Residual Voting in Florida[1]

Paul Gronke
Reed College and
Early Voting Information Center

Charles Stewart III
Massachusetts Institute of Technology

James Hicks
Early Voting Information Center

[1] October 2010. This research was funded by The Pew Charitable Trusts. The views expressed are those of the authors and do not necessarily reflect the views of The Trusts. The authors would like to thank the Director of Elections of the State of Florida for providing assistance and support for this project. We also thank Reed College for providing infrastructure support for the Early Voting Information Center. Eva Galanes-Rosenbaum (Reed), Tony Hill (MIT), and B.K. Song (MIT) also worked on portions of this project.

Executive Summary

The residual vote rate is a metric that can be used to evaluate the comparative performance of election systems, particularly voting technology and ballot design. If collected at the appropriate levels of disaggregation (county- or more preferably precinct-level) and by different modes of balloting (in person, absentee, and early voting), the residual vote rate can identify ways that voter demographics, voting technology, and the time and place of casting the ballot may influence varying levels of voting errors.

This report describes a project that examined the residual vote rates in the state of Florida's 2008 presidential preference primary, taking advantage of a state law that, for a time, required all jurisdictions to report over- and under-votes at the precinct level.

The report contains these sections:

A description of and rationale for the residual vote measure;

A step by step description of how the information needed to create and analyze residual vote rates are collected;

Analyses of residual vote rates by mode and by race/ethnicity;

Appendices containing data code books and programming files.

The major hurdle the project faced was the ever-changing legal climate in Florida. The state had expected to collect over- and under-vote rates at the precinct level starting in the 2006 general election, yet the law requiring these data was not fully implemented until the 2008 presidential primary. Soon after that, the state changed its laws again, and no longer required jurisdictions to report this information. This severely handicapped our ability to examine improvements (or erosions) in system performance statewide. The one election for which we had good data, the 2008 presidential primary, was unusual after substantial controversy about the timing of the Democratic election led national party leaders to declare that Florida's delegates would not be seated at the summer convention. The uncertainty about effects of this announcement hampered our ability to make generalizations.

As a result, the project refocused toward providing a clear road map for interested parties who may want to collect and analyze residual vote rates in the future, as well as building an argument as to why providing detailed precinct-level voting returns is important and helpful for election administrators, third party observers, as well as scholars and election analysts.

Our **findings** are these:

1. Precinct-level data significantly enhance the ability for elections officials and researchers to evaluate system performance and identify potential trouble spots in the elections ecosystem.
2. Though comprehensive, low-level elections data are generally *available*, local jurisdictions do not always make them *accessible*. Finding the right person to speak to –– not always the Supervisor or Director –– is typically crucial to obtaining the desired data.
3. The wide range of data formats significantly hampers research at this level.
4. In keeping with past research, over- vote rates are highest where ballots are counted at a central location (i.e., absentee voting).
5. In parts of Florida, a high concentration of African-American, Hispanic, and in some cases senior voters, is correlated with higher residual vote rates.

Our **recommendations** are these:

1. States, localities, and system vendors should work together to establish and adopt consistent, standardized data formats for elections data. Such standards will substantially enhance the ability of states and localities to compare and improve the performance of elections management.
2. State legislatures and state elections offices should work to create coherent and consistent legal and administrative requirements for election results reporting.

3. Data output standards should be required as part of a state's contractual process with voting systems vendors, thus reducing or eliminating the cost and technical burden of reporting mandates on local jurisdictions.
4. Monitoring systems should be put in place to ensure that local jurisdictions are following state mandates about recording and reporting data.

Contents

Introduction

An important goal of the intensive efforts over the past decade to improve election administration has been to make every vote count. This, in turn, has led reformers of all stripes to search for ways to measure how many votes have gone uncounted, owing to shortcoming in how elections are run in the United States. This document is a report of an effort to extend the use of a major metric that can diagnose problems with voting equipment, the **residual vote rate**, to a finer degree of granularity than ever attempted before.

The residual vote rate is based on observing the number of ballots that fail to contain a legitimate vote In the privacy of the voting booth, even the most informed of voters often fail to vote in every race. Americans have adapted to the information demands imposed by the longest ballots in the world by adopting a number of strategies, one of which is simple abstention. It is easy to argue that if a voter has no information about a particular race, it is understandable, rational, and even moral to abstain.

Prior to the 2000 presidential election, few doubted that that when a voter failed to cast a vote in a particular race, it was simply an act of free choice. The election of 2000 taught us otherwise. Problems with pregnant and hanging chads demonstrated that sometimes a "blank" ballot is caused by the failure of a voting technology to properly record a vote that was cast intentionally.[2] Conversely, the "butterfly ballot" demonstrated that poor ballot designs could confuse voters, misleading them to over-vote unintentionally.[3]

In short, we now recognize that over- and under-votes occur through a combination of pure voter intention and machine-induced voter error. A core feature of the Help

---

[2] "Chad" is defined as the small pieces of card stock punched out of a punch card ballot. Many of the controversies surrounding the 2000 recount in Palm Beach County, Florida derived from the fact that the punches did not always break free of the ballot card, leading to ambiguities in interpreting voter intent. "Pregnant" chad referred to cases where chad was not punched clear of the ballot card, but instead was bowed outward, the result of the stylus meeting resistance behind the card. "Hanging" chad referred to cases where chad was still attached to the card at one or more points. On the physical properties of punch cards that lead to these problems, see, Douglas W. Jones, "Chad --- From Waste Product to Headline," http://www.cs.uiowa.edu/~jones/cards/chad.html, accessed September 26, 2010.

[3] A "butterfly ballot" refers to a ballot design that is uniquely associated with Votomatic punch cards. Such a ballot places the names for an office on facing pages, with the associated punch locations in the middle. This design ends up confusing some voters, because the sequence of the names on the ballot does not correspond with the sequence of the punch-positions on the ballot card. On the effects of the butterfly ballot in Florida see Jonathan N. Wand et al, "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida," *American Political Science Review,* 2001, vol. 95(4):793-810.

America Vote Act (HAVA) and many state-level efforts at voting system reform early in the decade was a push to eliminate over-votes altogether and to limit under-votes to those clearly abstaining.

Analyzing the 2000 Florida election also led to the development of an important measure of the performance of voting systems, the residual vote rate. This report discusses efforts to calculate the residual vote rate at a highly disaggregated level in one state, Florida. As a preface to examining the challenges to calculating this simple measure, we begin by defining it and discussing how it has been used over the past decade.

Residual vote rate defined

The number of residual votes is simply the sum of over- and under-votes in a particular race. While the residual vote can be calculated for any race, it is most commonly calculated for the race at the top of the ballot, such as president, in order to avoid contaminating this measure with legitimate abstentions induced by "voter fatigue" further down the ballot.

The residual vote rate is the number of residual votes divided by the total number of voters (i.e., turnout), multiplied by 100 to convert it to a percentage. Expressed as a formula,

$$Residual\ vote\ rate = \frac{Overvotes + Undervotes}{Turnout}$$

It is rare for jurisdictions to report over- and undervotes separately, so that the number of residual votes must be calculated by subtracting the total number of votes counted in a race (including write-ins) from turnout. For instance, in the 2008 presidential election, the state of Florida reported that 8,456,329 voters showed up, either on Election Day, in early voting, or through absentee ballots.[4] The official returns report that 4,282,074 were cast for Barack Obama, 4,045,624 for John McCain, and 63,046 for all other candidates.[5]

[4] Turnout statistics were taken from the Florida Secretary of State's official turnout report, URL: <https://doe.dos.state.fl.us/elections/resultsarchive/Index.asp?ElectionDate=11/4/2008>, last accessed March 12, 2010.

[5] The election returns were taken from the Florida Secretary of State's official election returns, URL: <https://doe.dos.state.fl.us/elections/resultsarchive/Index.asp?ElectionDate=11/4/2008>, last accessed March 12, 2010.

The sum of these votes cast is 8,390,744. Therefore, the number of residual votes is 8,456,329 – 8,390,744 = 65,585. The residual vote rate is 65,585/8,456,329 = 0.78%.

The term "residual vote rate" was chosen because there is no standardized way of talking about over- and under-votes across all the states. Very few states explicitly acknowledge residual votes, regardless of the precise nomenclature they use, when they report election returns. One state, Massachusetts, calls residual votes "blank ballots," even though the ballots counted as "blank" include both over- and under-votes. A few states, notably Florida, mandate the explicit accounting for over- and under-votes separately at the county level, at least for selected races.[6] In the midst of this diversity of terminology and reporting practice, a single term, residual vote, was called for.

The residual vote rate was first used in a white paper issued by the Caltech/MIT Voting Technology Project in the spring of 2001.[7] Since then, the measure has been used in dozens of books, academic journal articles, and conference papers. (See Appendix D for a selective bibliography of these publications.) HAVA charges the Election Assistance Commission (EAC) with periodically conducting and publishing research into election administration issues. One of the issues the EAC is expected by law to report about regularly is the "best methods for establishing voting system performance benchmarks, expressed as a percentage of residual vote in the Federal contest at the top of the ballot."[8]

The component parts of the residual vote rate measure — total turnout and votes cast for all candidates, including write-ins — are surprisingly elusive. The most frequently missing residual vote component is turnout. For instance, a review of state-published vote tallies following the 2008 election revealed that thirteen states did not publish official turnout figures. A few states publish unofficial turnout figures that are collected without the benefit of statewide standards governing the collection and

[6] See Section 101.595, Florida Statutes.
[7] Caltech/MIT Voting Technology Project, "Residual Votes Attributable to Technology: An Assessment of the Reliability of Existing Voting Equipment," version 2, March 30, 2001. URL: <http://vote.caltech.edu/drupal/files/report/residual_votes_attributable_to_tech.pdf> last accessed March 12, 2010.
[8] Help America Vote Act of 2002, Public Law 107-252, sec. 241(b)(17). Whether the EAC is honoring this mandate is questionable; the 2004 and 2006 Election Administration and Election Day surveys asked states to report information that would allow the calculation of the residual vote rate, but these items were removed from the 2008 and forthcoming 2010 surveys.

auditing of this information. Finally, a few states publish turnout but fail to account for all write-in votes, which inflates the residual vote rate.[9]

Despite the fact that the residual vote rate can be calculated for only about three-fourths of America's election jurisdictions, it has become the most common metric of voting system performance. Its popularity is a mixed blessing, and probably reflects as much the difficulty in finding other good measures of election system performance as it does the value of the measure itself.

Like all indicators, it is very good at measuring some things and not so good at measuring others. Its greatest value is in assessing the performance of the machines themselves, since the core of the behavior being measured is a human-machine interaction. Nonetheless, even in its most valid uses, it can be unclear what aspects of voting machine performance — mechanical challenges, ballot design confusion, precinct workflow pressures, etc. — are responsible for variation in residual vote rates across time and space.

The presence of high residual vote rates draws our attention and makes us more confident that something has gone wrong with the voting technology. Yet, when we observe high residual vote rates, we cannot be certain what caused the numbers to spike. Was there a problem with the voting equipment itself, as when Palm Beach County, Florida struggled with hanging and pregnant chads in 2000? Was the problem with ballot design, as when Duval County voters were confused by the "caterpillar ballot?" Did election officials just fail to count ballots that had been cast, as might have happened in Bradford, New Hampshire in 2002?[10] Or, was there no problem with election administration, with voters simply abstaining in droves? The residual vote rate cannot directly adjudicate across these, and other, possible explanations. But, as we discuss below, by using the residual vote rate in analysis that takes advantage of

---

[9] C.f. R. Michael Alvarez, Stephen Ansolabehere, and Charles Stewart III, "Studying election: Data quality and pitfalls in measuring the effects of voting technologies," *The Policy Studies Journal,* 33 no. 1 (2005): 15–24 and Geralyn M. Miller, "Methodology, statistics, and voting error: An exploration of 2000 presidential election data in two states," *The Policy Studies Journal* 33 no. 1 (2005): 1–13.

[10] Stephen Ansolabehere and Andrew Reeves, "Using recounts to measure the accuracy of vote tabulations: Evidence from New Hampshire elections, 1946–2002," *VTP Working Paper* # 11 (2004). URL: <http://www.vote.caltech.edu/drupal/files/working_paper/vtp_wp11.pdf>, last accessed March 12, 2010.

variation in election administration across space and time, we can eliminate some explanations in favor of others.[11]

Evidence of the power of the residual vote rate

The power of the residual vote rate as an indicator of voting machine problems was readily demonstrated when the residual vote rate was calculated in the context of the obvious mechanical failures. The 2000 presidential election provided significant evidence that the physical failings of punch card-based voting equipment could produce an unusual number of under-votes. For instance, if a voter pushed his or her stylus through the appropriate hole to vote for his favored candidate, but the chad was only bent, not dislodged, then the voter would have produced a "pregnant chad," which would not register on the punch card reader as a vote. Research reported by Douglas Jones, a computer scientist at the University of Iowa who has studied the physical qualities of punch card voting technologies, provided more systematic evidence about how voters could vote on a punch card in the prescribed manner, and yet a machine malfunction would lead to an over-vote.[12]

If Palm Beach County's chad problems were a more general problem with voting machines, then we should have expected that localities that used punch cards to vote would have higher residual vote rates than communities that used other equipment. The Florida case bore out this prediction. This is illustrated in the following table, which reports the average residual vote rate for all counties in Florida for the 1992, 1996, and 2000 presidential elections, separating the counties according to whether they used punch cards to vote.

---

[11] The residual vote rate is a diagnostic tool, and analogous to diagnostics in medicine, it identifies potential problems but may not specifically identify a disease. Walter Mebane explains the use of election diagnostics and "forensics" in some depth in a series of paper on the use of Benford's Law to identify possible election fraud. Walter Mebane, "Election Fraud or Strategic Voting?", paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 2010; "Election Forensics: Statistical Interventions in Election Controversies," paper presented for presentation at the Annual Meeting of the American Political Science Association, August 2007.

[12] See, for instance, Douglas W. Jones, "Douglas W. Jones's chad page," available from <http://www.cs.uiowa.edu/~jones/cards/chad.html>. Internet; accessed 12 March 2010. Also see Douglas W. Jones, "On optical Mark-Sense scanning," D. Chaum et al., eds., *Towards Trustworthy Elections*, Heidelberg, IAVOSS/Springer: 175–190.

