## MIT Open Access Articles

## Losing Votes by Mail

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation: Stewart III, Charles. "Losing Votes by Mail." New York University Journal of Legislation and Public Policy, Symposium Issue - Helping America Vote: The Past, Present, and Future of Election Administration, 13.3 (Fall 2010): 573-601.

As Published: http://www.law.nyu.edu/ecm_dlv3/groups/public/ @nyu_law_website_journals_journal_of_legislation_and_public_policy/documents/ documents/ecm_pro_068045.pdf

Publisher: New York University School of Law
Persistent URL: http://hdl.handle.net/1721.1/71845
Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike 3.0

## Losing Votes by Mail

Charles Stewart III ${ }^{1}$<br>The Massachusetts Institute of Technology

Draft of June 10, 2010

[^0]
## Losing Votes by Mail

## Charles Stewart III

The 2000 election was a wake-up call for America, demonstrating the vulnerability of the democratic process to breakdowns of voting technology, election law, and election administration. It shamed states and the federal government into action, yielding, in its most expansive (and expensive) manifestation, the Help America Vote Act (HAVA) of 2002. ${ }^{1}$ HAVA had many provisions; the one that most `concretely addressed the Florida recount controversy required states to phase out mechanical lever machines and punch card voting. This requirement was underwritten with the authorization of hundreds of millions of federal dollars.

The implementation of HAVA funds yielded equipment upgrades that, in turn, led to the recovery of at least a million votes in the 2004 and 2008 presidential elections - votes that would have otherwise been lost because of the decrepitude of punch card and mechanical lever voting machines. HAVA increased the likelihood that a voter who wakes up on Election Day intending to vote, and then does everything required of him to cast a ballot, will have his vote counted as intended.

HAVA solved one set of problems but failed to address others. In particular, HAVA, which has been so effective in reducing the "lost votes" problem due to voting technology failures, has been less effective in strengthening all the ties that bind a citizen's desire to vote to the successful completion of the act. Technology failures are but one reason why votes are lost. Other reasons, such as registration problems or poor polling place practices, were also addressed by HAVA. However, with the exception of the requirement that states maintain centralized voter registration lists, HAVA only addressed these reasons indirectly.

[^1]In retrospect, the biggest shortcoming of HAVA may have been its virtual lack of attention to voting by mail. Compared to in-person voting, either in traditional precincts on Election Day or in early-voting centers, vote-by-mail is very decentralized. It relies on millions of people who are unschooled in election law and out of sight of election administrators to perform a series of clerical tasks they otherwise rarely encounter. The chain-of-custody of ballots is less exacting. Opportunities to correct mistakes or clarify how to mark the ballot are harder to access. Finally, the technological safeguards mandated by HAVA to guard against unintended over- and under-votes do not exist.

As legislators respond to calls to make voting more convenient, and public officials respond to demands to make elections less costly, voting by mail is becoming more prevalent. This trend naturally raises the question about whether the gains of HAVA, which have cured many of the ills of in-person voting, may be undercut by the inexorable shift to voting by mail.

The answer to this question is mixed and preliminary. It is mixed because the best evidence suggests that the pipeline that moves mail ballots between voters and election officials is very leaky. On the other hand, the rise of voting-by-mail has not caused a precipitous rise in the residual vote rate, despite the lack of technological safeguards against over- and undervoting. The answer is preliminary because the quality of the best evidence we have is highly variable, and reliant on reports from state election officials who have fifty different ways of defining and gathering data about mail-in ballots. Some of the evidence also relies on the recalled memories of voters who may have psychic incentives to blame others (i.e., election administrators) for their failures to vote.

The larger purpose of this paper is not to argue that voting methods that rely on the mail, whether they are mail-in absentee ballots or Oregon's statewide vote-by-mail system, do or must
result in an inordinate number of lost votes. ${ }^{2}$ Rather, the purpose is to show that we should be monitoring the lost-votes problem in the context of voting by mail, and that the current state of post-election data gathering is insufficient to identify where the biggest problems with vote-bymail exist.

The remainder of this article proceeds as follows. First, I frame the problem of lost votes by introducing the notion of a "voting pipeline," which is inspired by the 2001 report of the Caltech/MIT Voting Technology Project (VTP), which articulated a holistic perspective of the lost-votes problem in the context of voting in-person on Election Day. ${ }^{3}$ Second, I apply that metaphor to the vote-by-mail system and demonstrate that the voting pipeline in this context has many more points of weakness. Third, having framed the issue, I discuss the rise of vote-bymail over the past four decades and identify the regions of the country in which the practice has become more prevalent. Fourth, relying on data from a unique public opinion survey, the Survey of the Performance of American Elections (SPAE) and the Election Assistance Commission's 2008 Election Administration and Voting Survey (EAVS), I show that the number of lost votes through the vote-by-mail system in 2008 may have been as large as 7.6 million, or approximately one-in-five individuals who attempted to vote by mail. ${ }^{4}$ These lost votes happened largely because of problems in the distribution system of mail-in ballots. There is little evidence that the vote-by-mail system is prone to excessive residual vote rates, despite the lack of feedback

[^2]mechanisms that are supposed to alert in-precinct voters that they have over- or under-voted their ballots.

I conclude by considering some objections to the analysis I provide, which helps to suggest an agenda for addressing the lost vote problem along the vote-by-mail path. Three points are emphasized. First, addressing the lost-votes problem in the vote-by-mail context depends critically on improving the data gathering and -analysis capacity in the domain of election administration. Second, progress will not be made in addressing the problems identified in this paper without more careful attention to the normative position that voting-by-mail occupies in American elections. Third, the empirical investigation of problems associated with voting-by-mail will be assisted by making sharper distinctions between situations in which voters are allowed to vote by mail, as opposed to required to vote by mail.
I. The Voting Pipeline and Lost Votes, 2000 to 2008

In their 2001 report Voting: What Is/What Could Be?, the Caltech/MIT Voting Technology Project (VTP) argued that we will significantly underestimate the size of the "lost vote" problem if we focus only on voter confusion (illustrated by the butterfly ballot) and equipment malfunctions (illustrated by hanging chad) and fail to grasp the process of voting more holistically. ${ }^{5}$ Figure 1 illustrates this line of thinking, using the metaphor of a pipeline.
[Figure 1 about here]
In thinking about the voting pipeline, it is helpful to imagine a representative voter who wakes up on Election Day intending to vote for her favored candidate for President. "Success" to this voter constitutes having her vote recorded as she intended when the final tally is complete. To achieve success, four steps in a stylized election administration system must be navigated

[^3]successfully. First, the voter must find the polling place, get there, and get to the front of the check-in line. Second, the voter must identify herself and have that identity verified. Third, she must use the equipment and it must function flawlessly. Fourth, the vote must be counted accurately. ${ }^{6}$

Using the 2000 Voting the Registration Supplement (VRS) of the Current Population
Survey, the VTP estimated that almost one million would-be voters had their efforts to vote thwarted by polling place practices (like long lines), and that between 1.5 and 3.0 million wouldbe voters had registration problems that kept them from voting. Relying on statistical analysis of residual votes from 1988 to 2000, they also estimated that between 1.5 and 2.0 million votes were lost because of machine-related problems. The VTP study did not originally identify the problem of accurately counting votes as part of the pipeline, and so failed to estimate the number of lost votes here. ${ }^{7}$

Approximately 105.4 million votes were recorded in 2000. Working backward along the pipeline, we can back-out an estimate of how many people "woke up on Election Day intending

[^4]to vote" in the range of 109.4 to 111.4 million people. All told, between 4.0 and 6.0 million votes were "lost" when one link of the voting chain broke in 2000.

