TECHNOLOGY AND VOTER INTENT: EVIDENCE FROM THE CALIFORNIA RECALL ELECTION

Thomas S. Dee*

Abstract—Conventional evaluations of voting systems focus on ballots for which no vote can be recorded (that is, "residual" votes). However, recorded votes that misrepresent voter intent are another potentially important, but less easily measured, source of error. I present evidence that a nontrivial number of voters in the recent California recall election mistakenly voted for one of the four candidates positioned next to the two major candidates on the ballot. I also find that punch-card systems significantly increased the frequency of these errors. These results indicate that future assessments of voting technologies should consider their effects on both recorded and residual votes.

I. Introduction

THE turmoil that surrounded the two most recent presidential elections has motivated a broad and ongoing national dialogue about the potentially decisive effects of alternative voting technologies. Most of the resulting research and discussion has focused on an explicit measure of ballot performance: the share of voters for whom no vote could be recorded (that is, the "residual vote" rate).[1] For example, the National Commission on Federal Election Reform, which included Presidents Ford and Carter as honorary cochairs, recently recommended that states establish residual vote rates of 2% or less as benchmarks for voting-system performance. And empirical studies routinely assess the comparative effectiveness of voting systems by analyzing their effects on residual vote rates (for example, Brady et al., 2001; Alvarez et al., 2001).

Any voting system that systematically increased the prevalence of uncounted votes should unquestionably be viewed as an affront to the important and universally shared norms of electoral fairness. However, the understandable focus on reducing residual votes has obscured another potentially important source of voting errors: mistakenly cast votes that are legitimately recorded. The controversy over the unusual "butterfly ballot" used in Florida's Palm Beach County in 2000 provided dramatic evidence that inferior voting technology can promote such errors and even change the outcomes of closely contested elections (Wand et al., 2001). However, there is virtually no evidence on whether the broad types of voting systems used in the United States (such as punch cards, optical scan, touch screens) influence the prevalence of miscast, but correctly recorded, ballots.

The absence of information on this aspect of ballot performance is undoubtedly due to the fundamental difficulties of establishing whether a voter's recorded vote was not for the candidate that they intended. In this study, I present new empirical evidence on these issues by exploiting some unique aspects of California's gubernatorial recall election that occurred on October 7, 2003. Specifically, I present evidence that a surprisingly large number of voters mistakenly voted for one of the four gubernatorial candidates who were assigned ballot positions *next to* the two major candidates (Bustamante and Schwarzenegger).[2] I then examine whether the cross-county variation in recorded support for these four diverse candidates was related to each county's voting technology.

Like conventional evaluations of residual votes, these results suggest that punch cards led to a substantial increase in voting errors. Specifically, I find that "Votomatic" punch-card systems (and, to a lesser extent, "Datavote" punch-card systems) increased the recorded support for these four candidates by at least a third (and, in most cases, much more). These errors could not have influenced the outcome of the California election. However, these results do provide evidence that punch-card systems could distort the outcomes of more closely contested contests by misrepresenting voter intent. And, more generally, these results suggest the need to evaluate voting systems by how accurately a voter's true intent is recorded in addition to the current practice of simply considering whether it is recorded (that is, minimizing residual votes).

II. Voting Technology and Errors

In the United States, the choice of a voting system is largely left to the discretion of local governments, typically counties. There are five broad types of voting systems currently in use: hand-counted paper ballots, punch cards, optical-scan ballots, lever machines, and direct-record electronic (DRE) machines. A survey of election jurisdictions (EDS, 2004) suggested that, for the general election in 2004, less than 1% of registered voters (that is, those in small, rural communities) relied on hand-counted paper ballots. Punch-card and optical-scan systems also rely on paper but they automate the process of counting. Optical-scan systems, which were widely adopted in the 1990s, require voters to indicate their choice by filling in a circle or arrow as in a standardized test. Voters using punch-card systems

Received for publication May 16, 2005. Revision accepted for publication August 3, 2006.

* Department of Economics, Swarthmore College, and NBER.

[1] Residual votes can occur for several reasons. For example, a voter may attempt to vote but fail to make a valid mark ("unintentional undervoting"). Validly marked ballots may also be miscounted. A ballot could also be spoiled when a voter expresses support for more than the allowable number of candidates ("overvoting"). However, voters who simply choose not to cast a ballot in a particular election ("intentional undervoting") are another source of residual votes.

[2] The four "bookend" candidates were John Christopher Burton (I), Cheryl Bly-Chester (R), Lawrence Steven Strauss (D), and George Schwartzman (I). However, it should be noted that two of these candidates (Burton and Schwartzman) may have benefited from name confusion as well as ballot placement. The possibility that voters confused Schwartzman with Schwarzenegger was widely noted (e.g., Kershaw, 2003). And John Burton was the name of the president of the California Senate during this period (Rainey & Hoffman, 2003).

The Review of Economics and Statistics, November 2007, 89(4): 674–683

indicate their choice by perforating a thick paper card. In 2004, a third of registered voters used optical-scan ballots and 19% used punch cards (EDS, 2004).

The remaining two systems rely on machines to record the votes. Voters using mechanical-lever machines indicate their preferences by flipping small levers arranged with a large display of the entire ballot. These machines were first introduced in the late ninteenth century and are no longer manufactured. However, 13% of registered voters used them in 2004 (EDS, 2004). Electronic systems are the newest voting technology and were used by nearly 30% of voters in 2004 (EDS, 2004). Older electronic systems are similar to lever machines but require that voters push a button instead of flipping a lever, while newer systems use touch screens.

