

Residual Votes Attributable to Technology

An Assessment of the Reliability of Existing Voting Equipment

The Caltech/MIT Voting Technology Project¹

Version 2: March 30, 2001²

American elections are conducted using a hodge-podge of different voting technologies: paper ballots, lever machines, punch cards, optically scanned ballots, and electronic machines. And the technologies we use change frequently. Over the last two decades, counties have moved away from paper ballots and lever machines and toward optically scanned ballots and electronic machines. The changes have not occurred from a concerted initiative, but from local experimentation. Some local governments have even opted to go back to the older methods of paper and levers.

The lack of uniform voting technologies in the US is in many ways frustrating and confusing. But to engineers and social scientists, this is an opportunity. The wide range of different voting machinery employed in the US allows us to gauge the reliability of existing voting technologies. In this report, we examine the relative reliability of different machines by examining how changes in technologies within localities over time explain changes in the incidence of ballots that are spoiled, uncounted, or unmarked – or in the lingo of the day the incidence of “over” and “under votes.” If existing technology does not affect the ability or willingness of voters to register preferences, then incidence of over and under votes will be unrelated to what sort of machine is used in a county.

We have collected data on election returns and machine types from approximately two-thirds of the 3,155 counties in the United States over four presidential elections, 1988, 1992, 1996, and 2000. The substantial variation in machine types, the large number of

¹ The Caltech/MIT Voting Technology Project is a joint venture of the two institutions. Faculty involved are Michael Alvarez (Caltech), Stephen Ansolabehere (MIT), Erik Antonsson (Caltech), Jehoshua Bruck (Caltech), Steven Graves (MIT), Nicholas Negroponte (MIT), Thomas Palfrey (Caltech), Ron Rivest (MIT), Ted Selker (MIT), Alex Slocum (MIT), and Charles Stewart (MIT). The principal author of this report is Stephen Ansolabehere; communications about this report can be directed to him at sda@mit.edu. We are grateful to the Carnegie Corporation for its generous sponsorship of this project.

²This version updates our initial report in three ways. First, we have expanded the data set considerably: increasing the number of valid cases from roughly 5500 to 8000. We have added complete data for several states, such as Kentucky, Massachusetts, and Vermont, and nearly complete coverage of the available data from the 2000 election. Second, we present more detail about the data, such as yearly averages, and examine possible technology curves and other hypothesized relationships. Third, we incorporate more speculation about the performance of DREs. The next version of the report will integrate data from 1980 and from the 1980, 1990, and 2000 censuses, which will allow us to examine possible interactions between machine performance and demographic characteristics of county populations.

observations, and our focus on presidential elections allows us to hold constant many factors that might also affect election returns.

The central finding of this investigation is that manually counted paper ballots have the lowest average incidence of spoiled, uncounted, and unmarked ballots, followed closely by lever machines and optically scanned ballots. Punchcard methods and systems using direct recording electronic devices (DREs) had significantly higher average rates of spoiled, uncounted, and unmarked ballots than any of the other systems. The difference in reliabilities between the best and worst systems is approximately 1.5 percent of all ballots cast.

We view these results as benchmarks for performance. It is our hope that the information here is helpful to manufacturers as they improve equipment designs and to election administrators who may wish to adopt new equipment. Our results apply to broad classes of equipment; the performance of specific types of equipment may vary. Where possible we test for possible differences (such as different types of punch cards).

We do not attempt to isolate, in this report, the reasons for differential reliability rates, though we offer some observations on this matter in the conclusions. Our aim is measurement of the first order effects of machine types on the incidence of votes counted.

Machine Types and their Usage

We contrast the performance of five main classes of technologies used in the US today. The technologies differ according to the way votes are cast and counted.

The oldest technology is the paper ballot. To cast a vote, a person makes a mark next to the name of the preferred candidates or referendum options and, then, puts the ballot in a box.³ Paper ballots are counted manually. Paper ballots enjoyed a near universal status in the US in the 19th Century; they remain widely used today in rural areas.

At the end of the 19th Century, mechanical lever machines were introduced in New York state, and by 1930 every major metropolitan area had adopted lever machinery. The lever machine consists of a steel booth that the voter steps into. A card in the booth lists the names of the candidates, parties, or referenda options, and below each option is a switch. Voters flick the switch of their preferred options for each office or referendum. When they wish to make no further changes, they pull a large lever, which registers their votes on a counter located on the back of the machine. At the end of the voting day, the

³ How we mark ballots has changed over time. In the middle of the 20th Century, many states required that the voter cross out the options not chosen. See for example, The Book of the States, 1948.

election precinct workers record the tallies from each of the machines. Lever machines automate both the casting of votes and the counting of votes through mechanical devices.

