sarasota, the first display was taken up entirely by Senate candidates, including several "No Party Affiliation" candidates and a write-in. The next display had a small heading for the CD-13 race, the Republican and Democratic candidates, followed by a large teal rectangle announcing the state-level races, and the rest of the screen was filled with candidates for the Governor's race. It is hypothesized that the CD-13 race was overlooked by many voters, who also failed to notice that the summary screen had no CD-13 vote. This explanation could account for the lower CD-13 undervote rates in neighboring counties and on Sarasota's absentee ballots, because this ballot design layout was used for the CD-13 race only on iVotronics in Sarasota County. This theory is consistent with the many voters who complained of missing the CD-13 race entirely (e.g. Gluck, Allen, and Saewitz 2006). As Frisina et al. (2006) point out, the Florida Attorney General's race in Charlotte and Lee counties (but not Sarasota) had a similar ballot layout issue on the iVotronics—and even higher undervote rates than the CD-13 race.

It seems likely that ballot design contributed to higher CD-13 undervote rates in Sarasota County, but the ballot does not explain all of the high undervotes. Stewart observes that the amount of CD-13 undervoting some attribute to the ballot design is very high compared to ballot-related undervote effects reported in the political science literature (Stewart 2006). The theory cannot account for the many voters who complained of apparently selecting a candidate for the CD-13 race, only to discover that the vote had disappeared when the summary screen was displayed (e.g. Gluck et al. 2006). The explanation also does not account for the large difference between the undervote rates in early and election day voting. News reports state that Supervisor of Elections Kathy Dent "sent an e-mail to those manning the precincts after complaints during early voting made it clear that people were reading over the District 13 race," but most of those

[4]The last five dates on which "Terminal clear and test" events occur are 10/23/2006 through 10/26/2006 and 11/05/2006. Of the 169 machines with "Terminal clear and test" events on those dates, 159 are machines that have "Normal ballot cast" events for PEB 0, which we surmise indicates the machine is accessibility configured.

who called the Sarasota Herald-Tribune to describe their experience on election day "said they were not warned" (Gluck et al. 2006).[5]

The ballot design argument cannot in any simple way explain why different kinds of voters seem to have responded to the design very differently, nor why the same voters responded differently on different occasions. **The primary features of the ballot design that supposedly affected the CD-13 undervote rate seem not to have induced similar rates or patterns of undervoting when they appeared elsewhere in the Sarasota ballot pages**. The race for Hospital Board Southern District Seat 1 had several of those features: the two candidates running for that office were listed at the top of the page followed by a bold heading in a teal box and then the choices for other offices. But **the same voters who undervoted significantly more often for the CD-13 race undervoted significantly less often for the Hospital Board race**. To explain such variations in terms of the ballot design, one would need to believe the same stimulus provoked some voters to make errors on one page but not on another page.

Our analysis shows several other factors that are significantly related to large variations in the undervote rate, such as the details of statewide partisan voting patterns. Since it is unclear why there should be a strong correlation between these factors and response to ballot layout, we conclude that ballot format is not a complete explanation of CD-13 undervoting.

Minority precincts. Early news coverage pointed out that the CD-13 undervote rate was high in precincts that had above average proportions of black and Hispanic voters (Marsteller 2006). We find that the estimated relationships among the CD-13 undervote rate, statewide office voting patterns and voting machine characteristics are unchanged if we omit from the analysis the 32 precincts that have percentages of black and Hispanic registered voters greater than the county average.

Voter revulsion. One theory advanced by the local Supervisor of Elections, the Republican candidate, and others was that the undervotes represented protests by voters disgusted by a negative campaign. A sharp version of this explanation focuses on Republican voters who may have been reacting to "a bitter Republican primary. Buchanan won the GOP nomination with 32 percent of the vote in a five-candidate race, handily taking Manatee but losing in Sarasota" (Marsteller 2006). Susan MacManus was quoted as saying, "It's not hard to understand why the undervote would be greater in Sarasota than in Manatee, because Manatee voters more heavily supported Buchanan, whereas, in Sarasota, Republicans supported a lot of different Republicans" (Marsteller 2006).

As several others have noted (Frisina et al. 2006), this explanation is highly implausible in light of the much lower CD-13 undervote rates on Sarasota's absentee ballots, which were cast in the same time interval as Sarasota's electronic early voting, and on paper or electronic voting technology in neighboring counties that included parts of CD-13. We did not give this claim further consideration.

[5]The set of people the newspaper interviewed was not a random sample. "The Herald-Tribune interviewed and compiled responses from more than 300 voters who contacted the newspaper about their problems at the polls" (Gluck et al. 2006).

Data

Our analysis uses information from two key sources, namely the *event logs* and the *vote image logs* produced from the voting machines. The ES&S iVotronic voting machines used in Sarasota for the 2006 election maintain redundant electronic copies of important audit information. One record is the event log, which contains a sequence of time-stamped events for each voting machine. The iVotronic machines use Personalized Electronic Ballots (PEBs), and the event log records the PEB that was being used in the machine at the time each event occurred. The event log is not especially detailed. It records events such as clearing and testing the machine before the election, opening it to receive votes, printing zero tapes, casting votes, cancelling votes, changing the date and time on the machine and use of invalid PEBs. Table 1 shows all the different types of events that occur in the event log file and the frequency with which each type of event occurs. Another record is the vote image log, which contains a record of the combination of votes cast by each voter. The vote records are not actual screen images, but ASCII text records of the individual votes, in randomized order to protect the secrecy of the ballots. We obtained them from the Orlando Sentinel newspaper, which obtained them from Sarasota County under Florida's open records laws.[6]

The audit information from each machine can be collected and uploaded by elections office staff to ES&S's Unity election management system, which runs on a server computer. Unity can generate reports in various formats, including separate reports containing the event logs from all of the machines and vote records from all machines. These data are useful for diagnosing machine problems, printing out paper records of the votes for hand recounts to meet the "manual audit capacity" mandate for voting equipment under HAVA, and, occasionally, recovering votes that would otherwise have been lost due to serious hardware, software, or administrative error.

The analysis discussed here is based on individual voted ballots extracted from the vote image log files and on information about voting machines and PEBs taken from the event log file. The date information in the event file is used to distinguish voting machines that were used during the early voting period from machines used on election day. Vote records are shuffled in the vote image log file, so it is not possible to associate a particular voted ballot with any particular vote-casting event. The vote image log file indicates the voting machine on which each ballot was cast, as well as the precinct. The total number of votes cast in the vote image log file (119,919) equals the number of votes cast during early voting, on election day and using provisional ballots described in an "Un-official General Election Report" dated November 17, 2006, 6:24 pm.[7]

Key questions

The motivation for the analysis is to find evidence for or against various explanations for the high CD-13 undervote rate. Frisina et al. (2006) use comparisons with votes cast for other offices in Sarasota and for various offices in other Florida counties to make a strong case that a principal reason for the high undervote rate is the format of the ballot. But the format of the ballot, which for the CD-13 race was constant for all early voting and election day voters in the county, cannot

[6] The event log file is EL152All.txt, with run date 11/20/06 11:02 AM. The vote image log files are EL155pct1_25.txt, EL155pct26_50.txt, EL155pct51_75.txt, EL155pct101_125.txt, EL155pct126_999.txt and EL155pct76_100.txt, with run date 11/20/06 10:37 AM.

[7] The report is contained in a file named "EL45A Complete 111706.txt," run date 11/17/06 06:24 PM.

Table 1: Events Recorded in the Sarasota 2006 Event Log File

Code	Description	Count
01	Terminal clear and test	1,503
02	Terminal screen calibrate	20
04	Enter service menu	70
05	Service password fail	2
06	Enter ECA menu	34
08	Date/time change	57
09	Terminal open	1,503
10	Terminal close	1,503
13	Print zero tape	366
14	Print Precinct results	5
17	Votes recollect	1
18	Invalid vote PEB	308
19	Invalid super PEB	48
20	Normal ballot cast	119,772
21	Super ballot cast	184
22	Super ballot cancel	225
25	Open with super votes	6
26	Terminal left open	1
27	Override	21
28	Override fail	1
36	Low battery lockout	66
37	Nonmaster PEB collection	8
51	PEB/CF Election ID mismatch	1

Notes: Code and frequency of events in the event log file `EL152All.txt`.

explain variations that may be observed in the undervote rate across different sets of the county's voters. Undervotes may be related to two classes of factors that vary from voter to voter: variations in the voting propensities of different voters; and variations in the conditions under which votes are cast.

Voter propensities include considerations such as each voter's commitment, neutrality or general skepticism regarding the major political parties. Different voters may also be more or less likely to make mistakes or to abstain from voting in particular races. Ballots that have different patterns of voting for the statewide offices reflect voters who have different preferences and other characteristics. We cannot observe the attributes of the voters that led them to cast their statewide votes in the different patterns, so we use the patterns as proxies for the various kinds of voters. The CD-13 undervote rate varies significantly across the statewide office voting patterns.

Variations in the CD-13 undervote rate are also correlated with differences and irregularities in the operation of different voting machines that were used to record votes cast in the election. The principal circumstance that can vary from voter to voter is whether the voting machine the voter is assigned to use performs correctly. The number of undervotes associated with apparent variations in voting machine performance is considerably greater than the difference in votes recorded for the respective candidates in the CD-13 race. Hence the number of undervotes that may have been caused by voting machine deficiencies is large enough to have changed the outcome of the election.

The remainder of this paper presents the detailed results that support the conclusions about undervoting in the Sarasota CD-13 race we summarized above. First we consider differences among voters, classifying voters by their statewide office voting patterns. Then we look at differences among voting machines, focusing on selected aspects of the machines' use that can be observed using information in the event log file. Then we consider interactions between the observable differences among voters and the attributes of the voting machines. An important question in this latter analysis is whether observable characteristics of the machines somehow enhance or trigger the voter effects. Finally we show that the systematic variations we identify in the CD-13 undervote do not depend on the precincts in Sarasota that have above average proportions of black or Hispanic voters.

CD-13 Undervote Differences Among Voters

Using information in the vote image log files, we can distinguish several sets of ballots that exhibit substantially different rates of undervoting in the CD-13 race. We key these discriminations to the votes recorded on the ballots for the five statewide offices on the Sarasota 2006 ballot, namely U.S. Senator, Governor, Attorney General (AG), Chief Financial Officer (CFO) and Commissioner of Agriculture (CA). As Table 2 indicates, for Senator and for Governor some ballots had votes for third-party, nonparty or write-in candidates. For the other offices only votes for the Democrat or the Republican were recorded, or no vote was recorded. As Table 2 shows, the CD-13 undervote rate tends to be substantially higher for ballots that have a third-party, nonparty or write-in vote, and in most instances the CD-13 undervote rate is even higher for ballots that do not have a vote recorded for any one of the statewide offices. Table 3 shows that the CD-13 undervote rate does not vary monotonically with the number of statewide offices that have undervotes. The CD-13 undervote rate varies somewhat across the statewide offices for the ballots that have a vote for the Democrat or the Republican, moreso for the former

Table 2: Sarasota 2006, Votes for Statewide Offices and CD-13 Undervotes

Election Day

Vote	Senator rate	Senator count	Governor rate	Governor count	Atty Genl rate	Atty Genl count	CFO rate	CFO count	Comm Ag rate	Comm Ag count
Dem	.151	49,676	.155	36,675	.137	39,141	.144	43,214	.150	33,912
Rep	.106	36,548	.114	47,795	.104	45,643	.105	41,659	.105	50,347
Other	.286	1,687	.259	3,316	—	0	—	0	—	0
Under	.444	1,114	.299	1,239	.535	4,241	.436	4,152	.428	4,766

Early Voting

Vote	Senator rate	Senator count	Governor rate	Governor count	Atty Genl rate	Atty Genl count	CFO rate	CFO count	Comm Ag rate	Comm Ag count
Dem	.199	18,786	.207	14,766	.185	15,531	.192	16,436	.203	13,451
Rep	.127	11,417	.138	15,110	.124	13,972	.126	13,194	.125	15,912
Other	.348	411	.330	584	—	0	—	0	—	0
Under	.432	259	.257	413	.612	1,370	.502	1,243	.481	1,510

Notes: Rates are the proportion of the ballots in each category that have a CD-13 undervote. The counts show the number of ballots having the respective number of votes for the five statewide offices. The "Other" category includes ballots with votes for a third-party, nonparty or write-in candidate. Ballots that have no votes recorded for any offices are not included.

than the latter. The CD-13 undervote rate is always higher for ballots that have a Democrat vote for a statewide office than for ballots with a Republican vote.