After the high-profile difficulties with punch-card systems during the 2000 election, media reports frequently suggested that such systems were more likely to be used in resource-constrained, high-poverty communities (e.g., *The Economist,* 2001). However, the available evidence suggests that this was not actually the case (e.g., Knack & Kropf, 2002). A recent study by Garner and Spolaore (2005) argues that these results can be reconciled by a model where higher-income communities are more likely to adopt a new technology but where the possession of a new technology reduces the likelihood of adopting subsequent innovations. In other words, higher-income communities were early adopters of punch cards when they were introduced in the 1960s. However, as a result, they were less likely to adopt the optical-scan and electronic systems developed during the 1980s and 1990s.[3]

The apparent shortcomings of punch-card systems have been a frequent theme of the extant research literature on voting systems (such as Saltman, 1988; Shocket, Heighberger, & Brown, 1992).[4] In particular, most of the criticism has focused on the widely used Votomatic punch cards. In a Votomatic system, a voter receives a punch card with as many as several hundred prescored and numbered perforations. They insert the punch card into a device that has the ballot (typically several pages) attached to its face. Using a stylus, the voter then punches out the chads that correspond to their choices. Errors can occur when the punch card is not seated correctly, when the stylus is misplaced, when the stylus fails to remove a chad effectively (an undervote), and when a voter perforates too many chads (an overvote). Punch cards can also compound these errors by making it difficult for voters to know when they have made mistakes. More specifically, the Votomatic punch card does not have candidates' names, so to check for alignment errors and overvotes, a voter would have to remove the card from the device and compare the numbers of the perforated chads to the corresponding numbers on the ballot.

By contrast, in Datavote punch-card systems, voters use a stapler-like punching device instead of a stylus and the cards include the candidates' names. These features appear to reduce the prevalence of spoiled ballots relative to Votomatic punch cards (e.g., Brady et al., 2001). However, nationally, this type of punch card has not been widely used. For example, in 2004, only 1.8% of registered voters used a Datavote punch-card system (EDS, 2004). However, over 6% of voters in the California recall election used Datavote punch cards while 43% used Votomatic-style punch cards. The remaining voters used either optical-scan (41%) or touch-screen systems (10%). The empirical results presented here allow for effects unique to each of these four systems.

A small experiment conducted by Shocket et al. (1992) suggests that Votomatic-style punch cards do significantly increase the prevalence of spoiled ballots. Specifically, they randomly assigned voters ($n = 265$) participating in a simulated election for nine city-council positions to groups using punch cards, paper ballots, and a DRE machine. They found that, because of increases in undervotes and overvotes, the share of punch-card voters who voted for the full number of candidates was almost 50% lower than in the two other groups. Analyses of actual election data generate similar results. For example, Alvarez et al. (2001), using data from the 1988–2000 presidential elections (table 1), found that in communities using punch cards, the prevalence of residual votes is at least 50% (in other words, 1 percentage point) higher than in counties using optical-scan systems. Brady et al. (2001) and Knack and Kropf (2003) came to similar conclusions after conducting multiple-regression analyses that controlled for a variety of county-level observables (such as race, income, educational attainment, and political observables).

These results suggest that the voting systems currently in use can have an appreciable effect on the prevalence of uncounted ballots. However, we know relatively little about whether these voting systems also influence the probability that a *counted* ballot accurately reflects a voter's intent. There are a number of reasons to suspect that punch cards increase the likelihood that a voter mistakenly votes for a candidate or position other than the one they intended. For example, voters can mistakenly vote for the wrong candidate if their punch cards are not seated correctly in the holding device. Votomatic voters with unsteady hands or poor eyesight may also be more prone to errors since they need to manipulate a stylus accurately over punch cards with prescored perforations that are concentrated densely.

[3] This may explain some of the heterogeneity in voting systems used in California during the recall election. Auxiliary regressions suggest that the probability of using a punch-card system during the recall election was unrelated to most observed demographic and socioeconomic traits (as well as a measure of per capita income in 1969). However, higher levels of educational attainment were associated with the probability of using a punch-card system.

[4] Another more recent concern is that touch-screen voting systems may be amenable to fraud. A recent study by Card and Moretti (2005) suggests that there have not been "voting irregularities" with touch-screen systems. However, they found that the introduction of these systems did have modest effects on voter turnout.

TABLE 1.—VOTING SYSTEMS BY COUNTY: OCTOBER 7, 2003

Punch Card–Votomatic/Pollstar	Punch Card–Datavote	Optical Scan	Touch Screen
Los Angeles	Alpine	Amador	Alameda
Mendocino	Calaveras	Butte	Plumas
Sacramento	Del Norte	Colusa	Riverside
San Diego	El Dorado	Contra Costa	Shasta
Santa Clara	Glenn	Fresno	
Sierra	Imperial	Humboldt	
Solano	Inyo	Kern	
	Monterey	Kings	
	San Benito	Lake	
	Tehama	Lassen	
	Ventura	Madera	
	Yolo	Marin	
	Yuba	Mariposa	
		Merced	
		Modoc	
		Mono	
		Napa	
		Nevada	
		Orange	
		Placer	
		San Bernardino	
		San Francisco	
		San Joaquin	
		San Luis Obispo	
		San Mateo	
		Santa Barbara	
		Santa Cruz	
		Siskiyou	
		Sonoma	
		Stanislaus	
		Sutter	
		Trinity	
		Tulare	
		Tuolumne	

Source: California Secretary of State (http://www.ss.ca.gov/elections/voting_systems_2003.pdf).

And, as noted above, it is relatively difficult for voters using Votomatic punch cards to be aware that they have made an alignment error since the cards do not have any candidates' names.

A small experiment conducted by Roth (1998) suggests that these concerns may be valid. Roth (1998) provided 32 subjects with instructions to vote for specific candidates and issues, using punch cards under slightly varying conditions and found that approximately 15% of the votes were incorrectly cast.[5] In several cases, these voters mistakenly cast a ballot for the candidate or issue next to the intended one. And, in subsequent surveys, the subjects also indicated that they were often unsure which holes were correct and that they had difficulty verifying their progress.