Punch card machines automated the counting process using the computer technology of the 1960s. Upon entering the polling place the voter is given a paper ballot in the form of a long piece of heavy stock paper. The paper has columns of small, perforated rectangles (or chads). There are two variants of the punch card – one, the DataVote, lists the names of the candidates on the card; the other (VotoMatic) does not. In the booth (for VotoMatics), the voter inserts the card into a slot and opens a booklet that lists the candidates for a given office. The voter uses a metal punch to punch out the rectangle beside the candidate of choice. The voter then turns the page, which lists the options for the next office and shifts the card to the next column of rectangles. When finished, the voter removes the card and puts it in the ballot box. At the end of the day, the election workers put the cards into a sorter that counts the number of perforations next to each candidate.

Optically scanned ballots, also known as “marksense” or “bubble” ballots, offer another method for automating the counting of paper ballots. The form of the optically scanned ballot is familiar to anyone who has taken a standardized test. The voter is given a paper ballot that lists the names of the candidates and the options for referenda, and next to each choice is small circle or an arrow with a gap between the fletching and the point. The voter darkens in the bubble next to the preferred option for each office or referendum, or draws a straight line connecting the two parts of the arrow. The ballot is placed in a box, and, at the end of the day, counted using an optical scanner. Some versions of this technology allow the voter to scan the ballot at the polling place to make sure that he or she voted as intended.

Direct recording electronic devices, DREs for short, are electronic versions of the lever machines. In fact, the first widely used electronic machine (the Shouptronic 1242) was modeled on the lever machine and developed by one of the main lever machine manufacturers. The distinguishing feature of a DRE is that an electronic machine records the voter’s intentions, rather than a piece of paper or mechanical device. To the extent that there is a paper trail it is generated by the machine, not the voter. Electronic machines vary along a couple of dimensions, having to do with the interface. First, there are many devices used to register the vote: the interfaces are either push button (e.g., the Shouptronic) or touch screen (e.g., Sequoia Pacific’s Edge or Unilect’s Patriot) or key pads (see the Brazillian machine). Second, the ballot design is either full-faced or paginated. With full-faced ballots, common among push button equipment, the voter sees the entire ballot at once. With paginated systems, common among touch screen devices, the voter views a page for each office or question on the ballot. A voting session goes roughly as follows. Upon entering the polling place, the voter is given a card that is inserted into the machine to activate the individual voting session. When finished the voter touches the name on the screen to register his or her preference and, typically, the voter may review the entire session (or ballot) to check the vote. Like lever machines it is not possible to vote twice for the same office (i.e., over vote). Each electronic machine tallies the votes locally and the tallies, usually on a disc, are sent to a central location.

Each type of technology involves many variations based on specifications of manufacturers, ballot formats, and implementation. Our focus is on the five main types of machines, as we hope to learn which mode of voting looks most promising. In almost all states county election officials decide which machinery to use, so counties are, almost everywhere, the appropriate unit of analysis. Some counties do not have uniform voting technologies. In these situations, municipalities and, sometimes, individual precincts use different methods. These counties are called Mixed Systems. They occur most commonly in Massachusetts, Michigan, Maine, New Hampshire, and Vermont, where town governments usually administer elections.

We examine the variation in usage across counties and over time. Our data on voting equipment come from the Election Data Services and from state, county, and municipal election officials. We appreciate the helpfulness of election administrators and the EDS in our data collection efforts.

The data do not distinguish centrally counted and precinct counting of ballots sufficiently well that we could estimate with confidence the difference in performance between central and precinct counting. Some states provide information about which administrative units count the ballots for some machine types. Precinct and central counting of optically scanned ballots became quite controversial in the Florida 2000 election.

Even without this additional level of detail, the pattern of equipment usage across the United States looks like a crazy quilt. Americans vote with a tremendous array of types of equipment. Table 1 displays the wide variation in machines used in the 1980 and 2000 elections. The first two columns present the average number of counties using various types of equipment in each year. The last two columns report the percent of the population covered by each type of technology in the 1980 and 2000 elections.

In the most recent election, one in five voters used the “old” technologies of paper and levers – 1.3 percent paper and 17.8 percent levers. One in three voters use punch cards – 31 percent of the VotoMatic variety and 3.5 percent of the DataVote variety. Over one in four use optically scanned ballots. One in ten use electronic devices. The remaining 8 percent use mixed systems.

Within states there is typically little uniformity. In some states, such as Arkansas, Indiana, Michigan, Pennsylvania, and Virginia, at least one county uses each type of technology available. The states with complete or near uniformity are New York and Connecticut with lever machines; Alaska, Hawaii, Rhode Island and Oklahoma with scanners; Illinois with punch cards; Delaware and Kentucky with electronics.