Tests show that even though ballots that include a third-party, nonparty or write-in vote for Senator or Governor are more likely to have undervotes for the remaining statewide offices, the third-party voting and statewide undervote variables are each independently related to the CD-13 undervote rate. Loglinear analyses (Agresti 1990, 143–149, 174–176) of the three-way contingency table relating the CD-13 undervote to a variable that counts the number of statewide office undervotes and a variable that measures whether there is a third-party, nonparty or write-in vote for either Senator or Governor shows there is no significant three-factor interaction. Respectively for election day and early voting, the likelihood-ratio chi-squared statistics for the three-factor interaction are 8.1 and 5.3, each with five degrees of freedom. The upper tail probabilities of the chi-squared distribution with five degrees of freedom are .15 and .39, which indicates the three-factor interaction is insignificant.

The relationship between third-party voting and undervoting in the CD-13 race may reflect the distaste third-party votes may feel regarding major party candidates in general. Third-party voters who feel that way are more likely to abstain when no third-party option is available. The relationship between undervoting for the statewide offices and undervoting in the CD-13 race may reflect individual voter propensities or it may reflect problems with voting machines that for some voters affected multiple offices. There is no indication, however, that the relationship between undervoting for the statewide offices and undervoting in the CD-13 race is affected by the

Table 3: Sarasota 2006, Undervotes for Statewide Offices and CD-13 Undervotes

		Number of Undervotes for Statewide Offices					
		0	1	2	3	4	5
Election day	CD-13 undervote rate	.11	.37	.58	.31	.51	.80
	total ballot count	80,824	3,612	2,491	1,597	378	140
Early voting	CD-13 undervote rate	.15	.44	.67	.32	.50	.85
	total ballot count	28,223	1,261	796	459	105	33

Notes: Rates are the proportion of the ballots in each category that have a CD-13 undervote. The total ballot counts show the number of ballots having the respective number of undervotes for the five statewide offices. Ballots that have no votes recorded for any offices are not included.

measures of voting machine characteristics that we consider further below. In any event, the demonstrably higher probabilities that such voters cast a CD-13 undervote means that to try isolate how voting machine variations may have affected the CD-13 undervote rate, it is better to focus on ballots that do not have third-party, nonparty or write-in votes or statewide office undervotes. For the remainder of this paper, the analysis we report includes only such ballots.

The ballots that remain in the analysis all have votes for either the Democrat or the Republican for each of the five statewide offices. There are 32 possible patterns of those statewide votes. We use the pattern of the votes for the statewide offices to put each ballot into one of 32 subsets. Assuming that the votes recorded for the statewide offices are the same as the votes each voter actually cast, the pattern of Democrat and Republican choices for those offices indicate each voter's basic partisan predisposition. Knowing only the votes cast for the five statewide offices, one can predict with great accuracy the vote likely to be cast for the CD-13 race. Someone who votes for five Republican candidates for the statewide offices is extremely likely to vote for the Republican U.S. House candidate, and someone who votes for five Democrats is extremely likely to vote for the Democrat U.S. House candidate. Ballots that have a mix of Democrat and Republican votes for the statewide offices reflect voters who have a mix of concerns tied to particular offices and particular candidates that they apparently do not believe are well expressed by a homogeneous partisan choice.

Each distinct pattern of such votes may reflect a distinctive collection of local or national concerns, and voters who cast the same pattern of statewide office votes may be more similar to one another than they are to voters who cast different patterns. Such similarities and differences may also relate to similarities and differences in voters' propensities to undervote in the CD-13 race. The CD-13 undervote rate may vary across the different patterns of statewide office voting either because the voters have divergent tastes and habits, or because some feature of the voting machine technology treats the votes cast by different voters differently. Bugs or malicious hacks placed in the voting machines' software, firmware or hardware may be keyed to act differently when presenting the ballot or when recording the vote cast in the CD-13 race, depending on the vote choices entered for the five statewide offices.

Before we examine the relationship between the CD-13 undervote rate and various observable features of the voting machines, let's look at how the CD-13 undervote rate varies in relation to the pattern of Democrat and Republican votes cast for the five statewide offices. Table 4 shows the number of election day and early voting ballots that have each of the 32 different patterns of Democrat and Republican votes for the statewide offices, along with the CD-13 undervote rate for the ballots that have each pattern.

The CD-13 undervote rate varies considerably across the sets of ballots. Two substantial patterns are readily apparent. First, the CD-13 undervote rate is usually higher during early voting than on election day. For those statewide office voting patterns that have a large enough number of ballots for the comparison to be statistically reliable, the early voting CD-13 undervote rate is always greater than the election day CD-13 undervote rate. Usually the percentage of early voting CD-13 undervotes is on the order of two to five percent greater than the election day percentage. Second, the CD-13 undervote rate is much higher for ballots that voted Democrat for all five statewide races than it is for ballots that voted Republican for all five statewide races. The difference between the CD-13 undervote percentages is 5.2 percent in the election day data and 8.3 percent in the early voting data.

There are other systematic differences among the patterns that may not so readily apparent. For example, in many instances statewide office voting patterns that differ only in the vote for CA have a higher CD-13 undervote rate for the pattern that has a vote for the Democrat CA candidate: compare, e.g., patterns 2 and 10, 3 and 13, and 4 and 15.

To put the relationship between the CD-13 undervote rate and the patterns of partisan voting for the statewide offices in some perspective, it is worthwhile to compare it to the relationship between these same partisan patterns and the undervote rate in the race for Hospital Board Southern District Seat 1. This is a useful comparison to consider, because the Hospital Board race was a contested, partisan race that appeared on the electronic ballot with visual features similar to the ones that supposedly distracted many voters away from voting for the CD-13 race. On page two of the ballot, the CD-13 alternatives—a Republican and a Democrat—appear at the very top of the page. Below those alternatives the capitalized word "STATE" appears in a teal box, followed by the seven alternatives available for choice for Governor. On page six of the ballot, the Hospital Board alternatives—a Republican and a Democrat—appear at the very top of the page. Below those alternatives the capitalized word "NONPARTISAN" appears in a teal box, followed by questions asking whether two Justices of the Supreme Court shall be retained. Plainly the offices differ and the context of placement is not exactly the same for the CD-13 race and the Hospital Board race. But if there is a predominant ballot format effect that is triggered by the placement of the alternatives on the electronic screen and by the labeling and color scheme, we might expect the rates and perhaps the patterns of undervoting to be similar for the two races.

Inspection of Table 5, which presents the Hospital Board undervote rate in relation to the pattern of Democrat and Republican votes cast for the five statewide offices, using the same format as in Table 4, reveals that in basic ways the distribution of undervotes for the Hospital Board is not the same as the distribution for CD-13. For instance, the Hospital Board undervote rate is frequently higher on election day than it is during early voting. In the election day data the Hospital Board undervote rate is greater for ballots that voted Republican for all five statewide races than it is for ballots that voted Democrat for all five statewide races. And in many instances statewide office voting patterns that differ only in the vote for CA have a higher CD-13 undervote rate for the pattern with the Republican vote for CA: compare, e.g., patterns 2 and 10, 3 and 13, and 4 and 15. In general, for each corresponding statewide office voting pattern, the undervote rate for the Hospital Board race is substantially lower than the CD-13 undervote rate.

Table 4: Sarasota 2006, Partisan Voting Patterns and CD-13 Undervotes

	Office					Election Day		Early Voting	
						Undervote	Ballot	Undervote	Ballot
Pattern	Sen	Gov	AG	CFO	CA	Rate	Count	Rate	Count
1	D	D	D	D	D	.131	25,326	.185	10,764
2	R	D	D	D	D	.179	330	.204	108
3	D	R	D	D	D	.129	2,035	.182	742
4	D	D	R	D	D	.110	801	.172	204
5	D	D	D	R	D	.158	330	.161	155
6	D	D	D	D	R	.085	4,141	.108	1,580
7	R	R	D	D	D	.138	246	.125	72
8	R	D	R	D	D	.192	52	.167	12
9	R	D	D	R	D	.161	31	.091	11
10	R	D	D	D	R	.139	137	.133	30
11	D	R	R	D	D	.109	521	.118	161
12	D	R	D	R	D	.136	236	.169	89
13	D	R	D	D	R	.093	1,158	.129	387
14	D	D	R	R	D	.163	141	.174	46
15	D	D	R	D	R	.100	769	.119	219
16	D	D	D	R	R	.130	338	.217	143
17	R	R	R	D	D	.118	153	.231	39
18	R	R	D	R	D	.120	83	.143	14
19	R	R	D	D	R	.101	278	.103	87
20	R	D	R	R	D	.114	35	.111	18
21	R	D	R	D	R	.105	133	.178	45
22	R	D	D	R	R	.156	64	.000	17
23	D	R	R	R	D	.113	300	.124	89
24	D	R	R	D	R	.103	1,654	.104	451
25	D	R	D	R	R	.121	552	.129	232
26	D	D	R	R	R	.139	595	.223	175
27	R	R	R	R	D	.081	284	.062	65
28	R	R	R	D	R	.075	1,144	.118	314
29	R	R	D	R	R	.077	439	.140	143
30	R	D	R	R	R	.140	371	.245	98
31	D	R	R	R	R	.117	5,067	.131	1,553
32	R	R	R	R	R	.079	29,364	.102	9,455

Office abbreviations: U.S. Senator (Sen), Governor (Gov), Attorney General (AG), Chief Financial Officer (CFO) and Commisioner of Agriculture (CA). Rates are the proportion of the ballots in each category that have a CD-13 undervote.

Table 5: Sarasota 2006, Partisan Voting Patterns and Hospital Board Undervotes

			Office			Election Day		Early Voting	
Pattern	Sen	Gov	AG	CFO	CA	Undervote Rate	Ballot Count	Undervote Rate	Ballot Count
1	D	D	D	D	D	.040	25,326	.042	10,764
2	R	D	D	D	D	.045	330	.037	108
3	D	R	D	D	D	.066	2,035	.081	742
4	D	D	R	D	D	.065	801	.098	204
5	D	D	D	R	D	.124	330	.077	155
6	D	D	D	D	R	.074	4,141	.078	1,580
7	R	R	D	D	D	.069	246	.028	72
8	R	D	R	D	D	.154	52	.000	12
9	R	D	D	R	D	.000	31	.091	11
10	R	D	D	D	R	.080	137	.067	30
11	D	R	R	D	D	.084	521	.087	161
12	D	R	D	R	D	.055	236	.045	89
13	D	R	D	D	R	.115	1,158	.119	387
14	D	D	R	R	D	.106	141	.043	46
15	D	D	R	D	R	.120	769	.114	219
16	D	D	D	R	R	.107	338	.084	143
17	R	R	R	D	D	.098	153	.077	39
18	R	R	D	R	D	.084	83	.071	14
19	R	R	D	D	R	.068	278	.103	87
20	R	D	R	R	D	.114	35	.056	18
21	R	D	R	D	R	.105	133	.067	45
22	R	D	D	R	R	.047	64	.118	17
23	D	R	R	R	D	.100	300	.045	89
24	D	R	R	D	R	.141	1,654	.131	451
25	D	R	D	R	R	.100	552	.095	232
26	D	D	R	R	R	.108	595	.103	175
27	R	R	R	R	D	.095	284	.062	65
28	R	R	R	D	R	.088	1,144	.089	314
29	R	R	D	R	R	.084	439	.063	143
30	R	D	R	R	R	.094	371	.051	98
31	D	R	R	R	R	.099	5,067	.092	1,553
32	R	R	R	R	R	.046	29,364	.039	9,455

Office abbreviations: U.S. Senator (Sen), Governor (Gov), Attorney General (AG), Chief Financial Officer (CFO) and Commisioner of Agriculture (CA). Rates are the proportion of the ballots in each category that have a CD-13 undervote.

Clearly the visual features of placement, labeling and color that the CD-13 race and the Hospital Board race have in common do little or nothing to induce similar patterns of undervoting between them. Something else significantly affected the votes cast for the CD-13 race, something that greatly increased and altered the partisan distribution of the undervotes recorded in that case.

CD-13 Undervote Differences Among Voting Machines

Now we consider features of the voting machines and of the ways the voting machines were used. We highlight three aspects of the machines' use that can be observed using information in the event log file: which machines had at least one occurrence of event 18, "Invalid vote PEB"; which machines were used to record votes with a PEB that had a distinctive identifying code; and which machines were used with the same PEBs.