These provocative results suggest that Votomatic-style punch-card systems may promote miscast (but legitimately recorded) ballots in addition to the spoiled and uncounted ballots that have been the focus of conventional analyses. However, the extent of this potential problem is difficult to assess in most real-world settings because it is difficult to establish credibly whether recorded votes differ from voters' intent.[6] In the next section, I present evidence that the unusual gubernatorial recall election that occurred in California on October 7, 2003, provides a novel and compelling measure for these sorts of voter errors. I then assess the comparative effects of the voting systems used in California on these measures of miscast ballots.

III. Voting Errors and the "Bookend" Candidates

California's recall election presented voters with 135 candidates to replace Governor Grey Davis if he were recalled. The election was almost postponed because of a legal challenge. The American Civil Liberties Union (ACLU) argued that the election would be unfair because the voters in communities using Votomatic-style punch-card systems were significantly more likely to have their votes spoiled or uncounted (Kleffman, 2003). In California, voting systems vary at the county level. And, for the special recall election, each of California's 58 counties used a punch-card, an optical-scan, or a touch-screen system (table 1). The ACLU argued that the election should be postponed until the March primary when the seven counties still using Votomatic-style punch cards (table 1) would have switched to other systems. Studies of the election outcomes indicate that the ACLU's concerns were well founded (Brady, 2003; McDonald, 2003). More specifically, Brady (2003) concludes that at least 176,000 votes on the recall question were lost due to the poor performance of punch cards.[7]

However, as in most discussions of voting systems, the controversy in California did not focus on the issue of recorded votes that misrepresent a voter's intent. The basic motivation for this empirical study is that the gubernatorial election provides a unique and compelling opportunity to measure these sorts of errors and assess the comparative effects of voting systems.[8] More specifically, the two major candidates in the election, Schwarzenegger (R) and Bustamante (D), received over 80% of the 8.6 million votes cast. However, the four candidates who were positioned in front of and behind these two candidates on virtually all the ballots ("bookend" candidates) also did surprisingly well. A maintained assumption of this study is that a substantial number of the votes cast for these four candidates actually reflected voter error.

The popular discussion of mistaken votes for these candidates focused almost exclusively on George Schwartzman

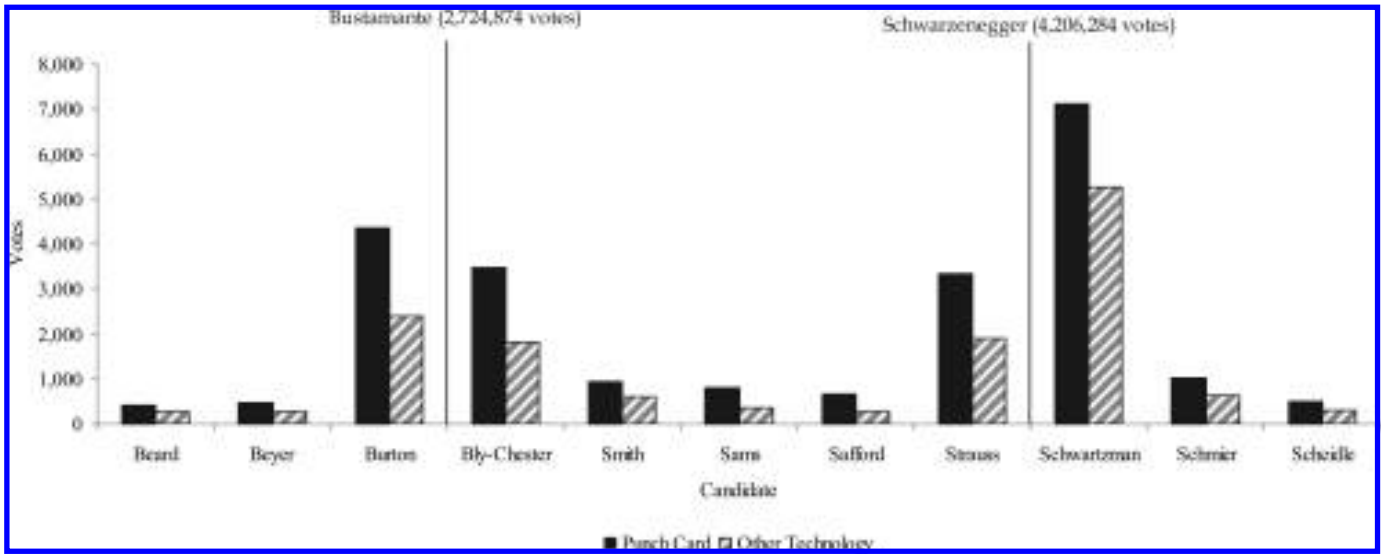
[5] Roth (1998, p. 7) also found that eight of the subjects were responsible for 92% of the errors. However, even if we disregard these errors, the implied error rate is still 1.2%, which is comparable to residual vote rates.

[6] One potential approach would be to compare recorded vote shares with those suggested by exit polling. However, most exit polls are not designed to be representative at the level of the variation in voting systems (typically, the county level). Furthermore, the small sample sizes of exit polls suggests that would have weak power for detecting effect sizes that would be relevant from a policy perspective.

[7] Fortunately, since the closest item on the ballot (in other words, the recall question) was decided by 711,000 votes, the ballots that were lost due to Votomatic-style punch cards would not have been consequential and no postelection lawsuit occurred (Egelko, 2003).

[8] However, the inferences based on California's recent experiences should be generalized with some caution since the recall election was relatively unique in its complexity (e.g., Alvarez et al. 2004).