As impressive and dramatic have been the changes in technology over time. The third column of the table reports the percent of the 2000 electorate that would have used each machine type had the counties kept the technologies they used in 1980. The data are

pretty clear: out with the old and in with the new. Optically scanned ballots and DREs have grown from a combined 3.2 percent of the population covered to 38.2 percent of the population covered. There has been little change in the mixed and punch card systems. Paper ballots have fallen from 9.7 percent of all people in 1980 to just 1.3 percent in 2000. Lever machines, by far the dominant mode of voting in 1980, covered 43.9 percent of the electorate. Today, only 17.8 percent of people reside in counties using lever machines.

A somewhat different distribution of voting technology across counties holds, owing to the very different population sizes of counties. Punch cards and electronic devices tend to be used in more populous counties, and paper ballots tend to be used in counties with smaller populations.

Table 1

Usage of Voting Equipment in the 1980 and 2000 Elections

	Percent of Counties Using Technology		Percent of 2000 Population Covered by Technology	
	1980	2000	1980	2000
Paper Ballots	40.4	12.5	9.8	1.3
Lever Machines	36.4	14.7	43.9	17.8
Punch Card				
“VotoMatic”	17.0	17.5	30.0	30.9
“DataVote”	2.1	1.7	2.7	3.5
Optically scanned	0.8	40.2	9.8	27.5
Electronic (DRE)	0.2	8.9	2.3	10.7
Mixed	3.0	4.4	10.4	8.1

Three comments about the change in equipment are in order. First, this is an industry in flux. Between 1988 and 2000, nearly half of all counties adopted new technologies (1476 out of 3155 counties), and over the twenty-year period between 1980 and 2000, three out of five counties changed technologies. These changes have occurred without any federal investment.

Second, there is a clear trend toward electronic equipment, primarily scanners but also electronic voting machines. This trend, and the adoption of punch cards in the 1950s and 1960s, reflects growing automation of the counting of votes. Punch cards, optical scanners, and DREs use computer technology to produce a speedy and, hopefully, more reliable count. An influential 1975 report sponsored by the General Accounting Office

and subsequent reports by the Federal Elections Commission called for increased computerization of the vote counts and laid the foundation for methods of certification.⁴

Third, voting equipment usage has a strongly regional flavor. The Eastern and Southeastern United States are notable, even today, for their reliance on lever machines. Midwestern states have a penchant for paper. And the West and Southwest rely heavily on punch cards. In 1980, almost all eastern and southeastern states used levers, and levers were rare outside this region. Notable exceptions were the use of paper in West Virginia and punch cards in Ohio and Florida. In 1980, Midwestern counties used hand counted paper ballots. Illinois was a notable exception with its use of punch cards. And in 1980, almost all counties along the pacific coast and in the Southwest used punch cards. Notable exceptions to the pattern were the use of levers in New Mexico.

This historical pattern of usage evidently had a legacy. As counties have adopted newer technologies over the last twenty years, they have followed some distinctive patterns. Counties tend to adopt newer technologies that are analogous to the technology they move away from. Optical scanning has been most readily adopted in areas that previously used paper, especially in the Midwest. Where counties have moved away from lever machines, they have tended to adopt electronic machines -- for example, New Jersey, Kentucky, central Indiana and New Mexico. These tendencies are strong, but they are not iron clad. In assessing the performance of technology, we will exploit the changes in election results associated with changes in technology. This allows us to hold constant features of the states, counties, and their populations.

Residual Votes: A Yardstick for Reliability

Our measure of reliability is the fraction of total ballots cast for which no presidential preference was counted. We call this the “residual vote.”

A ballot may show no presidential vote for one of three reasons. Voters may choose more than one candidate – commonly called an over vote or spoiled ballot. They may mark their ballot in a way that is uncountable. Or, they may have no preference. The latter two possibilities produce under votes or blank ballots. The residual vote is not a pure measure of voter error or of machine failure, as it reflects to some extent no preference. Consequently we prefer the term residual vote instead of error rate or uncounted vote.

The residual vote does provide an appropriate yardstick for the comparison of machine types, even though it is not purely a measure of machine error or voting mistakes. If

⁴See, Roy Saltman, Accuracy, Integrity and Security in Computerized Vote-Tallying, NBS SP 500-158, August 1988, NIST, Gaithersburg, MD. The report is available online at www.nist.gov/itl/lab/specpubs/500-158.htm.

voting equipment has no effect on the ability of voters to express their preferences, then the residual vote should be unrelated to machine types. To measure such effects, we estimate the average residual vote associated with each machine type, and we assess whether these averages differ significantly across machine type. Averaging guards against idiosyncratic results, and measures what we expect to happen in a typical case.⁵

In our data, the residual vote in the average county equaled 2.3 percent.⁶ In other words, in the typical US county from 1988 to 2000 2.3 percent of ballots casts did not register a presidential preference, for whatever reason. Because county populations vary dramatically, this does not equal the fraction of people who cast an under or over vote for president in these years. This figure is somewhat smaller: 2.1 percent of people who cast ballots did not register a presidential preference. There is considerable variation around this average. Our aim in this report is to assess whether machine types explain a statistically noticeable amount of the variation around this national average residual vote.