Residual vote rates in Florida presidential elections by voting equipment type, 1992–2000

Year	Punch cards	All other equipment
1992	2.45%	1.69%
1996	2.66%	1.94%
2000	3.79%	1.43%

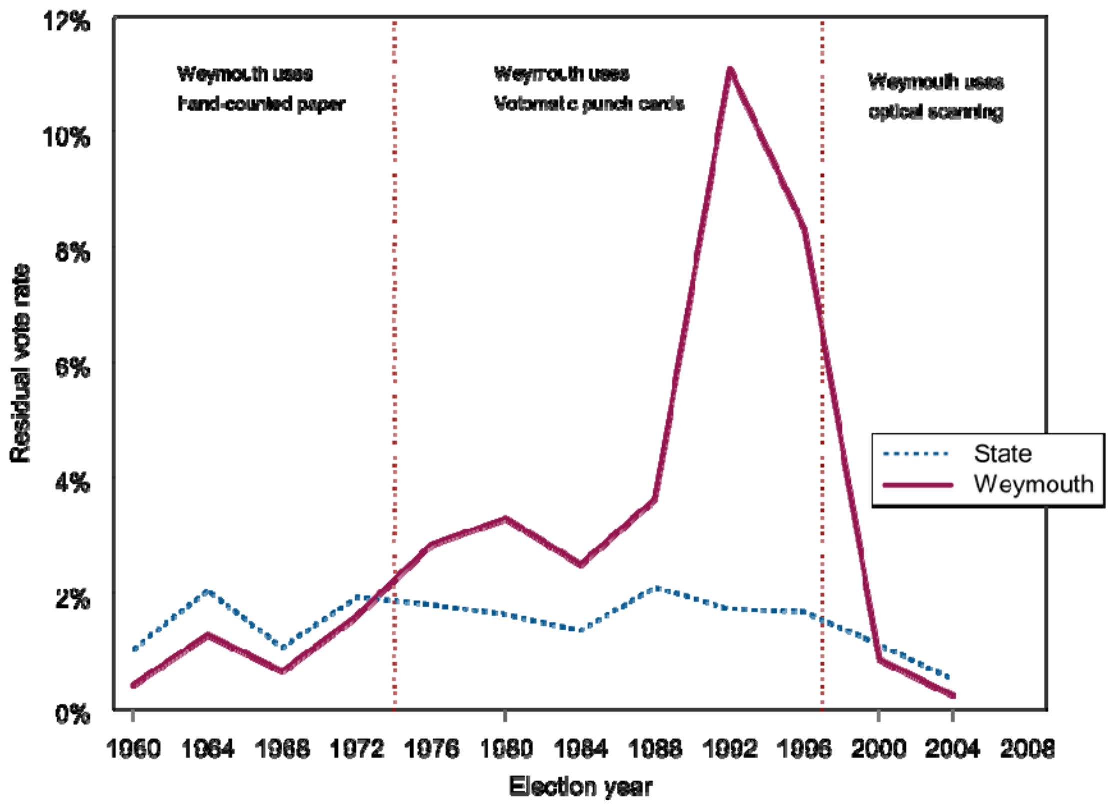
One of the notable patterns in these data is that while the residual vote rate in Florida was essentially staying constant, the residual vote rate among counties that used punch cards was growing. Palm Beach County, the epicenter of the 2000 election controversy, had a residual vote rate of 6.43% in 2000, roughly four times the statewide average.

Had the residual vote rate been regularly calculated in Florida before 2000, as a diagnostic tool, it is likely that problems with punch card voting machines would have been spotted earlier.

Florida was not the only state that was sitting on evidence––unexamined residual vote rates––of problems with punch card voting machines. Similar problems had surfaced in Massachusetts just four years earlier and, in retrospect, the pattern was even more striking. The problem of missing votes due to voting technology malfunction emerged in the 1996 Democratic primary for the tenth congressional district of Massachusetts. The original vote count, which was accepted by the Secretary of State, gave Phil Johnston a 110-vote lead over William Delahunt. Delahunt demanded a recount and, after the recount still showed him behind, brought a challenge to state court. The judge assigned to the case, Elizabeth Donovan, took nearly 1,000 disputed ballots home with her, to examine personally. Based on the marks on the ballots she observed — chad that was deformed, chad that was partially dislodged, etc. — she was able to allocate most of the ballots to one of the candidates. Delahunt ended up being the recipient of many more additional votes than Johnston, and he was declared winner by 110 votes. Delahunt went on to defeat his Republican opponent and serve fourteen years in the House, recently announcing his retirement.

Weymouth, a town that used the same Votomatic punch card voting equipment that was later discredited in Palm Beach County, was one of the towns whose high numbers of blank ballots drew the attention of the Delahunt campaign when they challenged the outcome. Apparently unnoticed by anyone before, the overall residual vote rate in Weymouth had been rising since the town adopted punch card voting in the mid-1970s. This is illustrated in the accompanying figure (1.1), which graphs the residual vote rate in presidential elections for Weymouth, starting in 1960. (For comparison the overall residual vote rate for the state is also shown.)

Figure 1.1: Residual vote rate in Weymouth, Massachusetts



Massachusetts has published turnout and a complete accounting of all votes cast in state and federal elections for over a century, so it is possible to use the Weymouth case to illustrate the usefulness of the residual vote rate as a diagnostic measure. Note, first, that Weymouth's residual vote rate was slightly below the statewide rate when it still used traditional hand-counted paper ballots. When Weymouth adopted the Votomatic system, the rate increase by about 1 percentage point, compared to the rest of the state. The most telling pattern, though, concerns the elections of 1992 and 1996, when the residual vote rate in Weymouth skyrocketed. This spike is an indicator of inadequate maintenance of the Votomatic devices.[13] In all likelihood, chad had built up within some of the devices that held the punch cards for voting, making it impossible in some circumstances for the voter to produce a clean punch.

Finally, also note that after Weymouth abandoned its Votomatic system, in favor of optical scanning, the residual vote rate returned to a level just below the statewide average.

Prior to 2000, election officials, the press, or the public rarely calculated the residual vote rate. Incidents of unusually large numbers of blank ballots seemed rarely to enter into investigations of voting equipment. Within academic research, blank ballots were attributed almost entirely to "voter fatigue" and "roll-off," not to voting equipment failure.[14] Within the election administration community, the famous "Saltman Report," which detailed investigations into computerized voting systems as of 1988, documented many disputes over-vote totals, but virtually none that focused on what we now call residual votes.[15] Ironically enough, the one case Roy Saltman highlights involving an unusual number of uncounted ballots was Palm Beach County, Florida in 1984.

As the Weymouth example illustrates, the residual vote rate is most useful as a comparative indicator of voting system performance. If the data are available, comparisons can be made along two dimensions. First, the same community can be studied across time. When a community changes voting systems, long-term changes in the residual vote rate can be an indicator of the overall efficacy of the equipment. Communities, especially those with identical voting equipment and similar demographics, can also be compared with each other, for evidence about the relative quality of local administrative practices.

The time series (Figure 1.1) above for Weymouth illustrates how the residual vote rate can be used to diagnose problems (or successes) with changing voting equipment.[16]

---

[13] Jones, *op cit.*

[14] See Jack L. Walker, "Ballot forms and voter fatigue: An analysis of the office block and party column ballots," *Midwest Journal of Political Science* 10, no. 4 (1966): 448–463; Stephen M. Nichols and Gregory A. Strizek, "Electronic voting machines and ballot roll-off," *American Politics Research* 23, no. 3 (1995): 300–318.

[15] Roy G. Saltman, "Accuracy, Integrity, and Security in Computerized Vote-Tallying," NBS Special Publication 500-158, Institute for Computer Sciences and Technology, National Bureau of Standards, August 1988.

[16] Whether the measure is used as a diagnostic tool depends on the attitudes of local officials and citizens. In Weymouth, after the abnormally large number of blank ballots was brought to his attention in 1996, the town clerk Franklin Fryer insisted that the problem was with voters, not the equipment, and that the blank ballots were caused

Of course, the experience of just one town may simply be the story of one town, and therefore it is important to analyze as many towns as possible in order to draw general conclusions about machines, per se. The more general failings of punch machines can be illustrated again from Massachusetts, by examining changes in the residual vote rate among a collection of towns that switched to punch cards from hand-counted paper at the same time, and then to see what happened when they were all required to abandon punch cards after 1996.[17]

Two hundred ninety Massachusetts towns used hand-counted paper ballots in 1972, accounting for 1.6 million voters. Four years later, 274 of these towns still used paper ballots, while 16 (representing 230,000 voters) switched to punch cards. In 1972, very little distinguished these two types of towns: the ones that would stay with hand-counted paper had an average residual vote rate of 1.7%, while those that would switch over to punch cards had an average rate of 1.9%. In 1976, there were significant differences. The 274 towns that kept their paper ballots had a collective residual vote rate of 1.5%, while the 16 towns (including Weymouth) that adopted punch cards had a rate of 3.6%. The 1.7 percentage-point increase in the residual vote rate among towns that switched from paper to punch cards represents almost 4,000 votes that were "lost" due to the adoption of this equipment.

Turning the analysis around, by 1996, 48 towns had adopted punch cards for voting, affecting 545,000 voters. The collective residual vote rate in all these towns in 1996 was 3.1%, compared to the collective residual vote rate of towns that used optical scanners of 1.2%. In 2000, when the punch card towns were now required to use some other paper-based methods, those that adopted optical scanners saw their rates drop to 1.0%. Towns that used optical scanners in both 1976 and 2000 had a collective residual vote rate in 2000 of 1.1%. The 2.1 percentage-point drop in the residual vote rate among towns that abandoned punch cards in favor of optical scanning resulted in the "finding" of over 11,000 votes in Massachusetts — votes that would not have been counted if punch cards had not been banned.

---

by voters simply deciding to abstain. See Carolyn Ryan, "State will keep watch over Election Day punch-card vote," Quincy (Mass.) *Patriot Ledger,* Oct. 9, 1996, p. 11, and "Weymouth clerk blames voters not ballots," Quincy (Mass.) *Patriot Ledger,* Oct. 8, 1996, p. 8.

[17] After the Weymouth debacle, the Massachusetts Secretary of State William Galvin pulled certification from punch cards, so that they had virtually disappeared from the state in time for the 2000 presidential election.

Towns can also be compared one with the other, though this comparison is much more difficult than the cross-time comparison, owing to the fact that localities vary so much, in terms of their demographics, political attitudes, etc. It is also the case that localities often choose particular voting equipment because of these differences. For instance, very small towns might rely on hand-counted paper, while a large city is more likely to rely on a system that is heavily computerized and automated. As a consequence, cross-sectional differences tend to be handled through statistical techniques, in which demographic and political factors are controlled explicitly or implicitly.

Articles written by Ansolabehere and Stewart[18] and Stewart[19] provide two example of how implicit controls were used. They employed "fixed effects regression" on two datasets of American towns and counties that had been constructed to simultaneously study cross-sectional and cross-time differences in residual vote rates. The primary purpose of these articles was to quantify the relative performance of different voting machines, separating out political and demographic effects. Ansolabehere and Stewart's research, published after the 2000 election, verified the inferior performance of punch cards and quantified the superior performance of optical scanners and DREs. Stewart's research, which was conducted after 2004 election, put hard numbers on how many votes were "found" due to the HAVA requirement that localities abandon antiquated equipment in favor of newer technologies. Overall, he found that the HAVA requirement added nearly one million votes to the totals in 2004, compared to what would have occurred had equipment upgrades not occurred.

Shortcomings/limitations of residual vote analysis

While it is a powerful measure of voting system performance, the residual vote rate has limitations. One misunderstanding of the measure is that it is based on an assumption that all over- and under-votes are due to machine error. As the discussion above makes clear, nothing can be farther from the truth. Over- and under-votes are caused by a combination of voter abstention, voter error/confusion, and system failure. Election officials may have no responsibility for reducing rates of voter abstention, but

[18] Stephen Ansolabehere and Charles Stewart III, "Residual votes attributable to technology," *Journal of Politics* 67 no. 2 (2005): 365–389.

[19] Charles Stewart III, "Residual vote in the 2004 election," *Election Law Journal* 5 no. 2 (2006): 158–169.

they are responsible for seeing that residual votes due to confusion and system failure are as low as possible. The greatest power of the residual vote rate measure is when statistical techniques are used to analyze changes or differences in the residual vote rate to assess whether efforts made by election officials to improve the performance of election systems have been successful.

Because the residual vote rate has been suggested as a measure of the quality of election administration and voting system performance, it is important to understand if there is a floor below which the residual vote rate cannot fall. If it is at least conceivable that a voting system could drive the number of over- and under-votes due to system failure and voter confusion to zero, then this question is reduced to asking how many voters abstain in presidential and other top-of-the-ticket races. Because of the secret ballot, it is impossible to know precisely how many people deliberately abstain from voting, but there are some ways to estimate what this number might be.

Nevada is a good place to start, since it allows voters to choose the response "none of these candidates" in the presidential race. With an explicit place on the ballot for voters to record abstention, Nevada provides a direct measure of the practice, at least in that state.

The following table reports the percentage of Nevada voters who chose "none of these candidates" in the 2000, 2004, and 2008 presidential election, along with the gubernatorial races in 1998, 2002, and 2006.

"None of these candidates" and residual vote rates in Nevada, 1998–2008

Year	Turnout	"None of these candidates" pct.	Residual vote pct.	Voting system
Presidential elections				
2000	613,360	0.54%	0.72%	Punch
2004	831,833	0.44%	0.27%	DRE
2008	970,019	0.65%	0.22%	DRE
Gubernatorial elections				
1998	440,042	2.87%	1.46%	Punch

“None of these candidates” and residual vote rates in Nevada, 1998–2008

Year	Turnout	“None of these candidates” pct.	Residual vote pct.	Voting system
2002	512,433	4.62%	0.17%	DRE
2006	586,274	3.53%	0.70%	DRE

The first thing to note is that the percentage of voters choosing “none of these candidates” was fairly stable across time among presidential and gubernatorial voters—approximately 0.5% for the former and 3.5% for the latter. These figures provide a rough estimate of deliberate abstention rates in these races in Nevada. Second, note that when Nevada switched from punch cards to electronic machines, the residual vote rate dropped considerably.

From the perspective of the question, “what is the lowest practical residual vote rate?” Nevada provides a provisional answer. If residual votes are entirely due to voter abstention, and a state does not have a “none of these candidates” option on the ballot, then that rate is approximately ½ percentage point in presidential elections. Even with such a provision on the ballot, some voters do not record a choice. Whether this is due to a system shortcoming that may be remedied or voter characteristics that are impossible to remedy through policy or administrative practice is open to further analysis.

A second way to estimate an empirical floor on the residual vote rate is to rely on public opinion surveys that ascertain how many voters abstain in particular elections, even though they have gone to the polls. Commercial surveys rarely report the fraction of respondents who say they did not vote for president, but academic surveys do. In the 2008 American National Election Study (ANES), which is the longest-running academic study of American elections, 0.65% of respondents who stated that they voted also stated that they did not vote for President. Another major academic survey, the Cooperative Congressional Election Study (CCES), estimated a much lower 0.14%. The ANES estimate is consistent with its surveys in previous years; the CCES is a relatively new project with only one presidential election under its belt. The ANES result suggests that the Nevada estimate may apply nationwide, while the CCES estimate suggests it may actually be somewhat lower.