FIGURE 1.—VOTES RECEIVED BY BALLOT PROXIMITY TO MAJOR CANDIDATES AND VOTING TECHNOLOGY



(I), who was behind Arnold Schwarzenegger on nearly all of the ballots. Schwartzman received the ninth most votes (12,382), placing him between celebrity Gary Coleman and Mary Cook, a pornography actress whose campaign had been publicized relatively widely. Schwartzman's candidacy undoubtedly benefited from voters who confused his name with Schwarzenegger's (for example, Green, 2003). However, he may have also benefited from miscast ballots simply because his ballot position was next to a major candidate who received over 4.2 million votes (in other words, alignment errors).

The importance of a bookend position on the ballot is clearly suggested by the surprising (and little-noticed) success of all the other candidates who had been next to Schwarzenegger and Bustamante. For example, the candidate in front of Schwarzenegger on the ballot, Lawrence Steven Strauss (D), also had an unusually high number of votes (5,245 for a rank of 18). Similarly, the candidates placed in front of and behind Bustamante, John Christopher Burton (I) and Cheryl Bly-Chester (R), ranked 14th and 17th with 6,748 and 5,297 votes, respectively. As suggested earlier, Burton's candidacy may have also benefited from some name confusion.[9] However, it should be noted that whether the extent to which support for these candidates is due to ballot position or name confusion is not important for this study. The relevant assumption is that, because they largely reflect miscast ballots, they provide a novel measure of ballot performance.

Were a substantial number of the votes cast for these bookend candidates really mistakes? The surprising number of votes received by these four candidates relative to the other candidates clearly suggests so. Specifically, the median number of votes received by a gubernatorial candidate was 1,077.[10] Each of these bookend candidates received at least *five times* as many votes. More formally, a simple nonparametric test based on rankings of these four candidates suggests that their joint electoral success was extremely unlikely to have occurred by chance.[11]

Figure 1 provides a graphical illustration of the bookend phenomenon by showing the vote counts for two main candidates and eleven minor candidates surrounding them as they were positioned on most ballots. The ballot positions for all the candidates were determined by the state through a randomized ordering of the alphabet. This initial ballot ordering was used for voting precincts in the first of California's eighty assembly districts. Then, in each subsequent district, candidates were moved one position up on the ballot (and the candidate who had been at the top of the ballot moved to the bottom). Figure 1 illustrates that the bookend candidates did substantially better than the other minor candidates with close ballot positions. By comparing the electoral support for the bookend candidates across counties with and without punch cards, figure 1 also illustrates the apparent effect of punch-card voting systems on recorded support for the four bookend candidates. Roughly 49% of the votes cast for a governor were in counties using Votomatic or Datavote punch cards. However, all four of the bookend candidates received more votes in counties using punch cards than in counties using touch-screen or optical-scan systems. The regression analysis discussed in the next section generalizes these comparisons.

Another interesting piece of evidence, which is consistent with viewing much of the recorded vote for bookend

[9] John Burton was also the name of the president of the California Senate (Rainey & Hoffman, 2003).

[10] This is based only on the 135 candidates who appeared on the ballot. Including the write-in candidates reduces the median to 823.

[11] A Wilcoxon rank sum test based on the null hypothesis that the population locations of the four candidates and the other candidates are the same leads to a test statistic with a p-value of 0.005.

TABLE 2.—WLS ESTIMATES, PERCENT VOTES FOR "BOOKEND" CANDIDATES IN THE CALIFORNIA RECALL ELECTION

Independent Variable	Dependent Variables: Vote Share (100×) for			
	Strauss (D)	Schwartzman (I)	Burton (I)	Bly-Chester (R)
Percent for Schwarzenegger	.059*	.208***	—	—
	(.032)	(.060)		
Percent for Bustamante	—	—	.200***	.105***
			(.038)	(.036)
R^2	.0492	.1786	.3346	.1305
Dependent mean (100×)	.061	.143	.078	.061

Standard errors are reported in parentheses. The weight is the number of voters in the gubernatorial election.
*Statistically significant at the 10% level.
**Statistically significant at the 5% level.
***Statistically significant at the 1% level.

candidates as mistakes, is that there appears to have been a smaller but similar phenomenon for the two candidates whose ballot position placed them near the major candidate who placed *third* (Alvarez et al., 2004).[12] These two candidates also did surprisingly well, each garnering over 2,000 votes. The fact that these candidates did noticeably better than the median candidate—but substantially worse than candidates positioned next to more popular candidates—is exactly consistent with the hypothesis that much of the recorded support for these candidates reflects miscast ballots.

At an anecdotal level, the available evidence from press coverage of the election clearly indicates that none of these four candidates had achieved a notoriety or popular support consistent with their strong electoral performance. For example, Schwartzman—the leading vote-getter of these four—was described after the election as an "utterly obscure" (Kershaw, 2003) or "unknown" (Green, 2003) candidate who spent very little and campaigned through a Web site and a few small-scale personal appearances. Furthermore, all four of these candidacies appeared to involve no political professionals and had little or no party support. George Schwartzman is a San Diego–area businessman involved in providing "urological services to hospitals." Lawrence Steven Strauss (D) is a personal-injury lawyer in the Los Angeles area. John Christopher Burton is a civil-rights attorney in the Los Angeles area and a self-described Socialist candidate. Cheryl Bly-Chester (R) is an environmental engineer who described herself as a pro-business, fiscal conservative. Lexis-Nexis searches of California newspapers also suggest that these candidates were mentioned in few, if any, preelection news articles.[13]

Another important piece of evidence that these candidates benefited from a substantial number of mistakenly cast ballots comes from the cross-sectional variation in their vote shares. More specifically, simple regressions indicate that support for the bookend candidates is positively and significantly related to support for the major candidate next to them on the ballot (table 2). For example, though Schwarzenegger was surrounded on the ballot by two candidates who did *not* share his party affiliation, regression results (table 2) indicate that counties with higher support for Schwarzenegger also cast significantly more votes for Strauss (D) and Schwartzman (I). Similarly, Bustamante's ballot position was also surrounded by two candidates who did *not* share his party affiliation, Burton (I) and Bly-Chester (R). Yet, counties that had higher support for Bustamante also cast significantly more votes for these neighboring candidates (table 2). Interestingly, the point estimates from these auxiliary regressions are larger in the two cases (Burton and Schwartzman) where the bookend candidates may have benefited from name confusion as well as ballot position.[14]