We examine the residual vote instead of just the over vote because technology can enable or interfere with voting in many ways. Some technologies seem to be particularly prone to over voting, such as the punch card systems implemented in Florida in the 2000 election. Lever machines and DREs do not permit over voting. Some technologies may be prone to accidental under votes. Lever machines either lock out a second vote or register no vote when the person switches two levers for the same office. Also, paper ballot are sometimes hard to count owing to the many ways that people mark their ballots. Finally, some technologies might intimidate or confuse voters. Many Americans are unaccustomed to using an ATM or similar electronic devices with key pads or touch screens, and as a result DREs might produce more under voting. Also, it may be the case that we react differently to paper than to machines. We are trained in school to answer all of the questions as best as possible, especially on standardized tests similar to the format used for optically scanned voting. Improper installation or wear and tear on machines may lead to high rates of under voting. In Hawaii in 1998, 7 of the 361 optical scanners failed to operate properly.

In depth study of particular states and of contested elections may provide insight into the components of the residual vote or more specific problems related to voting equipment. A number of papers published on the Internet examine the effects of machine types on over votes and on under votes separately for the Florida 2000 election, and several

⁵ Some analyses focus on extreme cases – under and over votes in specific elections in particular counties. Indeed, much of the analysis of Florida falls into this category. Such case studies can be misleading, especially if they reflect outcomes peculiar to a locale, or a local machine failure. Another advantage of averaging is that it washes out the effects of typographical errors, which are inevitable in data, even official government reports.

⁶ We exclude from the analysis all cases in which the official certified report shows more presidential votes cast than total ballots cast, that is, cases with negative residual vote rates. We have tried to resolve all of these cases. They do not appear to be due to absentee votes or other votes being excluded. Instead, they appear to be typographical errors in the data reported by the counties and secretaries of state. This affects about 2 percent of the counties in our analysis. Including these cases changes the numbers reported, but does not affect the pattern of results that we observe.

Secretaries of State and state Election Divisions or commissions present analyses of their own state.

One important caveat is in order in this analysis. There are errors that we cannot count. There is no way to measure whether voters accidentally cast ballots for the wrong candidate. We know of no statistically acceptable measures of fraud. And we know of no studies that attempt to measure the incidence and magnitude of errors in the counting of votes produced by transcription errors or programming errors. Residual votes provides the best available measure of the extent to which technology enables or interferes with the ability of voters to express their preferences.

Many other factors may explain under and over voting beside machine types. Other prominent offices on the ballot, such as senator or governor, might attract people to the polls who have no intention to vote for president. A large turnout might make it difficult for election administrators to tend to voter education at the polls. Demographic characteristics of the county's electorate might explain the incidence of people prone to make mistakes. The wealth of the county might account for expenditures on election administration. New machinery might produce elevated levels of voter confusion, simply because people make mistakes more with unfamiliar tasks.

We examine total ballots cast and ballots cast for President in the 1988, 1992, 1996, and 2000 elections. The data cover approximately 2800 counties and municipalities, though not for all years. All told, there are approximately 7800 counties and municipalities for which we have been able to identify the machines used and to collect data on total ballots and presidential ballots cast. As with the voting equipment data, our data on elections returns come from the Election Data Services and from the relevant election commissions of particular states, counties, and municipalities. The large number of observations produces high levels of precision in estimating average residual vote rates associated with each machine type. Studies of one election in one state may not have yielded sufficiently large samples to determine whether there are significant differences across voting equipment.

We examine the presidential vote in order to hold constant the choices voters face. Within each state one might also examine residual votes in Senate and governor races, with the caveat that these offices have higher "no preference" and thus higher residual votes.

We examine the data at the level of the county or municipality that reports the information. Within each of these jurisdictions, the same voting equipment is used and the administration of the election is under the same office (e.g., has the same budget, etc.). Counties and municipalities are a useful level of analysis because they allow us to hold constant where the equipment is used when we measure which equipment is used. This is of particular concern because equipment usage today is correlated with factors such as county size. We do not want to attribute any observed differences in reliability to equipment, when in fact some other factor, such as county demographics, accounts for the pattern.

