

Probable Winners:

A Probability Model and Empirical Analysis of Reversals and Accuracy in Election Recounts¹

Stephen Ansolabehere² Andrew Reeves³ Stephen C. Graves⁴

October 17, 2007

¹We wish to thank the Carnegie Foundation and the Knight Foundation for their support of this research and New Hampshire Assistant Secretary of State Anthony Stevens for his helpful comments.

²Corresponding author. Elting Morrison Professor of Political Science, Massachusetts Institute of Technology (Building E53-449, Cambridge, MA 02139; sda@mit.edu; (617)253-5236).

³Ph.D. Candidate, Department of Government, Harvard University (Institute for Quantitative Social Science, Harvard University, 1737 Cambridge St., Cambridge, MA 02138, <http://www.people.fas.harvard.edu/~reeves>, reeves@fas.harvard.edu, (617)496-8300).

⁴Abraham J. Siegel Professor of Management Science, Professor of Mechanical Engineering and Engineering Systems, Massachusetts Institute of Technology (Building E53-347, Cambridge, MA 02139; sgraves@mit.edu; (617)253-6602)

Abstract

This paper examines the probability that recounts reverse initial election outcomes. We develop a counting probability model of election reversals from recounts, which leads to a normal approximation for assessing the likelihood that a recount reverses an initial count. We confirm the model by testing two basic empirical implications of the model using data from New Hampshire towns. We further show how the model may be adapted to more specific circumstances, including variations in the mode of counting and possible asymmetries between candidates in the rate of recovery of votes. Based on an analysis of recounted ballots from 1946 through 2004 in the state of New Hampshire, we find that the discrepancies between initial counts and recounts is nearly twice that for hand counted ballots as for machine counted ballots (1.55 compared to .78). We also find that election recounts rarely change the winners.

The 2000 presidential election recount in Florida exposed potentially serious flaws in the tabulations of ballots and raised doubts about the certainty of election counts. The succeeding years have witnessed intense academic, legal, and public scrutiny to election administration in the United States and throughout the world with the aim of lowering the rate of errors in vote counting. The task is daunting. In a national election in the United States, over 100 million voters are processed on a single day, often by town and county clerks, and those town and county election officials must also count the votes not just for President, but for thousands of local, state, and federal offices. Mistakes will surely occur.

Most academic research, legal debate, and public policy about election reform since 2000 has rightly focused on problems of voter access to the polls and system failures. *Bush versus Gore* raised a more immediate and primitive question: who won? Or to put the matter more generally, how indicative is the initial vote count of the ultimate winner? When an election is a landslide the answer is trivial, but in very close elections where state law, judicial intervention, or just common sense demand a recount, the answer is not so obvious.

In this paper, we consider when counting errors in the initial tabulation of votes are likely to affect the ultimate determination of an election. Our focus is not on the the errors arising in registration databases or polling place operations or even in marking ballots, which have been studied extensively. Rather, our interest is in the final outcome determined by the legal process that leads to the certification of the election results. Specifically, we examine the theoretical and empirical odds that a careful recount of ballots would uphold an initial election. The model and data presented here provide political scientists and legal scholars a theoretical model for gauging the magnitude of election problems in the American and comparative elections and can guide public policy-makers concerned about when recounts ought to occur. The analysis also provides a baseline for assessing when other sorts of problems, such as vote buying, tampering, or other corrupt practices, actually matter.

We present a counting model of the likelihood that a recount reverses an initial tabulation of the vote. Given the large number of ballots typically involved in an election recount, a

normal curve approximates the distribution of votes in a recount well and simplifies the analytical modeling of the probability of a reversal.¹ Analysis of the counting model leads immediately to several insights that can be verified using actual election results.

First, the model implies a new measure of vote counting reliability based on the discrepancy between the initial counts and recounts of ballots in contested elections. We call this measure the *tabulation invalidation rate*, and use it as a metric to compare the reliability of counting methods for paper based systems. In an analysis of all recounts in New Hampshire from 1946 to 1962, we establish a historical tabulation invalidation rate for hand counted paper ballots of .8 percent. We then examine New Hampshire recounts from the 2000, 2002, and 2004 elections and establish invalidation rates for hand counted (1.7 percent) and optically scanned ballots (.7 percent). Controlling for the total town vote, total candidate vote, and the office of the race, we find that the tabulation invalidation rate is approximately eight-tenths of one percentage point lower in towns using scanners.