On its face, the occurrence of event 18 for a voting machine would appear to signal that something untoward occurred. Consider the first occurrence of such an event in the event log file, as shown the excerpt from the file that appears in Table 6. The columns indicate respectively identifying numbers for the voting machine and the PEB in use for each event, the type of the PEB ("SUP" for supervisor and "VTR" for voter), the date and time of the event, and the code and description of the event. The dates recorded for this particular voting machine mark it as a machine that was used during early voting. When the voting machine number or PEB number and type for an event are the same as for the preceding event printed on the same page of the event log file, the information is not repeated and the respective space is left blank. Notice that about midway down the Table, the PEB is shown with identifying number "0" and there is an occurrence of event 18. In this instance, the "Invalid vote PEB" event occurred about 93 minutes before the first vote was recorded on the machine on that particular day. In the event log file more events are recorded for this particular machine after that event, including many "Normal ballot cast" events and three more "Invalid vote PEB" events. One of these additional occurrences of event 18 appears near the bottom of Table 6. Each of the three subsequent occurrences of event 18 for this voting machine happened at approximately the same time of day (specifically, on 10/31/2006 at 08:13:12, on 11/01/2006 at 08:11:16 and on 11/02/2006 at 08:10:14).

Table 7 shows an excerpt from the event log file for a machine that was used on election day. There is one occurrence of event 18 for this machine, happening early in the morning after a few ballots had been cast on the machine but before most. In all, there were 42 "Normal ballot cast" events on this particular machine after the event 18 occurrence, the last "Normal ballot cast" event happening on 11/07/2006 at 19:00:30.

Table 8 shows the number of voting machines having at least one occurrence of event 18 respectively during early voting and on election day. More than half of the early voting machines had such an event, compared to about 12 percent of the election day machines that did so. Table 9 shows the hours during the day when occurrences of event 18 are recorded. On election day there is a slight concentration of the events between 700 and 800 hours, but otherwise the events are scattered relatively evenly throughout the day. During early voting, 80 percent of the events occur between 800 and 900 hours, and the remaining events are scattered between 1000 and 1800 hours.

PEBs with an identifying number of "0" sometimes occur in conjunction with the seemingly normal casting of a ballot. Nonetheless, such occurrences are distinctive and we flag them. This is the second aspect of voting machine usage that we highlight. Table 10 shows an excerpt from the event log file with an example of such occurrences on an election day machine. All the ballots

Table 6: Sarasota 2006, Event Log File Entries for the First Voting Machine

Votronic	PEB#	Type	Date	Time	Event
103416	107394	SUP	10/18/2006	08:37:09	01 Terminal clear and test
	109213	SUP	10/23/2006	08:04:50	09 Terminal open
	109462	SUP	10/23/2006	09:45:37	20 Normal ballot cast
			10/23/2006	11:02:01	20 Normal ballot cast
	105061	SUP	10/23/2006	11:50:30	20 Normal ballot cast
	109462	SUP	10/23/2006	14:29:00	20 Normal ballot cast
	107042	SUP	10/23/2006	15:30:14	20 Normal ballot cast
			10/24/2006	09:16:18	20 Normal ballot cast
	109462	SUP	10/24/2006	10:24:08	20 Normal ballot cast
			10/24/2006	12:58:50	20 Normal ballot cast
	105061	SUP	10/24/2006	15:56:16	20 Normal ballot cast
	107042	SUP	10/24/2006	16:13:16	20 Normal ballot cast
	109462	SUP	10/25/2006	09:31:27	20 Normal ballot cast
	107042	SUP	10/25/2006	09:58:13	20 Normal ballot cast
			10/25/2006	10:49:57	20 Normal ballot cast
			10/25/2006	13:16:48	20 Normal ballot cast
	105061	SUP	10/25/2006	14:12:08	20 Normal ballot cast
	107042	SUP	10/25/2006	15:31:45	20 Normal ballot cast
	109462	SUP	10/26/2006	11:13:27	20 Normal ballot cast
			10/26/2006	13:17:18	20 Normal ballot cast
	105061	SUP	10/26/2006	15:48:39	20 Normal ballot cast
	0	VTR	10/27/2006	08:28:35	18 Invalid vote PEB
	107042	SUP	10/27/2006	10:01:17	20 Normal ballot cast
	105061	SUP	10/27/2006	10:16:48	20 Normal ballot cast
	109462	SUP	10/27/2006	14:25:04	20 Normal ballot cast
	107042	SUP	10/28/2006	09:07:11	20 Normal ballot cast
			10/28/2006	10:53:15	20 Normal ballot cast
	105061	SUP	10/28/2006	11:31:09	20 Normal ballot cast
	107042	SUP	10/28/2006	12:27:19	20 Normal ballot cast
	109462	SUP	10/29/2006	13:25:41	20 Normal ballot cast
			10/30/2006	08:43:25	20 Normal ballot cast
	107042	SUP	10/30/2006	10:28:41	20 Normal ballot cast
	109462	SUP	10/30/2006	11:04:45	20 Normal ballot cast
	107042	SUP	10/30/2006	12:54:34	20 Normal ballot cast
	109462	SUP	10/30/2006	13:23:55	20 Normal ballot cast
			10/30/2006	14:03:56	20 Normal ballot cast
	107042	SUP	10/30/2006	15:21:49	20 Normal ballot cast
	109462	SUP	10/30/2006	15:51:23	20 Normal ballot cast
	105061	SUP	10/30/2006	16:46:18	20 Normal ballot cast
	0	VTR	10/31/2006	08:13:12	18 Invalid vote PEB
	107042	SUP	10/31/2006	08:40:48	20 Normal ballot cast
	109462	SUP	10/31/2006	10:06:59	20 Normal ballot cast

Table 7: Sarasota 2006, Event Log File Entries for an Election Day Voting Machine

103565	108961	SUP	09/26/2006 09:40:10	01 Terminal clear and test
	107219	SUP	11/07/2006 06:30:45	09 Terminal open
	107729	SUP	11/07/2006 07:12:03	20 Normal ballot cast
	107456	SUP	11/07/2006 07:18:15	20 Normal ballot cast
	107387	SUP	11/07/2006 07:23:08	20 Normal ballot cast
			11/07/2006 07:28:35	20 Normal ballot cast
			11/07/2006 07:41:55	20 Normal ballot cast
			11/07/2006 07:51:52	20 Normal ballot cast
			11/07/2006 07:58:09	20 Normal ballot cast
			11/07/2006 08:07:21	20 Normal ballot cast
			11/07/2006 08:22:02	20 Normal ballot cast
			11/07/2006 08:29:08	20 Normal ballot cast
	0	VTR	11/07/2006 08:34:01	18 Invalid vote PEB
	107387	SUP	11/07/2006 08:39:14	20 Normal ballot cast
	107456	SUP	11/07/2006 08:45:53	20 Normal ballot cast
			11/07/2006 08:56:57	20 Normal ballot cast
			11/07/2006 09:00:16	20 Normal ballot cast
	107387	SUP	11/07/2006 09:10:27	20 Normal ballot cast
			11/07/2006 09:43:59	20 Normal ballot cast
	109208	SUP	11/07/2006 09:50:52	20 Normal ballot cast
	107387	SUP	11/07/2006 09:54:19	20 Normal ballot cast
	107456	SUP	11/07/2006 10:03:15	20 Normal ballot cast
	109208	SUP	11/07/2006 10:14:16	20 Normal ballot cast
	107387	SUP	11/07/2006 10:21:24	20 Normal ballot cast

Table 8: Sarasota 2006, Voting Machines with Event 18 (“Invalid Vote PEB”) by Election Period

	Event 18	
	No	Yes
Early Voting	39	45
Election Day	1,242	173

Table 9: Sarasota 2006, Occurrences of Event 18 (“Invalid Vote PEB”) by Hour During the Day

	Hour													
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
Early Voting	0	48	40	0	2	2	8	1	4	1	3	1	0	0	110
Election Day	6	29	10	18	18	11	16	16	15	15	15	16	10	3	198

Table 10: Sarasota 2006, Event Log File Entries for an Election Day Voting Machine with Ballots Cast using PEB 0

Votronic	PEB#	Type	Date	Time	Event
103609	116292	SUP	10/25/2006	14:18:38	01 Terminal clear and test
	100828	SUP	11/07/2006	06:40:43	09 Terminal open
			11/07/2006	06:43:45	13 Print zero tape
			11/07/2006	06:45:06	13 Print zero tape
	0	VTR	11/07/2006	07:10:41	20 Normal ballot cast
			11/07/2006	07:22:17	20 Normal ballot cast
			11/07/2006	08:45:52	20 Normal ballot cast
			11/07/2006	08:57:12	20 Normal ballot cast
			11/07/2006	09:05:03	20 Normal ballot cast
			11/07/2006	09:09:41	20 Normal ballot cast
	100828	SUP	11/07/2006	09:13:09	22 Super ballot cancel
	0	VTR	11/07/2006	09:23:44	20 Normal ballot cast
			11/07/2006	09:38:03	20 Normal ballot cast
			11/07/2006	09:47:42	20 Normal ballot cast
			11/07/2006	09:54:05	20 Normal ballot cast
	100828	SUP	11/07/2006	09:58:50	22 Super ballot cancel
	0	VTR	11/07/2006	10:12:56	20 Normal ballot cast
			11/07/2006	10:20:48	20 Normal ballot cast
			11/07/2006	10:31:16	20 Normal ballot cast
			11/07/2006	10:42:07	20 Normal ballot cast
			11/07/2006	10:51:08	20 Normal ballot cast

cast on this machine are associated with a PEB that has an identifying number of “0.” This is usually the situation with the machines that have “Normal ballot cast” events for PEB 0—all the ballots are cast using PEB 0—but a few machines that have such events also have a few occurrences of event 21, “Super ballot cast,” associated with a “SUP” PEB that has a non-zero identifying number. Table 11 shows an example of an election day machine that has such usage.

As of this writing we lack a list to show which voting machines used for the 2006 election in Sarasota were specially equipped to meet the accessibility requirements of HAVA, but we strongly suspect that the PEB 0 machines largely or entirely comprise the set of such machines. Sarasota used 156 precincts for the 2006 election, as as Table 12 shows, on election day there were 156 machines that were used to cast ballots with a PEB having an identifying number of “0.” During early voting there were nine PEB 0 machines, and according to a press release dated October 19, 2006, there were seven early voting locations.[8] All told, about 11 percent of the voting machines both during early voting and on election day are PEB 0 machines. If the PEB 0 machines were indeed accessibility configured machines, then we know they were running software or using

[8]The URL for the press release is http://www.srqelections.com/press_releases/press_release126.htm, accessed December 2, 2006.

Table 11: Sarasota 2006, Event Log File Entries for an Election Day Voting Machine with Ballots Cast using PEB 0 and a “SUP” PEB

106111	108031	SUP	10/25/2006 10:03:56	01 Terminal clear and test
	133071	SUP	11/07/2006 06:49:10	09 Terminal open
			11/07/2006 06:51:55	13 Print zero tape
			11/07/2006 06:53:21	13 Print zero tape
	0	VTR	11/07/2006 07:10:15	20 Normal ballot cast
			11/07/2006 07:23:45	20 Normal ballot cast
			11/07/2006 09:02:27	20 Normal ballot cast
			11/07/2006 09:08:40	20 Normal ballot cast
			11/07/2006 09:23:23	20 Normal ballot cast
	107591	SUP	11/07/2006 09:25:28	22 Super ballot cancel
	0	VTR	11/07/2006 09:32:33	20 Normal ballot cast
			11/07/2006 09:42:41	20 Normal ballot cast
			11/07/2006 10:02:56	20 Normal ballot cast
	107072	SUP	11/07/2006 10:04:39	22 Super ballot cancel
	0	VTR	11/07/2006 10:11:10	20 Normal ballot cast
			11/07/2006 10:22:03	20 Normal ballot cast
			11/07/2006 10:46:51	20 Normal ballot cast
			11/07/2006 10:50:49	20 Normal ballot cast
			11/07/2006 11:00:28	20 Normal ballot cast
			11/07/2006 11:13:02	20 Normal ballot cast
	107591	SUP	11/07/2006 11:20:53	21 Super ballot cast
	0	VTR	11/07/2006 11:25:31	20 Normal ballot cast
			11/07/2006 11:32:39	20 Normal ballot cast
			11/07/2006 11:40:22	20 Normal ballot cast
			11/07/2006 12:12:19	20 Normal ballot cast
			11/07/2006 12:26:40	20 Normal ballot cast
			11/07/2006 12:54:19	20 Normal ballot cast
			11/07/2006 13:23:43	20 Normal ballot cast
			11/07/2006 13:38:44	20 Normal ballot cast
			11/07/2006 14:17:28	20 Normal ballot cast
	118996	SUP	11/07/2006 14:59:56	22 Super ballot cancel
	0	VTR	11/07/2006 15:47:06	20 Normal ballot cast
			11/07/2006 16:18:23	20 Normal ballot cast
			11/07/2006 17:21:48	20 Normal ballot cast
			11/07/2006 17:37:05	20 Normal ballot cast
			11/07/2006 18:03:20	20 Normal ballot cast
			11/07/2006 18:20:18	20 Normal ballot cast
			11/07/2006 18:26:48	20 Normal ballot cast
	133071	SUP	11/07/2006 19:17:39	10 Terminal close

Table 12: Sarasota 2006, Voting Machines with Ballots Cast using PEB 0 by Election Period

	PEB 0 Ballots	
	No	Yes
Early Voting	75	9
Election Day	1,259	156

Table 13: Sarasota 2006, PEB Clusters by Election Period

	Exact Match Clusters		Network Clusters	
	Singleton	Multiple	Singleton	Multiple
Early Voting	12	8	8	7
Election Day	191	159	150	146

software options that differed from that on the voting machines that were not similarly capable.