To hold constant the many factors that operate at the county level, we exploit the natural experiment that occurs when locales change machinery. We measure how much change in the residual vote occurs when a county changes from one technology to another. The average of such changes for each technology type provides a fairly accurate estimate of the effect of the technology on residual voting, because the many other factors operating at the county level (such as demographic characteristics) change relatively slowly over the brief time span of this study.

To guard against other confounding factors, we also control for contemporaneous Senate and gubernatorial races on the ballot, total turnout, and year of the election.

Results

Typical Counties and Typical Voters

A simple table captures the principle results of this investigation. Table 2 presents the average residual vote rate for each type of voting equipment. The first column of numbers is the average; the second column is the margin of error associated with this estimate; the third column is the median residual vote rate; and the final column is the number of observations (counties and years) on which the estimate is based. The average is the arithmetic mean residual vote across counties. The median is the residual vote such that half of all counties have lower values and half of all counties have higher values. A lower median than mean reflects skew in the distribution of the residual vote produced by a few cases with exceptionally high rates of under and over votes. These averages do not control for other factors, but they reveal a pattern that generally holds up to statistical scrutiny.⁷

Two clusters of technologies appear in the means and medians. Paper ballots, lever machines, and optically scanned ballots have the lowest average and median residual vote rates. The average residual voting rates of these technologies are significantly lower than the average residual voting rates of punch card and electronic voting equipment. The differences among punch card methods and electronic voting equipment are not statistically significant. Punch cards and electronic machines register residual voting rates for president of approximately 3 percent of all ballots cast. Paper ballots, lever machines, and optically scanned ballots produce residual voting rates of approximately 2 percent of all ballots cast, a statistically significant difference of fully one percent. Or to put the matter differently, the residual voting rate of punch card methods and electronic devices is 50 percent higher than the residual voting rate of manually counted paper ballots, lever machines, and optically scanned ballots. This pattern suggests that simply

⁷ The data in the table only include counties with positive residual vote rate. Approximately 2 percent of counties report negative numbers; these are the figures in the official certified vote. Including counties with negative residual vote rates changes the numbers slightly but does not change the results.

changing voting equipment, without any additional improvements, could lower the incidence of under and over voting substantially.

Table 2

Average Residual Vote By Machine Type
In US Counties, 1988-2000 Presidential Elections

Machine Type	Residual Vote				
	County Average	Standard Deviation	Median	Percent of All Ballots	N
Paper Ballot	1.9	2.1	1.5	1.9	1,540
Lever Machine	1.9	1.7	1.4	1.7	1,382
Punch Card					
“VotoMatic”	3.0	1.9	2.5	2.6	1,893
“DataVote”	2.9	2.7	2.0	2.4	383
Optically scanned	2.1	2.8	1.3	1.6	1,821
Electronic (DRE)	2.9	1.8	2.7	2.2	494
Mixed	2.2	1.8	1.7	1.5	283
Overall	2.3	2.2	1.8	2.1	7,796

Another take on the average reliability of equipment is the percent of all ballots cast for which no presidential vote was registered. This is displayed in the fourth column of numbers: this is the weighted average of the county residual vote, in which we weight by total ballots cast in the county. All of the figures shrink toward zero but the same general pattern holds. In fact, optical scanning seems to do particularly well by this measure. Only 1.6 percent of all ballots cast with optical scanners showed an over vote or no vote over the years 1988 to 2000. Approximately, 1.8 percent of voters cast an over vote or no vote using paper ballots or lever machines. Slightly more than 2 percent of voters cast an over vote or no vote with punch cards or electronics.

To explore the robustness of the pattern further, we isolate specific years. Table 3 presents the residual vote rates for each year of our data.⁸ The bottom row of the table presents residual vote as a fraction of all ballots cast in each year. The entries in the table are the residual vote as a fraction of all ballots cast using each type of technology in each year. It should be noted that year-to-year one expects more random variation in the numbers simply by chance. Every time someone votes on a machine they have a small

⁸ We also present these yearly analyses to set the record straight. A story on cnn.com reports that different people looking at the same data can reach different conclusions. The story cites a separate analysis of the EDS data which suggests that electronics did particularly well in 1996. We have contacted EDS and have confirmed that the pattern of results in Table 3 is consistent with their data. Our data for 1996 come mainly from EDS. When we analyze just the EDS data, we arrive at the same pattern of means, with electronics producing a relatively high average residual vote.

chance of making a random error. Taking averages over many cases gives us a more precise measure of the typical behavior. This is especially true for categories of equipment for which there are relatively small numbers of observations, namely DataVote and Electronics.

Even with this statistical caveat, the yearly averages bear out the same general pattern as the overall averages. In each year, except perhaps 2000, paper ballots and lever machines on the whole have lower residual vote rates than the other technologies. In 2000, paper and levers had relatively low residual vote rates, but so too did scanners and electronics.