A final way to estimate a floor on the residual vote rate is just to look at the distribution of the residual vote rates across all counties in the United States. In the 2,000 counties in which it was possible to calculate the residual vote rate in 2008, 90% had a residual vote rate of 0.39% or higher, and 95% had a residual vote rate of 0.26% or higher. Using this standard, we might conclude that a county with a residual vote rate of 0.25% in the presidential vote would be reaching the practical limits of how low the rate could possibly go.

The residual vote rate and the evolution of election administration

The residual vote rate was developed following the 2000 presidential election as an indicator of the performance of voting technologies. The original analysis by the Caltech/MIT Voting Technology Project correlated changes in voting equipment with changes in the residual vote rate to estimate which types of voting equipment were associated with more (or fewer) "lost votes."

One shortcoming of that analysis, which has only grown over time, is the fact that it fails to distinguish voters who vote in person from those who vote absentee. In jurisdictions that use DREs or mechanical lever machines, the voting method is different, depending on whether one votes in person or by mail. Over the past decade, the fraction of voters casting ballots by mail has grown nationwide, from 7.9% in 1996, to 10.2% in 2000, to 16.4% in 2008.[20] This nationwide growth masks tremendous variability in individual states, with the West particularly enjoying a much more rapid growth of mail-in balloting than the rest of the country. In California, for instance, 42% of voters cast mail-in ballots in 2008, compared to only 25% in 2000.[21] In Colorado, 78.6% of ballots were cast by mail in 2008, a growth of 65 percentage points in four short years.

There are reasons to suspect that the residual vote rates of mail-in voters will differ from in-person voters. Most importantly, mail-in ballots do not have the same protections as do many precinct-based voting machines, which warn voters when they have cast an over- or under-vote.

---

[20] Paul Gronke, James Hicks, and Daniel Krantz Toffey, "N=1? The Anomalous 2008 Election and the Lessons for Reform," paper presented at the 2009 meeting of the American Political Science Association. Figures are reported on pg. 6 and are drawn from the Current Population Survey Voting and Registration Supplement.

[21] California Secretary of State, *Statement of the Vote*, November 2000 and 2008 general elections.

Differences between the two modes are important to capture, owing to the centrality of residual votes in the 2000 presidential elections. Billions of dollars have been spent to allow localities to buy new equipment for their in-precinct voting; the residual vote rate is an important indicator of whether these expenditures have been successful. To the degree that residual vote rates are a product of in-person and mail-in voting, the causes of changes to the residual vote rates over time in a jurisdiction will be obscured.

While most, but not all, states and localities report the necessary data elements to calculate the residual vote rate (more on this below), very few disaggregate this information by voting mode.[22] Therefore, it is becoming increasingly difficult to assess whether changes in voting practices are leading to an increase, or decrease, in lost votes.

One highly visible example of example where reporting residual vote rates by voting mode was particularly instructive came in 2006, following the November general election in the thirteenth congressional district of Florida. This was a highly contested race between Vern Buchanan (R) and Christine Jennings (D). At the time, it was the most expensive congressional race in American history. When the votes were counted, it was discovered that the residual vote rate in this race was 12.9% in the precincts that were located in Sarasota County, compared to 2.5% in the precincts located in the other counties that comprised the district.[23] Upon closer examination, the residual vote rates among absentee voters in Sarasota County was 2.5% (a rate similar to that in the other counties in the district) but 17.6% among early voters and 13.9% among Election Day voters. Because early and Election Day voting was conducted on ES&S iVotronic electronic voting machines, whereas absentee balloting was conducting using optical scanners, attention was drawn to the performance of the electronic machines.

The Sarasota case led to a court case, and eventually an election challenge that was brought to the House of Representatives. The case was a cause célèbre among those concerned about the role of electronic voting machines in American elections. The important point here is the fact that Sarasota County and most of the surrounding counties reported precinct-by-precinct election returns in such a way that the residual vote rate could be calculated for early voters, Election Day voters, and absentee voters

[22] Also known as "group" in Florida, mode refers to the method of balloting used by a voter –– e.g., in person, absentee (by mail), early, or provisional.
[23] See Charles Stewart III, "Declaration of Charles Stewart III on Excess Over-Votes Cast in Sarasota County, Florida for the 13th Congressional District Race," November 20, 2006.

in most precincts of the district. This allowed precise estimates of not only the effect of the iVotronic malfunction, but of the efforts by the Sarasota Supervisor of Elections to mitigate these problems when they were first manifested in the early voting on these machines. Absent this detailed data, the election challenge would still have gone forward, but it would have provided almost nothing of a diagnostic nature, for those more interested in making sure that the mistakes made in this election, whatever they were, were not repeated.

The case of Florida's 13th congressional district in 2006 provides a natural transition to discussing the calculation of residual vote rates at a highly disaggregated level, for the purposes of auditing election performance and improving them in the future. Florida, which had passed several laws in the 2000s mandating the reporting and analysis of residual votes, would seem the natural focus of a study that shows how disaggregated election returns could be used to help improve elections. As the rest of the report documents, however, even with the best of intentions and the law on the side of very precise reporting, gathering residual vote data remains a significant challenge.

In summary, the residual vote rate is a primary indicator of the health of election systems. It is relatively easy to calculate, can be compared in absolute value across diverse electoral contexts, and has been collected and reported over many election cycles. Comparatively high residual vote rates are a diagnostic tool, flagging a potential problem with voting technology or the voting system, but may also indicate voter disinterest in a particular election. While residual vote is not the only performance indicator available, residual vote rates should be part of an index that measures election performance.[24]

This report details an attempt to move beyond past work, collecting and analyzing residual vote rates at the precinct (not county) level and across different modes of balloting (early in person, absentee, Election Day) in Florida. Using precinct-level data confers substantial advantages to our desire to identify the potential sources of voting errors, whether they be a function of machine / technology, ballot style, political context or campaigns, or voters themselves.

[24] Heather Gerken, *The Democracy Index: Why our Election System is Failing and How to Fix It,* Princeton, NJ: Princeton University Press, 2009.

In the next section, we describe the data collection process we followed in Florida. This section summarizes our data collection efforts, highlights problems that we encountered, and acts as a handbook for future analysts interested in collecting and calculating residual vote rates. It is important to note the data challenges we faced, since these help to identify changes that states and localities can make so that future analysts can more easily understand the performance of their elections systems. We hope that the handbook helps other election advocates, observers, and scholars to understand how the residual vote measure is created, and possibly encourage them to collect residual vote measures on their own. To our knowledge, this is the only easily accessible users' guide to the use of the residual vote rate.

Data collection for calculating the residual vote rate

Summary

The greatest problem for the residual data project has been collecting *good data*. For the purposes of this project, this comprised three basic elements:

- coverage of all Florida counties;
- collection of data formats that were readily convertible to a standard format; and
- coverage of multiple elections.

Florida's Division of Elections has been a willing and crucial partner in this project, and has been responsive to all our data requests. Unfortunately, most of the data we need (if it is available *at all*) continues to reside at the county level, or in the memory cards of election machines. In short, the data elements exist in most states and localities, but the correct output format or report is either not produced, or is produced and not disseminated. This remains the greatest hurdle to developing standardized measures by which we can evaluate the performance of elections systems over time––not just within a single state such as Florida, but nationwide.

Our original plan was to collect precinct-level data on over- and under-votes for all 67 Florida counties from 2006 to 2008. Unfortunately, between the time of initial funding and project initiation, the state changed its laws such that election jurisdictions were no longer required to submit precinct-level election results, including under- and over-votes. Thus, the only election for which there are statewide data at the precinct level that include under- and over-votes is for the 2008 presidential primary. For some counties, data prior to the 2008 presidential preference primary and subsequent to the primary continues to be available, and we have reported on these counties in this report and on an external website: <http://www.earlyvoting.net/resvote>.

Project staff found that the pertinent data are often *available*, but are not always easily *accessible*. [25] The issue of data accessibility remains the key challenge in performing residual vote analysis at the precinct level.

---

[25] It would be possible, of course, to contact each of the 67 Florida counties in an attempt to collect the information. At points in this project, we did so in order to address limitations in the dataset provided by the state. However, if elections data metrics are to be produced in states with large numbers of counties, and certainly in more than a

Election results

Our first objective was to obtain and process election results files from each of Florida's 67 counties. We planned to compile single datasets for the 2006 general, 2008 presidential primary, and 2008 general elections, and merge these into a single, pooled cross-sectional time-series dataset.

The genesis of the project was a change in state law that Florida planned to implement prior to the 2006 general election (though unfortunately, the law did not go into effect until after the 2006 general election). Though the counties are the primary repositories of these election data in Florida (as in most states), new state legislation required them to report to the state precinct-level breakdowns of election results, including over- and under-votes.[26] The Secretary of State's Division of Elections collected these data, and the Division was able to provide us with CDs containing these results for 2006 and 2008.

In many cases, however, we found these CDs lacked the data requisite for our study, and we had to turn directly to the individual counties. We found many of them provided additional data and formats on their websites, and we compiled (and in some cases cross-referenced) data from both sources to build our datasets.[27]

There were clear differences between years. It was substantially easier to find complete data for the 2008 presidential preference primary. In the 2006 general election, we were able to collect complete precinct-level residual vote data for only 8 of 67 counties. The elections division at Miami-Dade (by far Florida's largest county) provided excellent quality data for both elections, meaning that we are able to cover about 25% of Florida's population in 2006. Counties that provided usable data in 2006 and 2008 were: Okaloosa and Miami-Dade (complete computer-readable files) and Bay, Charlotte, Clay, Hamilton, Orange, and Santa Rosa (full EL30 or EL30A files).

For the 2008 presidential preference primary, we were able to include 51 of 67 counties in our dataset.

---

number of states, it is not reasonable to have researchers contacting each county or local jurisdiction individually. In addition, LEO's are already burdened with reporting requirements and data requests; automating the reporting of this information to the state would release them from responding to individual data requests.

[26] § 101.573, Fla. Stat. (2007), since repealed after lobbying by the county clerks.

[27] As an aside, this was not the first time that we discovered that the counties were not following state-mandated reporting, with little apparent consequence.

The following counties provided high quality data for this election: Alachua; Baker; Bay; Bradford; Broward; Calhoun; Charlotte; Citrus; Clay; DeSoto; Dixie; Flagler; Franklin; Gadsden; Gilchrist; Glades; Gulf; Hamilton; Hardee; Hendry; Hernando; Highlands; Holmes; Indian River; Jackson; Jefferson; Lake; Lee; Leon; Levy; Madison; Manatee; Marion; Martin; Miami-Dade, Monroe; Nassau; Okaloosa; Okeechobee; Orange; Putnam; Sarasota; St Johns; St Lucie; Sumter; Taylor; Union; Volusia; Wakulla; and Walton.

The following counties did not provide usable data for our project: Brevard; Collier; Columbia; Duval; Escambia; Hillsborough; Lafayette; Liberty; Osceola; Palm Beach; Pasco; Pinellas; Polk; Santa Rosa; Seminole; Suwannee; and Washington. The many counties that made available incomplete data for 2006 most commonly did not provide a breakdown by mode (at the precinct level).[28]

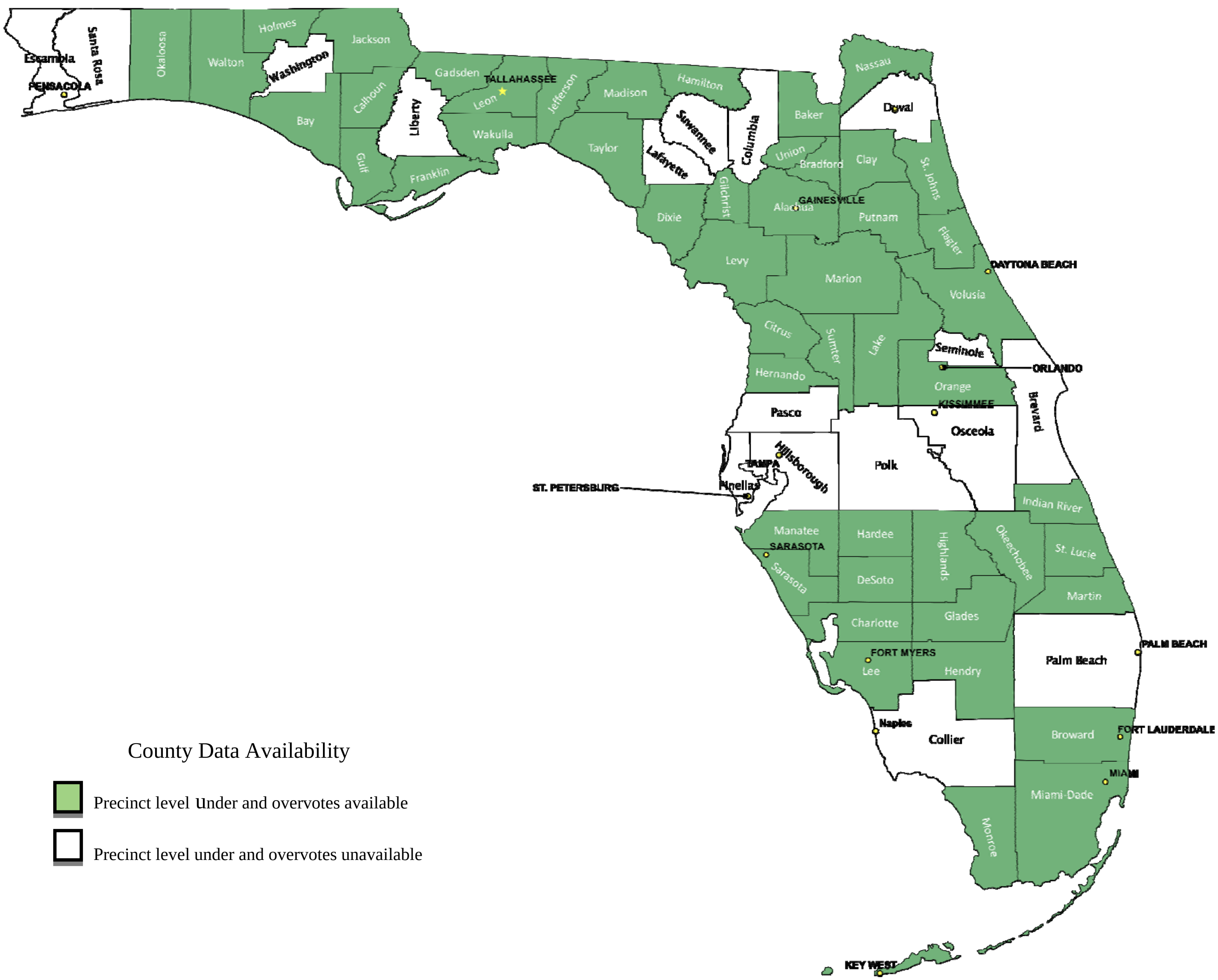
Figure 2.1 shows this graphically.[29] It may seem from the map that the populous, urban, and in most cases wealthy areas in Florida were able to provide full data, while the less urban, poorer counties were unable to supply data. This is generally true in Florida, but Hillsborough (Tampa), Pinellas (St. Petersburg), and Palm Beach, among other wealthier counties, have substantial missing data. Many counties simply do not have the funding or the personnel available to respond to demands for high quality data, regardless of whether the demands come from the state, scholars, the press, or the public. Rather than cease data reporting mandates, as happened in Florida, states should provide either funding, technical support, or both to their local election offices.