We can update these estimates using subsequent analysis. Later research estimated that approximately one million fewer votes were lost in 2004 because of machine problems, compared to 2000 ; it is reasonable to conclude that an additional half million votes were recovered because of the next generation of upgrades to voting machines between 2004 and 2008. ${ }^{8}$ Fewer respondents mentioned registration problems as a reason for not voting in 2004 and 2008, although the percentage of non-voters who blamed polling place practices was virtually unchanged across the decade. Therefore, if conditions facing voters in 2008 had obtained in 2000, there would have been one million fewer lost votes due to equipment problems and about half million fewer lost votes due to registration problems. ${ }^{9}$ This represents an overall reduction in lost votes in the range of twenty-to-third percent, most of which can be attributed to efforts associated with the implementation of HAVA.

## II. The Vote-by-Mail Pipeline

One flaw with the description of voting pipeline is that for many voters the decision to vote is not made on Election Day, but may be made days ahead of time, when the voter decides whether to vote by mail (usually absentee) or wait until Election Day. If the voter decides to wait until Election Day, then the pipeline illustrated in Figure 1 still applies. However, if the voter decides to use the mail route, a completely new pipeline is involved.

[^5]Figure 2 provides a schematic voting pipeline that includes the possibility of vote-bymail. The first decision is whether to vote in-person or by mail. ${ }^{10}$ If the decision is to vote inperson, then the thinking proceeds as before. If the decision is to vote by mail, then the reasoning is illustrated by the bottom track.
[Figure 2 about here]
With the exception of voters in Oregon, Washington, and places that have permanent absentee voting, the vote-by-mail process begins when the voter requests a ballot. ${ }^{11}$ The first leak in the pipeline can occur if the request is never received. If it is received, then losses can occur at the point of verifying the voter's identity and eligibility to receive a ballot. If there is a registration problem, then the attempt to vote is thwarted at this point. If the ID is validated, then a ballot is sent to the voter to be cast. If the mailed ballot is not received by the voter, there is another pipeline leak. If the ballot is marked by the voter, it is then returned - but it could get lost in the return mail, thus frustrating the attempt to vote. If the ballot is returned for counting, it must again undergo identification verification. If there has been a registration error that was not manifest earlier, then the pipeline loses another vote. The final step is the accurate tabulation of the returned ballot.

Before considering the likelihood that problems would emerge at each step, a few things leap out at us when we compare the top and bottom tracks of Figure 2. First, there are simply more ways to lose votes along the bottom than along the top. It could be argued that this is because the bottom track is described at a greater degree of granularity than the top, but the

[^6]answer to this argument brings us to the second major point: The reason the bottom path has more ways to lose votes is that many steps of the vote-by-mail process have to be accomplished twice. The voter sends two pieces of mail to the central elections office, (1) the request for the absentee and (2) the absentee ballot itself. ${ }^{12}$ A miniscule fraction of the mail is actually lost, but navigating mail channels involves more than simply surviving the U.S. Postal Service, such as the handling and sorting process at both ends of a letter's journey. ${ }^{13}$ We simply do not know how reliable the system is, once a ballot has left the hands of a postal worker.

Below, I estimate how many votes are lost at every step along the way. For the moment, it is sufficient to note that the opportunities to lose votes appear to be greater along the mail route than along the in-person route.

## III. The Rise of Vote-by-Mail since 1972

The logistics of requesting and delivering absentee ballots introduce more opportunities for lost votes. In order to estimate the potential magnitude of the problem, I start by reviewing the rise of voting-by-mail over the past four decades.

Figure 3 shows the percentage of ballots cast by mail in federal elections from 1972 to 2008. The estimates are provided by the VRS and are based on recall in a national survey. ${ }^{14}$ (The Census Bureau did not gather information about voting mode in 1988.) The prevalence of mail-in ballots has been growing exponentially since 1972. The pace of growth has quickened

[^7]over the past decade, as more states have relaxed their "for cause" absentee laws, developed permanent absentee ballot databases, and mandated the use of the mails for an increasing number of voters. Approximately $16 \%$ of all ballots cast for president in 2008 were sent through the mails. ${ }^{15}$

## [Figure 3 about here]

Figure 4 shows the percentage of ballots cast by mail in each state in 2008, estimated using the VRS. ${ }^{16}$ Self-reported mail-in ballots ranged from $1.9 \%$ in West Virginia to $97.5 \%$ in Oregon. The laws and practices of the states have a significant influence over these absentee rates. In the 27 states that allowed "no-excuse" absentee voting by mail in $2008,22 \%$ of ballots were cast by mail, compared to $6.0 \%$ of ballots in the 21 states and the District of Columbia that still require an excuse to vote absentee. ${ }^{17}$ Focusing only on the no-excuse states, the twelve with permanent absentee voting saw $39.8 \%$ of their ballots cast by mail, compared to an $11.2 \%$ mail rate in no-excuse states without permanent absentee voting.
[Figure 4 about here]
Although mail is currently the minority mode for voting, it is growing, and growing faster in some places than others. Over twenty percent of voters use the mails in nine states. Because of concerns about the costs and logistical headaches associated with in-precinct voting, election officials in many parts of the country feel compelled to respond to citizen demands for greater

[^8]convenience by expanding vote-by-mail options. All this raises questions, though, about whether the voting pipeline is becoming more and more fragile for voters in these states.

## IV. An Estimate of Lost Votes in the 2008 Vote-by-Mail System

Estimating the strength of the vote-by-mail pipeline requires that we know how many registered voters attempted to use the mails to vote in 2008 and how many were successful. Estimating the strength of the vote-by-mail system requires us to have solid figures pertaining to all the possible sources of vote loss that were identified in Figure 2. Unfortunately, although election data for the purposes of diagnosing problems with the election system are better than they used to be, they are still a work-in-progress. Therefore, this section takes a first stab at measuring lost votes along the by-mail channel, but these estimates should be taken with a huge grain of salt, and considered illustrative, not definitive.