Second, we consider the probability that an election is reversed due to unreliable initial count. We show that the probability of a reversal drops precipitously with the increase in the total number of ballots cast and the margin of victory in the initial count. For instance, the probability of a reversal is a meager .04 in an election of 50,000 votes decided by a margin of eight tenths of a percentage point. While this margin would launch an automatic recount in many states,² it is extremely unlikely to result in a new victor. We also explore any systematic relationships between the initial count and the true distribution of the votes observed in the recount. We find little evidence of systematic partisan effects; however, we do find that candidates for whom the recount is potentially decisive (those on the cusp of winning or losing) tend to receive a *smaller* vote share in the recount than they do in the initial count.

¹Poisson models of errors have also been proposed, and might be appropriate for very small elections. Poisson models are difficult to adapt to recounts because the total number of ballots counted are not constant between the initial and recount; it is easier to analyze proportions.

²For example, Maine, Nebraska, and South Carolina have laws that require automatic recounts if the margin between winner and loser is less than 1%.

1 Using recounts to measure accuracy

A number of important studies of the performance and accuracy of voting technologies have sought to measure the error rate of vote tabulations. The main metric that emerges from these evaluations uses “residual votes” — the discrepancy between total ballots cast and votes cast for a particular office, such as president or governor. The incidence of residual votes should be unrelated to the type of technology used, and the difference in residual votes across technologies measures the extent of errors in the casting or tabulation of votes that are attributable to specific technology. Similar jurisdictions using different technologies ought to have the same residual vote rate, on average. By this metric, hand-counted paper ballots and optically scanned paper ballots have shown better overall performance than punch cards, lever machines, and electronic voting machines.³

Residual votes offer the best available general measure of the overall error rate of the voting system, but they are at once too expansive and too narrow to ascertain whether a recount would reverse an initial count. The discrepancy between the overall vote and the vote for any office reflects voters’ confusion as well as tabulation errors, and voter errors, such as arose with the butterfly ballot, cannot be corrected in a recount.⁴ Also, residual votes will miss some tabulation errors that affect recounts. Specifically, tabulation errors that incorrectly give a candidate a vote that should have gone to another candidate will make the count deviate from the true vote. However, since the votes are counted these errors are not reflected in the discrepancy between total ballots cast and votes cast for a given office. A second sort of measure is needed to gauge the magnitude of such problems.

The agreement between initial counts and recounts in disputed elections offers a direct measure of the magnitude of tabulation errors. We term this the *tabulation validation rate*.

³See Caltech / M.I.T. Voting Technology Project (2001) and Ansolabehere and Stewart (2005) for further definition and discussion of residual votes. Studies using similar measures include Knack and Kropf (2003), whose results are consistent with the Caltech/MIT study, and Wand et al. (2001), who find similar problems with punch cards but suggest somewhat better performance for electronic equipment.

⁴For a discussion of the complexity of errors possible see Niemi and Herrnson (2003) and Wand et al. (2001).

Likewise, the discrepancy between the initial count and the recount is the tabulation invalidation rate. When there is a legal challenge to an election or a mandated recount, state or local election offices conduct an audit of votes cast. For paper ballot systems, election officials reexamine the ballots to determine voter intention. Tabulations may change from the initial count to the recount for a variety of reasons: ballots may be mishandled; machines may have difficulty reading markings; people and machines may make tabulation errors. Because recounts are used to certify the vote, greater effort is taken to arrive at the most accurate accounting of ballots cast. The initial count of ballots, then, is treated as a preliminary count, and the recount as the official count. The recount, then, validates the initial tabulation.

Two caveats accompany the use of recounts to measure the validity of tabulations. First, this measure only concerns tabulation. It ignores other factors, such as voter confusion about how to mark a machine-readable ballot, which might lead people to vote accidentally twice for an office or not at all. Such phenomena concern the overall quality of the election system, rather than the likelihood of a reversal, and are captured by residual votes.