---

[28] As noted above, the project was initiated with the expectation that precinct level data would be submitted to the state starting with the 2006 general. This turned out to have been a misinterpretation of the law by then-Director Amy Tuck, and statewide collection did not begin until 2008. We discovered that some counties were able to provide us the necessary data elements in 2006, and in some cases before, but only learned this after communicating with individual counties.

[29] Interested users may wish to browse the project website –– http://www.earlyvoting.net/resvote –– that provides both the 2006 and 2008 data in a common format.

Figure 2.1: 2008 Presidential Preference Primary coverage



Data formats

Florida law contained no provision specifying the format of the elections data sent to the state, nor did the Secretary of State's office require a common format.[30] The diverse voting technologies and vendors used by different counties meant that there was (and is) no single standard for the election results data. In fact, in many cases, our different sources (Secretary of State, county offices, and websites) provided a number of different formats for the same data.

We found the data were largely available in four formats:

1. Flat file

A flat-file database records data in such a way that a single line of text is a "record," for instance, the election data for a single precinct. Within a record, different data fields — such as the number of votes for the candidates, the number of over-votes, etc. — are defined in a standardized way that is consistent throughout the data file. Flat files generally use only formatting and characters that can be created with a typical computer keyboard.[31] These files may be viewed with simple text editors (e.g., Windows Notepad, Apple TextEdit) available on all modern computers. Their simple format makes them ideally suited for containing standard, machine-readable data.

1.1. CSV

"Comma separated values" (CSV) files report data values, using a comma as a way to distinguish different data elements (each "column"). Sometimes the individual values are also enclosed in quotations (most typically when the variables have alphanumeric values that could themselves have a comma embedded, such as "Charles H. Stewart, Jr."). Ideally, the first line in these files will be a "header" that contains the variable names. Without a header line, a separate list of variable names and the order in which they appear on each line

[30] In a cursory search of state election codes, we have not found a single state that specifies data formats.

[31] Spreadsheets and statistical packages generally represent databases in files that use "binary" characters that are usually not readable to the human eye. A well-known file of this sort is the Microsoft Word .xls file. Spreadsheet and statistical packages use binary files because they were more efficient for the program they are associated with to use. A major disadvantage of binary files, in addition to the fact that they cannot be read directly by humans, is that the program they are associated with can only read them. For instance, if one wishes to use an .xls data file in the statistical program SPSS, it is necessary to convert the Excel file either to a flat file, which SPSS could read, or to the binary file associated with SPSS.

is needed (as was the case in many of the files we received). Traditional spreadsheet and statistical programs can read CSV files. They are straightforward to process.

1.2. Fixed width

A "fixed-width" file is a special type of flat file, the most common format for election return reports produced by standard reporting software systems. Every "record" in the file has precisely the same variables in the same columns. Suppose, for example, that you know that the ZIP+4 is contained in columns 10-19 of each data record. This would mean that the 5-digit ZIP code is in columns 10-14, column 15 will have a dash, and columns 16-19 contain the "+4" code. Fixed width files were very common prior to about 1980, when many computer programs could not handle files with line lengths greater than 80 characters. Today, fixed length files can have very long line lengths. Fixed width files are extremely efficient computationally. However, a major problem is that one has to create a fixed length ahead of time for fields such as "Last Name" that can vary significantly in length. These files contain similar information to an ES&S machine reporting format called an EL30 (described below), but without most of the whitespace and extraneous labeling.

The following page contains an example observation from Miami-Dade's fixed-width file:

This line describes Rudy Giuliani in Precinct 001, in the 2008 presidential preference primary:

```
110100100100000850000520000160000170000000000000REP       PRESIDENT                             Rudy Giuliani                     PRECINCT 1
```

The numeric string is actually composed of several sets of values (of varying widths). The challenge with these files is identifying which set of columns corresponds to a particular data element. To illustrate, the fields of interest in this example are shown below:

```
110100100100000850000520000160000170000000000000REP       PRESIDENT                             Rudy Giuliani                     PRECINCT 1
----   --| -----|-----|-----|-----|            ---       -------------------------------|------------------------------|-------------
 1-4 8-10 12-17 18-23 24-29 30-35             48-50             58-113                          114-151                         152-206

 1-4     Contest Code (1101 is the Republican presidential primary)
 8-10    Precinct number
 12-17   Total number of votes cast
 18-23   Number of votes cast at the polling place
 24-29   Number of votes cast early in-person
 30-35   Number of votes cast no-excuse absentee
 48-50   Party identifier
 58-113  Contest
 114-151 Candidate name, or data description (“over votes”, “ballots cast”, etc.)
 152-206 Precinct text name
```

The example above provides a mapping of variable names (Contest Code, Precinct number, etc.) onto the columns of the data record. A vexing complication of election data reports is that the codes and labels used to identify the contests are rarely consistent across counties, even among counties that use the same software to report election returns. While the major federal and state contests are sometimes identified in the same way, numerous county-level and local contests have unique identifiers. The analyst needs to decide whether he or she is interested in residual votes for these lower level contests, or just statewide and higher contests.

These files contain only alphanumeric data without any other identifying information.[32] Though importing the data string is straightforward, the files require a data "map," a codebook, or a set of formatting statements, so that a computer program can reconcile particular columns with particular variables. This is not a complicated procedure but must be precise. Furthermore, at least in the files provided by the counties to the Florida Division of Elections, the location of the data elements in the columns was not the same from county to county. Each county uses a slightly different ordering and different sets of data elements (e.g. some report returns broken down by mode, some report over and under-votes aggregated by race only, and some report neither). Consequently, without a full EL30 for comparison, or some other form of data map, it is not possible to know what columns contain what data elements.

On the other hand, the advantage of these files is that once we know which column is which, they are very simple to process. Using statistical software such as *Stata* or *R*, one can easily convert these data, given a suitable "map".

The following is a sample programming fragment (for *Stata*)[33]:

---

32 The line length in our illustration is 151 characters long. In order to inspect flat files (CSV or fixed format) you need a text editor that will not "wrap" long lines. Do not use Word to view these files. On a Macintosh, you can view these files in editors such as TextWrangler or BBEdit. On the PC, Windows Notepad does not wrap by default.

33 The full data processing programs are available on the project website <http://www.earlyvoting.net/resvote>.

```
 /* this program processes fixed-width files derived from EL30s ::
be sure to
 pass county, year, path to file, and the appropriate modes in the
order they
 appear in the results file (see codebook) */

 capture program drop fixedwidth
 program define fixedwidth
 version 10
 args cou year path mode1 mode2 mode3 mode4 mode5
 capture log using "logs/fixedwidth/`year'`cou'.log", replace

 #delimit ;
 infix contest_code 1-4       /* unique numerical identifier for
each
 contest (NOT consistent across counties) */
   precinct_code 8-11         /* numerical precinct identifier */
   mode1 18-23 mode2 24-29 mode3 30-35 mode4 36-41 mode5 42-47
                         /* breakdown of the total by mode */
   str party 48-50            /* party identifier */
   str contest_text 58-113         /* human-readable contest
identifier */
   str element 114-151        /* candidate/response identifier */
   str precinct_text 152-206  /* precinct text identifier */
   using "`path'", clear;
 #delimit cr
```

2. **XLS**

   Some files came in various forms of Excel/HTML tables, which required a different conversion for each file. These files are easier to convert into an analysis-friendly data format than some files below, but these conversions must be done manually, file-by-file.

3. Formatted Text

3.1. ELXX (*Figure 2.3*)

ES&S machines, when instructed properly, will produce an EL30 or EL30A (less commonly, EL45 or EL52) report. These are widely used, human-readable text files. We found this format to have two main advantages: first, its clear labeling generally avoids any confusion; and second, those counties that provided EL30s tended to provide complete and detailed data.

However, importing these into our dataset was more of a challenge. The header information on these files is not standardized and may be customized for each county by the vendor. Professor Stewart was able to convert the files by hand, but that was a labor-intensive process that would not be replicable across a large number of counties, states, or years.

PREC REPORT-GROUP DETAIL — PRESIDENTIAL PREFERENCE PRIMAR — OFFICIAL RESULTS
BAY COUNTY, FLORIDA
JANUARY 29, 2008
RUN DATE:02/08/08 12:47 PM — STATISTICS

0001 1 HEISLER HALL

	TOTAL VOTES	%	POLL M100	POLL TS	ABSENTEE	EV M100	EV TS
REGISTERED VOTERS - TOTAL	1,783						
REGISTERED VOTERS - REPUBLICAN	670	37.58					
REGISTERED VOTERS - DEMOCRAT	709	39.76					
REGISTERED VOTERS - NONPARTISAN	404	22.66					
BALLOTS CAST - TOTAL.	576		474	1	39	61	1
BALLOTS CAST - REPUBLICAN	291	50.52	244	1	17	29	0
BALLOTS CAST - DEMOCRAT.	221	38.37	176	0	19	25	1
BALLOTS CAST - NONPARTISAN.	64	11.11	54	0	3	7	0
VOTER TURNOUT - TOTAL		32.31					
VOTER TURNOUT - REPUBLICAN.		43.43					
VOTER TURNOUT - DEMOCRAT		31.17					
VOTER TURNOUT - NONPARTISAN		15.84					

********** (REPUBLICAN) **********

PRESIDENT
(Vote for) 1

	TOTAL VOTES	%	POLL M100	POLL TS	ABSENTEE	EV M100	EV TS
Rudy Giuliani	25	8.62	17	0	2	6	0
Mike Huckabee	60	20.69	46	0	3	11	0
Duncan Hunter	0		0	0	0	0	0
Alan Keyes	0		0	0	0	0	0
John McCain.	101	34.83	87	1	6	7	0
Ron Paul.	10	3.45	7	0	0	3	0
Mitt Romney.	90	31.03	85	0	4	1	0
Tom Tancredo	0		0	0	0	0	0
Fred Thompson	4	1.38	1	0	2	1	0
Over Votes	0		0	0	0	0	0
Under Votes	1	.34	1	0	0	0	0

Figure 2.3: EL30 Sample (Bay County, 2008 Presidential Preference Primary)

We considered whether it would be possible to write an "intelligent" computer program or widget which would analyze the headers of some section of the EL30, discern the pattern by which the results were being reported, and output a standardized set of results. While we did not pursue this option in this project, we believe this is quite feasible and would go a long way toward making election results in their current form more widely accessible. The problem is identifying the most commonly used "formatted text" styles that are used nationwide. In an ideal world, the program would also know where the files are stored and could act as a "web crawler" and data aggregator, but simply providing a program or set of programs and letting individual users process files would be a boon to the elections community.

We are not the first to suggest some sort of common data format for elections data. More information on this point is contained in the recommendations section at the end of this section and in Appendix E.

4. **Adobe PDF**

   A few counties submitted their election returns data as Adobe PDF files. This was, by far, the most problematic format. Some PDF files present the data in a simple tabular layout, making it easy to read the data into a statistical package. Many others however are structured in way that makes data difficult to extract (in the worst cases, they were scans of paper documents). Some attempts were made to convert these files using optical character recognition (OCR) software but this is both processor intensive and error prone. Manual data entry is another option, but this suffers from human error.

Data quality

The data were not always of the quality required to build our dataset. The three general problems were:

1. **Different unit of analysis**

   While counties were required to submit precinct-level results to the Secretary of State, the laws did not specify how complete those results had to be. So, while every county reported basic vote totals, and most reported mode breakdowns (at the precinct-level), not all published the crucial residual vote data at this same level.

2. **Missing data**

   The files that we have for several counties in 2008 (and many in 2006) are simply missing data of some form. In some cases, early and absentee voting totals are provided at the county level (rather than for individual precincts). In others, neither specific residual votes, nor ballots cast in each contest, were given –– making it impossible to impute residual vote rates.

3. **Lack of labeling**

   See the discussion of fixed-width files (1.2) above. This was a fatal problem, where relevant.

Demographics

Our second task was to obtain and process files containing demographic information and ballot history for Florida voters. There are two obvious sources of demographic information for this type of analysis: (1) the U.S. Census; (2) voter registration files. Because Florida falls under the provisions of Section V of the Voting Rights Act, the registration file does contain racial information, along with age, gender, party affiliation, and place of residence. However, the much richer trove of information from the Census would be much preferred.

Census

The obvious source for demographic data is the U.S. Census Bureau, which collects information on a regular basis through the American Community Survey (ACS), in addition to its decennial responsibility. However, since our unit of observation in this project was the electoral precinct, it was vital to obtain demographic data grouped by that unit.

If a state's precinct lines do not conform to Census geography, then it is very difficult to produce valid population estimates within a few years of the decennial census, especially in a state that experiences population shifts, growth, or decline. One way for a state's lines to correspond closely to Census divisions is if the participates in the voting district / block boundary suggestion program (VTD/BBSP).[34] This program was initiated by the Census Bureau as a way to improve the accuracy and currency of the TIGER files disseminated by the Census Bureau, and to improve the usability of these files for state and local officials drawing political boundaries. The program provided an opportunity for states and local jurisdictions to suggest Census block definitions that corresponded to known geographical features and existing political districts. It required participating states to conform to Census block boundaries when drawing their federal and state legislative districts.

[34] This 2010 version of the program is described here: http://www.census.gov/rdo/program_phases/2010_census_program_phases.html. A nationally successful VTD/BBSP program would aid election and voting rights analysts tremendously.

Unfortunately, Florida did not participate in the 2000 program, and obtaining valid population estimates via other channels proved cost-prohibitive.[35] Therefore, we relied on the more limited information available via the voter registration files.

Voter Files

With more success, we worked with two files provided by the Florida Division of Elections. The first was the "voter history" file (approximately 73 million observations). The file holds a record for each individual voter––from around 1980––that contains the date and mode of balloting (absentee, polling place, etc.) for each election. Second, we used the "voter extract" file (about 11 million observations), which records contact information and demographic details for current, registered voters.

These two files contained a common voter identification code, which allowed us to merge them. Combining the files yielded a single dataset of current registered voters that allows the vote mode to be examined alongside a variety of demographic variables, such as age, sex, race and party affiliation.