Electronics did relatively poorly in 1988, 1992, and 1996. 2000 was the banner year for electronics, but in that year paper ballots and optically scanned ballots had even lower average residual vote rates.

Votomatic punch cards have consistently high average residual vote rates. In 1988, 1996 and again in 2000, punch cards had substantially higher rates of over and under votes than other available technologies. This is of particular concern because approximately one in three voters use punch cards. If election administrators wish to avoid catastrophic failures, they may heed the warning contained in this table and the last. It is the warning that Roy Saltman issued in his 1988 report. Stop using punch cards.

Electronic machines look similarly prone to high residual vote rates, except for 2000, which offers a glimmer of promise for this technology.

Table 3

Residual Vote as a Percent of Total Ballots Cast By Machine Type and Year
US Counties, 1988-2000 Presidential Elections

Residual Votes as a Percent of All Ballots				
Machine Type	1988	1992	1996	2000
Paper Ballot	2.2	1.4	2.1	1.3
Lever Machine	2.0	1.5	1.7	1.7
Punch Card				
"VotoMatic"	2.9	2.2	2.6	3.0
"DataVote"	3.7	2.4	2.1	1.0
Optically scanned	2.5	2.4	1.5	1.2
Electronic (DRE)	3.5	2.5	2.9	1.6
Mixed	2.1	1.4	1.5	2.7
Overall	2.5	2.0	2.1	2.0

Effects of Technology Adoption on Residual Vote Rates

Of course many other factors might explain the observed pattern, including features of the counties and specific elections. The difference between the county and population-weighted averages suggests that county size strongly affects residual vote rates: larger counties typically have lower residual vote rates than smaller counties. We clearly need to hold constant where equipment is used in order to gauge accurately the effects of equipment types on residual vote rates. There are certainly many other factors, such as county literacy rates, education levels, election administration expenditures, other candidates on the ballot, years in which shifts in technology occur.

We hold constant turnout, shifts in technology, other statewide candidates on the ballot, and all factors at the county and state level that do not change dramatically over the period of study, such as literacy rates. To hold these other factors constant we performed a multiple regression of changes in the residual voting rate at the county level on changes in the machine used at the county level, controlling for the year of the election, whether there was a switch in technology in a specific year in a given county, and the total vote in the county. This approach removes the effects of all factors that distinguish the counties, changes in turnout levels within counties, and some features of the election in the state.

In essence, our statistical approach is that of a “natural experiment.” We observe within each county how residual votes change when counties change machine technologies. Between 1988 and 2000, slightly more than half of all counties changed their voting equipment.

The effect of specific technologies on residual votes is expressed relative to a baseline technology. We chose lever machines to serve as this baseline for the contrasts, because levers were the modal machines in 1988. The observed effects contrast the change in residual vote associated with a specific technology compared to a baseline technology. By making multiple comparisons (e.g., paper to lever, scanners to lever, etc.), we measure the relative performance of existing equipment.

We omit counties with Mixed Systems, as it is unclear exactly what technologies are in use. The exceptions are Massachusetts and Vermont, where equipment is uniform within towns: we have collected the information at the town level for these states.

Table 4 reports the observed difference between lever machines and other machine types, along with the “margin of error” (i.e., a 95 percent confidence interval) associated with the observed differences. The complete regression analyses are available upon request. Positive numbers mean that the technology in question has higher average residual vote than lever machines and negative numbers mean that the technology in question has lower average residual vote than lever machines. The wider the margin of error, the less certainty we have about the observed difference. A margin of error in excess of the actual effect means that the observed effect could have arisen by chance.

Table 4 presents results from two separate analyses. One analysis, presented in the first two columns, contains all valid cases. A second analysis, presented in the last two columns, trims the data of extreme cases. To guard against outliers and typographical errors, we omit the cases with lowest 5 percent of residual vote and highest 5 percent of residual vote.

Table 4 bears out the same patterns as Tables 2 and 3. After introducing considerable statistical controls, we reach the same conclusions about the relative performance of different equipment types.

Two clusters of technologies appear in Table 3. Paper ballots, optically scanned ballots, and lever machines appear to perform noticeably better than punch card methods and electronic devices. Paper might even be an improvement over lever machines and scanners.