Second, the metric differs for paper ballot systems and elections in which votes are cast on lever machines or electronic voting machines (also known as Direct Recording Electronics, or DREs). Mechanical and electronic voting machines do not retain a separate record of the voters' intention, so it is impossible, at least with current technology, to compare voters' intentions with the machines' recordings. Recounts with mechanical and electronic voting machines merely capture whether the election office made a recording error or whether the machine is functioning. With direct recording mechanical and electronic machines it is impossible to gauge the degree of malfunction or the disparity between voter intentions and machine recordings.

2 Recounts and Reversal Rates

Two fundamental questions about the reliability of elections have emerged since 2000. First, who actually won? Second, can we make the system better? We consider the first of these questions through the lenses of recounts. Comparing recounts with initial counts reveals the rate of failures or errors in initial counts and the incidence with which reliance on just the initial counts would lead election officials to declare the wrong candidate the winner. To this end, we examine the probability that an election is reversed due to an unreliable initial count.

2.1 A Model of Recounts and Reversals

We seek to compare two distributions — the true distribution votes and the distribution with tabulation errors. We will assume that these errors occur at random, though that assumption can be relaxed (see below). The discrepancies between these may then be viewed as a random variable and analyzed appropriately. We treat the recounts accepted by the courts or Secretary of States' offices as the true or official count. That is, the recount is the true distribution of the votes among the candidates. In every election we observe the proportion of the votes that each candidate receives (p_i), but only rarely are the votes recounted so that the *true* distribution is actually observed (ρ_i).

While other research has focused on votes counts (e.g. Wand et al. (2001)), we focus here on proportions of votes received by the candidates.⁵ The probability of a reversal is based on the margin of victory in the initial count ($p_1 - p_2$), the error in the initial count ($\rho_i - p_i$), and the total number of ballots cast (N). Using the normal approximation of the binomial distribution, we can calculate the probability of a reversal in a two candidate election as

⁵Pure counting models in this area encounter the complication that the number of valid ballots increases between the initial and recount. As a result it is easier to model proportions.

$$P(\textit{Reversal}) = 1 - F\left(\frac{p_1 - \rho_1}{2\sqrt{\frac{\rho_1(1-\rho_1)}{N}}}\right). \quad (1)$$

In the case where $\rho_1 = .5$, the model simplifies to

$$P(\textit{Reversal}) = 1 - F\left(\frac{2p_1 - 1}{\sqrt{\frac{1}{N}}}\right). \quad (2)$$

This model allows us to derive comparative statics for the two parameters are key – the number of ballots and the margin. The probability of a reversal drops precipitously as the size of the margin and especially the total number of valid ballots increases.

Figure 1 graphically illustrates the probability of a reversal for varying margins of victory and for elections of different sizes. This figure uses Equation 2 and assumes that the true underlying division of votes is split between candidate 1 and candidate 2 equally (i.e., $\rho_1 = .5$ and $\rho_2 = .5$). Each plot also includes the margin for which the probability of a reversal is less than .05. For instance, in an election of one hundred votes determined by a margin of more than eight percent (8.3%), candidate 2 is very unlikely to win in a recount — the probability of a victory is less than .05. In an election of 50,000 votes, a reversal is extremely unlikely with a margin of approximately four tenths of one percent.⁶ As shown in Equation 1, when $\rho_1 = .5$, the probability of a reversal declines in proportion to the size of the election at the rate of $\sqrt{\frac{1}{N}}$.

2.2 Reversals in New Hampshire

We examine data on recounts for New Hampshire towns for several reasons. First, New Hampshire has a lot of recounts. Although recounts are generally rare, New Hampshire recounts about one in four ballots at each election — this provides a large number of cases

⁶Consider the 2006 contest between George Allen and Jim Webb for U.S. Senate in Virginia. Webb beat Allen 1,175,606 to 1,166,277 — a margin of less than four tenths of one percent. The probability of a reversal in a recount is approximately 4.3×10^{-10} .