As an aside, the regression results in table 2 provide a straightforward way to estimate what share of the officially recorded votes for the bookend candidates actually were mistakes. More specifically, this can be done by predicting the vote share each bookend candidate would have received if the main candidate had only received the median vote share (that is, 1,077 votes out of the 8.6 million cast). A comparison of these predicted values to their actual vote shares suggests that 50% to 80% of the votes recorded for bookend candidates were actually intended for a main candidate. One caveat to this exercise is that the regression-adjusted predictions rely on a main-candidate vote share that is well outside the range observed in the data (in other words, in no county did a main candidate receive such a small vote share). A simple alternative is to assume that, absent a bookend effect, each candidate would have received the median number of votes. A comparison of this value to their recorded votes implies that 79% to 91% of the votes recorded for bookend candidates were actually intended for one of the two main candidates.

[12] The Republican candidate, Tom McClintock, placed third with over 13% of the vote (less than half the amount received by Bustamante).

[13] And the articles where they were mentioned tended to be negative. In particular, a few articles (for example, Rainey, 2003) discussed how two of these candidates (Bly-Chester and Schwartzman) joined a bus trip with roughly a dozen other candidates that shadowed a Schwarzenegger caravan through the state. Rainey (2003) described these candidates as "pilot fish on the flanks of the Schwarzenegger leviathan."

[14] The association between support for a main candidate and for their bookend candidates was also significantly larger in punch-card counties, a result consistent with the reduced-form evaluations presented in the next section.

TABLE 3.—WLS ESTIMATES, PERCENT VOTES FOR "BOOKEND" CANDIDATES IN THE CALIFORNIA RECALL ELECTION

Independent Variable	Dependent Variables: Vote Share (100×) for							
	Strauss (D)		Schwartzman (I)		Burton (I)		Bly-Chester (R)	
Punch card–Votomatic/Pollstar	.040***	.038***	.043***	.053***	.059***	.052***	.045***	.045***
	(.007)	(.009)	(.015)	(.015)	(.008)	(.009)	(.007)	(.007)
Punch card–Datavote	.026*	.019	.071**	.046**	.003	.015	.004	.011
	(.014)	(.014)	(.029)	(.023)	(.015)	(.013)	(.014)	(.012)
Touch screen	.013	.007	−.011	−.038*	.017	.007	−.007	−.014
	(.012)	(.014)	(.024)	(.022)	(.013)	(.013)	(.011)	(.011)
Additional controls?	no	yes	no	yes	no	yes	no	yes
R^2	.3818	.5102	.2085	.6276	.5313	.7456	.4857	.7196

Standard errors are reported in parentheses. The weight is the number of votes cast in the gubernatorial election. The nine additional control variables are percent black, percent Hispanic, percent white non-Hispanic, percent aged 65 or over, percent in poverty, median household income, percent that are high school graduates, percent with some college, and percent that are college graduates.

*Statistically significant at the 10% level.

**Statistically significant at the 5% level.

***Statistically significant at the 1% level.

It should be noted that the electoral success of these candidates in the few instances when the ballot design positioned them away from the two major candidates could also provide some supporting evidence for the effects of a bookend position on the ballot. More specifically, as noted above, ballot positions were determined through a randomized ordering of the alphabet for precincts in the first of California's assembly district. However, in each subsequent district, candidates were moved one position up on the ballot (and the candidate who had been at the top of the ballot moved to the bottom). This arrangement implied that, within parts of Orange County, the candidates who surrounded Bustamante on the ballot were separated because a candidate rolled from the top to the bottom of the ballot. Similarly, the candidates surrounding Schwarzenegger were separated by this rollover within parts of San Diego County. Because both of these counties reported voting results by the assembly-district boundaries, we can assess whether these separations seemed to influence the popularity of these candidates.

The comparative results with respect to the two candidates who *preceded* a major candidate (Burton and Strauss) clearly suggest that a bookend position on the ballot made a substantial contribution to the votes received by these candidates. For example, on ballots in Orange County's 68th assembly district, Burton (last position) was separated from Bustamante (first position). In this district, Burton received only 0.010% of the vote, while in the remainder of the county, his vote share was three times larger. Similarly, Strauss received only 0.014% of the vote in San Diego County's 74th assembly district where his position at the bottom of the ballot separated him from Schwarzenegger. But his vote share was over six times larger in the remainder of San Diego County, where his ballot position was next to Schwarzenegger's. These sharp differences do not appear to merely reflect the consequences of being at the bottom of the ballot. In both of these cases, most voters used ballots on which these candidates were positioned at or near the bottom.[15]

However, these comparisons are less dispositive with respect to the two bookend candidates who were typically positioned *behind* a major candidate on the ballot (Bly-Chester and Schwartzman). Both of these candidates had basically similar within-county results in communities where they had moved to the top of the ballot and a major candidate had moved to the bottom.[16] These results do not necessarily imply that bookend effects are not broadly important. These *trailing* bookend candidates may have done relatively well when at the top of the ballot simply because some voters would select them, rather than make an effort to find the top candidates who were toward the bottom of the ballot (e.g., Darcy, 1986). This is particularly so with respect to votes for Schwartzman, who undoubtedly generated more name confusion as well. Another caveat associated with this particular check is that these highly localized comparisons may be subject to some omitted variable biases.[17]

IV. Voting Systems and the Bookend Candidates

In table 3, I present the key results from baseline models that examine the association between a county's voting technology and its degree of support for each of these four candidates. These point estimates are based on weighted least squares (WLS) applied to a linear probability model.[18] The results indicate that the vote shares for each of the four candidates were significantly higher in the counties with Votomatic-style punch cards relative to optical-scan counties. More specifically, these results suggest that Votomatic-style punch cards led to statistically significant

[15] Furthermore, these comparative results are similar when the comparisons are based only on the outcomes from districts where both candidates were near the bottom of the ballot.