The third aspect of voting machine usage that we highlight relates to what we call *PEB clusters*: clusters of voting machines that were used with the same PEBs. We have two different methods for identifying such clusters. For *exact match* clusters, machines are members of the same cluster if and only if exactly the same set of PEBs were used to cast ballots on the machines, where PEBs are identified using the identifying numbers in the event log file. The ballot casting events that are considered here are events 20 ("Normal ballot cast") and 21 ("Super ballot cast"). Each PEB that appears in the event log file with an identifying number of "0" is artificially assigned a unique code to link it only to the voting machine with which it is being used. So no machine that was used to cast ballots with a "0" PEB is clustered with any other machine. The other clustering method identifies *network* clusters. For this, machines are members of the same cluster if there is any PEB that was used to cast ballots on both of the machines. So two machines may belong to the same network cluster even though they do not have any PEB in common, as long as there is a third machine with which each shares a PEB, or there is a chain of machines that have PEBs in common that ultimately connect the two machines. Table 13 shows the number of clusters each method produces, with separate counts for the clusters that contain only one machine (*singleton*) and the clusters that contain more than one machine (*multiple*). The distinction is motivated by the idea that malicious software may be designed to defeat parallel testing by including features to detect whether the PEBs being used with it are being shared among multiple machines[9]. Table 14 shows the number of voting machines in each subset of clusters, with a separate category shown to distinguish the machines in clusters that contain or are networked with a PEB 0 machine.

Each of these three aspects of how the voting machines were used in the 2006 election in Sarasota relates to the CD-13 undervote rate. Table 15 shows that machines that have at least one occurrence of event 18 have a significantly higher CD-13 undervote rate in the election day data and a significantly lower CD-13 undervote rate in the early voting data. Table 16 shows that the CD-13 undervote rate is somewhat higher on PEB 0 machines in the election day data, but there is

[9]This idea was proposed by the second author in answer to a question during testimony before the U.S. Senate Committee on Rules and Administration, June 21, 2005

Table 14: Sarasota 2006, Machines in PEB Clusters by Election Period

	Exact Match Clusters			Network Clusters		
	PEB 0	Singleton	Multiple	PEB 0	Singleton	Multiple
Early Voting	9	3	72	15	0	69
Election Day	156	35	1,224	288	10	1,117

Table 15: Sarasota 2006, CD-13 Undervote Rate by Occurrence of Event 18 ("Invalid Vote PEB") by Election Period

		Event 18		Independence Test	
		No	Yes	Chi-squared	prob.
Election day	CD-13 undervote rate	.104	.112	5.6	.018
	total ballot count	67,137	9,971		
Early voting	CD-13 undervote rate	.155	.134	23.5	$< .0001$
	total ballot count	12,621	14,897		

Notes: Rates are the proportion of the ballots in each category that have a CD-13 undervote. Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

not a significant difference between the PEB 0 machines and the other machines in the early voting data.

Table 17 shows the distribution of the CD-13 undervotes across the various kinds of PEB clusters. The results shown for the exact match clusters refine the preceding results for the distinction between PEB 0 machines and the other machines. In the election day data the CD-13 undervote rate is significantly lower on machines in singleton clusters than on machines in multiple clusters or on PEB 0 machines (the difference between machines in multiple clusters and PEB 0 machines is not significant). In the early voting data the CD-13 undervote rate is larger for machines in singleton clusters than for the others, but the number of ballots on such machines is too small for the differences to be statistically significant. In the early voting data the CD-13 undervote rate is virtually the same for the PEB 0 machines and the multiple exact cluster machines. For the network clusters we find clusters that contain a PEB 0 machine have significantly higher CD-13 undervote rates in the early voting data. In the election day data, the differences across PEB 0, singleton and multiple clusters are slightly smaller using the network clusters than they are using the exact match clusters, and the decrease in the number of ballots on singleton cluster machines helps make the differences not statistically significant. This suggests that in the election day data there may be an effect that is tied to particular PEBs and particular machines, such that the kind of merging of clusters that happens when we move to the network clusters muddies the signal. In the early voting data, on the other hand, the PEB 0 machines seem not to be special on their own, but all the machines with which those machines interact have a distinctively higher CD-13 undervote rate.

Table 16: Sarasota 2006, CD-13 Undervote Rate by Occurrence of PEB 0 by Election Period

		PEB 0 Ballots		Independence Test	
		No	Yes	Chi-squared	prob.
Election day	CD-13 undervote rate	.104	.113	3.1	.078
	total ballot count	72,303	4,805		
Early voting	CD-13 undervote rate	.144	.139	.3	.568
	total ballot count	25,803	1,715		

Notes: Rates are the proportion of the ballots in each category that have a CD-13 undervote. Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

Table 17: Sarasota 2006, CD-13 Undervote Rate by PEB Clusters by Election Period

		Exact Match Clusters			Independence Test	
		PEB 0	Singleton	Multiple	Chi-squared	prob.
Election day	CD-13 undervote rate	.113	.091	.105	6.0	.049
	total ballot count	4,805	1,377	70,926		
Early voting	CD-13 undervote rate	.139	.218	.143	5.6	.060
	total ballot count	1,715	133	25,670		
		Network Clusters			Independence Test	
		PEB 0	Singleton	Multiple	Chi-squared	prob.
Election day	CD-13 undervote rate	.110	.086	.104	4.7	.096
	total ballot count	12,327	290	64,491		
Early voting	CD-13 undervote rate	.170	—	.139	23.5	$< .0001$
	total ballot count	3,085	0	23,713		

Notes: Rates are the proportion of the ballots in each category that have a CD-13 undervote. Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

The CD-13 undervoting effects associated with event 18 and the effects associated with PEB 0 and the PEB clusters are statistically independent of one another. Loglinear analyses of the three-way contingency table relating the CD-13 undervote to the event 18 (Table 15) and PEB cluster (Table 17) variables shows there is no significant three-factor interaction.

Plainly the correlates of the highlighted aspects of voting machine usage differ between the early voting and election day periods. Interactions between PEBs and machines are associated with variations in the CD-13 undervote rate, but evidently the underlying mechanism differs between the two time periods. Machines that have event 18 have a significantly higher CD-13 undervote rate in the election day data but a significantly lower CD-13 undervote rate in the early voting data. But the incidence of event 18 on election day machines also differs significantly from the incidence on early voting machines. Only about 12 percent of election day machines have event 18, but more than half of the early voting machines do. Moreover, many of the early voting machines have several occurrences of event 18, typically one on each of several but not all of the days the machine was being used during early voting. It appears that the occurrence of event 18 is signalling that different things were happening to the voting machines, or being done to them, during the two voting periods.

The effects associated with the PEB 0 indicator also vary between early voting and election day. In the election day data each machine that was used to cast ballots with a PEB having an identifying number of "0" tends to have a higher CD-13 undervote rate, but in the early voting data the cluster of machines that are connected to a PEB 0 machine through other PEBs have a higher CD-13 undervote rate.

If we use the instances where we observe significant differences in the CD-13 undervote rate across different configurations of machines, we can construct what-if counterfactual estimates to give a sense of the magnitudes of the effects in terms of the number of votes that may have been lost because of aberrant voting machine operations. The simplest method is, in each instance, to imagine that all the ballots had been cast with a CD-13 undervote rate corresponding to the lowest that is observed. So, referring to Table 15, we may suppose that all of the 77,108 election day ballots included in that analysis were cast with a CD-13 undervote rate of .104 (more precisely, the rate is .1039665). This gives a counterfactual estimate of 8,017 CD-13 undervotes, which we may compare to the actual value of 8,095 CD-13 undervotes and conclude that event 18 was associated with an extra 78 CD-13 undervotes in the election day data. Similarly we may suppose that all of the 27,518 early voting ballots included in Table 15 were cast with a CD-13 undervote rate of .134. This gives a counterfactual estimate of 3,687 CD-13 undervotes, which compares to 3,947 actual CD-13 undervotes, supporting a conclusion that event 18 was associated with an extra 260 CD-13 undervotes during early voting. Referring to Table 17, for the election day ballots we may focus on the exact match cluster results and suppose that all of the election day ballots there were cast with a CD-13 undervote rate of .091. The result is a counterfactual estimate of 7,000 CD-13 undervotes, which is a reduction of 1,095 from the number of CD-13 undervotes that actually occurred. Using the network cluster results for the early voting ballots, we may suppose that all of the early voting ballots in Table 17 were cast with a CD-13 undervote rate of .139. This gives a counterfactual estimate of 3,832 CD-13 undervotes and a counterfactual reduction of 115 undervotes. Because the event 18 and PEB cluster effects on the CD-13 undervote are statistically independent, it is reasonable to add these estimates together to find a counterfactual reduction of $78 + 260 + 1,095 + 115 = 1,548$ in the number of CD-13 undervotes if the lowest CD-13 undervote rates had been realized in each of the circumstances in which they are applicable.

CD-13 Undervote Differences Among Voters and Voting Machines

Two questions call for a combined analysis that considers the differences among voters together with the differences among voting machines. First is a skeptical question that essentially asks whether the effects we have identified are spurious. Do the effects on the CD-13 undervote rate that we find to be associated with each class of factors remain when we consider them together? Because the current analysis is correlational, it is always possible that one class of effects is nothing more than a proxy for the other. It may be, in particular, that the effects we have identified with aspects of voting machine usage simply reflect that voters who had different patterns of statewide office voting happened to use different configurations of the voting machines. In that case, the voting machine effects should disappear once we take the statewide voting differences between the ballots into account.

The second question runs in the other direction, asking whether there is evidence that the machine effects somehow enhance or trigger the voter effects. We noted that the relationship between the statewide office voting patterns and the CD-13 undervote rate differs substantially from the relationship between the statewide office voting patterns and undervotes in the Hospital Board race. In particular, the CD-13 undervote rates are comparatively higher among ballots that have more votes for Democrat statewide candidates. Is there any evidence that such increased CD-13 undervoting among Democrat-favoring ballots depends on observable aspects of the voting machines' performance?

To conduct the combined analysis we use binary logistic regression estimation assuming, where applicable, a binomial likelihood (McCullagh and Nelder 1989, 98–120, Agresti 1990, 79–90). The binary outcome variable (y) is whether ($y = 1$) or not ($y = 0$) each ballot contains a CD-13 undervote. This regression model is a standard and well-established method of analysis for binary observations. Loglinear analysis such as we have previously reported with reliance on the chi-squared distribution for making statistical inferences is a special case of this regression model (Agresti 1990, 152–153).

In a binary logistic regression model, we compute estimates for the coefficients $b_0, b_1, \ldots, b_k$ in a linear predictor such as $z_i = b_0 + b_1 x_{1i} + \cdots + b_k x_{ki}$. For a sample size of n observations, the observed data are indexed by $i = 1, \ldots, n$. The variables $x_{1i}, \ldots, x_{ki}$ measure the factors we believe may be associated with a higher or lower probability that $y_i = 1$ for each observation. The linear predictor connects to the probability that $y_i = 1$ through the logistic function:

$$\text{prob}(y_i = 1) = \text{logistic}(z_i) = 1/(1 + \exp(-z_i))\,.$$

This function satisfies $\text{logistic}(0) = .5$. As z_i increases, $\text{logistic}(z_i)$ asymptotically approaches the value 1. As z_i becomes increasingly negative, $\text{logistic}(z_i)$ asymptotically approaches 0. We use the statistical programming environment **R** (R Development Core Team 2005) and specifically the function `glm` to compute estimates for the coefficients, along with estimates of the statistical uncertainty in the estimates, derived from the binomial loglikelihood function. This function is

$$L(y) = \sum_{i=1}^{n} (y_i \log \text{prob}(y_i = 1) + (1 - y_i) \log(1 - \text{prob}(y_i = 1)))\,.$$

As the sample size n gets large, the parameter estimates are asymptotically equivalent to the values that maximize this function given the observed values of y and x. We compute standard errors for the estimates and conduct tests for hypotheses regarding the coefficient values using standard methods.