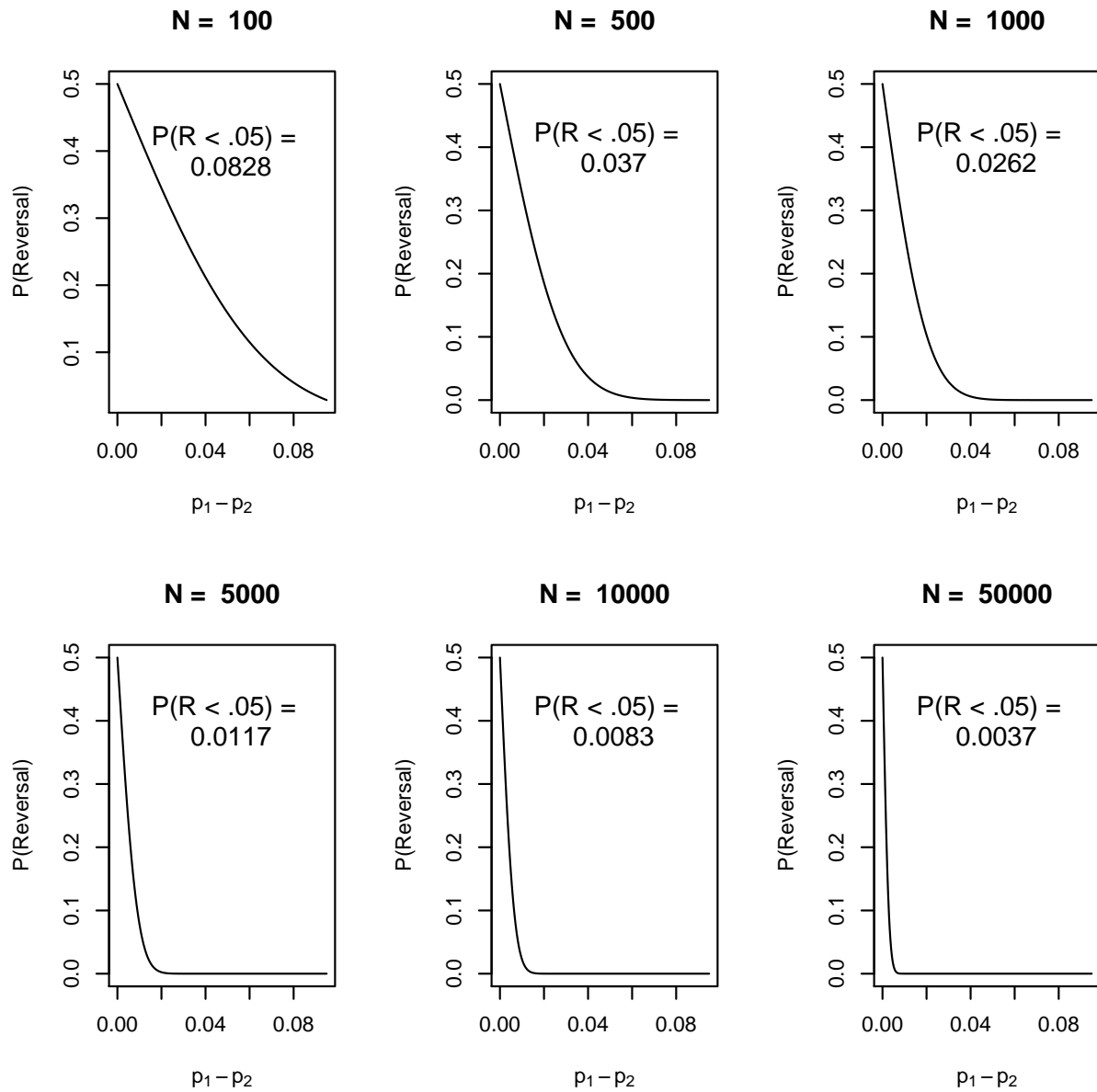


Figure 1: The probability of a reversal given the margin of victory for candidate 1 in the initial vote count ($p_1 - p_2$) for varying election sizes. The true proportion of candidate 1's votes is ρ_1 , which, in this example, we assume is .5. The margin at which the probability of a reversal is less than .05 is also given for each N . The probability that a recount reverses an initial count drops precipitously as p_1 and N increase.

to study.⁷ Second, New Hampshire has a uniform reporting system for recounts dating back at least to the 1940s. The data on recounts are, then, comparable and the historical record allows us to establish a solid baseline against which to contrast new tabulation methods. Third, New Hampshire uses hand-counted paper ballots extensively as well as optically scanned ballots. This allows us to contrast new and old methods of counting directly. Fourth, the number of ballots and margins of victory involved in recounts varies, allowing us to see directly the implications of the probability model.