[16] More specifically, Bly-Chester actually did relatively better in the part of Orange County where she was at the top of the ballot and Bustamante had moved to the bottom (the 69th assembly district where she received 32 votes). Furthermore, Schwartzman only did about 10% worse in that part of San Diego County where he was at the top of the ballot and Schwarzenegger had moved to the bottom (the 75th assembly district).

[17] For example, Schwartzman happened to be at the top of ballot used in his own congressional district (Green, 2003).

[18] The weight used in these regressions (the number of gubernatorial votes cast in the county) is intended to provide efficient estimates in the presence of heteroskedasticity. However, I present evidence on the robustness of my results across a variety of estimation techniques below.

TABLE 4.—ESTIMATES OF PUNCH-CARD EFFECT ON PERCENT VOTES FOR "BOOKEND" CANDIDATES IN THE CALIFORNIA RECALL ELECTION, ALTERNATIVE SPECIFICATIONS

Model Specification	Dependent Variables: Vote Share (100×) for			
	Strauss (D)	Schwartzman (I)	Burton (I)	Bly-Chester (R)
Weighted least squares (WLS)	.038***	.053***	.052***	.045***
	(.009)	(.015)	(.009)	(.007)
Grouped logit	.044***	.095***	.053***	.051***
	(.014)	(.025)	(.014)	(.012)
OLS, White standard errors, full sample	.021	.058*	.037*	.028
	(.015)	(.033)	(.022)	(.018)
OLS, White standard errors, counties with 20,000+ voters	.044***	.069***	.056***	.044***
	(.014)	(0.22)	(.017)	(.014)
OLS, bootstrapped standard errors, full sample	.021	.058	.037	.028
	(.017)	(.039)	(.024)	(.020)
OLS, bootstrapped standard errors, counties with 20,000+ voters	.044**	.069**	.056***	.044**
	(.018)	(.030)	(.021)	(.020)

Standard errors are reported in parentheses. The weight is the number of gubernatorial votes cast. All models include the nine additional control variables listed in table 3.
*Statistically significant at the 10% level.
**Statistically significant at the 5% level.
***Statistically significant at the 1% level.

increases in votes for these candidates that ranged from roughly one-third to two-thirds of their respective means.

These strikingly large effect sizes imply that a substantial amount of the recorded support for these candidates is due to errors induced by this particular punch-card technology. For example, these results indicate that Votomatic-style punch cards increased the probability of casting a ballot for Strauss by roughly 0.04 percentage points, an effect equal to two-thirds of his vote share. There were 3,688,357 gubernatorial voters in the seven counties using this technology. Therefore, these estimates imply that, relative to optical-scan ballots, these punch cards induced at least 1,401 mistakenly cast ballots for this candidate (in other words, over 27 percent of his total votes). Similar calculations suggest that 1,954 of the votes cast for George Schwartzman (over 16% of his total) were simply due to the mistakes caused by this punch-card technology. It should be emphasized that this does not imply that the remaining votes cast for these candidates reflected true voter intent. In other counties, recorded ballots undoubtedly misrepresented the true intent of voters (see figure 1). These estimates identify the *relative* impact of punch-card technology.

Interestingly, the Datavote-style punch cards were also uniformly associated with higher votes for these candidates. However, these estimates were only consistently large and statistically significant for Schwartzman, which suggests that this ballot technology only promoted significant voter errors in the presence of rather severe name confusion. The results in table 3 also suggest that, generally, touch-screen systems had small and inconsistently signed effects relative to optical-scan ballots.

The key results in table 3 suggest that Votomatic-style punch cards significantly increased the prevalence of recorded but miscast votes. However, an obvious—and critically important—concern with this basic inference is that these results may reflect spurious correlations driven by the unique tastes and characteristics of voters in the punch-card counties. The diverse nature of the seven counties that used Votomatic-style punch cards (table 1) suggests that this sort of bias is unlikely.

Nonetheless, table 3 provides additional evidence on this issue by presenting the key results from regression models that are saturated with nine additional control variables for county-level traits like race, ethnicity, income, poverty, and educational attainment. These controls include four demographic variables (percent black, percent Hispanic, percent white non-Hispanic, percent aged 65 and over), which are based on 2002 Census data. They also include the percentage in poverty and median household income and three variables indicating the percentage of county residents over age 25 whose highest educational attainment was high school graduate, some college, or college graduate. These five variables are based on 2000 data from the Census Bureau. The inferences about Votomatic-style punch cards (table 3) are robust to the introduction of these nine controls, suggesting that omitted variable biases are not problematic.

However, perhaps the most convincing evidence that these sorts of bias are not problematic is that *all four* candidates did significantly better in these punch-card counties. More specifically, the hypothesis that these results are biased by unobserved county traits is not easily tenable in light of the fact that voters in these counties consistently cast an unusually high number of votes for four very different candidates: a Republican, a Democrat, and two Independents (one of whom is a Socialist).