I rely on two data sources. The first was the EAC's 2008 Election Administration and Voting Survey (EAVS). The second was a national survey conducted by a team of researchers who were associated with the Caltech/MIT Voting Technology Project and funded by the Pew Charitable Trusts, the 2008 Survey of the Performance of American Elections (SPAE). ${ }^{18}$ The EAVS data give us insights into most of the internal processes involving mail ballots. The SPAE data are useful for estimating how many individuals actually requested an absentee ballot.

EAVS was conducted by sending a survey to election officials in each state, who were asked to provide a large amount of data about election administration in each county for the November 2008 election. ${ }^{19}$ Data gathered included quantities like the total number of ballots cast, the number of ballots cast by domestic absentee ballot or by overseas (UOCAVA) ballot,

[^9]the number of precincts in the county, etc. Although the response rate for the 2008 survey was significantly improved over the 2004 and 2004 versions of this survey, the dataset is still incomplete. For instance, three states (Alabama, Massachusetts, and South Carolina) did not provide a breakdown of ballots by in-precinct, absentee, etc. ${ }^{20}$ Furthermore, in six states (Connecticut, D.C., Hawaii, Indiana, Nebraska, and Texas) the sum of all votes cast in the different categories used to report ballots exceeded the total number of ballots cast in the state.

Although the EAVS data have their limitations, they are the best data available for assessing most of the details of election administration. Problems with missing and inconsistent data in the domestic absentee portion of the survey are minimal, so any conclusions we draw should be robust with reference to decisions about how to impute missing data.

With cautions in mind about the quality of the data, I estimate that 27.9 million ballots were cast by mail in 2008, based on the EAVS data, including data that need to be imputed because of missing values. ${ }^{21}$ This is the estimated number of ballots included in the count. How many ballots were transmitted to voters and returned to be counted? This, too, can be estimated using the EAVS dataset.

## Estimating the number of mail-in ballots transmitted to voters

I start by estimating the number of mail-in ballots that were transmitted to voters. EAVS contains a variable for the number of domestic absentee ballots transmitted to voters (Question C1a), which is used as the estimate for forty of the states. Two states reported precisely zero absentee ballots transmitted, Alabama and Washington. To create an estimate for Alabama, I

[^10]multiplied the Census Bureau's estimate of the percentage of ballots casts via the mails times the number of ballots counted for President to create an estimate. Washington, which conducted allmail balloting in every county except one, did not report the number of mail ballots that were transmitted to voters in the 38 counties with mail ballots. Because Washington mailed ballots to all registered voters, except those in Pierce County, my estimate of the number of transmitted ballots is equal to the number of registered voters in the state, excluding Pierce County.

Oregon, which conducts all its elections by mail, did not report the number of transmitted ballots, but did report a small number $(19,782)$ of ballots transmitted to voters through a separate absentee procedure. To estimate the number of transmitted mail ballots in Oregon, I used the number of registered voters in that state, as well.

In the remaining eight states, at least one county failed to report the number of mail ballots transmitted to voters, resulting in an under-reporting in these states. In these states, I imputed the number of transmitted ballots by taking the Census Bureau estimate of the percentage of ballots cast through the mails for the entire state and multiplying this by the number of ballots counted in the county with the missing data. This will produce an underestimate of the number of transmitted ballots in these states, but the error is likely to be small, because the missing counties are few and tend to have small voting populations.

With all these corrections, I estimate that 31.6 million ballots were transmitted from election officials to voters via the mails in 2008.

## Estimating the number of mail ballots returned for counting

Next, I estimate the number of ballots that were returned to election officials for counting. This number is captured in EAVS Question C1b, which was answered in full by thirty-eight states.

For Alabama, Oregon, and Washington - states for which I had to impute the number of transmitted ballots from scratch and that did not report the number of ballots returned for counting - I assumed that the return rate was the same as the return rate for all the states that did report the number of returned ballots, which was $90.8 \%$. I then multiplied this percentage by the number of transmitted ballots that were imputed for these three states to establish the estimate of the number of returned ballots.

Connecticut reported more ballots returned than transmitted. Therefore, I set the number of returned ballots equal to the number transmitted.

The nine remaining states were missing data from at least one county. For these nine states, I calculated the return rate in the counties that had reported a full set of data. Then, for each state, I took the state's overall return rate (for the counties with complete data) and multiplied it by the number of transmitted ballots reported for any county with missing data.

After making all the imputations, I estimate that 28.7 million ballots were returned for counting in 2008.

## Estimating the number of ballots counted

The next step is to estimate the number of ballots counted. The beginning point for this estimation is EAVS Question C4a, which records the number of domestic absentee ballots that were counted.

Forty-two states reported complete data for Question C4a, leaving values for nine states to be imputed. The values for Alabama and Washington were imputed by first calculating the overall rate of counting returned ballots for the states with a full set of data. This rate was $97.4 \%$. I then multiplied this percentage by the imputed number of returned ballots for these two
states to produce the estimates. The number of counted mail ballots for Oregon was set at the reported turnout for the state.

The remaining states had data missing for one or more counties. For each of these states, I calculated the "counting rate" within the state, using the counties that had the requisite data. I then multiplied the state counting rate by the number of returned ballots in the county to fill in the missing values for these counties.

Using this method, I estimate that 27.9 million mail ballots were counted.

## Estimating the number of requested ballots

Finally, I turn to the number of people who actually requested an absentee ballot. This is not an issue probed in the EAVS, but it is an issue examined by the 2008 Study of the Performance of American Elections (SPAE). The SPAE was a nationwide post-election survey of 10,000 registered voters in November 2008 that focused on a series of election administration issues, from the perspective of voters. The survey contained questions that can help to quantify the number of initial mail ballot requests, particularly the number of requests that were unfulfilled.

The SPAE asked registered voters whether they voted in the 2008 November general election. It then asked voters which mode they used to vote - in-person on Election Day, inperson before Election Day, or by mail. Registered voters who reported they did not vote were then asked several follow-up questions. One follow-up was the following:

Sometimes when voters can't get to the polls on Election Day, they vote using an absentee ballot. Please indicate which of the following statements most closely describes why you did not vote absentee in the November 2008 General Election.

Among the response categories was the answer, "I requested an absentee ballot, but it never came," which $6.8 \%$ of non-voters chose. ${ }^{22}$ If we use the best estimate of the number of registered non-voters in 2008, 57.6 million, then the estimated number of requests for absentee ballots that went unfulfilled is $6.8 \% \times 57.6$ million $=3.9$ million. ${ }^{23}$

Above, I estimated that 31.6 million mail ballots were transmitted from election offices to voters who requested them (or who lived in states that had all-mail voting). If we add the 3.9 million estimate of unfulfilled mail ballot requests to the number of estimated mail transmissions, we get an estimate of 35.5 million requests for absentee ballots in 2008.