Because this file contains address data, we were able to calculate estimates of the demographic characteristics of each precinct, at least in terms of registered voters. This method is not ideal, for a number of reasons. Most importantly, the voter registration file lacks a number of useful demographic variables (income, religion, and so on), leaving us with only gender, age, and race. This limits the depth of our analysis. Still, this file gives us a basic profile of each precinct, and allows us to begin to leverage the power of lower-level analysis; we show some examples of this below. It does not help us, obviously, understand how the demographic characteristics of *non-voters* may affect residual vote rates.

Organizing data

We started with a number of data sources for the 2008 presidential preference primary: a single voter registration file (provided by the Florida State Office of Elections) and a precinct-level election results file from each county. We merged the county files into a complete statewide results dataset (with the attendant problems that

[35] We investigated the possibility of obtaining inputted demographic figures for election districts in the state, something available from firms that produce such measures for political campaigns, but the costs were prohibitive (upwards of $30,000 for just one year).

we outlined in the *Data Collection* section above). We then aggregated the demographic data contained in the voter registration file by precinct. Finally, we joined these two datasets together, allowing us to make demographic inferences about the residual vote rates.

Appendix A contains a codebook for the final dataset. The data and codebook can also be downloaded at the project website: http://www.earlyvoting.net/resvote.

There are some variables that deserve particular explanation, because they illustrate some of the difficulties than arise when dealing with election data across many jurisdictions.

precinct_merge / precinct_code

precinct_code is a short alphanumeric string copied directly from the county results files. The format of these codes is standardized within, but not across, counties. Indeed, it occasionally denotes *sub*-precinct units (such as where a precinct contains several different congressional districts). These subdivisions are typically denoted by a numeric suffix on the standard precinct code (0, 1, 3, 6, etc) – see Example A below – but sometimes the numbering scheme is altered altogether – see Example B.

Example A: Numeric Suffix		Example B: Sequential Numbering	
10	precinct 1, no subdivision	1	precinct 1, no subdivision
23	precinct 2, congressional district 3	2	precinct 2, congressional district 3
106	precinct 10, congressional district 6	3	precinct 2, congressional district 6

This lack of standardization is not necessarily a problem, since we are typically interested in aggregated statistics, rather than any single precinct. However, to be able to attach demographic information from the voter registration file, we need to be able to correctly match the precincts in both datasets. Unfortunately, the precinct coding used by the state in the voter registration file did not always match that used by the counties, and the state file contained only whole precincts (never subdivisions).

These problems necessitated the creation of a new variable in the results dataset — *precinct_merge* — to match the state format. The descriptive *precinct_text* variable was helpful in this process, but it still had to be done manually for each county.

contest_code

This is a contest identifier, created by the project team. (A "contest" is the office up for election, the referendum being voted on, etc.) There was no consistency (and, in fact, wide variation) in the contest labeling schemes used by each county, meaning that the first task was to create a single, uniform system of codes.

First, we manually standardized the text labels of the three contests common to all counties (the two party primaries and a constitutional amendment question). Simple pattern recognition helped to speed this process up. Next, we generated a list of all the contests in our dataset (this included many councilors, mayors, school boards, and so on). We assigned unique codes to each of these, and then created this variable.

mode / modes, tech / techtype

The county results files typically disaggregated the election returns by different modes (i.e. polling place, absentee, etc). Some counties also went further, breaking these modes down by technology type ("Polling Place iVotronic", "Polling Place M100", etc.).[36] Although these were not consistent across counties, it was straightforward to produce a standardized variable (*modes*).

To allow us to make comparison across both modes and technologies, we created several further variables:

- *mode* is simply a labeled-numeric variable containing *only* the method of balloting. It takes one of five values: "Polling Place"; "Early Voting"; "Absentee"; "Provisional"; or "Other".
- *techtype* is a generic description of the type of technology used in voting. It takes one of three values: "Optical Scan/Precinct Count"; "Optical Scan/Central Count"; or "DRE".
- *tech* contains the specific machine used, where available. We gleaned this information from two sources: (1) the county results files, if available; and (2) a spreadsheet from the Secretary of State that listed the machine types used in each county.

[36] Bay County's EL30 results file provides a good example of this (appendix C).

Lessons learned during data collection

Many of the lessons learned in a long and sometimes-frustrating data collection process were detailed above, but we outline them here for the convenience of the reader.

- **The need for data standardization**
  Without some sort of interoperability standards for state and local elections data, it is extremely difficult for even technically savvy users to analyze election returns, election reports, voter history, and voter registration files. There is the possibility of a creating a set of computer programs that convert commonly used reporting formats into usable numerical data, but this is a poor substitute for consistent, standardized data reporting.

  Of course, we are not the first to make this call. There is a vibrant, ongoing discussion in research and policy circles over ways to encourage standards for national election data. Stakeholders as diverse as the National Academy of Sciences, in its recent report on voter registration database,[37] and Google, through its Voting Information Project,[38] are pioneering new models for data standards across the elections ecosphere. The National Institute of Standards and Technology (NIST) held a "Workshop of a Common Data Format for Electronic Voting Systems" in late 2009.[39] In his whitepaper, Paul Lux (Okaloosa County, FL) captures a central problem faced by researchers:

  > Of course each of the four major vendors of election tabulation software (identified by this author as ES&S, Hart InterCivic, Premier, and Sequoia) has their own individual sizes and data formats for each of these [election results] fields. Some limit certain fields to numeric characters, while others may allow alpha characters or even free form fields. When this happens, sharing such important information at any central level becomes problematic.[40]

  In Appendix E, we provide a list of reports and websites that provide rationales and tools for data standardization.

---

[37] Committee on State Voter Registration Databases. Improving State Voter Registration Databases. Washington, D.C.: National Academy of Sciences, 2010. Available at <http://www.nap.edu/catalog.php?record_id=12788>.
[38] Voting Information Project: <http://www.votinginfoproject.org>.
[39] See NIST's website for information and papers: <http://www.nist.gov/itl/vote/cdf-workshop-papers.cfm>.
[40] Paul Lux, "Data Normalization in Electronic Voting Systems: A County Perspective," presented at the NIST Workshop of a Common Data Format for Electronic Voting Systems, October 29-30, 2009, Gaithersburg, MD. Last accessed Jun 10, 2010, <http://www.nist.gov/itl/vote/upload/lux-white-paper-okaloosa.pdf>.

- **The need for consistent legal standards**
  Relatedly, this project was severely handicapped by the rapid changes made by the Florida legislature in their election code. The 2006 requirements were removed at the urging of local election officials, yet we discovered that most of the requirements of the law would have been easily met and that all election machines in the state were capable of producing the needed reports.

- **The need for partnerships between state and local elections officials and elections analysts**
  Much as we may be unhappy with the end to the 2006 mandate, it remains the case that fruitful partnerships between elections officials and scholars can provide benefits to both. At a minimum, such partnerships are critical because obtaining the right data means contacting the right person. We found that, in many cases, this was not the local election official but a key individual in their technical or computer support group.

- **The need for geospatial conformity between census and voting districts**
  Reconciling the Census Bureau's vast repository of demographics with individual voting districts would streamline the analysis of electoral performance, and the diagnosis of problems. This highly localized data would enable researchers and stakeholders to control for various non-administrative factors. Indeed, this is a long-desired goal for election officials, state legislators, political organizations, and elections analysts.

- **The value of metrics for improving election performance**
  As our simple example showed, residual vote rate outliers can alert election administrators to places where greater attention to voter assistance can help reduce the problem of "lost votes" within counties.

- **The value of data sharing**
  This project began with a promise to publicly disseminate our data as well as a guide for future elections analysts. We hope future projects take a similar stance.

Data Analysis

Introduction

Our project goal was to produce a more comprehensive and disaggregated analysis of residual votes than has been conducted previously. Extending past considerations of the impact of varied election technologies, we wanted to compare residual vote rates across precincts and important subgroups of the electorate. As noted in our introduction, the rapid growth of new and varied technologies and methods of casting a ballot may cause a reversal of the slow decline in residual vote rates that has been evident since the passage of HAVA and the elimination of punch card, lever, and other outdated election technology.

Due to the limitations of the data that we describe above, our analyses include data for only 51 of the 67 counties in Florida. Our data come from one election––the presidential preference primary in January 2008, which included three major contests: the Democratic and Republican presidential primaries, and a constitutional amendment question related to property tax.[41] Each voter was given the option to vote in one of the two partisan primaries and on the constitutional amendment. Ultimately our analysis is only suggestive of the kinds of questions that could be answered, assuming valid and reliable data were available. In all analyses to follow, we dropped very small precincts—those with fewer than 50 reported ballots—so as to avoid the potential for misleadingly high residual vote values.

Basic statistics

For the statewide constitutional amendment, we calculated an overall residual vote rate of 2.62%. Under-votes were registered on 2.60% of the ballots while over-votes were registered on .02% of the ballots. These rates varied dramatically by county. Miami-Dade had the highest rate of under-votes (6.0%), while Union had just 0.83%. Gilchrist had a notably high rate of over-votes (1.93%), while many counties reported none at all.[42]

---

[41] In some counties, there were additional local contests (such as school board), which are included in our dataset. However, as discussed earlier, these low-interest contests are of less use when calculating residual vote rates, since so many people voluntarily abstain.

[42] See Appendix B for the full table.

Because voting machines can give instant feedback, allowing voters to correct errors, we generally expect absentee voters to have higher rates of error than in-person voters. Indeed, this is the case in our data, as shown below: absentee ballots show residual vote rates of twice or more those of early in-person and polling place voters, comparing most closely with the error rates in provisional ballots.

Residual rates by voting mode
(statewide constitutional amendment)

	Under-votes	Over-votes	Combined
Polling Place	2.11%	0.01%	2.11%[43]
Early Voting	1.85%	0.00%	1.85%
Absentee	6.25%	0.08%	6.33%
Provisional	5.52%	0.00%	5.52%

As noted in the introduction to this report, a great advantage of the Florida data was that many counties reported the under- and over-vote rates not only by mode, but also by the type of voting technology. The residual vote data broken down by technology type supports indicates a clear difference between precinct-counted ballots, and those processed centrally. DREs (direct recording electronic machines) generally do not allow over-votes at all (hence the 0% rate below) and had an under-vote rate of 2.37%. Centrally-counted optical scan absentee ballots, on the other hand, display a residual rate of 6.33%.

[43] The combined rate is 2.11% after rounding.

Residual rates by voting technology
(statewide constitutional amendment)

	Under-votes	Over-votes	Combined
DRE	2.37%	0.00%	2.37%
Optical Scan/Central Count	6.25%	0.08%	6.33%
Optical Scan/Precinct Count	1.83%	0.01%	1.84%

The major contests on the ballot in this election were the presidential primaries for both the Republican and Democratic parties. There are some notable differences in the residual rates between the parties; the following table shows that in these races, Democratic voters were slightly more likely to under- or over-vote than Republicans.

Residual rates by mode
(Presidential party primaries)

	Under-votes		Over-votes		Combined	
	Dem.	**Rep.**	**Dem.**	**Rep.**	**Dem.**	**Rep.**
Polling Place	2.11%	1.48%	0.01%	0.00%	2.12%	1.48%
Early Voting	1.60%	1.33%	0.00%	0.00%	1.61%	1.33%
Absentee	1.49%	1.05%	0.16%	0.17%	1.65%	1.22%
Provisional	1.56%	1.61%	0.00%	0.00%	1.56%	1.61%

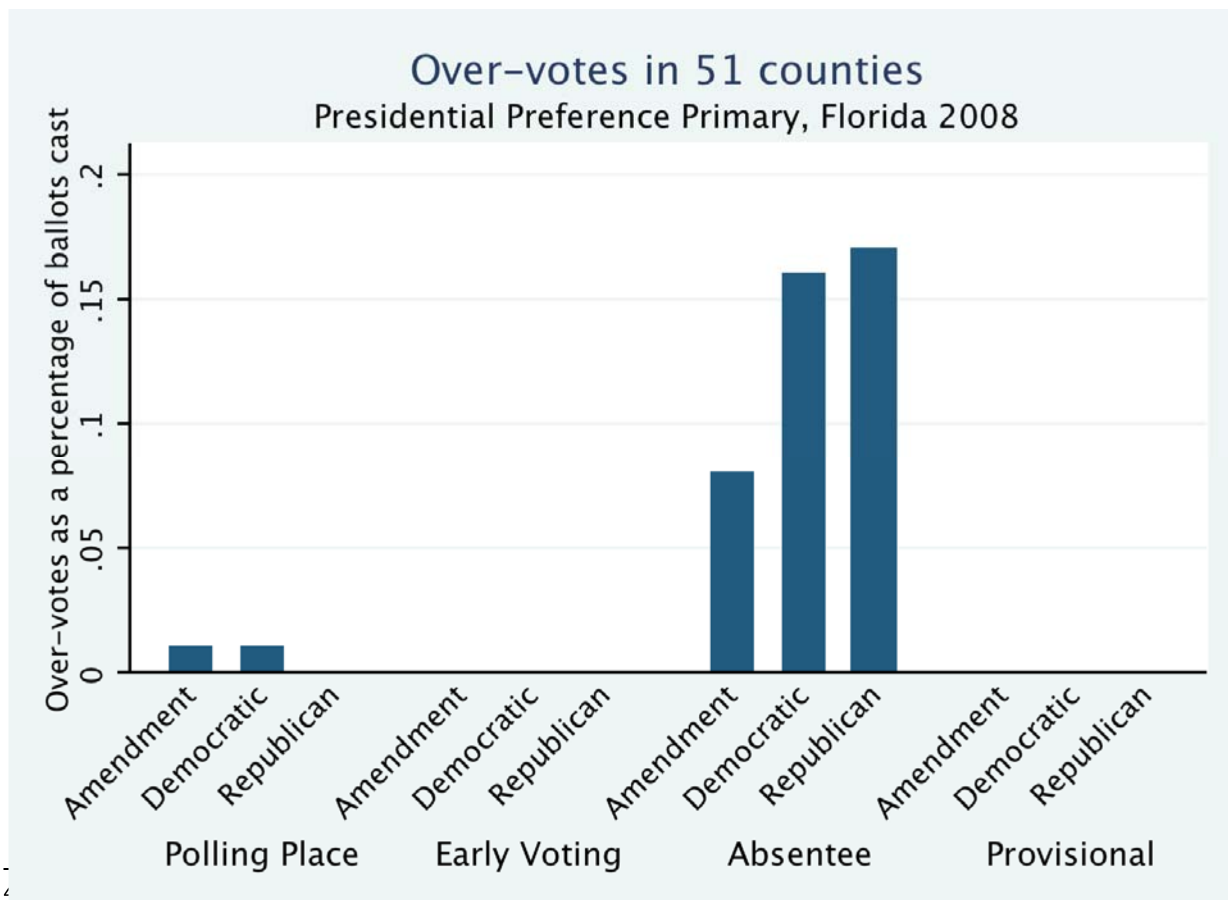
Democrats produced more residual votes than Republicans in every voting mode used. For example, at the polling place, Democratic ballots had a 2.11% under-vote rate, compared with 1.48% of Republican ballots; similarly, absentee voters had a combined residual vote rate of 1.49% in the Democratic primary, but just 1.05% in the Republican race. Provisional ballots for the Republican primary show a slightly higher residual vote rate (1.61% versus a Democratic 1.56%).