Table 4

Which is Best?
Residual Vote Attributable to Machine Type Relative to Lever Machines
US Counties, 1988-2000 Presidential Elections

Machine Contrast	All Counties		Excluding Extremes	
	Estimated Difference In % RV	Margin of Error (a)	Estimated Difference In % RV	Margin of Error
Paper Ballot v. Levers	-0.55	+/- 0.37	-0.19	+/- 0.19
Punch Card "VotoMatic" v. Levers	1.32	+/- 0.38	1.11	+/- 0.20
"DataVote" v. Levers	1.24	+/- 0.52	0.97	+/- 0.28
Optically scanned v. Levers	0.11	+/- 0.35	-0.05	+/- 0.19
Electronic (DRE) v. Levers	0.90	+/- 0.30	0.67	+/- 0.16
Number of Cases	7513		7078	

(a) This is the 95 percent confidence interval for the estimated effect; the half-width of the confidence interval equals $1.96 s/\sqrt{n}$, where s is the estimated standard error of the estimated coefficient for each machine type.

First consider the contrast between Paper and Levers. Looking at all counties (the first two columns of the table), the estimated effect of using paper ballots rather than lever machines of is to lower the residual vote rate by approximately one-half of one percent of all ballots cast (i.e., and estimated effect of -0.55). This effect is larger than the margin of error of .37, so the effect is unlikely to have arisen by chance. Omitting extreme cases, the evident advantage of paper ballots over lever machines shrinks: the effect becomes two-tenths of one-percent of ballots cast and this is not statistically different from zero difference between levers and paper.

Second consider optical scanning. The difference in the residual vote rate between scanners and levers is trivial once we hold constant where equipment is used, how many people voted, the year, other statewide candidates on the ballot, and technological changes. In both analyses, the difference between optically scanned ballots and lever machines is quite small and statistically insignificant. Levers and paper and scanned ballots appear to offer similar rates of reliability, at least as it is measured using the residual vote.

The third contrast in the tables is of punch cards to lever machines. Punch card methods produced much higher rates of residual voting. The VotoMatic variety of punch cards produced residual vote rates more than one-percentage point higher than what we observe with lever machines. In our examination of all cases, punch cards recorded 1.3 percent of all ballots less than lever machines did. The estimated effect remains in excess of one-percentage point even after we exclude the extreme cases. The DataVote variety of punch cards looks extremely similar to the Votomatic variety. Because DataVote punch cards have the candidate's names on the card, they were widely believed to be superior to the VotoMatic cards. We find no evidence to support this belief.

A final contrast in the table is between DREs and lever machines. Electronic machines registered significantly higher residual vote rates than lever machines (and, by extension, paper ballots and optically scanned ballots), but DREs do not do as badly as punch cards. Direct Recording Electronic devices had a residual vote rate that was almost one percentage point higher than lever machines, holding constant many factors, including the county. In other words, a county that switches from Levers to DREs can expect a significant rise in residual votes of approximately one percent of total ballots cast. Excluding extreme observations, the effect is somewhat smaller, two-thirds of one percent of all ballots cast. But that is still highly significant from a statistical perspective, and we find it to be a substantively large effect.

One final note about the estimated effect of the DRE performance is in order. Because this machine does not permit over voting, the observed difference in residual vote rates is due to a very significant rise in under voting attributable to electronic devices.

We checked the robustness of our results in a variety of ways. We tried various transformations of the dependent variable, and we split the data into counties of different sizes (under 5000 votes, 5000 to 100,000 votes, and over 100,000 votes). The pattern of results is always the same.

Perhaps the most instructive check on the robustness of our analysis comes when we track changes in equipment usage over time. What happened in the counties that used levers in 1988 in the subsequent three presidential elections? Some of those counts continued to use their lever equipment over the succeeding three presidential elections. Approximately half decided to adopt other technologies and almost all of those that changed went to either electronics or scanners. How did the residual vote rates in these counties compare to 1988?

Table 5

Counties Using Levers in 1988

From 1988 to Current year (92, 96 or 2000) ...

	Change in Residual Vote As % of All Ballots	Avg. Change in County Residual Vote	Median Change in County Residual Vote	N
Kept Levers	-0.21	-0.13	-0.25	520
To Scanners	-0.62	-0.18	-0.32	137
To DREs	0.55	0.73	0.83	250

Baseline Residual vote rate is 1.8 percent in 1988 for counties with lever machines.

Standard Deviation is approximately 0.16 for each group in the first column and 0.17 for each group in the second column.

The rows of Table 5 present three different sorts of counties. The first row shows counties that used lever machines in 1988 and stayed with levers in 1992, 1996, and 2000. The second row represents counties that had lever machines in 1988, but switched to optical scanning in one of the succeeding elections. The third row represents counties that had lever machines in 1988, but switched to DREs in one of the succeeding elections.

The columns of the table present the average change in the residual vote rate from 1988 to the current year. We then average over all years. Consider, for example, a county that had levers in 1988 and 1992, but scanners in 1996 and 2000. The first row includes the observed change in the residual vote rate from 1988 to 1992 for such a county. The second row contains the average change in the residual vote rate from 1988 to 1996 and from 1988 to 2000, the two elections in which the county used scanners.