## Summarizing the calculations

Figure 5 provides a summary of all these calculations. In words:

- 35.5 million requests were made for a mail ballot in 2008 , of which
- 3.9 million requests were unfulfilled, leaving
- 31.6 million mail-ballots transmitted to voters. Of these,
- 2.9 million were not returned for counting, leaving
- 28.7 million mail-ballots returned for counting. Of these,
- 0.8 million mail-ballots were not counted, leaving a total of
- 27.9 million absentee ballots counted in 2008.
[Figure 5 about here]

[^11]If we add together the estimated number of people whose ballot requests were unfulfilled, the number of ballots not returned for counting, and the returned ballots that were not counted, we get a total of 7.6 million voters who left the mail-ballot pipeline somewhere between requesting a ballot and ballots being counted, which amounts to $21 \%$ of all ballot requests.

What should we make of these estimates? The first thing to say is to caution against taking the 7.6 million figure (or the $21 \%$ rate) as a firm measure of the number of potential mail voters whose votes were lost in 2008. If our goal is to calculate the number of lost votes through the mail-ballot route, these estimates are probably too high. The estimated number of people who requested ballots but never received them is especially subject to skepticism, because nonvoters are well known for rationalizing their failure to vote. Because voting is a socially desirable behavior, there are strong psychological pressures prompting non-voters to blame the actions of others, including election administrators, for their failure to vote.

However, even if the estimated number of unfulfilled requests for a mail ballot is off by an order of magnitude - that is, the correct estimate is closer to 390,000 than 3.9 million - the resulting lost vote estimate total is still around 4.1 million, or a lost vote rate of $13 \%$.

Turning our attention to the other estimates, the other "leaks" in the pipeline are not due entirely to errors beyond the voter's control. A ballot that is mailed out but not returned may reflect a voter deciding against casting a vote after all, or deciding to go to the polls on Election Day instead. Still, if we consider all the non-returned ballots to reflect the conscious decision of these voters to abstain after they had gone through the additional effort to obtain an absentee ballot, this implies that the abstention rate among mail-in voters is $12 \%$, an implausibly high rate.

The failure of nearly one million returned mail-in ballots to be counted remains a bit of a mystery, since few states record why absentee ballots are rejected. Among the states that do
record the reasons why absentee ballots are rejected, the top category is the failure to return the ballot before the deadline, but close behind are a variety of reasons involving missing signatures.

Caveats noted, the number of potentially lost ballots through the vote-by-mail channel is significant, especially compared to estimates that focus on in-person voting. The $22 \%$ lost vote rate through the vote-by-mail channel is significantly larger than the overall rate of $4 \%$ estimated for all voters by the VTP in 2000. Even if further research narrows the gap, there is no doubt that the size of the phenomenon demands attention.

## V. Vote-by-Mail and Residual Votes

The analysis thus far has focused on the distribution channel of ballots as the major source of lost mail votes. Ballots can also be lost if the ballot-marking and -counting process fails. This is a peril shared with more traditional in-person voting, which is the context in which this problem has generally been studied. ${ }^{24}$ Analogous problems in the mail ballot domain have gone almost entirely unstudied, even though there are reasons to believe they may be more acute.

Research following the 2000 presidential election demonstrated that votes can be lost when voters are confused by the layout of the ballot or when the technology malfunctions. Voter confusion or machine malfunction can lead to unintended over- and under-votes, both of which are treated as blank ballots for the purposes of counting votes. The problem of over- and undervotes together is sometimes termed the "residual vote" problem. ${ }^{25}$

[^12]HAVA mandated that in-person voting technologies inform voters that they over- or under-vote their ballot, and provide a way to correct these. Mail-in ballots are exempt from this second chance feedback requirement, which raises the possibility that mail ballots will more likely contain unintentional residual votes than in-person ballots. Unfortunately, the research on this topic has been minimal. Research has been minimal, in large part, because very few states break down residual vote rates by voting mode - in-person $v s$. absentee, for instance.

Furthermore, even if residual vote rates were reported for different voting modes, it is unclear what we would make of these comparisons, since voters do not randomly distribute themselves into the in-person and vote-by-mail categories. In states where absentee balloting is primarily used for the convenience of a few voters, mail-in voters tend to be better educated and have higher incomes than in-person voters. ${ }^{26}$ This population tends to produce relatively fewer residual votes. Therefore, the residual vote rate of mail-in ballots could be lower than in-person ballots simply for demographic reasons.

Keeping this caveat in mind, there is no statistical evidence yet that a rise in vote-by-mail has led to an increase in the residual vote rate at the top of the ticket. The simplest way to test this relationship is to calculate the correlation across states of the changes in the vote-by-mail

[^13]rate and the residual vote rate from 2000 to 2008. This correlation is an anemic -.06 , which is no different from random chance.

A more sophisticated way to test whether the rise of vote-by-mail has led to an increase in the residual vote rate is to include a measure of the vote-by-mail rate in a larger regression model that tests for changes in the residual vote rate at a lower degree of aggregation (such as the county), controlling for factors such as changes in voting technology, turnout, and other local conditions. Such analysis has been reported by Ansolabehere and Stewart and could be adapted here. ${ }^{27}$ Unfortunately, we do not have county-by-county vote-by-mail rates for all counties across the country, and so a comprehensive test is currently impossible to conduct. As a first cut, we can augment the Ansolabehere and Stewart models by including the state-level measure of the vote-by-mail rate for each election year as an additional control, and add 2008 election returns to the analysis, in order to gain greater statistical leverage.

The appendix contains the details of the regression analysis. The most important finding is that in states that have seen the biggest increase in vote-by-mail rates over the past decade, the residual vote rates have fallen the most. The size of the effect predicts that a one standard deviation change in the vote-by-mail rate in a state would reduce the residual vote rate of a county by $1.0 \%$ points, on average. Therefore, the initial evidence is that a rise in vote-by-mail has not led to a rash of new residual votes at the top of the ticket.

There is another confounding factor in examining the residual vote rate of mail-in ballots - some voters do not return all the ballot cards, which infrequently occurs when voters vote in person. San Francisco, California provides a cautionary tale on this score from 2008.

Typical of California counties, the ballot in San Francisco in the November 2008 general election had to accommodate numerous races. Not only were there races for U.S. President, U.S.

[^14]House, Superior Court, Bay Area Rapid Transit (BART) directors (for some precincts), County Board of Education (vote for 4), County Board of Supervisors (rank-choice), and Community College Board (vote for 4), there were also twenty-two county-wide propositions and twelve state-wide propositions. To accommodate all these races, the optical scan ballot was distributed across four ballot cards, which were printed on the front and back. Card 1 contained the federal, state, and local offices, printed on the front and back. Card 2 contained the state propositions, with propositions 1A-5 printed on the front and 6-12 printed on the back. Card 3 contained the county measures, with measures $\mathrm{A}-\mathrm{J}$ on the front and $\mathrm{K}-\mathrm{V}$ on the back. Card 4 contained the rank-choice ballot for supervisors.