During the election, Democrats were engaged in a controversy over Florida's timing of the primary. National party leaders declared that Florida's delegates would not be seated at the convention, and that the state would lose its votes in the national party primary. While we cannot know for certain, we suspect that these differences are due to some voters skipping the Democratic contest.

Over-votes

Unlike under-votes, a large proportion of which are due to ballot roll-off, over-votes are a comparatively pure measurement of voter error. Unless a voter was intentionally trying to spoil his or her ballot, an over-vote represents a mistake on a particular contest. Past research has shown a consistent relationship between the frequency of over-votes and specific voting technology.[44]

The first thing to note is that over-voting is extremely rare. There is, however, a clear disparity between absentee balloting and other modes. As Figure 3.1 shows, while the over-votes for absentee ballots vary between 0.08% and 0.17%, the other modes all



44 Stephen Ansolabehere and Charles Stewart III, "Residual votes attributable to technology," *Journal of Politics* 67 no. 2 (2005): 365–389; Charles Stewart III, "Residual vote in the 2004 election," *Election Law Journal* 5 no. 2 (2006): 158–169; Caltech/MIT Voting Technology Project, "Residual Votes Attributable to Technology: An Assessment of the Reliability of Existing Voting Equipment," version 2, March 30, 2001. URL: <http://vote.caltech.edu/drupal/files/report/residual_votes_attributable_to_tech.pdf> last accessed March 12, 2010.

display negligible rates.

Figure 3.1: Over-votes in Florida's presidential primary, 2008

The lack of real-time feedback on absentee ballots probably causes the high over-vote rate. When voters use touchscreen machines during early or precinct-place voting, the software does not allow over-votes, preventing this type of error (or, at least, it may force the voter to acknowledge the over-vote).

Where in-person voting is conducted on paper, ballots are usually fed into an optical scan machine immediately as the voters submit them. The software notes an over-vote as an error and (depending on how it is programmed) will often reject the ballot, providing the voter an opportunity to make a correction. However, since absentee ballots are generally delivered by mail to the elections office, and counted in a central location, there is no mechanism to alert a voter to an over-vote.

The implications of this finding range more broadly than just absentee ballots cast by mail. Any voting technology that does not provide the voter immediate feedback about voting errors is prone to a higher residual vote rate. At the same time, even the most flawed technology (such as the much-maligned punchcard systems) could have low residual vote rates if the voter was able to validate the ballot.

A few lessons are evident given our current use of voting technology:

- *Any central count system holds the potential for substantially higher over-vote rates*. For example, in California, citizens can show up at county offices to cast an absentee ballot "early in-person" rather than by mail (most counties set up small privacy booths where citizens can fill out the ballot). However, if these ballots are not immediately run through an optical scanner, then "early voting" in California will display higher levels of over-votes.
- *All "vote-by-mail" systems (absentee, no-excuse absentee, and pure vote-by-mail) are prone to comparatively high over-vote rates.*
- *Whether "vote-by-mail" systems are prone to comparatively high* under-*vote rates has not been established.* The evidence here that compares voting in the highly salient presidential primary to the less salient constitutional amendment shows that mail-in ballots can actually have lower under-vote rates, but this is likely due to the

motivations of absentee voters who overcome limits of technology. This is a point that, at the very least, requires further research.

- *Technology matters*, but technological problems can be overcome by providing a system to validate the accuracy of the ballot.

Demographic patterns

A key goal of this project was to demonstrate the value of precinct-level data for residual vote analysis. In contrast to county data, precinct-level data covers geographic units that are far more homogeneous with respect to a wide variety of political and demographic data. This means that we can be much more confident about the relationship between characteristics of the electorate and residual vote rate.

As an illustration, refer to Figure 3.2. In this chart, we show a scatterplot of the residual vote rate in the constitutional amendment against the percentage of African-American voters in each precinct. This chart could be produced at the county level in Florida, but with just 67 counties, and much lower levels of variation in proportion of African-American population, such a chart would be meaningless.

Figure 3.2 shows a slight but statistically significant relationship between higher concentrations of African-Americans and higher residual vote rates. Figure 3.3 shows a similar relationship for Hispanic population. In this case, we see an even more marked rise in the residual vote rates as the minority population predominates in a precinct. Particularly in precincts whose voters are more than 80% Hispanic, there are a substantial number showing residual vote rates exceeding 10%, a level high enough to flag scrutiny by local elections officials.[45]

An additional note: an examination of our dataset confirms that the precincts showing high residual vote rates but low levels of Hispanic population are actually those high-population African-American precincts in figure 3.2 (and vice-versa). Discerning these patterns would not be possible without precinct-level data.

Our final two charts in this section further illustrate the unique leverage provided by precinct-level data. Figure 3.4 shows the residual vote rates in the Republican and

45 All interpolated lines on the following graphs were estimated via lowess regression.

Democratic presidential primary contests plotted against African-American population in a precinct. Figure 3.5 shows the same for Hispanics.

The patterns are broadly in line with our expectations. In the Democratic contest, residual vote rates actually *drop* as African-American populations increase. But African-American voters are reliably Democratic, and had a high-salience black candidate to support (Barack Obama). Obama notwithstanding, the Democratic ballot was unusual because of the DNC's threat to disregard the results; the slightly wider spread of Democratic residual vote rates in low-Hispanic areas is likely a reflection of this. The Republican residual vote rates, on the other hand, appear to have a more even spread across precincts.

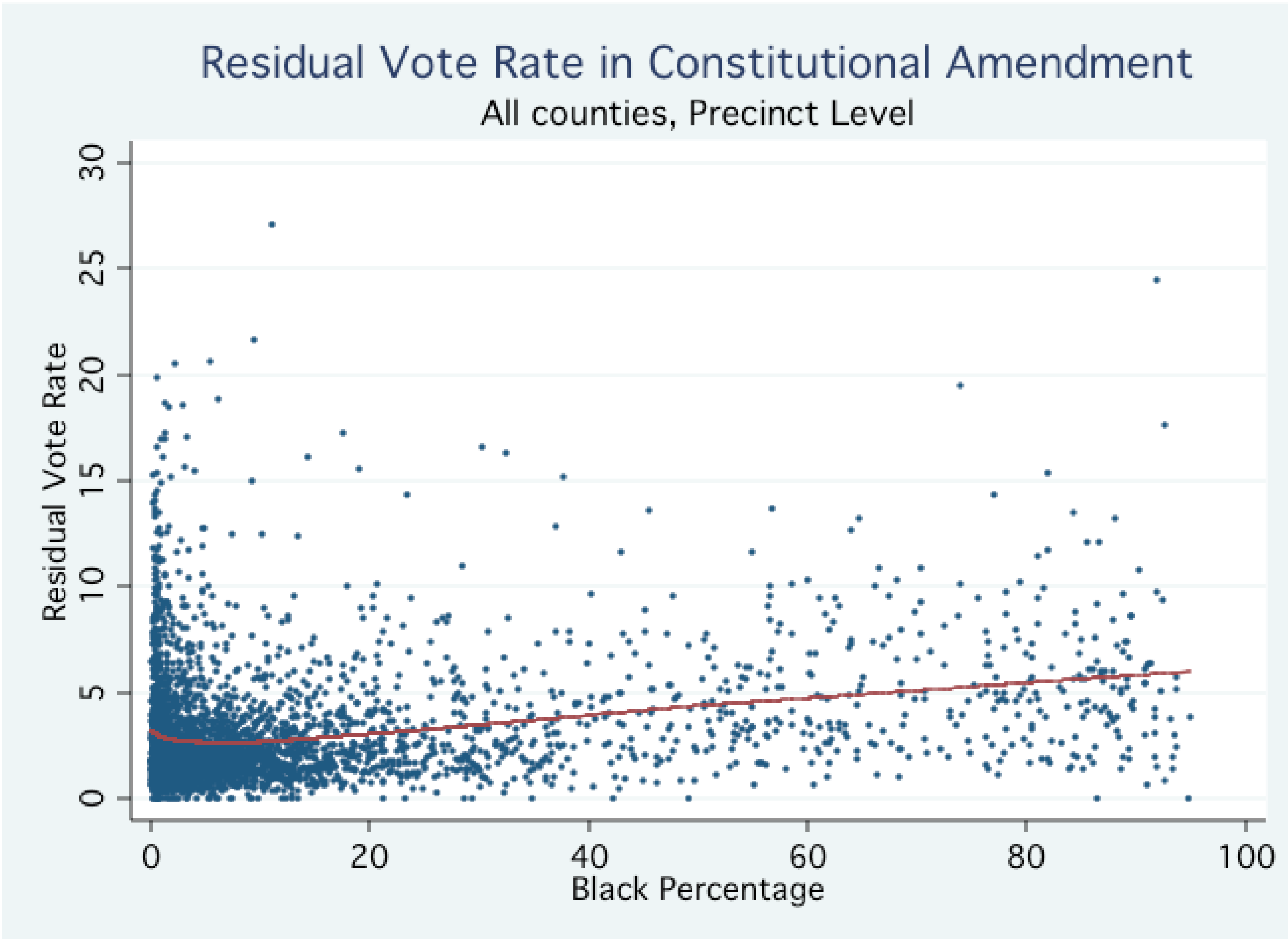


Figure 3.2: Constitutional Amendment by Level of ***African-American*** *Population*