What happened in these histories? On average, counties that kept their lever machines saw a slight improvement in their residual vote rates from 1988 to 1992, 1996, and 2000. On average, counties that switched to scanners had their residual vote rates fall by even more than the counties that stuck with levers. On average, counties that switched to DREs saw their residual vote rates increase above the residual vote rate that they had in

1988. The difference between the increment in residual vote rate for counties that changed to scanners and counties that changed to DREs is fully one percent of total ballots cast.

What Explains the High Residual Vote Rate of DREs?

We were very surprised by the relatively high residual vote rate of electronic equipment. When we began this investigation we expected the newer technologies to outperform the older technologies. Considering some of the glowing reports about electronics following the 2000 election, we expected the DREs to do well. They did not, especially compared optically scanned paper ballots.

We are not pessimistic about this technology, however. It is relatively new, and we see this as an opportunity for improvement. In this spirit we offer six possible explanations for the relatively high residual vote rates of electronic voting machines.

First, the problems may reflect existing interfaces and ballot designs. The results might stem from differences between touch screens and push buttons or between full-face and paginated ballots (paper and levers are full faced).

Second, there may be a technology curve. As the industry gains more experience with electronics they may fix specific problems.

Third, we may be still low on the voter learning curve. As voters become more familiar with the newer equipment errors may go down. As more people use electronic equipment in other walks of life, such as ATM machines for banking, residual votes may drop.

Fourth, electronics may require more administrative attention, especially at the polling place, and thus be more prone to problems under the administrative procedures used in most counties.

Fifth, electronic equipment may be harder to maintain and less reliable than a piece of paper or a mechanical device. Power surges, improper storage, and software errors may affect DREs.

Sixth, the problem may be inherent in the technology. One speculation is that people behave differently with different technologies. Electronic machines may be simply a less human friendly technology.

There is simply too little data from existing equipment usage to say with confidence what exactly accounts for the relatively high residual vote rate of DREs that we observe. We observe approximately 480 instances of electronic machine usage. When we divide the

cases according to features of the interfaces, there are too few cases to gain much leverage on the questions of interface design. Half of the observations in our data are Shouptronic 1242 machines; another one-quarter are Microvote machines. These are push button, full faced machines. One-in-six are Sequoia AVC Advantage machines. There is not enough variety in machines used or enough observations to accurately measure whether some features of the interface explain the results. Careful, systematic laboratory testing may be required to identify the importance of the interface.

Year-by-year analysis casts some doubt on the notion that there is a voter learning curve. The residual vote rate does not fall steadily for counties using DREs, but jumps around. This variation may owe to the small number of observations in each year. Again, to resolve questions of possible learning or technology curves more detailed analyses and information beyond what we have collected will be required.

Conclusions

Paper ballots, lever machines, and optically scanned ballots produce lower residual vote rates on the order of one to two percent of all ballots cast over punch card and electronic methods over the last four presidential elections.

Lever machines serve as a useful baseline: they were the most commonly used machines in the 1980s, the starting point of our analysis. The incidence of over and under votes with Lever machines is approximately two percent of all ballots cast. The incidence of such residual votes with punch card methods and electronic devices is forty to seventy percent higher than the incidence of residual votes with the other technologies.

We have not analyzed why these differences in residual votes arise. We believe that they reflect how people relate to the technologies, more than actual machine failures. State and federal voting machine certification tolerate very low machine failure rates: no more than 1 in 250,000 ballots for federal certification and no more than 1 in 1,000,000 ballots in some states. Certification serves as an important screen: machines that produce failure rates higher than these tolerance levels are not certified or used. We believe that human factors drive much of the “error” in voting, because the observed differences in residual voting rates that are attributable to machine types are on the order of 1 to 2 out of 100 ballots cast. Given the stringent testing standards for machinery in use, these differences are unlikely to arise from mechanical failures.

We have also not examined many details about the implementation of the machinery, such as manufacturer or precinct versus central counting of ballots or specific ballot layouts.

A final caveat to our findings is that they reflect technologies currently in use. Innovations may lead to improvements in reliability rates. In particular, electronic voting

technology is in its infancy during the period we are studying, and has the greatest room for improvement. It seems the most likely technology to benefit significantly from new innovations and increased voter familiarity.

In the wake of the 2000 election, many state and local governments are reconsidering their choices of and standards for voting equipment. Many manufacturers are seeking to develop or improve machinery. This report identifies a performance standard in practice – an average residual vote not in excess of 2 percent of total ballots cast. With this benchmark in mind, we wish to call attention to the excellent performance of the optically scanned ballots, the best average performance of the newer methods, and especially to the older methods of voting – lever machines and paper ballots.