Historical recounts provide an important baseline. Over a long period of time across many offices, we can examine the percent difference between initial counts and recounts.

We examine historical recount data in the state of New Hampshire from 1946 to 1962. The first use of punch card machinery in the United States occurs in 1964. In the general elections in this span, 108 races had recounts. The offices involved range from registrar of voters to representative of Congress as well as town questions. The majority of the recounts (69) occurred in races for state representative. All recounts are aggregated to the town level except those for Congressional races and state senate races which were aggregated to district levels.

These data provide complete and consistent information about the accuracy of hand counted votes for 9 elections over sixteen years. Since all ballots during this time were hand counted once in the initial count and again in the recount we are able to evaluate accuracy rates of the counts independent of residual votes.

Our analysis consists of 415 cases where a case is a reported ballot tabulation of a town or (district) for each candidate running in a race. We compute the invalidation rate for each candidate for each office in each jurisdiction. That is, we compute the percent difference between the initial count and the recount for the votes recorded in each jurisdiction (i.e., town or ward) for each candidate seeking an office.

Two different average invalidation rates are of interest. First, the simple average invali-

⁷We wish to thank New Hampshire Assistant Secretary of State Anthony Stevens for providing us with this information.

dation reflects what happens in a typical jurisdiction. In a state with many small towns and a handful of larger towns, the typical jurisdiction will be a small one. Second, the weighted average, weighting by total ballots cast, equals the percent of ballots cast that differed between the counts. This measure guards against the occasional aberrant tabulation in a small community. The weighted average measures the frequency with which one's vote is counted in the initial count with each type of equipment. We will focus on the latter, but will report both.

Historically, hand-counted paper tabulations have an invalidation rate just under 1 percent. The weighted average of the discrepancy between the initial count and the recount is approximately 0.83 percentage point with a standard error of 0.11, and that tabulation error rate in hand counts remains fairly constant over time.

More recent elections use a mix of technologies, but offer more detailed data on recounts. For now we will ignore that complexity and consider what can be learned about the probability of reversal from the pooled election data. Sixty races were recounted in the general and primary statewide elections of 2000, 2002, and 2004.⁸ The majority of the recounts (47) were for state representative races. Recounts were also conducted for state senate (7) and county commissioner (5) races as well as a single recount for a county attorney contest. Data on the recount are reported at the town and sometimes ward level. Because each race typically involves many towns, the 60 recounted races yield 1,356 cases — where a case is a recounted vote total for each candidate in each town in the district. Again, we distinguish between the average percent difference and the percent of all ballots cast (regardless of size of jurisdiction).

Our analysis here considers the changes between the initial counts and recounts. In the next section we will distinguish methods of vote tabulation. Again, we compute a weighted mean of the average absolute percent difference between the initial count and the recount. We present the data with and without two exceptional cases. In the 2000 general election a

⁸We exclude the recounted race for District 1 Executive Council in 2004 in the Democratic primary election because one of the candidates received only write-in votes.

faulty machine missed nearly 2,000 votes in the third ward of the town of Dover in a Strafford County commissioner’s race — this error was discovered and corrected in the recount and did not affect the outcome of the race. Likewise, a recount in the town of Bradford yielded tabulations differing from 15 to 20 percent.⁹

How well does the data from New Hampshire conform to our model? We examine reversals in the sixty recounted races in the general and primary elections in New Hampshire from 2000, 2002, and 2004. First we examine how well our model does in predicting reversals and then examine factors that may effect how a vote is counted in the recount.

Because there are multiple candidates in most of the New Hampshire races, we can make a simple adaptation to our model to take into account the increased variance brought about as a result of more than two candidates in the race. The probability of a reversal is

$$\frac{(p_1 - p_2) + (\rho_2 - \rho_1)}{\sqrt{\rho_1(1 - \rho_1) + \rho_2(1 - \rho_2) + 2\rho_1\rho_2}} \quad (3)$$

Where p_1 is the vote share of the winning candidate with the lowest vote total in the initial count (the cusp winner) and p_2 is the vote share of the losing candidate with the highest number of votes in the initial count (the cusp loser). This correction must be made to take into account the covariance between the candidates Ansolabehere and Belin (1993).