San Francisco is not unique in having a complicated ballot. What makes it unique is that it publishes precinct-by-precinct statistics that record not only the number of votes for each candidate by election mode (in-person or absentee), but also publishes the number of cards that were physically returned for counting. Table 1 reports the basic statistics.
[Table 1 about here]
Overall, 388,122 first cards were returned to be counted, 209,527 via the in-person route and 178,585 by mail. These numbers are given by the San Francisco Elections Department as the total turnout for the city and county, so these numbers will be used as the denominators in calculating the non-return rate of the other cards. ${ }^{28}$ A very small fraction of Cards 2 and 3 were not returned to be counted in the in-person precincts, $0.13 \%$ and $0.32 \%$, respectively. A much larger fraction, $2.6 \%$ and $3.3 \%$, of Cards 2 and 3 failed to be returned among the absentee

[^15]ballots. Calculating the non-return rates of Card 4 is a little trickier, because races for the Board of Supervisors were not held in every supervisory district in 2008. Thus, the denominator is calculated by adding up the number of Card 1's returned in the precincts in which Supervisor elections were held. This leads to a count of 234,678 voters overall in the supervisory races held that day, 127,027 in person and 107,651 by mail. The in-person non-return rate of Card 4 was $1.4 \%$, compared to $5.5 \%$ for absentee ballots.

Stated another way, although absentee voters amounted to only $46 \%$ of ballots cast in San Francisco in 2008, they accounted for between $77 \%$ and $94 \%$ of the ballot cards that were not returned for counting in 2008.

The failure to return cards to be counted affects the residual vote rate significantly, especially for "down-ballot" races that appear on every card but the first. This is illustrated in Table 2, which calculates the residual vote rate of each race that appeared first on each of the four ballot cards, along with the residual vote rate for Proposition 8 (gay marriage), which drew national attention. Residual vote rates are calculated using two denominators. The first is the total number of voters for that mode, that is, total turnout. The second is the total number of cards returned for that mode.
[Table 2 about here]
The residual vote rate using all voters as the denominator is always higher than the rate using the number of returned cards. Because the number of non-returned cards is so much greater among absentee voters, the two rates diverge most significantly among these voters. Ironically enough, one of the races where the divergence may have been the most significant was Proposition 8, which passed statewide, restricting marriage in California to opposite-sex couples. Had absentee ballots not experienced the disproportionate "unreturned card" problem, San

Francisco County could have contributed another 1,845 net votes against the proposition to the statewide tally. ${ }^{29}$ These additional votes would not have swung the result the other way, but it illustrates the importance of ballot cards not being returned when voters use the mails in determining the outcome of close proposition contests in California. ${ }^{30}$

Not all states have ballots as complicated as California's, but America is known for its long ballot. San Francisco provides insights into how the rise of vote-by-mail may either exacerbate or ameliorate the lost-votes problem that emerges because of the length and complexity of American ballots.

## VII. Conclusions

If twenty percent, or even ten percent, of voters who stood in line on Election Day were turned away, there would be national outrage. Yet, the estimates provided by this paper suggest that the equivalent may be happening, without comment, among voters who seek to cast their ballots by mail.

Three major objections may be lodged against the gist of the analysis in this paper. The first is that it draws conclusions based on poor data. The second is that it treats mail voters as innocent victims of election administration and the postal service. The third is that even though

[^16]the schematic view of voting-by-mail presented in Figure 2 is complicated, it is not complicated enough.

Each of these objections contain elements of truth, although they are ultimately beside the point. The analysis is only a strong as the data backing it up. Close attention to the EAVS reveals important gaps and inconsistencies in this, the best source of nationally comparable data about election administration. While blame in past years could be laid at the feet of missing data, the 2008 study is not shot full of holes, at least as far as voting by mail is concerned. Now, the problems are due to more vexing and difficult-to-solve problems such as inconsistent definitions across states or difficulties in getting all local officials to comply with state officials' request for data. Similarly, the SPAE is based on the recall of voters who may feel pressured to give socially acceptable answers to questions about how they fulfilled their highest civic duty.

Still, even acknowledging the shortcomings of the data, efforts to quantify the general scope of the lost-votes problem in the vote-by-mail context are not called into question. Instead, the data shortcomings argue in favor of redoubling efforts to make the best data we have even better.

The question of whom to blame for lost votes naturally arises in an analysis such as this. Are lost votes fundamentally the fault of voters? Election administrators? The Postal Service? Over-zealous campaign staff? The clichéd, but true, answer is "all of the above." Unfortunately, the very existence of voting by mail is so normatively ambiguous that it is not always clear how to think about assigning responsibility, and therefore it is not always clear how to think about reducing the number of lost votes.

To take just the simplest of reasons for mail votes being lost as an example, consider ballots that are rejected because of a missing signature on the ballot's return envelope. Is this a
lost vote we should worry about? If it is because of the sloppiness or inattention of voters, maybe not. We are often reminded that the reason for having an educated electorate is so that we can guard our rights ourselves. On the other hand, if it is because of confusing instructions that were written in strict compliance with state law, maybe so. Research has consistently shown that election materials often seem designed to confuse voters and produce mistakes. ${ }^{31}$

The question about where to assign responsibility shifts attention away from the voters in situations where they are required to vote by mail. In these cases, it is not possible to shift blame onto voters by arguing that they must assume the risks in return for the added convenience. In these cases, if election officials do not address the added lost-vote perils when vote are cast by mail, they risk greater skepticism among the public.

There is already evidence that such skepticism may be present in the two states that require residents to vote by mail. In the SPAE, respondent were asked, "How confident are you that your vote in the General Election was counted as you intended?" Nationwide, $69 \%$ of respondents answered "very confident." In Oregon, the percentage was $63 \%$, in Washington it was $52 \%$. Interestingly, when asked whether they supported or opposed laws that mandate all votes be cast by mail, $34 \%$ of Oregon voters opposed, as did $47 \%$ of Washington voters. (The nationwide level of opposition was $84 \%$.) Among respondents who opposed vote-by-mail in Oregon and Washington, the percentages answering that they were very confident their ballots were counted as intended were $50 \%$ and $44 \%$, respectively; supporters were very confident at the $62 \%$ and $61 \%$ levels. This relationship between confidence and attitudes about voting by mail did not exist in the rest of the nation.