In the model specifications we use, the regressor variables $x_{1i}, \ldots, x_{ki}$ are *dummy variables* that take the values one or zero to represent the presence or absence of particular attributes for each ballot. There are 32 dummy variables to represent the 32 possible patterns for the statewide office votes. We include 31 of these variables in the model, omitting the variable that indicates that a ballot has five Republican votes for the statewide offices. With this specification each of the coefficients for the 31 dummy variables represents the difference in CD-13 undervoting between the referent voting pattern and the straight-Republican voting pattern. The straight-Republican statewide office voting pattern is the reference category to which all the other voting patterns are being compared. The difference between any two other categories is measured by the difference between their respective coefficients. We also include a dummy variable to indicate ballots that were cast on machines that have at least one occurrence of event 18. We use two dummy variables to represent the PEB cluster categories, one for the singleton clusters and one for the multiple clusters. This means that for these two dummy variables the PEB 0 clusters comprise the reference category.

We also consider interactions between the statewide office voting pattern variables and, respectively, the event 18 and PEB cluster dummy variables. Such interactions directly address the question of whether the relationship between the statewide office voting patterns and the CD-13 undervote rate vary depending on the observable condition of the voting machine being used.

Estimates for a basic regression model that includes the 31 statewide office voting pattern dummy variables, the event 18 dummy variable and the two PEB cluster variables show that the effects on the CD-13 undervote rate that we find when we examine the voter and the machine characteristics separately remain when we consider the two classes of factors together. Table 18 shows the estimates for the election day data, respectively using the exact match and the network PEB cluster definitions. Table 19 shows the estimates for the early voting data. The coefficient labeled "Constant" in the table represents the parameter denoted b_0 in the discussion above. The estimate for this parameter is significantly more negative in the election day data than in the early voting data, a difference that reflects the generally higher rate of CD-13 undervoting during early voting than on election day. We can use the logistic function to express this baseline difference on a probability scale. Using the exact match cluster estimates, and assuming a ballot that has straight-Republican votes for all five statewide offices, does not have an occurrence of event 18 and is cast on a PEB 0 machine, the estimated probability of having a CD-13 undervote in the election day data is $.084 = \text{logistic}(-2.39)$ while in the early voting data the probability is $.103 = \text{logistic}(-2.16)$. Using the network cluster estimates and assuming the same conditions, the estimated probability is $.082 = \text{logistic}(-2.42)$ in the election day data and $.126 = \text{logistic}(-1.94)$ in the early voting data.

Many of the coefficients for the statewide office voting pattern dummy variables are significantly positive, which means that ballots that have the referent statewide office voting pattern are significantly more likely to have a CD-13 undervote than are ballots that have the straight-Republican statewide office vote pattern. The statewide office voting pattern dummy variables are labeled using the pattern numbers listed in Table 4 and a sequence of the letters "D"

Table 18: Sarasota 2006, Binary Logistic Regression Models for the CD-13 Undervote, Election Day

Variable	Exact Match Clusters Coef.	SE	prob.	Network Clusters Coef.	SE	prob.
Constant	−2.39	.05	.000	−2.42	.03	.000
1 DDDDD	.56	.03	.000	.56	.03	.000
2 RDDDD	.93	.15	.000	.93	.15	.000
3 DRDDD	.54	.07	.000	.54	.07	.000
4 DDRDD	.36	.12	.002	.36	.12	.002
5 DDDRD	.77	.15	.000	.77	.15	.000
6 DDDDR	.08	.06	.172	.08	.06	.174
7 RRDDD	.62	.19	.001	.62	.19	.001
8 RDRDD	1.02	.35	.004	1.02	.35	.004
9 RDDRD	.79	.49	.104	.80	.49	.102
10 RDDDR	.63	.25	.011	.63	.25	.012
11 DRRDD	.35	.14	.012	.36	.14	.012
12 DRDRD	.60	.19	.002	.60	.19	.002
13 DRDDR	.18	.10	.082	.18	.10	.084
14 DDRRD	.82	.23	.000	.82	.23	.000
15 DDRDR	.26	.12	.035	.26	.12	.035
16 DDDRR	.55	.16	.001	.55	.16	.001
17 RRRDD	.44	.25	.081	.44	.25	.080
18 RRDRD	.46	.34	.171	.46	.34	.171
19 RRDDR	.26	.20	.199	.26	.20	.201
20 RDRRD	.41	.53	.439	.41	.53	.446
21 RDRDR	.31	.28	.276	.31	.28	.277
22 RDDRR	.76	.34	.028	.76	.34	.027
23 DRRRD	.39	.18	.032	.39	.18	.033
24 DRRDR	.28	.08	.001	.28	.08	.001
25 DRDRR	.47	.13	.000	.47	.13	.000
26 DDRRR	.63	.12	.000	.63	.12	.000
27 RRRRD	.02	.22	.920	.02	.22	.915
28 RRRDR	−.06	.11	.608	−.06	.11	.611
29 RRDRR	−.03	.18	.886	−.02	.18	.891
30 RDRRR	.64	.15	.000	.64	.15	.000
31 DRRRR	.43	.05	.000	.43	.05	.000
Event 18	.09	.03	.011	.08	.03	.014
PEB singleton	−.25	.10	.019	−.27	.21	.195
PEB multiple	−.08	.05	.113	−.05	.03	.125

Notes: $n = 77{,}108$. Loglikelihood: -25640.82 (exact match); -25642.00 (network). In each statewide voting dummy variable name, offices are ordered U.S. Senator, Governor, Attorney General (AG), Chief Financial Officer (CFO) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

Table 19: Sarasota 2006, Binary Logistic Regression Models for the CD-13 Undervote, Early Voting

	Exact Match Clusters			Network Clusters		
Variable	Coef.	SE	prob.	Coef.	SE	prob.
Constant	−2.16	.08	.000	−1.94	.05	.000
1 DDDDD	.70	.04	.000	.70	.04	.000
2 RDDDD	.81	.24	.001	.80	.24	.001
3 DRDDD	.67	.10	.000	.66	.10	.000
4 DDRDD	.60	.19	.001	.60	.19	.001
5 DDDRD	.53	.22	.016	.52	.22	.020
6 DDDDR	.08	.09	.355	.09	.09	.332
7 RRDDD	.19	.36	.599	.18	.36	.612
8 RDRDD	.58	.78	.456	.57	.78	.459
9 RDDRD	−.12	1.05	.913	−.13	1.05	.900
10 RDDDR	.32	.54	.551	.33	.54	.538
11 DRRDD	.16	.25	.512	.15	.25	.546
12 DRDRD	.57	.29	.044	.59	.29	.040
13 DRDDR	.28	.16	.074	.27	.16	.081
14 DDRRD	.64	.39	.099	.64	.39	.102
15 DDRDR	.17	.21	.418	.16	.21	.450
16 DDDRR	.90	.21	.000	.89	.21	.000
17 RRRDD	.96	.38	.012	.93	.38	.015
18 RRDRD	.35	.77	.650	.39	.77	.607
19 RRDDR	.01	.35	.978	.02	.35	.946
20 RDRRD	.13	.75	.861	.13	.75	.866
21 RDRDR	.65	.39	.096	.64	.39	.105
22 RDDRR	−10.43	78.71	.895	−10.43	78.66	.894
23 DRRRD	.23	.32	.483	.21	.32	.508
24 DRRDR	.03	.16	.857	.03	.16	.850
25 DRDRR	.27	.20	.169	.27	.20	.182
26 DDRRR	.92	.18	.000	.92	.19	.000
27 RRRRD	−.56	.52	.276	−.55	.52	.286
28 RRRDR	.16	.18	.360	.16	.18	.377
29 RRDRR	.36	.24	.143	.36	.24	.141
30 RDRRR	1.04	.24	.000	1.04	.24	.000
31 DRRRR	.29	.08	.001	.28	.08	.001
Event 18	−.18	.03	.000	−.17	.04	.000
PEB singleton	.46	.22	.040	—	—	—
PEB multiple	.08	.07	.292	−.18	.05	.000

Notes: $n = 27{,}518$. Loglikelihood: -11115.52 (exact match); -11111.08 (network). In each statewide voting dummy variable name, offices are ordered U.S. Senator, Governor, Attorney General (AG), Chief Financial Officer (CFO) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

and "R" that indicate which vote was cast for each of the five offices. In these abbreviated names, the offices are ordered Senator, Governor, AG, CFO and CA. Many of the statewide office voting pattern dummy variables have statistically significant coefficient estimates. Here we are judging statistical significance by observing whether the t-ratio—defined as the coefficient estimate divided by its estimated standard error—for each coefficient has an absolute value greater than 1.96. This corresponds to the conventional two-tailed test of the hypothesis that the true value of the coefficient is zero, using a test-level probability α of $\alpha = .05$. The "prob." values shown in the table correspond to the tail probabilities indicated by the observed value of the t-ratio for each coefficient. All of the estimates for the statewide office voting pattern dummy variables that have a tail probability less than .05 are positive. For each statewide office vote pattern the estimates for the exact match and the network method of defining PEB clusters are virtually identical. Moreover, for each statewide office vote pattern the election day and early voting estimates are statistically indistinguishable from one another.

The estimates in Tables 18 and 19 show clearly that the CD-13 undervote rate was significantly higher for many of the statewide office voting patterns than it was for the straight-Republican pattern, but the tables do not allow us to say whether the patterns represented in the tables differ from one another. For instance, it seems unlikely that the estimates for pattern 1 (the straight-Democrat pattern) and pattern 3 (Democrat except for a Republican vote for Governor) differ from one another, since the difference between the estimates is smaller than the standard errors of the respective estimates. But to assess formally whether two estimates differ significantly, it is necessary to take the covariance between the estimates into account.

The differences that are significant when the estimation uncertainty is correctly accounted for are reported in Table 20. As Table 20 shows, relatively few of the estimates for the coefficients for the statewide office voting patterns differ significantly from one another. Most notably, the coefficients for the straight-Democrat pattern do not differ significantly from the coefficients for any other pattern. In the election day data only 12 pairs of coefficient estimates differ from one another significantly, and those differences involve just 10 distinct coefficients. In the early voting data even fewer pairs of coefficient estimates show significant differences. That such small proportions of the statewide office voting pattern variables have distinctive effects on the CD-13 undervote rate suggests that not all of the 32 statewide office voting patterns may be necessary to capture the most substantial ways the undervote rate varies across voters. We return to this point below.

The effects estimated for event 18 in Tables 18 and 19 echo those found in Table 15. The occurrence of event 18 is associated with small but significant increases in the CD-13 undervote rate in the election day data and with somewhat larger decreases in the early voting data. We can use the logistic function to translate the positive coefficient estimates into a statement of the size of the differences on a probability scale. Using the election day exact match cluster estimates (the network cluster estimate is similar), for a ballot that has straight-Republican votes for all five statewide offices and is cast on a PEB 0 machine, the estimated probability of having a CD-13 undervote in the absence of event 18 is $.084 = \text{logistic}(-2.39)$, while on a machine that has an occurrence of event 18 the probability is $.091 = \text{logistic}(-2.39 + .09)$. Using the early voting exact match cluster estimates, the probability of having a CD-13 undervote in the absence of event 18 is $.103 = \text{logistic}(-2.16)$, while on a machine that has an occurrence of event 18 the probability is $.088 = \text{logistic}(-2.16 - .18)$. Using the early voting network cluster estimates, the probability of having a CD-13 undervote in the absence of event 18 is $.126 = \text{logistic}(-1.94)$,

Table 20: Sarasota 2006, CD-13 Undervotes: Statewide Office Voting Pattern Coefficient Estimates that Differ Significantly from One Another

Election Day Exact Match Clusters Pairs that Differ		Early Voting Network Clusters Pairs that Differ	
2 RDDDD	27 RRRRD	7 RRDDD	27 RRRRD
2 RDDDD	29 RRDRR	9 RDDRD	17 RRRDD
7 RRDDD	27 RRRRD	17 RRRDD	27 RRRRD
7 RRDDD	29 RRDRR	19 RRDDR	30 RDRRR
8 RDRDD	18 RRDRD	27 RRRRD	30 RDRRR
8 RDRDD	27 RRRRD		
8 RDRDD	28 RRRDR		
8 RDRDD	29 RRDRR		
9 RDDRD	28 RRRDR		
9 RDDRD	29 RRDRR		
19 RRDDR	30 RDRRR		
27 RRRRD	30 RDRRR		

Notes: Labels for statewide office voting pattern variables that have coefficient estimates that differ significantly from one another in the binary logistic regression models for the CD-13 undervote that are reported in Tables 18 and 19.

while on a machine that has an occurrence of event 18 the probability is $.108 = \text{logistic}(-1.94 - .17)$.