Figure 2 plots the predicted probability of a reversal along with the margins from each race between the winning and the losing candidates. Figure 2 also indicates the races reversed in the recount by plotting them as triangles. The findings from the observed empirical data generally agree with the predictions from the model in the previous section. As the margin between candidates shrink, the probability of a reversal plummets. In the New Hampshire case, reversals occur in just over thirteen percent of the recounted races (eight out of sixty). Based on our model, our predictions are relatively sound with no reversals occurring among cases with predicted probabilities of .20 or lower. If New Hampshire opted not to recount elections where the probability of a reversal is less than, say, .05, than over thirteen percent

⁹We have confirmed both of these irregularities with the New Hampshire Secretary of State’s office.

of its recounts would have been deemed unnecessary. Based on our analysis here, had this decision rule been in place, there would be nil effect on the election results — only the saved resources from thirteen percent of New Hampshire’s requested recounts.

Since the 2002 election the debate over voting technology has focused on the ability to hack into election system software (Kohno et al., 2004). Many advocates have pushed for hand counted paper ballots as a solution, and have alleged that the tabulation software of optical scanners and electronic voting machines have been tampered with. These arguments suggest that the tabulation error rates should vary across modes of counting, manual and machine. Tabulation validation rates provide direct evidence concerning these matters. In particular, we can discover whether optical scanners have unusually high reversal rates (due to malicious code in tabulation software) and whether hand counting offers superior performance in tabulations. The answers are no and no.

While most counties and towns continue to use hand-counted paper, many of the towns in three counties had shifted to optically scanned paper ballots. Two sorts of scanners were used: Sequoia-Pacific’s Optech and Global’s (now Diebold’s) Accuvote.¹⁰

Study of tabulation invalidation rates is of particular importance. Since the 2002 election a debate has raged in the United States over the ability to hack into election system software (Kohno et al., 2004). The tabulation software of optical scanners in particular has been questioned for its system security. This debate among political scientists, political activists, and computer scientists has proceeded on the basis of what might occur, but has occurred in a complete vacuum of hard empirical evidence. Tabulation validation rates provide some evidence concerning these matters.

Recounts allow us to see more precisely whether the introduction of voting machines has improved the tabulation of votes. Do machines have higher validation rates than hand-counted paper ballots? Independent of this study, Herron and Wand (2007) have studied the presidential recount in New Hampshire to measure whether there is partisan bias in

¹⁰Voting machinery of New Hampshire towns obtained from <http://www.nh.gov/sos/voting%20machines.htm>, accessed 22 May 2002 and 1 March 2006.