In table 4, I present evidence on the robustness of my key results to alternative estimation techniques. The first row in table 4 reiterates the estimated effects of Votomatic-style punch cards on the vote shares of each candidate based on weighted least squares. The estimates in the second row are marginal effects from "grouped logit" specifications that

TABLE 5.—WLS ESTIMATES OF PUNCH-CARD EFFECT ON PERCENT VOTES FOR "BOOKEND" CANDIDATES IN THE CALIFORNIA RECALL ELECTION, ALTERNATIVE SAMPLES

Sample	Dependent Variables: Vote Share (100×) for			
	Strauss (D)	Schwartzman (I)	Burton (I)	Bly-Chester (R)
Full sample	.038***	.053***	.052***	.045***
	(.009)	(.015)	(.009)	(.007)
Without candidate's home county	.045***	.039**	.038***	.045***
	(.012)	(.015)	(.010)	(.008)
Without Los Angeles County	.045***	.067***	.038***	.029***
	(.012)	(.018)	(.010)	(.008)
Without Mendocino County	.038***	.054***	.051***	.045***
	(.009)	(.015)	(.009)	(.008)
Without Sacramento County	.033***	.059***	.050***	.045***
	(.008)	(.014)	(.009)	(.008)
Without San Diego County	.042***	.039**	.068***	.061***
	(.010)	(.015)	(.007)	(.005)
Without Santa Clara County	.031**	.051**	.029***	.023***
	(.012)	(.019)	(.010)	(.008)
Without Sierra County	.038***	.053***	.052***	.045***
	(.009)	(.015)	(.009)	(.008)
Without Solano County	.039***	.051***	.052***	.045***
	(.009)	(.014)	(.009)	(.008)

Standard errors are reported in parentheses. The weight is the number of gubernatorial votes cast. All models include the nine additional control variables listed in table 3.
*Statistically significant at the 10% level.
**Statistically significant at the 5% level.
***Statistically significant at the 1% level.

explicitly acknowledge the binary nature of the underlying voter choice.[19] These estimates are larger than the WLS estimates but, in most cases, these differences are well within the implied sampling variation. The results of OLS models that rely on an unrestrictive correction for heteroskedasticity (White, 1980) are presented in the third row of table 4. These estimates imply that the punch-card effects are generally smaller and more imprecise (that is, larger standard errors). These differences are likely to reflect the relative unimportance of voting technology for voters in California's most sparsely populated counties.[20] To examine this issue directly, I estimated these OLS models excluding data from the 19 counties with fewer than 20,000 voters. The remaining 39 counties contained over 98% of Californian voters. The results of these evaluations are quite similar to the baseline results based on weighted least squares (table 3). An alternative and similarly unrestrictive approach to constructing standard errors is to rely on bootstrapping.[21] This approach also implied substantially larger standard errors. However, in models based on the 98% of voters in the largest counties, the punch-card effects are still statistically distinguishable from 0 with at least 5% significance.

A final set of specification checks are presented in table 5. The motivation for these checks is that the effects of sample exclusions can provide fairly compelling ad hoc evidence on the reliability of this study's main inferences. For example, it may be that the apparent errors caused by Votomatic-style punch-card technology are actually due to the fact that three of these candidates resided in counties that used those voting systems. In other words the apparent effects of punch-card technology could actually be due to these candidates being relatively well known in their own communities. To assess the empirical relevance of this issue, I reevaluated this study's main results in models that excluded the data from each candidate's home county.[22] The results, which are presented in the second row of table 5, suggest that this potential source of bias was not problematic. The remaining results in table 5 assess whether this study's main results were spuriously driven by any of the seven punch-card counties. More specifically, I reevaluated the punch-card effects in models that excluded each of these seven counties. The results were remarkably stable and uniformly suggested that Votomatic-style punch-card technologies significantly increased the recorded support for the diverse set of four bookend candidates.

V. Conclusions

The notion that voters can expect to have their preferences recorded accurately and fairly is fundamentally important in a democracy. However, the events surrounding recent elections have raised questions about how well

[19] More specifically, the dependent variable in these models is the natural log of the odds ratio. These models are estimated by weighted least squares where the weight reflects the explicit heteroskedasticity associated with this specification (Maddala, 1983). To facilitate interpretation, I have multiplied the estimated coefficients and standard errors by 100.

[20] Brady (2003) similarly analyzes data from the largest California counties, in part because of evidence that "election administration is more variable in smaller counties."

[21] In this context, bootstrapping refers to identifying the sampling distributions for the point estimates by repeatedly estimating each model (2,000 times) with 58 observations randomly drawn from the data set (with replacement).

[22] The excluded counties and their respective candidates are the following: Strauss (Los Angeles), Schwartzman (San Diego), Burton (Los Angeles), and Bly-Chester (Placer).

particular voting technologies perform. Comparative assessments of voting systems typically focus on the prevalence of "residual votes," that is, ballots that cannot be counted. However, relatively little attention has been paid to the comparative effects of voting technologies on a different and less easily measured type of error: recorded votes that do not accurately reflect a voter's true intent.

In this study, I presented evidence on this question by exploiting some unique aspects of the recent California gubernatorial election. Specifically, I presented evidence that a surprisingly large number of California voters mistakenly cast their gubernatorial votes for one of the four candidates positioned next to the two major candidates on the ballot. I then assessed the comparative effects of California's voting systems on the recorded support for these four diverse candidates. The results of these evaluations suggested that voting technologies can differ dramatically in the accuracy with which they record a voter's intent. In particular, this evidence indicated that the use of Votomatic-style punch-card systems substantially increased the likelihood that voters mistakenly voted for these candidates.

These technology-induced errors would not have been consequential for the California recall election because Schwarzenegger's margin of victory was so large and because punch cards drew away votes intended for both major candidates. Furthermore, this study's evidence that punch-card systems perform poorly in a previously unappreciated way now has relatively narrow policy relevance. The Help America Vote Act (HAVA), signed into law by President Bush in 2002, authorized new federal spending to help states meet a requirement to replace punch-card voting machines by 2006 (Pear, 2002). In response, the use of punch-card machines has declined precipitously. However, it should be noted that not all states are currently in compliance with HAVA (Seligson, 2006; Wood, 2006). In particular, one study (EDS, 2006) estimated that, during the general election in 2006, over five million registered voters (over 3% of the total) would be in communities still using punch-card systems.