The PEB cluster effects are similar to those found in Table 17. In the election day data, the CD-13 undervote rate is smaller on a machine that is in a cluster by itself than on a PEB 0 machine. The coefficient estimate has close to the same value using both the exact match and the network clusters, but the estimate is statistically significant only in the former case. Using the exact match cluster estimates, the estimated probability of having a CD-13 undervote on a ballot with straight-Republican statewide office votes and with no event 18 is $.067 = \text{logistic}(-2.39 - .25)$ on a singleton-cluster machine. Neither the difference between multiple-cluster machines and PEB 0 machines nor the difference between multiple-cluster machines and singleton-cluster machines is significant. In the early voting data, the CD-13 undervote rate is considerably higher on a machine that is in a cluster by itself than on a PEB 0 machine. In the early voting data, singleton-cluster machines exist only using the exact match clusters. Using the exact match cluster estimates, the estimated probability of having a CD-13 undervote on a ballot with straight-Republican statewide office votes and with no event 18 is $.154 = \text{logistic}(-2.16 + .46)$ on a singleton-cluster machine. In the early voting data using the exact match clusters, neither the difference between multiple-cluster machines and PEB 0 machines nor the difference between multiple-cluster machines and singleton-cluster machines is significant. Using the network clusters the CD-13 undervote rate is significantly lower for multiple-cluster machines than for PEB 0 machines. In this case the estimated probability of having a CD-13 undervote on a ballot with straight-Republican statewide office votes and with no event 18 is $.107 = \text{logistic}(-1.94 - .18)$ on a multiple-cluster machine.

To check whether the relationship between the CD-13 undervote rate and the statewide office voting patterns varies depending on event 18 or on the PEB clusters, we add a set of interaction terms to the basic models. These interaction terms are defined by multiplying the statewide office voting pattern dummy variables by, respectively, the event 18 dummy variable or each of the one or two dummy variables that identify the PEB cluster type. To test whether the added variables identify significant variations in the CD-13 undervote rate, we perform likelihood ratio tests that compare the models that include the interaction terms to the basic models. In such tests we subtract the value of the loglikelihood $L(y)$ for the basic model from the loglikelihood for a model that includes the interaction terms, using the parameters estimated for each model to evaluate the respective loglikelihoods. If the coefficients for the interaction terms are all truly zero, then that difference, multiplied by -2, should be distributed as a chi-squared random variable with degrees of freedom equal to the number of estimated interaction terms.

As the test statistics reported in Table 21 show, interaction terms that involve the event 18 dummy variable in the election day data have significant effects on the CD-13 undervote rate. The event 18 interaction terms are not significant in the early voting data. None of the PEB cluster interactions have significant effects on the CD-13 undervote rate. Note that the degrees of freedom associated with the PEB cluster interactions vary because some of the statewide office voting patterns do not occur for some of the types of PEB clusters.

Looking at the coefficient estimates for the election day models that include the event 18 interaction effects shows that only a few of the individual interaction terms are significant. Table 22 shows the coefficient estimates for the exact match clusters (the interaction terms using the network clusters are essentially identical). The interaction terms that involve statewide office voting patterns 19 ("RRDDR"), 27 ("RRRRD") and 29 ("RRDRR") are statistically significant,

Table 21: Sarasota 2006, Likelihood Ratio Tests for Interactions in Binary Logistic Regression Models for the CD-13 Undervote

		Exact Match Clusters			Network Clusters		
	Interactions	Difference	DF	prob.	Difference	DF	prob.
Election day	Event 18	51.5	31	.012	51.7	31	.011
	PEB clusters	62.7	58	.314	40.5	49	.801
Early voting	Event 18	23.0	31	.848	22.7	31	.859
	PEB clusters	38.0	44	.727	38.6	31	.164

Notes: Likelihood ratio tests for interaction terms added to the binary logistic regression models for the CD-13 undervote that are reported in Tables 18 and 19. Legend: Difference, -2 times the difference between the basic and interaction models' loglikelihood values; DF, degrees of freedom; prob., tail probability above the Difference value for the chi-squared distribution with DF degrees of freedom.

and all three of these estimated coefficients are positive and large. So on a PEB 0 machine on which event 18 did not occur, the estimated probabilities of a CD-13 undervote on ballots with, respectively, each of the three statewide office voting patterns are $.084 = \text{logistic}(-2.39 + 0.00)$, $.069 = \text{logistic}(-2.39 - 0.21)$ and $.063 = \text{logistic}(-2.39 - 0.31)$, while on a PEB 0 machine on which event 18 did occur the probabilities are $.201 = \text{logistic}(-2.39 + 0.00 + 1.01)$, $.190 = \text{logistic}(-2.39 - 0.21 + 1.15)$ and $.218 = \text{logistic}(-2.39 - 0.31 + 1.42)$. According to Table 4, these interactions with event 18 do not directly affect very many ballots. In the election day data, statewide office voting pattern 19 occurs on 278 ballots, pattern 27 occurs on 284 ballots and pattern 29 occurs on 439 ballots. But any sign that event 18 increases the degree to which the production of CD-13 undervotes depends on the particular choices recorded for the statewide races is worrisome.

Parsimonious Models of the CD-13 Undervote

Finding a highly significant interaction effect between the statewide office voting pattern variables and the event 18 variable that resolves to only a few of the 31 interaction terms reinforces the impression that a model that does not use all 32 of the statewide office voting patterns may capture how the CD-13 undervote rate varies across voters and machines as well as the 32-pattern model does. In particular it might be useful to discover that the major variations in the undervote depend on only a few of the statewide offices, as that might help winnow the list of mechanisms that may have helped generate the excessive CD-13 undervoting rate.

To that end, we organize a search for more parsimonious models around the idea that a subset of the statewide offices may be sufficient to capture most of the variation picked up in the large but in important ways sparse models of Tables 19 and 22. We consider the set of all possible models that can be constructed using such subsets, using the principle that for each subset we define dummy variables to represent all possible patterns of votes cast for the Democrat or Republican candidates for the offices. With five offices there are 31 nonempty subsets of offices.

Table 22: Sarasota 2006, Binary Logistic Regression Models for the CD-13 Undervote, Election Day, with Event 18 Interactions (Exact Match PEB Clusters)

Variable	Coef.	SE	prob.		Coef.	SE	prob.
Constant	-2.39	0.05	0.000	1 DDDDD × Event 18	-0.06	0.08	0.482
1 DDDDD	0.57	0.03	0.000	2 RDDDD × Event 18	-0.85	0.55	0.124
2 RDDDD	1.01	0.15	0.000	3 DRDDD × Event 18	-0.04	0.21	0.857
3 DRDDD	0.55	0.07	0.000	4 DDRDD × Event 18	0.20	0.32	0.523
4 DDRDD	0.33	0.13	0.009	5 DDDRD × Event 18	-0.14	0.45	0.746
5 DDDRD	0.79	0.16	0.000	6 DDDDR × Event 18	0.00	0.17	0.981
6 DDDDR	0.08	0.06	0.210	7 RRDDD × Event 18	0.13	0.47	0.779
7 RRDDD	0.59	0.21	0.004	8 RDRDD × Event 18	1.08	0.99	0.277
8 RDRDD	0.88	0.39	0.024	9 RDDRD × Event 18	-11.30	132.57	0.932
9 RDDRD	1.08	0.50	0.031	10 RDDDR × Event 18	0.72	0.72	0.321
10 RDDDR	0.55	0.27	0.041	11 DRRDD × Event 18	-0.04	0.43	0.919
11 DRRDD	0.36	0.15	0.017	12 DRDRD × Event 18	-10.97	60.30	0.856
12 DRDRD	0.77	0.19	0.000	13 DRDDR × Event 18	-0.64	0.38	0.096
13 DRDDR	0.24	0.11	0.024	14 DDRRD × Event 18	0.55	0.58	0.345
14 DDRRD	0.72	0.26	0.005	15 DDRDR × Event 18	0.05	0.35	0.887
15 DDRDR	0.25	0.13	0.057	16 DDDRR × Event 18	0.57	0.40	0.156
16 DDDRR	0.45	0.18	0.015	17 RRRDD × Event 18	0.23	0.82	0.774
17 RRRDD	0.42	0.27	0.118	18 RRDRD × Event 18	-0.46	1.11	0.679
18 RRDRD	0.52	0.36	0.146	19 RRDDR × Event 18	1.01	0.44	0.021
19 RRDDR	0.00	0.25	0.996	20 RDRRD × Event 18	1.02	1.31	0.433
20 RDRRD	0.24	0.61	0.694	21 RDRDR × Event 18	0.61	0.71	0.396
21 RDRDR	0.21	0.32	0.515	22 RDDRR × Event 18	0.25	0.88	0.773
22 RDDRR	0.71	0.38	0.063	23 DRRRD × Event 18	0.12	0.49	0.806
23 DRRRD	0.37	0.20	0.064	24 DRRDR × Event 18	0.18	0.22	0.411
24 DRRDR	0.25	0.09	0.006	25 DRDRR × Event 18	-0.05	0.39	0.908
25 DRDRR	0.48	0.14	0.001	26 DDRRR × Event 18	-0.09	0.37	0.806
26 DDRRR	0.64	0.13	0.000	27 RRRRD × Event 18	1.15	0.50	0.021
27 RRRRD	-0.21	0.26	0.419	28 RRRDR × Event 18	-0.35	0.37	0.337
28 RRRDR	-0.02	0.12	0.900	29 RRDRR × Event 18	1.42	0.41	0.001
29 RRDRR	-0.31	0.22	0.158	30 RDRRR × Event 18	0.45	0.43	0.295
30 RDRRR	0.58	0.16	0.000	31 DRRRR × Event 18	-0.16	0.15	0.279
31 DRRRR	0.45	0.05	0.000				
Event 18	0.11	0.06	0.090				
PEB singleton	-0.24	0.10	0.022				
PEB multiple	-0.08	0.05	0.112				

Notes: $n =$ 77,108. Loglikelihood: -25615.09. In each statewide voting dummy variable name, offices are ordered U.S. Senator, Governor, Attorney General (AG), Chief Financial Officer (CFO) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

For the election day data we examine all 31 of the consequent models both including and omitting the interaction terms produced by multiplying each subsets voting pattern dummy variables by the event 18 dummy variable. For the early voting data we examine only the 31 models that include no interaction terms.[10] To decide which models are the most parsimonious, we use the Bayesian information criterion (BIC) (Schwarz 1978). If a model includes m coefficients, then the BIC statistic for that model may be defined by BIC $= -2L(y) + m\log(n)$, where $L(y)$ is evaluated using the coefficient estimates. If one believed that one of the models we are considering is the model that truly describes how the CD-13 undervotes were produced in the referent body of data, then the model for which the BIC statistic is smallest would be the one that, given the data, is most probably that correct model. We do not have such strong beliefs about the descriptive accuracy of these models, so our use of the BIC is heuristic. We use it to discover whether the smaller models are indeed more parsimonious than the bigger ones (a smaller model that fits the data considerably worse is not more parsimonious but merely smaller). Also we'd like to know what the more parsimonious model specifications are.

The BIC statistics confirm that the 32-pattern model specification is far from parsimonious. In the list of all the models for the election day data, shown in Table 23, the 32-pattern models have the largest BIC values.[11] The models with the smallest BIC values include two or three offices, and two of the offices are always Senator and CA. Indeed, the model that includes only the four voting patterns among those two offices with no event 18 interactions has the smallest BIC value. If interactions involving event 18 are added to that model, the interactions are not significant. The same is true of the model that has the second smallest BIC value. That model includes the eight voting patterns among Senator, Governor and CA. The model that has the third smallest BIC value includes voting patterns for Senator, AG and CA along with significant event 18 interactions. The model with the fourth smallest BIC value is the same as the first model but including the insignificant event 18 interaction terms. Presumably the models that have larger BIC values than that model should be counted as sufficiently nonparsimonious that they are not worthwhile candidates to be the parsimonious reduction of the 32-pattern model.