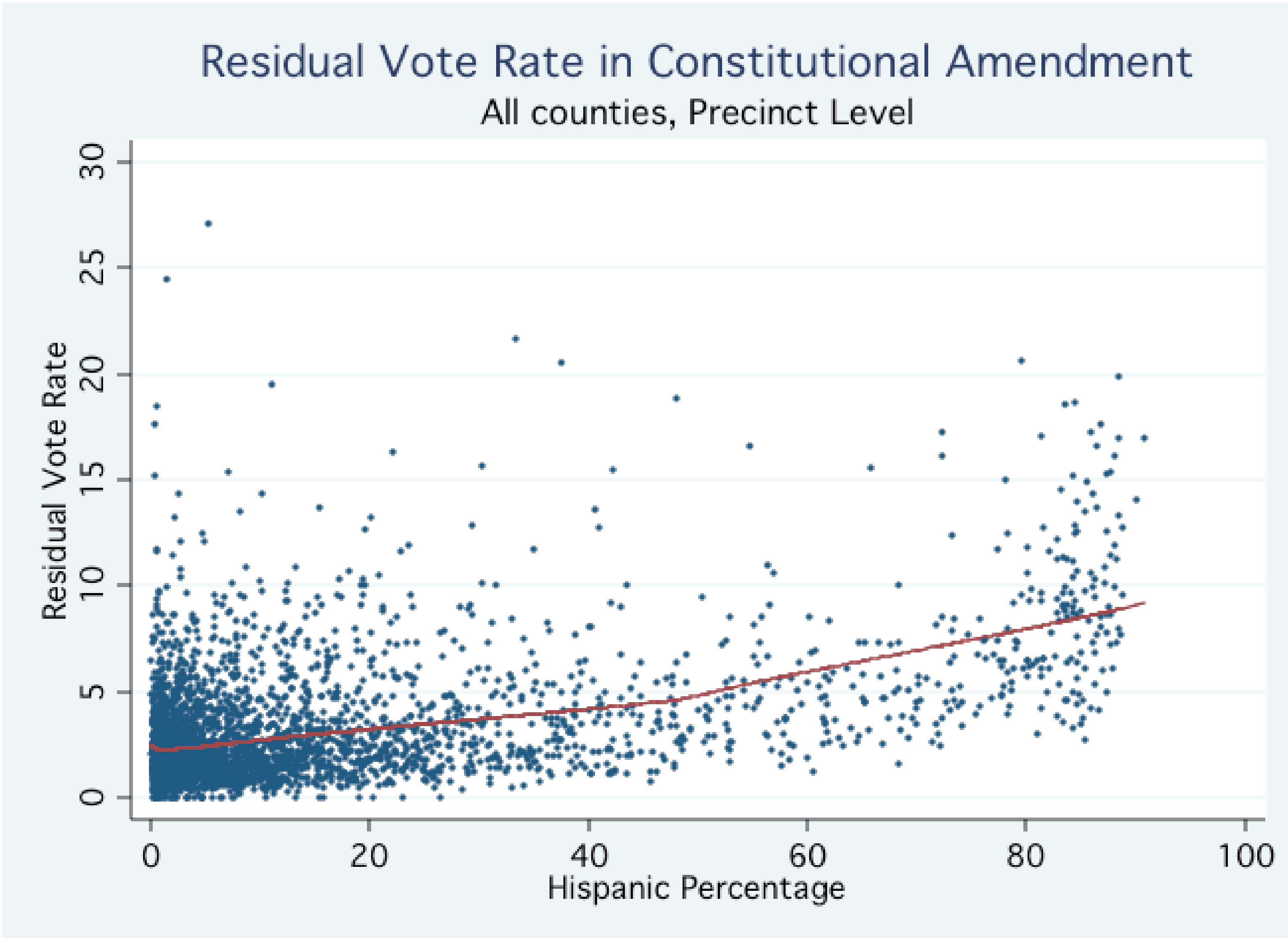


Figure 3.3: Constitutional Amendment by Level of ***Hispanic*** *Population*

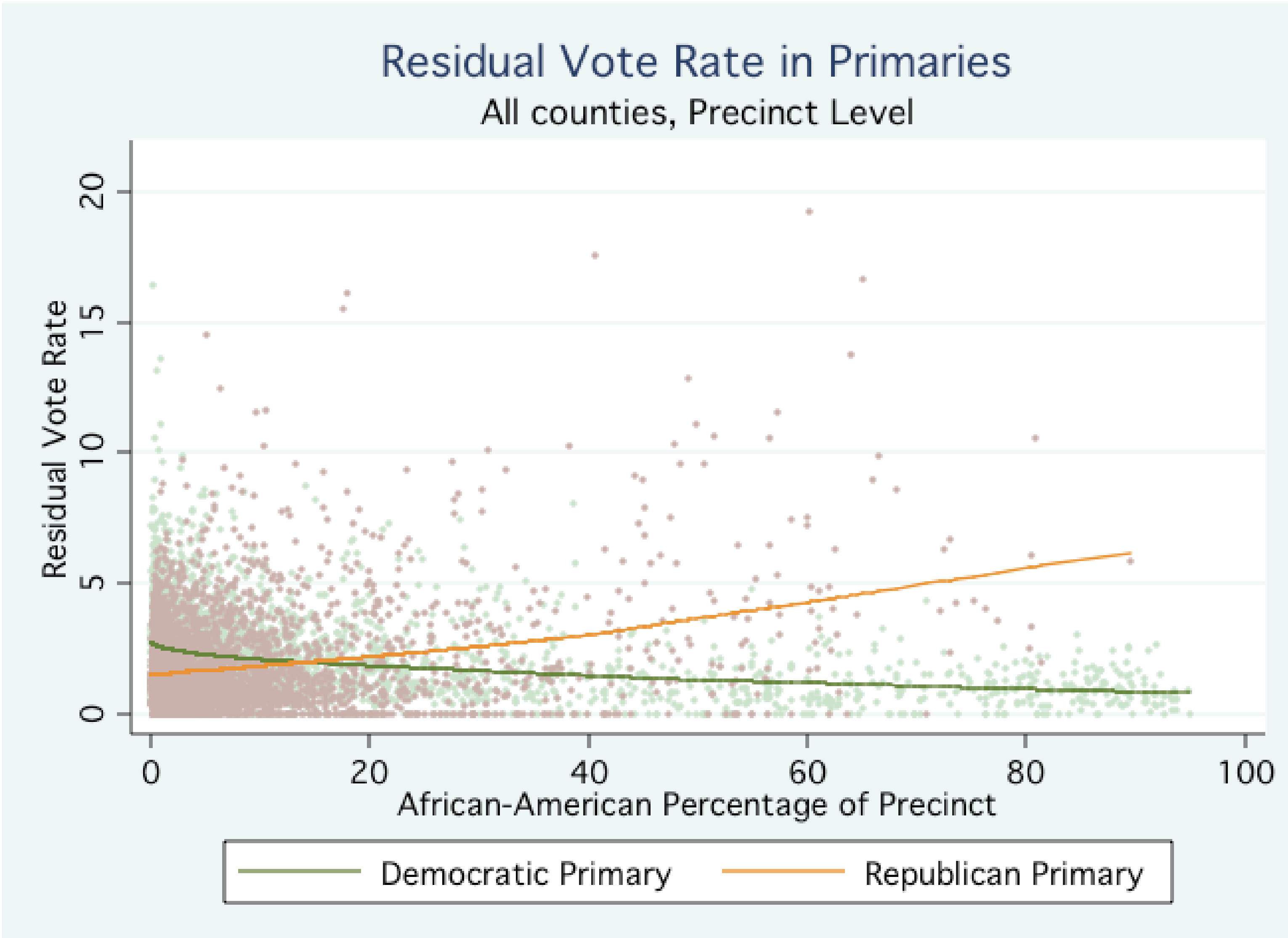


Figure 3.4: Presidential Primaries by Level of ***African-American*** *Population*

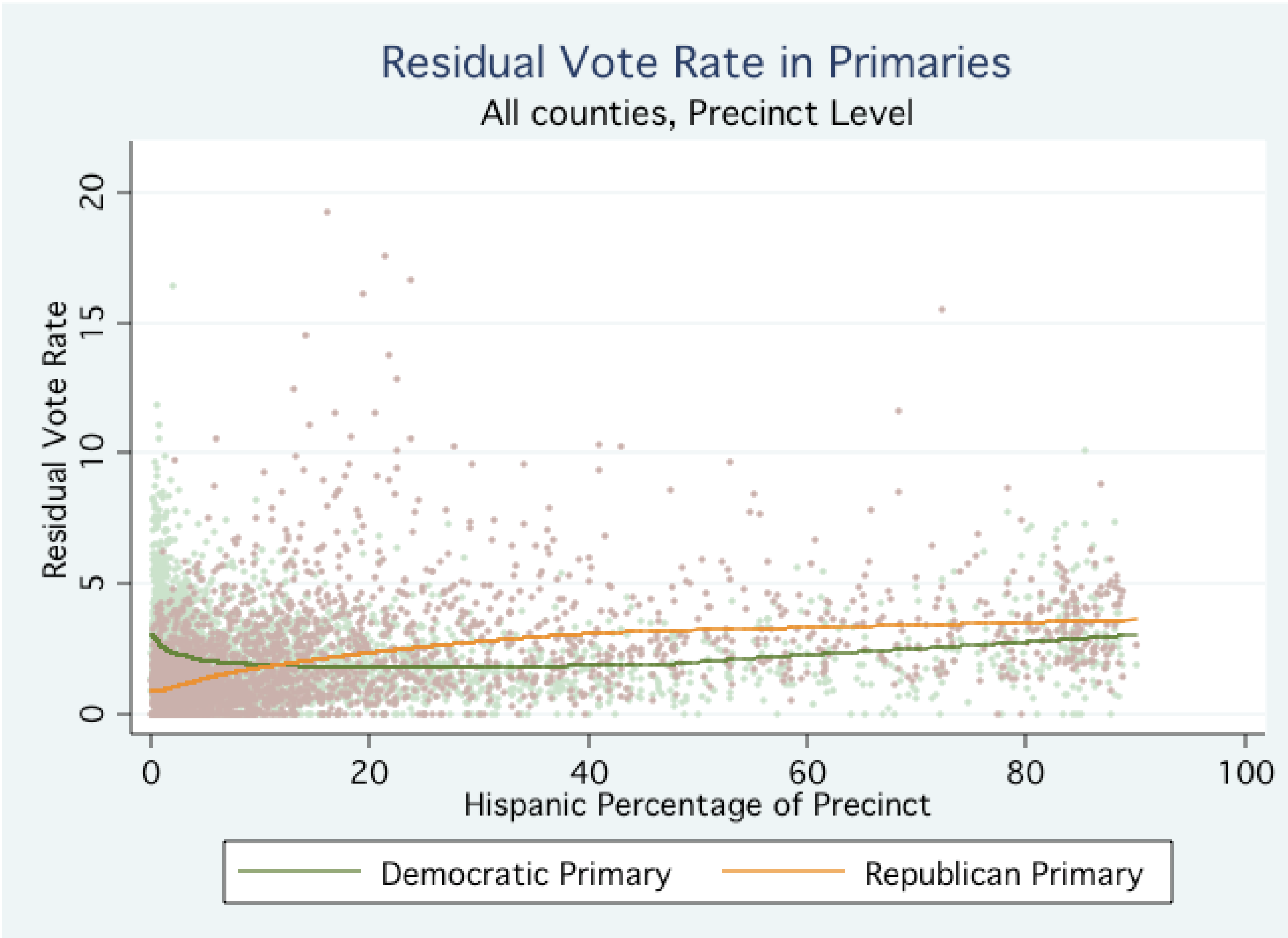


Figure 3.5: Presidential Primaries by Level of ***Hispanic*** *Population*

Regression analysis

Finally, to tie some of these demographic patterns together, we present in Table 3.1 the results of a linear regression analysis, focused on the three contests discussed so far. The dependent variable is the same as in the plots above: the residual vote rate in each precinct, expressed as a percentage.[46] The independent variables are several demographic characteristics of the precincts.[47]

As the results indicate, there are clear and significant changes in residual vote rates amongst different groups. Precincts with many senior citizens saw higher residual vote rates, particularly on the constitutional amendment. On the other hand, that same contest had lower residual vote rates in precincts with more registered Republicans.

In the Democratic primary, an increasingly African-American population is associated with a significant reduction in a precinct's residual vote rate: for each 10 percentage-point increase in the African-American population, our model predicts a 0.17 percentage-point decline in the residual vote rate. At the same time, African-Americans had a higher residual vote rate, in general, on the constitutional amendment.

Unfortunately, due to the limitations of voter registration files, the analysis suffers from a lack of information about income and education. Undoubtedly, some of our variables are standing as proxies for these hidden characteristics, and we would certainly expect both these factors to effect change in residual vote rates. We hope to include them in future analyses.

---

[46] The regression included a dummy variable for each county, which – Miami-Dade excepted – we have excluded from display for brevity. Given the high density of Hispanic precincts in Miami-Dade, we have included it here, along with an interaction term.

[47] To interpret the tables, consider that for each percentage-point increase in the relevant variable, the residual vote rate is expected to increase by the amount of the coefficient (also expressed as a percentage point).

Table 3.1: Multivariate Regression Results (Residual Vote Rate in Each Precinct)

	Constitutional Amendment		**Democratic Primary**		**Republican Primary**	
	Coef.	*t*	Coef.	*T*	**Coef.**	*t*
Percentage Hispanic	**0.018**	*3.2*	-0.005	*-1.4*	**0.037**	*7.2*
Miami-Dade Dummy Variable	-0.355	*-1.3*	**0.505**	*2.6*	**4.467**	*15.8*
Miami-Dade Dummy Variable × Percentage Hispanic	**0.049**	*7.2*	**0.011**	*2.8*	**-0.052**	*-8.4*
Percentage African-American	**0.031**	*11.69*	**-0.017**	*-14.1*	**0.069**	*31.3*
Percentage Over 60	**0.044**	*18.88*	**0.007**	*4.6*	**0.026**	*12.2*
Percentage Registered Republican	**-0.042**	*-8.64*	-	-	-	-
Constant	*3.213*	*10.73*	*1.464*	*8.9*	*-1.126*	*-4.9*
	Adj-R = 0.5097 *N = 3700*		*Adj-R = 0.2816* *N = 3735*		*Adj-R = 0.4656* *N = 3513*	

Precinct-level residual vote rates as an administrative tool

Throughout this section, we have focused on the benefits that residual vote rates provide to scholars and advocates who may wish discern to discern patterns of growth, decline, and variation in voting systems performance. However, residual vote rates can also serve as a management tool for election administrators, helping them identify key precincts where some aspect of the elections system is not performing as well as it could.

It is important to realize that residual vote as a performance indicator does not signal *why* a particular precinct may be experiencing a problem, but is can help identify *where* problems are occurring. As we wrote in the introduction, the actual source of over- and under-votes are diverse, only some of which are under the direct control of elections officials.

Below, we present two tables. In the first table, we display the ten "worst" precincts in the two counties with the highest overall residual vote rates––Miami-Dade and Calhoun. In the second, we display the "worst" (highest residual vote rate) precinct for the 51 counties which provided satisfactory data to the state, along with the reported residual vote rate in these precincts. We provide these without much interpretation because we view them mainly as an illustration of how a diagnostic device like the residual vote can be a productive management tool for election administrators.

If, for instance, Calhoun County wanted to attack its residual vote rate, one strategy would be to see if there are "hot spots" in the county and whether there are common characteristics in those precincts. Perhaps there is a flaw in the poll worker training system; for example, there may be a relatively higher proportion of citizens with language difficulties or otherwise citizens who are facing problems with the voting technology. Or it may instead be something completely innocuous. There may be a college or university in the precinct, and large number of students may choose to opt out of local races. Or the precincts may have substantially higher levels of no-excuse absentee balloting, systems where we have already shown that there will be a higher rate of residual vote. See the case study on Lake County as another example of how residual vote rates can aid elections management.

Indeed, if there is one take-away message from this report, this would be it: residual vote rates and precinct level returns are not a weapon by which election administrators will be attacked, but are a management tool with which elections performance can improve.

How Residual Vote Rates Can Help LEO's Target Policy Solutions

Lake County, Florida has one of the lower residual rates in the state (1.76% in the statewide amendment). Yet, as with every county, Lake has pockets of high residual votes. What might be causing these high rates? Here is a case where precinct level residual vote rates can help administrators diagnose a potential problem with the election and potentially identify a solution.

The highest residual vote rate in Lake County was in precinct 98, which had a rate of 13.48% on the statewide constitutional amendment. A quick investigation showed that 98% of the registered voters in the precinct were 71 or older. Surely the high residual vote rate and the age profile were not coincidental.

A bit of sleuthing reveals that the polling place for precinct 98 is The Lodge at Waterman Village, a senior community in Mt. Dora, FL. The boundaries of the precinct are completely congruent with the boundaries of the retirement community. The second highest residual rate in Lake (8.59%) was in Precinct 77. Like Precinct 98, Precinct 77 is wholly contained within a retirement community, Lake Port Square in Leesburg.

There are other characteristics of Precincts 98 and 77 that help explain the high residual vote rate and that can help election administrators in Lake target possible voting issues among the elderly.

The use of absentee ballots was substantially higher than statewide: 21.9% in P98 and 22.1% in P77 versus 12.9% in state, yet early voting rates were much lower (4.1% in P98, 3.3% in P77, and 15.9% statewide). The residual vote rates in both precincts were relatively high at the polling place, but were exceedingly high on the absentee ballots (a whopping 33.92% in P98).

	Total Ballots	**Polling Place Ballots**	**Residual Rate**	**Absentee Ballots**	**Residual Rate**	**Early In Person Ballots**	**Residual Rate**
Precinct 98	230	156	7.05%	56	33.92%	18	5.5%
Precinct 77	384	286	6.9%	85	15.29%	13	0%

Other precincts that showed comparatively high residual rates in Lake, such as precinct 64 (6.30%), were more typical of problematic precincts statewide, with relatively higher proportions of minority registrants.

Presumably, any policy solutions help minority registrants statewide will help those same populations in Lake County. For precincts 98 and 77, however, the solution to a high residual vote rate may lie elsewhere, in the unique demands of delivering ballots to populations at long-term residential facilities and nursing homes. This is a population, for example, that relies heavily on absentee ballots and is dissuaded from early in person voting (presumably because of mobility issues). This is a population that seems to have some problems with absentee ballots, at least in Lake.

There may be a ballot design solution, such as using a larger font size on absentee ballots. Alternatively, election officials may consider mobile in-person early voting stations, which have been used in some states to target residential facilities, or even opening up a satellite early voting station within these two facilities for a limited period of time.