The more general and policy-relevant import of this study is that future assessments of voting technologies should emphasize more than just minimizing residual votes. Put differently, a fair voting system should do more than "count every vote"; it should also represent voter intent accurately. This study provided evidence that voting technology can influence the prevalence of recorded votes that misrepresent voter intent. However, this study did not provide any consistent evidence on the relative effects of other increasingly prominent voting systems (such as touch screens). However, only four counties in California used touch screens for the recall election. Furthermore, the effects of such systems may depend critically on particular features of their voter interface. This, combined with the difficulty of knowing when a recorded vote misrepresents voter intent, suggests that small-scale randomized experiments may play a prominent role in future efforts to identify design features that minimize these errors. It should be noted that failing to consider such errors could be consequential in elections that are as tightly contested as several have been in recent years.

REFERENCES

Alvarez, R. Michael, Stephen Ansolabehere, Erik Antonsson, Jehoshua Bruck, Stephen Graves, Thomas Palfrey, Ron Rivest, Ted Selker, Alex Slocum, and Charles Stewart III. "Voting: What Is, What Could Be," Caltech/MIT Voting Technology Project (July 2001).

Alvarez, R. Michael, Melanie Goodrich, Thad E. Hall, D. Roderick Kiewet, and Sarah M. Sled. "The Complexity of the California Recall Election," *PS: Political Science and Politics* 37:1 (January 2004), 23–26.

Brady, Henry E., "Revised Memorandum on Residual Vote Rates," memo (October 9, 2003), available at: http://ucdata.berkeley.edu/new_web/recall/twenty.pdf.

Brady, Henry E., Justin Buchler, Matt Jarvis, and John McNulty, Counting All the Votes: The Performance of Voting Technology in the United States, Department of Political Science, Survey Research Center, Institute of Governmental Studies, UC Berkeley (September 2001).

Card, David, and Enrico Moretti, "Does Voting Technology Affect Election Outcomes? Touch Screen Voting and the 2004 Presidential Election," NBER working paper no. 11309 (May 2005).

Darcy, Robert, "Position Effects with Party Column Ballots," Western Political Quarterly 39 (1986), 648–662.

The Economist, "Unfair, Again" (June 9, 2001).

Egelko, Bob, "176,000 Bad Punch-Card Ballots, ACLU Says. Group Not Planning to Sue Over Lost Votes," *San Francisco Chronicle* (October 9, 2003).

Election Data Services, "New Study Shows 50 Million Voters Will Use Electronic Voting Systems, 32 Million Still with Punch Cards in 2004," 2004 Voting Equipment Study (Feb. 12, 2004), available at: http://www.electiondataservices.com.

———— "69 Million Voters Will Use Optical Scan Ballots in 2006; 66 Million Voters Will Use Electronic Equipment," 2006 Voting Equipment Study (Feb. 6, 2006), available at: http://www.electiondataservices.com.

Garner, Phillip, and Enrico Spolaore, "Why Chads? Determinants of Voting Equipment Use in the United States," *Public Choice* 123 (2005), 363–392.

Green, Kristen, "By George, Unknown Finishes in 9th Place," *San Diego Union-Tribune* (October 9, 2003).

Kershaw, Sarah, "Just Being in the Game Was Enough for Some Who Came Away as Losers," *New York Times* (October 9, 2003).

Kleffman, Sandy, "University of California-Berkeley Study Remains at the Heart of the Recall Date Debate," *Contra Costa Times* (September 23, 2003).

Knack, Stephen, and Martha Kropf, "Who Uses Inferior Voting Technology?" *PS: Political Science and Politics* 35:3 (September 2002), 541–548.

———— "Voided Ballots in the 1996 Presidential Election: A County-Level Analysis," *The Journal of Politics* 65:3 (August 2003), 881–897.

Maddala, G. S., "Limited-Dependent and Qualitative Variables in Econometrics," Econometric Society Monographs, Cambridge University Press (1983).

McDonald, Michael P., "California Recall Voting: Nuggets of California Gold for Voting Behavior," *The Forum* 1:4, article 6 (2003), available at: http://www.bepress.com/forum/vol1/iss4/art6.

Merzer, Martin, "Chads Still Hanging Around in 22 States," *Miami Herald* (February 15, 2004).

Pear, Robert, "Bush Signs Legislation Intended to End Voting Disputes," *New York Times* (October 30, 2002), section A, p. 22.

Rainey, James, "The Recall Campaign: Pursue Schwarzenegger; At Least a Dozen Gubernatorial Hopefuls Relish the Trip on the "Ex-Terminator' Bus in the Wake of the Front Runner," *Los Angeles Times* (October 4, 2003).

Rainey, James, and Allison Hoffman, "For 130, It's Not Whether You Win or Lose," *Los Angeles Times* (October 9, 2003).

Roth, Susan King, "Disenfranchised by Design: Voting Systems and the Election Process," *Information Design Journal* 9:1 (1998), 1–8.

Saltman, Roy G., "Accuracy, Integrity and Security in Computerized Vote-Tallying," National Bureau of Standards Special Publication 500-58, Gaithersburg, MD (1988), available at: http://www.itl.nist.gov/lab/specpubs/500-158.htm.

Seligson, Dan, "The HAVA Lawsuits," *Campaigns and Elections* (June 2006), 54.

Shocket, P. A., N. R. Heighberger, and C. Brown, "The Effect of Voting Technology on Voting Behavior in a Simulated Multi-Candidate City Council Election: A Political Experiment of Ballot Transparency," *Western Political Quarterly* 45:2 (1992), 521–537.

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane Jr., Michael C. Herron, and Henry E. Brady, "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida," *American Political Science Review* 95:4 (December 2001), 793–810.

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48:4 (May 1980), 817–838.

Wood, Daniel B., "As Election Season Nears, Efforts to Upgrade Voting Machines Bog Down," *Christian Science Monitor* (January 19, 2006), 3.