For the early voting data the 32-pattern model again has the largest BIC value. As Table 24 shows, the BIC favors models that include one, two or three offices, one of which is always CA.[12] The model with the lowest BIC that does not include CA is the model that includes the voting patterns for Senator and Governor, but that model has a demonstrably worse fit to the data than does the model that includes only CA. The model that includes the voting patterns for Senator, AG and CA ranks slightly better than the Senator-Governor model according to the BIC value, but according to a likelihood ratio test that model does not significantly improve on the model that includes the voting patterns for only Senator and CA. For all the models that have BIC values smaller than the Senator-AG-CA model, likelihood ratio tests reject the hypothesis that a smaller model is acceptable.[13] Strictly on grounds of parsimony, then, we might choose to reduce the

[10]It is harmless to omit consideration of the election day models that include interactions with the PEB cluster dummy variables and all of the interaction models for the early voting data. In none of those models are the interaction terms statistically significant.

[11]Table 23 shows results for the exact match PEB clusters. The results using the network clusters are virtually identical.

[12]Table 24 shows results for the network PEB clusters. The results using the exact match clusters are virtually identical.

[13]The step from the model that includes the voting patterns for AG and CA to the model that includes the voting

Table 23: Sarasota 2006, Bayesian Information Criterion Statistics for Binary Logistic Regression Models for the CD-13 Undervote, Election Day (Exact Match PEB Clusters)

Offices	Specification	BIC	Offices	Specification	BIC
2	Sen CA	51450.52	1	Sen	51525.67
3	Sen Gov CA	51451.66	2	Sen CFO	51526.45
3	Sen AG CA i	51452.94	3	Gov AG CA	51526.96
2	Sen CA i	51454.53	4	Sen AG CFO CA	51527.54
3	Sen CFO CA	51455.57	3	Gov CFO CA i	51528.31
3	Sen Gov CA i	51456.49	2	Sen CFO i	51530.15
3	Sen AG CA	51463.03	3	Gov AG CA i	51533.24
3	Sen CFO CA i	51464.43	3	Sen AG CFO	51537.59
2	Sen Gov	51480.98	3	Sen AG CFO i	51542.71
2	Sen Gov i	51483.85	3	AG CFO CA i	51542.93
1	CA	51498.14	3	AG CFO CA	51545.62
1	CA i	51499.38	2	Gov AG	51564.82
2	Gov CA	51505.17	2	Gov AG i	51568.40
2	Sen AG	51507.52	2	Gov CFO	51572.06
4	Sen Gov CFO CA	51507.95	2	Gov CFO i	51576.50
2	Sen AG i	51508.73	1	Gov	51584.25
2	AG CA i	51509.49	1	AG	51585.28
2	Gov CA i	51509.65	1	Gov i	51586.17
2	AG CA	51510.27	1	AG i	51586.76
3	Sen Gov AG i	51511.59	4	Sen Gov AG CFO	51590.67
3	Sen Gov AG	51516.18	3	Gov AG CFO	51595.08
3	Sen Gov CFO	51516.43	4	Gov AG CFO CA	51595.17
3	Gov CFO CA	51516.68	4	Sen Gov AG CFO i	51597.96
2	CFO CA	51519.25	2	AG CFO	51598.99
4	Sen Gov AG CA i	51519.82	2	AG CFO i	51602.92
4	Sen Gov AG CA	51523.27	4	Gov AG CFO CA i	51604.60
2	CFO CA i	51524.41	3	Gov AG CFO i	51605.69
3	Sen Gov CFO i	51525.23	1	CFO	51631.83
4	Sen Gov CFO CA i	51525.41	1	CFO i	51633.42
4	Sen AG CFO CA i	51525.56	5	Sen Gov AG CFO CA	51675.49
1	Sen i	51525.66	5	Sen Gov AG CFO CA i	51686.03

Notes: BIC statistics. $n = 77{,}108$. The symbol “i” in a specification’s descriptor indicates the model include interaction terms between the statewide office vote pattern dummy variables and the dummy variable for event 18.

32-pattern model to the model that includes only the votes for CA, but if we were concerned to some extent about losing information we might choose one of the eight models that have the next smallest BIC values.

In both the election day and early voting data the model that includes the voting patterns for Senator, AG and CA is included among the models that, according to the BIC values, it is reasonable to treat as parsimonious reductions of the 32-pattern model. To enhance comparing effects between election day and early voting, we focus on this model in both data sets. For the election day data the model includes interaction terms between the statewide office voting pattern variables and the event 18 variable. Estimates for the models appear respectively in Tables 25 and 26.

The estimates resemble the estimates for the 32-pattern models in that all the significant estimates for the statewide office voting pattern dummy variables are positive, and the estimates for event 18 and for the PEB cluster variables are the same as in the 32-pattern model. In the election day data five of the seven coefficients for the pattern variables are significant, and the two that are not significant have significant positive event 18 interaction terms. In the early voting data, six of the seven pattern variables have significant positive coefficients. With three offices being included in the model, a straight-partisan pattern means that the votes for Senator, AG and CA all go to the same party. Straight-Republican ballots, in this sense, are clearly the ones that tend to have the lowest CD-13 undervote rate.

The parsimony of the models implies greater efficiency in them, in the sense that a higher proportion of the coefficients for the pattern variables differ from one another. The coefficient estimates that differ significantly from one another are reported in Table 27. In the election day data, only the coefficients for the pattern variables are considered, not the interactions terms. In contrast to the result in the 32-pattern models, the coefficients for the straight-Democrat pattern now differ significantly from the coefficients for most of the other patterns.

In the early voting data, the straight-Democrat pattern differs significantly from all the other patterns. Because, in Table 26, the coefficient for the “DDD” pattern has the most positive estimate of all the pattern variables, the result here suggests in the early voting data the straight-Democrat ballots tend to have the highest CD-13 undervote rate, voting machine effects notwithstanding. On a PEB 0 machine on which event 18 did not occur, the estimates for the model with network PEB clusters implies that the estimated probability of a CD-13 undervote is $.128 = \text{logistic}(-1.92)$ for a straight-Republican ballot, $.223 = \text{logistic}(-1.92 + .67)$ for a straight-Democrat ballot. This model estimates that on the same kind of voting machines ballots with the “RDD” pattern (Republican vote for Senator, Democrat votes for AG and CA) have an estimated CD-13 undervote probability of $.196 = \text{logistic}(-1.92 + .51)$.

In the election day data, the straight-Democrat pattern differs significantly from all but one of the other patterns. On machines that do not have an occurrence of event 18, there is one pattern (“RDD”) that tends have a higher CD-13 undervote rate than the straight-Democrat pattern does. On a PEB 0 machine on which event 18 did not occur, the estimates for the model with exact match PEB clusters implies that the estimated probability of a CD-13 undervote is $.085 = \text{logistic}(-2.38)$ for a straight-Republican ballot, $.139 = \text{logistic}(-2.38 + .56)$ for a

pattern for only CA is in some respects not compellingly significant. The significant probability for that test is .032, which is less than .05. But in this set we are looking at 12 different likelihood ratio tests. If we control the false discovery rate (Benjamini and Hochberg 1995) for 12 tests, then the threshold for significance falls to $0.0042 = .05/12$ and the result from going from AG and CA to CA alone is no longer significant.

Table 24: Sarasota 2006, Bayesian Information Criterion Statistics for Binary Logistic Regression Models for the CD-13 Undervote, Early Voting (Network PEB Clusters)

Offices	Specification	BIC
1	CA	22352.68
2	Sen CA	22354.71
2	Gov CA	22356.01
2	AG CA	22362.08
3	Gov CFO CA	22362.52
2	CFO CA	22366.25
3	Gov AG CA	22374.05
3	Sen CFO CA	22376.03
3	Sen Gov CA	22378.50
3	Sen AG CA	22387.86
2	Sen Gov	22391.29
2	Gov AG	22392.35
3	AG CFO CA	22394.28
1	Gov	22401.30
2	Gov CFO	22401.98
2	Sen AG	22408.01
1	AG	22411.06
3	Sen Gov AG	22420.71
1	Sen	22422.59
2	Sen CFO	22424.08
3	Gov AG CFO	22425.15
2	AG CFO	22425.50
3	Sen Gov CFO	22427.09
4	Sen Gov CFO CA	22427.16
4	Gov AG CFO CA	22435.02
3	Sen AG CFO	22441.85
1	CFO	22441.99
4	Sen Gov AG CA	22442.28
4	Sen AG CFO CA	22450.67
4	Sen Gov AG CFO	22485.04
5	Sen Gov AG CFO CA	22569.73

Notes: BIC statistics. $n = $ 27,518.

Table 25: Sarasota 2006, Parsimonious Binary Logistic Regression Models for the CD-13 Undervote, Election Day, with Event 18 Interactions

	Exact Match Clusters			Network Clusters		
Variable	Coef.	SE	prob.	Coef.	SE	prob.
Constant	−2.38	.05	.000	−2.41	.03	.000
1 DDD	.56	.03	.000	.56	.03	.000
2 RDD	.82	.11	.000	.82	.11	.000
3 DRD	.37	.08	.000	.37	.08	.000
4 DDR	.17	.05	.002	.17	.05	.002
5 RRD	.17	.16	.296	.17	.16	.294
6 RDR	.03	.13	.807	.03	.13	.809
7 DRR	.40	.04	.000	.40	.04	.000
1 DDD × Event 18	−.07	.08	.399	−.07	.08	.401
2 RDD × Event 18	−.46	.33	.169	−.45	.33	.173
3 DRD × Event 18	.17	.22	.419	.18	.22	.417
4 DDR × Event 18	−.04	.14	.784	−.04	.14	.783
5 RRD × Event 18	.84	.37	.022	.85	.37	.021
6 RDR × Event 18	1.04	.26	.000	1.04	.26	.000
7 DRR × Event 18	−.06	.12	.591	−.07	.12	.583
Event 18	.10	.06	.085	.10	.06	.094
PEB singular	−.24	.10	.020	−.27	.21	.201
PEB multiple	−.08	.05	.109	−.05	.03	.113

Notes: $n = 77{,}108$. Loglikelihood: -25657.5 (exact match); -25658.5 (network). In each statewide voting dummy variable name, offices are ordered U.S. Senator, Attorney General (AG) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

Table 26: Sarasota 2006, Parsimonious Binary Logistic Regression Models for the CD-13 Undervote, Early Voting

	Exact Match Clusters			Network Clusters		
Variable	Coef.	SE	prob.	Coef.	SE	prob.
Constant	-2.14	.08	.000	-1.92	.05	.000
1 DDD	.69	.04	.000	.67	.04	.000
2 RDD	.66	.22	.002	.51	.19	.007
3 DRD	.55	.09	.000	.38	.13	.003
4 DDR	.24	.07	.001	.18	.07	.014
5 RRD	.20	.22	.356	.21	.26	.415
6 RDR	.71	.19	.000	.15	.19	.416
7 DRR	.23	.07	.001	.26	.07	.000
Event 18	$-.19$	.03	.000	$-.17$	.04	.000
PEB singular	.46	.22	.038	—	—	—
PEB multiple	.07	.07	.304	$-.18$	.05	.000

Notes: $n = 27{,}518$. Loglikelihood: -11143 (exact match); -11143 (network). In each statewide voting dummy variable name, offices are ordered U.S. Senator, Attorney General (AG) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

straight-Democrat ballot. On the same kind of voting machines this model estimates that ballots with the "RDD" pattern have an estimated CD-13 undervote probability of $.174 = \text{logistic}(-2.38 + .82)$. On machines that do have event 18, there are two patterns that have a higher CD-13 undervote rate than the straight-Democrat pattern does. Ballots that have the "RRD" pattern have an estimated CD-13 undervote probability of $.192 = \text{logistic}(-2.38 + .10 + .84)$, and ballots with the "RDR" pattern have an estimated CD-13 undervote probability of $.224 = \text{logistic}(-2.38 + .10 + 1.04)$.

Table 28 shows the table of frequencies and CD-13 undervote rates for the patterns using the votes for Senator, AG and CA, separating ballots on machines that have an occurrence of event 18 from ballots on machines that do not. The CD-13 undervote rates differ slightly from the estimated probability values computed using the estimates for the models of Tables 25 and 26 because Table 28 ignores the differences associated with the PEB cluster variables and because in Table 26 there are no interaction terms involving the event 18 dummy variable.