Finally, although some will object that the pipeline metaphor illustrated in Figure 2 is overly complicated, others will object that it is not complicated enough. Some will object that

[^17]each state handles voting-by-mail differently, and that each state's processes, definitions, and normative standards must be taken into account. Others will note that there are fundamental differences between situations involving "for cause" absentee voting (that is, the relatively small number of absentee ballots because the voter is sick, out of town, etc.), permanent absentee voting, and mandatory vote-by-mail. As suggested above, the normative approach to for-cause absentee voting is quite different from mandatory vote-by-mail. Turning to logistics, the success of the mail channel hinges more critically on the quality of the registration lists when voters are required to vote by mail or when states establish permanent absentee databases.

The first objection, about nationwide heterogeneity in voting by mail, implies that scientific analysis of the vote-by-mail phenomenon is impossible, a position fundamentally at odds with the underlying premises of this article. More challenging to this paper's analysis is the objection that there are three different types of voting by mail that must be analyzed separately. This challenge implies the need for more research, which is an idea I completely endorse. If this additional research helps to quantify better the quality of the voter-registration lists that are used across the country, all the better.


One of the supreme ironies of the butterfly ballot problem in Palm Beach County, Florida in 2000 is that the county Supervisor of Elections chose this unusual ballot layout deliberately to solve a problem she was facing. If the extremely long list of candidates on the ballot had been presented traditionally, in a single column, the font size necessary to present the candidates would have been so small that voters with limited eyesight could not have read the ballot. The butterfly ballot allowed the use of a larger, more readable font. The Supervisor was just trying to be helpful. The result was a disaster.

The drift toward more voting by mail is similarly benign. Busy, harried voters demand convenience. Put-upon election officials demand greater control over the process. Cashstrapped local governments demand greater efficiencies. These are all good reasons to encourage more voters to use the mails. But, what are the hidden costs and what are the tradeoffs? As Palm Beach County demonstrated, significant dangers lurk when we make changes to how we vote before we understand all the ramifications. As voting by mail seems destined to spread even further, it is incumbent upon us to understand better what this means in terms of lost votes.

## Appendix

This appendix reports a fixed effects regression that estimates the degree to which the residual vote rate for president is related to the prevalence of vote-by-mail in a state. The model is adopted from research conducted by Ansolabehere and Stewart, who studied the relationship between voting technology and residual vote rates for the 1988 to 2004 elections. Here, I start with a simple model that pools together observations from 2000, 2004, and 2008. The dependent variable is the residual vote rate in a county for a particular election year. The independent variables are (1) a series of dummy variables that indicate whether a county used a particular type of voting technology (punch card, paper, mechanical lever machine, optical scan, DRE, or mixed) in an election year, zero otherwise; (2) a dummy variable for each election year; and (3) the logarithm of turnout (i.e., votes cast). ${ }^{32}$ Added to this basic setup is a variable that is equal to the percentage of votes cast in the state the county belongs to in the election year in question. (The state is used, rather than the county, because we do not have nationwide estimates of the number of ballots cast by mail in each county.)

The results of the regression are reported in Table A1.
The results that are relevant to this paper are contained in the first line of the table, which reports the coefficient associated with the percentage of mail ballots cast in the state.

Substantively, the effect is relatively small. For instance, for a state that increased its vote-bymail percentage from $20 \%$ to $30 \%$ (from 0.20 to 0.30 ) would experience a drop in the residual vote rate in each county of 0.12 percentage points. However, because the same state rate is applied to each county in a state, this variable is full of measurement error, since it is being used as an indicator of the vote-by-mail rate in each county. Therefore, it is likely that if we could have measured the vote-by-mail rate at the county level, the effect would have been even greater.

[^18]Table A1. Residual vote rate and votes cast by mail, 2000-2008. (Standard errors in parentheses)

| Variable | Coefficient |
| :---: | :---: |
| \% state ballots cast by mail | $\begin{aligned} & \hline-0.012 \\ & (0.001) \end{aligned}$ |
| Voting technology |  |
| Optical scanning | Omitted category |
| Punch cards | $\begin{gathered} 0.0033 \\ (0.0007) \end{gathered}$ |
| Mechanical lever machine | $\begin{aligned} & -0.022 \\ & (0.001) \end{aligned}$ |
| Paper | $\begin{aligned} & -0.017 \\ & (0.005) \end{aligned}$ |
| DRE | $\begin{gathered} -0.0042 \\ (0.0007) \end{gathered}$ |
| Election year |  |
| 2000 | Omitted category |
| 2004 | $\begin{gathered} -0.015 \\ (0.0006) \end{gathered}$ |
| 2008 | $\begin{gathered} -0.017 \\ (0.0008) \end{gathered}$ |
| Log(turnout) | $\begin{gathered} 0.074 \\ (0.002) \end{gathered}$ |
| Intercept | $\begin{gathered} -0.86 \\ (0.03) \\ \hline \end{gathered}$ |
| N | 13,011 |
| $\mathrm{R}^{2}$ | . 89 |
| Fixed effects |  |
| (county) <br> Number of categories | 4,688 |
| $F$-test | $\begin{gathered} \mathrm{F}(4687,8314)= \\ 13,8(p<.0001) \end{gathered}$ |

Figure 1. The voting pipeline and lost votes in 2000


Source: Caltech/MIT Voting Technology Project, Voting: What $I s /$ What Could Be (2001)

Figure 2. The vote-by-mail pipeline.


Figure 3. The rise of voting-by-mail, 1972-2008.


Source: Census Bureau, Current Population Survey, Voting and Registration Supplement, various years.

Figure 4. State vote-by-mail in 2008.


Source: Census Bureau, Current Population study, Voting and Registration Supplement, 2008.

Figure 5. Lost votes from mail-in ballots.


Source: Calculated by author. See text.

Table 1. Return rates of ballot cards, San Francisco, California, November 2008 General Election.

|  | In-person, not returned |  | Absentee, not returned |  | Total, not returned |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | N | Pct. | N | Pct. | N | Pct. |
| Card 1 | - | - | - | - | - | - |
| Card 2 | 279 | 0.13\% | 4,697 | 2.6\% | 4,976 | 1.3\% |
| Card 3 | 674 | 0.32\% | 5,943 | 3.3\% | 6,617 | 1.7\% |
| N | 209,527 |  | 178,585 |  | 388,122 |  |
| Card 4 | 1,727 | 1.4\% | 5,906 | 5.5\% | 7,633 | 3.3\% |
| N | 127,027 |  | 107,651 |  | 234,678 |  |

Source: City and County of San Francisco, Department of Elections, Statement of Vote, URL: http://www.sfgov2.org/ftp/uploadedfiles/elections/ElectionsArchives/2008/november/SOV_0811 04.pdf.