Residual vote rates: Top ten precincts in top two counties
Constitutional Amendment

Miami-Dade

Precinct	Residual Rate
563	21.94%
501	20.80%
336	20.57%
361	20.24%
323	19.85%
332	19.46%
528	18.87%
27	18.61%
542	18.52%
503	17.60%

Calhoun

Precinct	Residual Rate
12	15.36%
13	7.35%
4	6.17%
3	6.08%
6	5.65%
2	5.61%
10	4.90%
1	4.80%
9	4.41%
7	4.10%

Highest rate of residual votes in a precinct, by county

Constitutional Amendment

County	Precinct	Residual Rate
Broward	Z073	27.0%
Marion	2	24.4%
Miami-Dade	542	21.6%
Sarasota	999	18.5%
Hendry	8	17.6%
Leon	9000	16.1%
Calhoun	12	15.4%
Lake	98	13.5%
DeSoto	5	12.1%
Volusia	620	11.7%
Gilchrist	1C	11.6%
Hardee	8	10.9%
Alachua	16	9.7%
Levy	6	9.1%
Okaloosa	48	8.9%
Orange	622	8.9%
Bay	37	8.7%
Manatee	30	8.6%
Indian River	310	8.0%
Monroe	4	7.9%
Lee	25	7.8%
Clay	608	7.1%
Walton	26	6.7%
Okeechobee	16	6.4%
St. Lucie	17	6.4%
Dixie	11	6.1%

County	Precinct	Residual Rate
Holmes	8	6.0%
Jefferson	14	5.9%
Hernando	11	5.7%
Franklin	5	5.7%
Putnam	28	5.7%
Flagler	1	5.5%
Jackson	24	5.4%
Madison	2	5.3%
Highlands	22	5.2%
Sumter	204	5.0%
Taylor	12	4.9%
Baker	6A	4.7%
Gulf	4	4.6%
St. Johns	210	4.6%
Martin	25	4.3%
Hamilton	1	4.1%
Citrus	409	3.1%
Wakulla	6	2.8%
Union	5A	2.6%
Bradford	7	2.3%
Glades	10	2.2%
Nassau	404	2.2%
Glades	3	2.4%
Nassau	503	2.4%

Appendix A: Data Codebook

county	Standard three-character abbreviation

```
        type: string (str3)

    unique values: 50            missing "": 0/67021

      examples: "CAL"
                "DAD"
                "GAD"
                "MRT"
```

county_full	Full name of county

```
        type: string (str12)

    unique values: 50            missing "": 0/67021

      examples: "Calhoun"
                "Manatee"
                "Miami-Dade"
                "Monroe"
```

precinct_merge	Precinct code (matches state VR file)

```
        type: string (str4)

    unique values: 1885           missing "": 0/67021

      examples: "210"
                "4"
                "587"
```

```
            "917"
```

precinct_code Precinct code (imported from raw results)

```
        type: string (str4)

    unique values: 1886          missing "": 0/67021

      examples: "20"
                "3340"
                "5041"
                "68"
```

precinct_text Precinct description

```
        type: string (str54)

    unique values: 3262          missing "": 1980/67021

      examples: "203"
                "D002"
                "PRECINCT 330"
                "PRECINCT 831"
```

contest_code Standardized contest code

```
        type: numeric (int)
        label: Contests

        range: [1,9011]          units: 1
    unique values: 56           missing .: 0/67021

      examples: 1   Constitutional Amendment (Tax Revision)
```

```
           2   Democratic President
           3   Republican President
           119  Commissioner District 6 Hollywood (BRO)
```

modes Mode as imported from raw data

```
         type: numeric (byte)
        label: Modes

        range: [1,41]            units: 1
unique values: 16                missing .: 0/67021

     examples: 3   Polling Place iVotronic
               12  Early Voting M100
               21  Absentee
               31  Provisional
```

mode Standardized modes

```
         type: numeric (byte)
        label: Mode

        range: [1,5]             units: 1
unique values: 5                 missing .: 0/67021

  tabulation: Freq.  Numeric Label
              18167     1 Polling Place
              16488     2 Early Voting
              15807     3 Absentee
              14004     4 Provisional
               2555     5 Other
```

tech Specific technology model

```
       type: numeric (byte)
      label: tech_models

      range: [1,11]              units: 1
  unique values: 7            missing .: 14087/67021

   tabulation: Freq.  Numeric Label
                20010     1 iVotronic
                 1217     2 Accuvote TSX
                 1371     5 M100
                 6066     7 Accuvote OS
                   24     8 Optech III-P Eagle
                23746    10 M650
                  500    11 Optech 400-C
                14087     .
```

techtype Technology type

```
       type: numeric (byte)
      label: techtype

      range: [1,3]               units: 1
  unique values: 3            missing .: 14087/67021

   tabulation: Freq.  Numeric Label
                21227     1 DRE
                29811     2 Optical Scan/CC
                 1896     3 Optical Scan/PC
                14087     .
```

over Over-votes

```
       type: numeric (byte)
```

```
      range: [0,22]            units: 1
  unique values: 17            missing .: 0/67021

       mean:  .027424
   std. dev:  .268332

 percentiles:    10%    25%    50%    75%    90%
                   0      0      0      0      0
```

under
Under-votes

```
       type: numeric (int)

      range: [0,745]           units: 1
  unique values: 162           missing .: 1200/67021

       mean:  3.25525
   std. dev:  11.4962

 percentiles:    10%    25%    50%    75%    90%
                   0      0      0      2      8
```

blank
Blank votes

```
            type: numeric (int)

      range: [0,252]           units: 1
  unique values: 91            missing .: 51712/67021

       mean:  2.16592
   std. dev:  7.35385
```

```
 percentiles:     10%     25%     50%     75%     90%
                   0       0       0       2       6
```

ballots Total ballots cast

```
        type: numeric (int)

       range: [0,2534]             units: 1
unique values: 1150               missing .: 0/67021

        mean:  100.35
    std. dev:  176.152

 percentiles:     10%     25%     50%     75%     90%
                   0       1      28     115     306
```

Appendix B: Residual Vote Rates in 2008

Residual Vote Rates in Florida Counties (Statewide Contests)

County	Combined	Under	Over
Miami-Dade	6.06%	6.00%	0.06%
Calhoun	5.99%	5.99%	0.00%
Hendry	5.72%	5.71%	0.01%
Monroe	3.70%	3.70%	0.00%
Holmes	3.68%	3.68%	0.00%
Alachua	3.15%	3.15%	0.00%
Madison	3.05%	3.05%	0.00%
Jackson	2.99%	2.99%	0.00%
Hardee	2.93%	2.93%	0.00%
Gilchrist	4.80%	2.86%	1.93%
Hamilton	2.66%	2.66%	0.00%
Franklin	2.63%	2.60%	0.03%
Taylor	2.60%	2.60%	0.00%
Bay	2.59%	2.59%	0.00%
Leon	2.55%	2.55%	0.00%
Dixie	2.50%	2.50%	0.00%
Orange	2.45%	2.44%	0.01%
Okeechobee	2.41%	2.41%	0.00%
DeSoto	2.33%	2.33%	0.00%
Baker	2.27%	2.27%	0.00%
Clay	2.25%	2.25%	0.01%
Broward	2.16%	2.16%	0.00%
Walton	2.11%	2.11%	0.00%
Volusia	2.05%	2.05%	0.00%
Gulf	2.03%	2.03%	0.00%
Levy	2.04%	2.03%	0.01%

Residual Vote Rates in Florida Counties (Statewide Contests)

County	Combined	Under	Over
Lake	2.02%	2.01%	0.00%
St. Johns	2.02%	2.01%	0.01%
Wakulla	2.01%	2.01%	0.00%
Flagler	1.97%	1.97%	0.00%
Marion	1.98%	1.96%	0.02%
Jefferson	1.91%	1.91%	0.00%
Indian River	1.89%	1.89%	0.00%
Okaloosa	1.89%	1.89%	0.00%
Lee	1.68%	1.68%	0.00%
Sarasota	1.65%	1.64%	0.00%
Manatee	1.57%	1.57%	0.00%
Martin	1.56%	1.56%	0.00%
Highlands	1.42%	1.39%	0.02%
Nassau	1.34%	1.34%	0.00%
Putnam	1.33%	1.33%	0.01%
Sumter	1.30%	1.30%	0.00%
Citrus	1.24%	1.24%	0.00%
Hernando	1.22%	1.22%	0.00%
St. Lucie	1.18%	1.18%	0.00%
Bradford	1.11%	1.10%	0.02%
Glades	0.91%	0.91%	0.00%
Union	0.83%	0.81%	0.02%

Appendix C: Bay County EL30 Sample

PREC REPORT-GROUP DETAIL PRESIDENTIAL PREFERENCE PRIMAR

BAY COUNTY, FLORIDA

JANUARY 29, 2008

RUN DATE:02/08/08 12:47 PM STATISTICS

0001 1 HEISLER HALL

	TOTAL VOTES	%	POLL M100	POLL TS
REGISTERED VOTERS - TOTAL		1,783		
REGISTERED VOTERS - REPUBLICAN	670	37.58		
REGISTERED VOTERS - DEMOCRAT	709	39.76		
REGISTERED VOTERS - NONPARTISAN	404	22.66		
BALLOTS CAST - TOTAL.	576		474	1
BALLOTS CAST - REPUBLICAN	291	50.52	244	1
BALLOTS CAST - DEMOCRAT.	221	38.37	176	0
BALLOTS CAST - NONPARTISAN.	64	11.11	54	0
VOTER TURNOUT - TOTAL		32.31		
VOTER TURNOUT - REPUBLICAN.		43.43		
VOTER TURNOUT - DEMOCRAT		31.17		
VOTER TURNOUT - NONPARTISAN		15.84		

********** (REPUBLICAN) **********

PRESIDENT

(Vote for) 1

	TOTAL VOTES	%	POLL M100	POLL TS
Rudy Giuliani	25	8.62	17	0
Mike Huckabee	60	20.69	46	0
Duncan Hunter	0		0	0
Alan Keyes	0		0	0
John McCain.	101	34.83	87	1
Ron Paul.	10	3.45	7	0
Mitt Romney.	90	31.03	85	0

Tom Tancredo	0		0	0
Fred Thompson	4	1.38	1	0
Over Votes	0		0	0
Under Votes	1	.34	1	0

Appendix D: Use of the Residual Vote in Policy Reports, Scholarly Articles, and Legal Pleadings

Policy Reports

Alvarez, R. Michael, et al. 2001. Voting: What Is, What Could Be. Caltech/MIT Voting Technology Project. Accessed at <http://www.vote.caltech.edu/media/documents/july01/July01_VTP_Voting_Report_Entire.pdf>

Brady, Henry E., Justin Buchler, Matthew Jarvis, and John McNulty. 2001. CountingAll the Votes:ThePerformance of Voting Technology in the United States. Berkeley: University of California.

Academic Articles

Alvarez, R. Michael, Stephen Ansolabehere, and Charles Stewart. 2005. Studying elections: Data quality and pitfalls in measuring the effects of voting technologies. *Policy Studies Journal* 33.1: 15-24.

Ansolabehere, Stephen and Charles Stewart. 2005. Residual votes attributable to technology. *Journal of Politics* 67.2: 365-389.

Buchler, Justin, Matthew Jarvis, and John E. McNulty. 2004. Punch card technology and the racial gap in residual votes. *Perspectives of Politics* 2.3: 517-524.

Frisina, Laurin, Michael C. Herron, James Honaker, and Jeffrey B. Lewis. Ballot formats, touchscreens, and undervotes: A study of the 2006 midterm elections in Florida. *Election Law Journal* 7.1: 25-47.

Herron, Michael C. and Jasjeet S. Sekhon. 2005. Black candidates and black votes: Assessing the impact of candidate race on uncounted vote rates. *Journal of Politics* 67.1: 154-177.

Herron, Michael C. and Jasjeet S. Sekhon. 2003. Overvoting and representation: An examination of over-voted presidential ballots in Broward and Miami-Dade counties. *Electoral Studies* 22.1: 21-47.

Kimball, David C., Chris Owens, and Katherine McAndrew. 2001. Who's afraid of an undervote? Paper presented at the annual meeting of the Southern Political Science Association, Atlanta, GA, November 9, 2001. Accessed at <http://www.umsl.edu/~kimballd/southern.pdf>.

Kimball, David C. and Martha Kropf. 2008. Voting technology, ballot measures, and residual votes. *American Politics Research* 36.4: 479-509.

Leib, Jonathan I. and Jason Dittmer. 2002. Florida's residual votes, voting technology, and the 2000 election. *Political Geography* 21.2: 91-98.

Mebane, Walter R. 2004. The wrong man is President! Over-Votes in the 2000 presidential election in Florida. *Perspectives on Politics* 2.3: 525-535

Mebane, Walter R. 2009. Machine errors and undervotes in Florida 2006 revisited. *William and Mary Bill of Rights Journal* 17.

Sinclair, D. E. "Betsy" and R. Michael Alvarez. 2004. Who over-votes, who undervotes, using punchcards?: Evidence from Los Angeles County. *Political Research Quarterly* 57.1: 15-25.

Stewart, Charles. 2006. Residual vote in the 2004 election. *Election Law Journal* 5.2: 158-169.

––. 2006. Changes in the residual vote rates between 2000 and 2004. *Election Law Journal* 5.2: 158–69.

––. 2006. Declaration of Charles Stewart III on excess undervotes cast in Sarasota County, Florida for the 13th congressional district race. Accessed at <http://moritzlaw. osu.edu/electionlaw/litigation/documents/declaration>.

Tokaji, Daniel P. 2005. The paperless chase: Electronic voting and democratic values. *Fordham Law Review* 73.4: 1712-1817.

Tomz, Michael and Robert P. Van Houweling. 2003. How does voting equipment affect the racial gap in voided ballots? *American Journal of Poltical Science* 47.1: 46-60.

Notable Legal Cases

Wexler v. Anderson, 452 F. 3d 1226 (Court of Appeals, 11th Cir. 2006).

Stewart v. Blackwell, 356 F. Supp. 2d 791 (Dist. Court, ND Ohio, Eastern Div. 2005).

Southwest Voter Registration Education v. Shelley, 278 F. Supp. 2d 1131 (Dist. Court, CD California 2003).

Black v. McGuffage, 209 F. Supp. 2d 889 (Dist. Court, ND Illinois, Eastern Div. 2002).

Appendix E: Data Standardization References

Committee on State Voter Registration Databases. *Improving State Voter Registration Databases*. Washington, D.C.: National Academy of Sciences; 2010. Available at: http://www.nap.edu/catalog.php?record_id=12788.

Paul Lux, "Data Normalization in Electronic Voting Systems: A County Perspective," presented at the NIST Workshop of a Common Data Format for Electronic Voting Systems, October 29-30, 2009, National Institute of Standards and Technology, Gaithersburg, MD. Last accessed Jun 10, 2010, <http://www.nist.gov/itl/vote/upload/lux-white-paper-okaloosa.pdf>

Paul Miller, "Common Data Language," presented at the NIST Workshop of a Common Data Format for Electronic Voting Systems, October 29-30, 2009, National Institute of Standards and Technology, Gaithersburg, MD. Last accessed Jun 10, 2010, <http://www.nist.gov/itl/vote/upload/miller-common-data-language.pdf>

John McCarthy, "Election Plumbing Standards: Data Format Requirements for Inter-Operability, Data Publication and Election Auditing," presented at the NIST Workshop of a Common Data Format for Electronic Voting Systems, October 29-30, 2009, National Institute of Standards and Technology, Gaithersburg, MD. Last accessed Jun 10, 2010, <http://www.nist.gov/itl/vote/upload/VVSG-EML-v21JMc.pdf>

John Sebes, "Common Data Formats For Digital Voting Systems," presented at the NIST Workshop of a Common Data Format for Electronic Voting Systems, October 29-30, 2009, National Institute of Standards and Technology, Gaithersburg, MD. Last accessed Jun 10, 2010, <http://www.nist.gov/itl/vote/upload/john-sebes-paper-open-source.pdf>

Websites:

The Voting Information Project (http://votinginfoproject.org) or http://code.google.com/p/election-info-standard/downloads/list