In Table 28, in the election day data, the interaction effects involving the "RRD" and "RDR" patterns are apparent: the CD-13 undervote rate is more than twice as large on machines with an occurrence of event 18 than on the other machines. What is very interesting is that in the early voting data, patterns "RRD" and "RDR" are the only ones for which the CD-13 undervote rate is higher on machines that have event 18 than for machines that do not. This is a hint that perhaps in the early voting data the same kind of interaction effect involving event 18 happens as happens in the election day data, but in the early voting data there are not enough ballots with the referent patterns to make the interaction statistically significant.

Table 27: Sarasota 2006, CD-13 Undervotes: Statewide Office Voting Pattern Coefficient Estimates that Differ Significantly from One Another in the Reduced Models

Election Day Exact Match Clusters Pairs that Differ		Early Voting Network Clusters Pairs that Differ	
1 DDD	2 RDD	1 DDD	2 RDD
1 DDD	3 DRD	1 DDD	3 DRD
1 DDD	4 DDR	1 DDD	4 DDR
1 DDD	5 RRD	1 DDD	5 RRD
1 DDD	6 RDR	1 DDD	6 RDR
2 RDD	3 DRD	1 DDD	7 DRR
2 RDD	4 DDR		
2 RDD	5 RRD		
2 RDD	6 RDR		
2 RDD	7 DRR		
3 DRD	6 RDR		
6 RDR	7 DRR		

Notes: Labels for statewide office voting pattern variables that have coefficient estimates that differ significantly from one another in the binary logistic regression models for the CD-13 undervote that are reported in Tables 25 and 26. In each statewide voting pattern dummy variable name, offices are ordered U.S. Senator, Attorney General (AG) and Commissioner of Agriculture (CA).

Table 28: Sarasota 2006, Reduced Partisan Voting Patterns and CD-13 Undervotes

				Election Day			
				Event 18		No Event 18	
		Office		Undervote	Ballot	Undervote	Ballot
Pattern	Sen	AG	CA	Rate	Count	Rate	Count
1	D	D	D	.134	3,601	.131	24,326
2	R	D	D	.120	100	.163	590
3	D	R	D	.141	227	.111	1,536
4	D	D	R	.097	802	.092	5,387
5	R	R	D	.207	58	.092	466
6	R	D	R	.217	120	.081	798
7	D	R	R	.117	1,059	.114	7,026
8	R	R	R	.087	4,004	.079	27,008

				Early Voting			
				Event 18		No Event 18	
		Office		Undervote	Ballot	Undervote	Ballot
Pattern	Sen	AG	CA	Rate	Count	Rate	Count
1	D	D	D	.169	6,589	.204	5,161
2	R	D	D	.120	92	.204	113
3	D	R	D	.135	259	.158	241
4	D	D	R	.111	1,335	.133	1,007
5	R	R	D	.129	62	.125	72
6	R	D	R	.124	137	.114	140
7	D	R	R	.125	1,204	.139	1,194
8	R	R	R	.099	5,219	.110	4,693

Office abbreviations: U.S. Senator (Sen), Attorney General (AG) and Commisioner of Agriculture (CA). Rates are the proportion of the ballots in each category that have a CD-13 undervote.

We can use the CD-13 undervote rates shown in Table 28 to construct what-if estimates for the number of undervotes that would have occurred under various counterfactual conditions for the operation of the voting machines. To do this we imagine that, in each instance, all the ballots for each voting pattern had been cast with a CD-13 undervote rate corresponding to the rate observed in either the presence or absence of event 18. For the election day data, this counterfactual exercise gives 8,018 CD-13 undervotes with event 18 completely absent and 8,626 undervotes if the rates with event 18 present are applied. The actual total is 8,095 CD-13 undervotes, so not having any occurrence of event 18 might have meant 77 fewer undervotes on election day. For the early voting data, the counterfactual total is 4,292 CD-13 undervotes if the rates in the absence of event 18 are applied and 3,670 undervotes with event 18 present. Since the actual total is 3,947 CD-13 undervotes, having the conditions that went with event 18 occur for all voters might have meant a reduction of 277 undervotes during early voting.

Minority Precincts

Early news coverage focused on the CD-13 undervote rate in "minority areas": "In the 32 precincts whose percentage of black and Hispanic voters was greater than the overall county average of 4.6 percent, the undervote rate averaged 15.3 percent. In the remaining 127 precincts, it was 12.4 percent" (Marsteller 2006).

Taking into account the proportion of registered voters in each precinct who are black or Hispanic does not materially change the relationships we have observed among the CD-13 undervote, statewide office voting patterns and voting machine characteristics. We present two sets of results to illustrate this point.[14] In Table 29 we add to the parsimonious model for the election day data variables to measure the proportion of registered voters in each precinct who are black and the proportion who are Hispanic. Both variables have significant coefficients. The CD-13 undervote rate is higher in precincts in which a higher proportion of registered voters are black and lower in precincts in which a higher proportion of registered voters are Hispanic. The coefficients estimated for the other variables are not significantly different from the estimates in the model that omits the precinct minority registration variables, however. In Table 30 we estimate the parsimonious model of Table 25 while omitting the ballots for the 32 precincts emphasized in (Marsteller 2006). Omitting these precincts does not significantly change the coefficient estimates.[15]

[14]Registration data are from "Sarasota County Supervisor of Elections Registration Tally, Sorted by Precincts, Reg Date: 01-01-1900 to 10-10-2006" (October 10, 2006, 08:59).

[15]Including the two registration proportion variables similarly does not change the coefficient estimates in the 32-pattern models.

Table 29: Sarasota 2006, Parsimonious Binary Logistic Regression Models for the CD-13 Undervote, Election Day, with Event 18 Interactions and Precinct Minority Composition

	Exact Match Clusters			Network Clusters		
Variable	Coef.	SE	prob.	Coef.	SE	prob.
Constant	−2.31	.05	.000	−2.34	.04	.000
1 DDD	.56	.03	.000	.56	.03	.000
2 RDD	.82	.11	.000	.81	.11	.000
3 DRD	.38	.08	.000	.38	.08	.000
4 DDR	.16	.05	.002	.16	.05	.002
5 RRD	.17	.16	.284	.17	.16	.284
6 RDR	.03	.13	.799	.03	.13	.802
7 DRR	.40	.04	.000	.40	.04	.000
1 DDD × Event 18	−.06	.08	.421	−.06	.08	.423
2 RDD × Event 18	−.44	.33	.187	−.44	.33	.191
3 DRD × Event 18	.17	.22	.433	.17	.22	.429
4 DDR × Event 18	−.04	.14	.794	−.04	.14	.792
5 RRD × Event 18	.85	.37	.021	.86	.37	.020
6 RDR × Event 18	1.04	.26	.000	1.05	.26	.000
7 DRR × Event 18	−.06	.12	.594	−.07	.12	.585
Event 18	.10	.06	.097	.10	.06	.106
PEB singular	−.25	.10	.017	−.35	.21	.099
PEB multiple	−.06	.05	.186	−.04	.03	.187
Black proportion	.44	.15	.003	.46	.15	.002
Hispanic proportion	−6.35	.93	.000	−6.38	.93	.000

Notes: $n = 77{,}108$. Loglikelihood: -25632.5 (exact match); -25633.5 (network). In each statewide voting dummy variable name, offices are ordered U.S. Senator, Attorney General (AG) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

Table 30: Sarasota 2006, Parsimonious Binary Logistic Regression Models for the CD-13 Undervote, Election Day, with Event 18 Interactions, Excluding Minority Precincts

	Exact Match Clusters			Network Clusters		
Variable	Coef.	SE	prob.	Coef.	SE	prob.
Constant	−2.38	.05	.000	−2.42	.04	.000
1 DDD	.54	.03	.000	.54	.03	.000
2 RDD	.85	.13	.000	.85	.13	.000
3 DRD	.31	.10	.002	.31	.10	.002
4 DDR	.20	.06	.000	.20	.06	.000
5 RRD	.12	.18	.500	.12	.18	.499
6 RDR	.05	.14	.729	.05	.14	.731
7 DRR	.38	.05	.000	.38	.05	.000
1 DDD × Event 18	−.07	.09	.427	−.07	.09	.436
2 RDD × Event 18	−.55	.40	.170	−.55	.40	.172
3 DRD × Event 18	.29	.24	.228	.29	.24	.226
4 DDR × Event 18	−.17	.16	.281	−.17	.16	.284
5 RRD × Event 18	1.01	.39	.010	1.02	.39	.010
6 RDR × Event 18	.88	.31	.004	.88	.30	.004
7 DRR × Event 18	−.03	.13	.840	−.03	.13	.835
Event 18	.08	.07	.262	.07	.07	.280
PEB singular	−.25	.12	.031	−.40	.24	.088
PEB multiple	−.09	.05	.088	−.06	.04	.111

Notes: $n = 63{,}616$. Loglikelihood: -20743.5 (exact match); -20743.5 (network). In each statewide voting dummy variable name, offices are ordered U.S. Senator, Attorney General (AG) and Commissioner of Agriculture (CA). Only ballots that have a vote for a Democrat or a Republican for all five of the statewide offices are included.

References

Agresti, Alan. 1990. *Categorical Data Analysis*. New York: Wiley.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57 (1): 289–300.

Coffey, Kendall and Mark Herron. 2006. "In the Circuit Court for the Second Judicial Circuit in and for Leon County, Florida Civil Division: Christine Jennings, nominee of the Democratic Party for Representative in Congress from the State of Florida's Thirteenth Congressional District, Plaintiff, v. Elections Canvassing Commission of the State of Florida, consisting of Governor Jeb Bush, Chief Financial Officer Tom Gallagher, and State Senator Daniel Webster; Sarasota County Canvassing Board, consisting of Supervisor of Elections Kathy Dent, Judge Phyllis Galen, and Commissioner Paul Mercier; Kathy Dent, as Sarasota County Supervisor of Elections; Sue M. Cobb, as Secretary of State of the State of Florida; Dawn K. Roberts, as Director of the Division of Elections of the State of Florida; and Vern Buchanan, as nominee of the Republican Party for Representative in Congress from the State of Florida's Thirteenth Congressional District, Defendants. Complaint to Contest Election." November 20, 2006.

Coffey, Kendall, Mark Herron, Donald B. Verrilli, Jr., Sam Hirsch, Brian Hauck, and Jessica Ring Amunson. 2006. "In the United States House of Representatives: Christine Jennings, Contestant, v. Vern Buchanan, Contestee. Notice of Contest Regarding the Election for Representative in the One Hundred Tenth Congress From Florida's Thirteenth Congressional District." December 20, 2006.

Frisina, Laurin, Michael C. Herron, James Honaker, and Jeffrey B. Lewis. 2006. "Ballot Formats, Touchscreens, and Undervotes: A Study of the 2006 Midterm Elections in Florida." Draft of December 1, 2006. URL http://www.dartmouth.edu/˜herron/cd13.pdf.

Gluck, Frank, Heather Allen, and Mike Saewitz. 2006. "Most callers report voting problems." *HeraldTribune.com*. November 19.

Gulliver, David, Heather Allen, and Frank Gluck. 2006. "Sarasota voters report touch-screen woes." *HeraldTribune.com*. November 9.

Herron, Michael C. and Jasjeet S. Sekhon. 2003. "Overvoting and Representation: An Examination of Overvoted Presidential Ballots in Broward and Miami-Dade Counties." *Electoral Studies* 22: 21–47.

Mahlburg, Bob and Maurice Tamman. 2006. "Dist. 13 voting analysis shows broad problem." *HeraldTribune.com*. November 9.

Marsteller, Duane. 2006. "Dist. 13 recount: state steps in." *HeraldTribune.com*. November 10.

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

R Development Core Team. 2005. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. URL `http://www.R-project.org`

Ruger, Todd. 2006. "Voting glitch prompts warning." *HeraldTribune.com*. November 9.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6 (2): 461–464.

Stewart, III, Charles. 2006. "Declaration of Charles Stewart III on Excess Undervotes Cast in Sarasota County, Florida for the 13th Congressional District Race." November 20, 2006.

Stratton, Jim. 2006. "Testers re-enact Election Day—results differ." *Orlando Sentinel*. November 29.

Tamman, Maurice and Matthew Doig. 2006. "Analysis suggests undervote caused by ballot design." *HeraldTribune.com*. November 15.

Wallace, Jeremy. 2006a. "Buchanan wins District 13 recount; legal action looms." *HeraldTribune.com*. November 18.

Wallace, Jeremy. 2006b. "Political operatives gather for recount." *HeraldTribune.com*. November 11.