Table 2. Residual vote rate due to non-returned ballot cards, San Francisco, California, November 2008 General Election.

|  | Top office on card | In person residual votes |  |  | Absentee residual votes |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Number | As pct. of cards returned | As pct. of all voters | Number | As pct. of cards returned | As pct. of all voters |
| Card 1 | President | 2,075 | 1.0\% | 1.0\% | 2,241 | 1.3\% | 1.3\% |
| Card 2 | Prop. A1 | 13,760 | 6.6\% | 6.7\% | 9,261 | 5.3\% | 7.8\% |
|  | (Prop. 8) | 5,671 | 2.7\% | 2.8\% | 4,438 | 2.6\% | 5.1\% |
| Card 3 | Meas. A | 14,547 | 7.0\% | 7.3\% | 8,304 | 4.8\% | 8.0\% |
|  | County |  |  |  |  |  |  |
| Card 4 | Bd. | 18,561 | 14.8\% | 16.0\% | 8,520 | 8.4\% | 13.4\% |

Source: City and County of San Francisco, Department of Elections, Statement of Vote, URL: http://www.sfgov2.org/ftp/uploadedfiles/elections/ElectionsArchives/2008/november/SOV_0811 04.pdf.


[^0]:    ${ }^{1}$ Kenan Sahin Distinguished Professor of Political Science, MIT, Cambridge, Massachusetts 02139. Email: cstewart@mit.edu

[^1]:    ${ }^{1} 42$ U.S.C. §15301 et seq.

[^2]:    ${ }^{2}$ The reader should be alerted to the fact that I generally treat as synonymous vote-by-mail and absentee voting. When the distinction is important, I try to make it clear in the text. Almost all absentee voting is by mail, but some absentee ballots are cast in-person in election departments. In states like Oregon that conduct all voting by mail, it is technically a misnomer to refer to this as absentee voting. Overall, though, issues of nomenclature are a distinction without a difference, as far as addressing the first order set of theoretical and empirical issues is concerned.
    ${ }^{3}$ Caltech/MIT Voting Technology Project, Voting: What Is/What Could Be, available at http://vote.caltech.edu/drupal/files/report/voting_what_is_what_could_be.pdf.
    ${ }^{4}$ The EAVS may be accessed at http://www.eac.gov/research/election_administration_and_voting_survey.aspx.. The SPAE may be accessed at http://www.vote.caltech.edu/drupal/node/231.

[^3]:    ${ }^{5}$ Ibid.

[^4]:    ${ }^{6}$ This fourth step, accurate counting of votes, was not included in the original VTP formulation. However, as scholars have continued to explore how to improve elections, the problem of making sure votes are counted accurately has become a great concern, especially among those who are worried about the functioning of "black box" electronic voting machines. A series of international vote-counting controversies in countries like Ukraine, Iran, and Afghanistan have also sensitized many to the reality of vote tampering after ballots have been cast. There is currently no way to estimate the extent of this problem in the U.S., but it is an important potential source of losses in the voting pipeline that deserves to be acknowledged. Efforts to quantify, or at last identify, tampering with vote totals include Mikhail Myagkov, Peter C. Ordeshook, and Dmitri Shakin, The Forensics of Election Fraud (New York: Cambridge University Press); Walter R. Mebane, "Fraud in the 2009 Presidential Election in Iran?" Chance, 23 (March 2010): 6-15; Mebane, Walter R., Jr., "Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections, in The Art and Science of Studying Election Fraud: Detection, Prevention, and Consequences, ed. R. Michael Alvarez, Thad E. Hall, and Susan D. Hyde (Washington, DC: Brookings).
    ${ }^{7}$ In all likelihood, the number of votes lost due to tabulation errors is already included in the number of machinerelated problems. An example is provided by the 2008 election in the Republic of Georgia, in which a car carrying ballots to the central vote-counting center in Tbilisi was involved in an automobile accident that resulted in nearly all the ballot papers being lost. If this had occurred in most American jurisdictions, the fact that voters had checked in at the polls would have established that they had voted, but the aggregate vote result would fail to record an actual vote. To the outside observer, a vote lost because paper ballots have been lost in a car accident before they have been counted is indistinguishable from a vote lost due to an under-voted ballot.

[^5]:    ${ }^{8}$ Charles Stewart III, "Residual Vote in the 2004 Election," Election Law Journal, 5 (2006): 158-169.
    ${ }^{9}$ In 2000, the percentage citing "registration problems" as a reason for not voting was $7.4 \%$. That number fell to $6.6 \%$ in 2004 and $5.5 \%$ in 2008. Those citing polling place practices amounted to $2.8 \%$ of non-voters in 2000 and $2.9 \%$ in 2004 and 2008. Below, I estimate that the total number of non-voting registered voters in 2008 was 57.6 million, compared to 40 million non-voters in 2000. Therefore, the absolute number of votes lost due to polling place practices and registration problems may have risen over the decade, but that is only because the number of non-voters has grown.

[^6]:    ${ }^{10}$ I assume that issues related to early in-person voting may be treated as simply a part of traditional in-precinct voting. This may or may not be reasonable, but it does not affect the line of reasoning that follows.
    ${ }^{11}$ According to the EAVS Report, 15 states maintained permanent absentee ballot databases in 2008. Thirty-seven percent of absentee ballots transmitted in 2008 were off these permanent lists. If we include the transmission of mail ballots from Oregon and Washington in this calculation, the percentage of automatic transmissions begins to approach $50 \%$.

[^7]:    ${ }^{12}$ Of course, the "first piece of mail" could actually be a phone call or a visit to the elections office, but the point remains that the voter has to navigate the elections office twice when it comes to mail-in ballots.
    ${ }^{13}$ The U.S. Postal Service apparently does not release estimates of the number of letters that are never delivered. However, in their most recent statistical report, the USPS did report that the average days-to-delivery of a pre-sorted piece of first-class mail was 2.3 , with 99.9 of mail delivered within ten days. See http://www.usps.com/financials/_pdf/QSR_FY10QT1.pdf.
    ${ }^{14}$ The EAC's EAVS collects reports from states about the actual number of ballots cast, including mode (in-person, civilian absentee, etc.). Although the response rate in 2008 was nearly $100 \%$, in previous years the EAC suffered under significant non-response problems. Because we cannot use EAC data to describe the long-term trend, I rely on the Census Bureau self-reports, which are cross-validated by the EAC data for 2008.

[^8]:    ${ }^{15}$ Based on the EAVS data, it appears that $4 \%$ of the mail-in ballots were from overseas, with the rest being either traditional absentee ballots or by-mail ballots from jurisdictions that mandate it.
    ${ }^{16}$ Inconsistencies across states in reporting statistics about absentee ballots preclude our use of the 2008 EAVS data. Forty-four states provided data about absentee voting that appear to be usable in the EAVS dataset. Focusing on these 44 states, the cross-state correlation between the EAC data, which are based on ballot counts from election officials, and the Census Bureau data, which are based on self-reports from voters, is .97 (weighting each state by total turnout). Overall, the Census Bureau estimate among these 44 states is (on aggregate) about 2.4 percentage points lower than the EAC count.
    ${ }^{17}$ Oregon and Washington are excluded from this analysis. I used the summary of absentee ballot laws produced by electionline.org: http://www.pewcenteronthestates.org/uploadedFiles/e\%20and\%20a\%20voting\%20laws.pdf.

[^9]:    ${ }^{18}$ The 2008 Survey of the Performance of American Elections was generously funded by a grant from the Pew Center of the States through their Make Voting Work initiative. The analysis presented here is solely the responsibility of the author.
    ${ }^{19}$ For New England, the survey asked that the data be broken down by municipality, not county.

[^10]:    ${ }^{20}$ Alabama did provide a report for the number of domestic absentee ballots cast, but otherwise did not provide an estimate about the number of ballots cast in precincts.
    ${ }^{21}$ This estimate exceeds the EAVS report of 25.6 million absentee ballots counted in 2008, for several reasons, which are detailed in the text that follows. The major reason for the deviation from the raw EAVS report is that I have added in the mail-in ballots cast in Oregon and Washington, which were not counted as absentee votes. I also imputed missing data for some counties and states that did not report the necessary statistics to the EAC. These imputations are less critical for the estimates that follow.

[^11]:    ${ }^{22}$ The other response categories were "I had no interest in voting in this election" ( $24.0 \%$ of responses), "It was too late to request an absentee ballot once I thought about it" (11.1\%), I requested an absentee ballot, but it never came (6.8\%), "I wouldn't have been allowed to vote absentee according to my state's election law" (4.1\%), "Requesting an absentee ballot requires too much effort" ( $2.2 \%$ ), "I didn't know how to request an absentee ballot" ( $21.8 \%$ ), and "Other" (9.8\%).
    ${ }^{23}$ The 57.6 million figure is arrived at this way: The 2008 EAVS Report (p. 8) estimates that 190.1 million Americans were registered to vote in 2008. David Leip's Atlas of U.S. Presidential Elections reports that 131.5 million votes were cast for president in 2008. Therefore, 190.1 million -132.5 million $=57.6$ million is the estimated number of registered non-voters in 2008.

[^12]:    ${ }^{24}$ See Caltech/MIT Voting Technology Project, Voting; Stephen Ansolabehere and Charles Stewart III, "Residual Votes Attributable to Technology," Journal of Politics, 67 (2005): 365-389; Stewart, "Residual Votes"; Jonathan N. Wand, Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Michael C. Herron and Henry E. Brady, "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida," American Political Science Association, 95 (2001): 793-810; Douglas W. Jones, "Chad --- From Waste Product to Headline," available at http://www.cs.uiowa.edu/~jones/cards/chad.html.
    ${ }^{25} \mathrm{~A}$ "residual vote" is simply a ballot that is untallied because it is over- or under-voted. The residual vote rate is calculated by dividing the number of residual votes by the number of ballots presented for counting, or alternatively, the number of voters who check-in at the polls. Residual vote analysis is usually done on the race at the top of the ballot, such as president, but in principle it can be applied to any race on a ballot, as the discussion of San Francisco

[^13]:    below demonstrates. For research involving residual votes see the following: Stewart, "Residual Vote"; Ansolabehere and Stewart, "Residual Votes"; Michael C. Herron and Jasjeet S. Sekhon, "Black Candidates and Black Voters: Assessing the Impact of Candidate Race on Uncounted Vote Rates," Journal of Politics, 67 (2005): 154-177; Stephen Ansolabehere, "Voting Machines, Race, and Equal Protection," Election Law Journal, 1 (2004): 61-70; Justin Buchler, Matthew Jarvis, and John E. McNulty, "Punch Card Technology and the Racial Gap in Residual Votes," Perspectives on Politics, 2 (2004): 517-524; Jonathan I. Leib and Jason Dittmer, "Florida's residual votes, voting technology, and the 2000 election," Political Geography, 21 (2002): 91-98; R. Michael Alvarez, Stephan Ansolabehere, and Charles Stewart III, "Studying Elections: Data Quality and Pitfalls in Measuring the Effects of Voting Technologies," The Policy Studies Journal, 33 (2005): 15-24; Donald P. Moynihan and Carol L. Silva, "What Is the Future of Studying Elections? Making the Case for a New Approach," The Policy Studies Journal, 33 (2005): 31-36.
    ${ }^{26}$ For instance, in the SPAE, the average household income of in-person voters was $\$ 65.1$ thousand, compared to $\$ 67.1$ thousand for in-person early voters and $\$ 72.3$ thousand for absentee and mail voters. The average education for in-person voters was 13.7 years, compared to 13.9 for early voters and 14.1 years for mail voters.

[^14]:    ${ }^{27}$ Ansolabehere and Stewart, "Residual Votes"; Stewart "Residual Votes".

[^15]:    ${ }^{28}$ San Francisco lists these numbers as the "number of voters" in their turnout report and in the State of California's Statement of Vote. See http://www.sfgov2.org/index.aspx?page=1670, last accessed June 1, 2010, and http://www.sfgov2.org/index.aspx?page=1670, last accessed June 1, 2010. It is likely that some (small) number of voters failed to return Card 1, while still returning one of the other voting cards. This number is unreported by the San Francisco Election Department, so I will follow their convention and treat the number of returned Card 1's as the number of voters.

[^16]:    ${ }^{29}$ The 1,845 figure is arrived at as follows. First, assume that the non-return rate for absentee Card 2 ballots in San Francisco had been the same as the in-person rate, $0.13 \%$. That would have resulted in an additional 4,465 Card 2's being returned. Second, assume that support for Proposition 8 among these additional cards would have been distributed the same way as other absentee voters - $28.1 \%$ yes, $63.4 \%$ no, $2.5 \%$ under-vote, and $0.05 \%$ over-vote. (The in-person percentages were less favorable to the Proposition - $20.9 \%$ yes, $76.4 \%$ no, $2.7 \%$ under-vote, and $0.5 \%$ over-vote.) This yields 1,253 votes yes and 3,098 votes no, with 112 ballots under-voted and 2 ballots overvoted. The net votes are 3,098 no $-1,253$ yes $=1,845$.
    ${ }^{30}$ As an aside, the residual vote rate is generally higher (regardless of the denominator) for absentee ballots than for in-person voting, except for the County Supervisor races, in which the absentee residual vote rate was actually lower, even when we account for the fact that many more ballot cards were simply not returned. This provides some evidence in favor of the proposition that mail-in ballots may help voters navigate complicated situations, such as the new ranked-choice option in San Francisco.

[^17]:    ${ }^{31}$ Marcia Lausen, Design for Democracy: Ballot and Election Design (Chicago, University of Chicago Press, 2007).

[^18]:    ${ }^{32}$ The omitted categories for the dummy variables are optical scan voting machines and the year 2000.