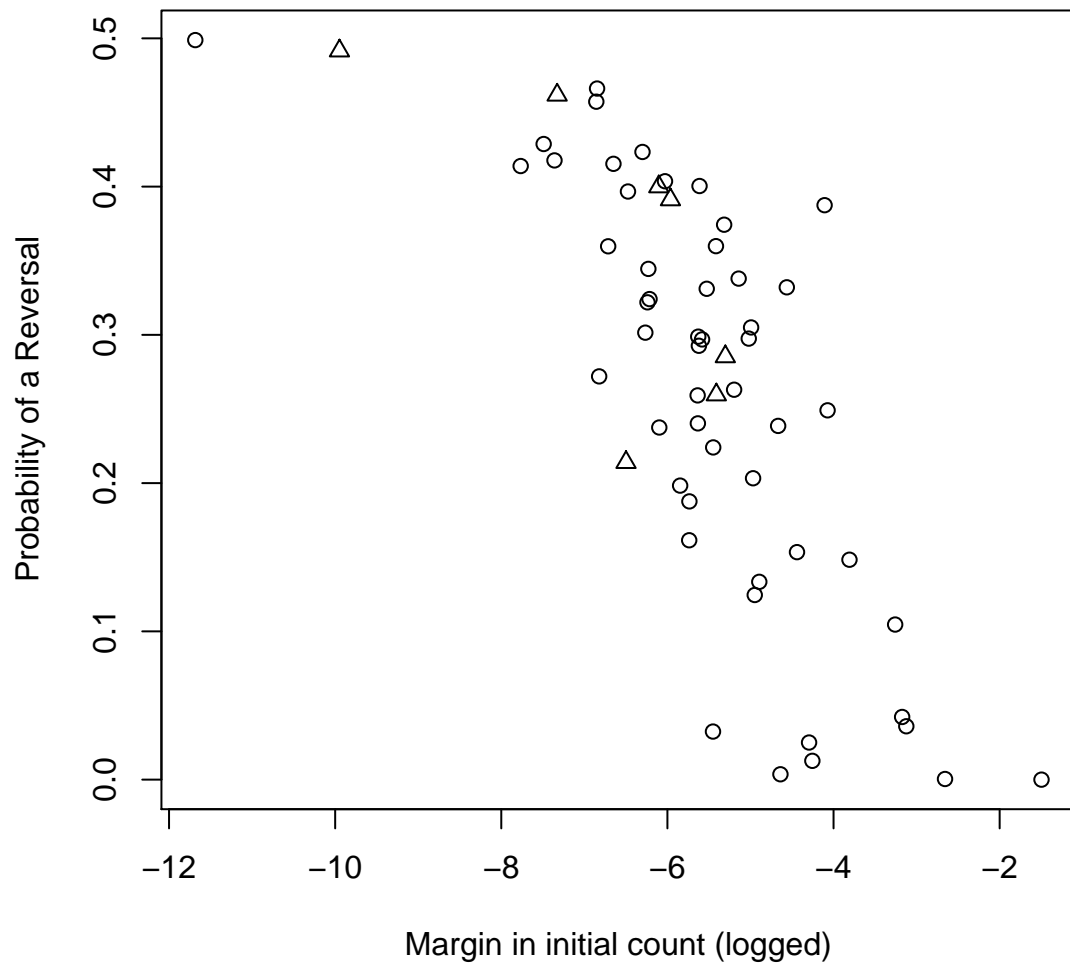


Figure 2: **The probability of a reversal and initial differences in the vote count.** Predicted probabilities of a reversal for 60 races in New Hampshire General and Primary elections. From the earlier discussion, the x axis is $p_1 - p_2$ (logged). Elections where recounts resulted in a reversal are indicated by a triangle (Δ). As the margin of victory increases, the probability of a reversal declines.

tabulation; their analysis also speaks to the question of “hacking,” as deliberate vote stealing would have a partisan slant. We think a prior and necessary question, is whether such tabulation errors are of a large magnitude, and whether machines show worse tabulation error rates than hand counting, or in fact are an improvement.

Recounts allow us to see more precisely whether the introduction of voting machines has improved the tabulation of votes. Do machines have higher validation rates (and lower invalidation rates) than hand-counts?

Our analysis here considers the changes between the initial counts and recounts for each method of vote tabulation. Again, we compute a weighted mean of the average absolute percent difference between the initial count and the recount. We present the data with and without two exceptional cases. In the 2000 general election a faulty machine missed nearly 2,000 votes in the third ward of the town of Dover in a Strafford County commissioner’s race — this error was discovered and corrected in the recount and did not affect the outcome of the race. Likewise, a recount in the town of Bradford yielded tabulations differing from 15 to 20 percent.¹¹

Table 1 presents the invalidation rates of hand-counted and scanned ballots in New Hampshire for the 2000, 2002, and 2004 general and primary elections. The invalidation rate for hand-counted paper ballots is 1.83 percent — higher than the historical average. The difference between hand counted and optically scanned paper was 1.16 percent. Scanned ballots have a significantly lower invalidation rate than hand-counted paper. Figure 1 presents these data graphically. This figure shows the similar invalidation rates among Accuvote and Optech technologies, both of which are lower than hand counted ballots. Additionally, the extent to which Bradford (indicated by triangles (\triangle) and Dover Ward 3 (indicated by pluses (+)) are outliers.¹²

The differences in the table may owe to the type of equipment or the size of the commu-

¹¹We have confirmed both of these irregularities with the New Hampshire Secretary of State’s office.

¹²The invalidation rates for hand counted ballots higher than Bradford are a result of very small initial votes tabulations greatly affected by small changes in the recount.

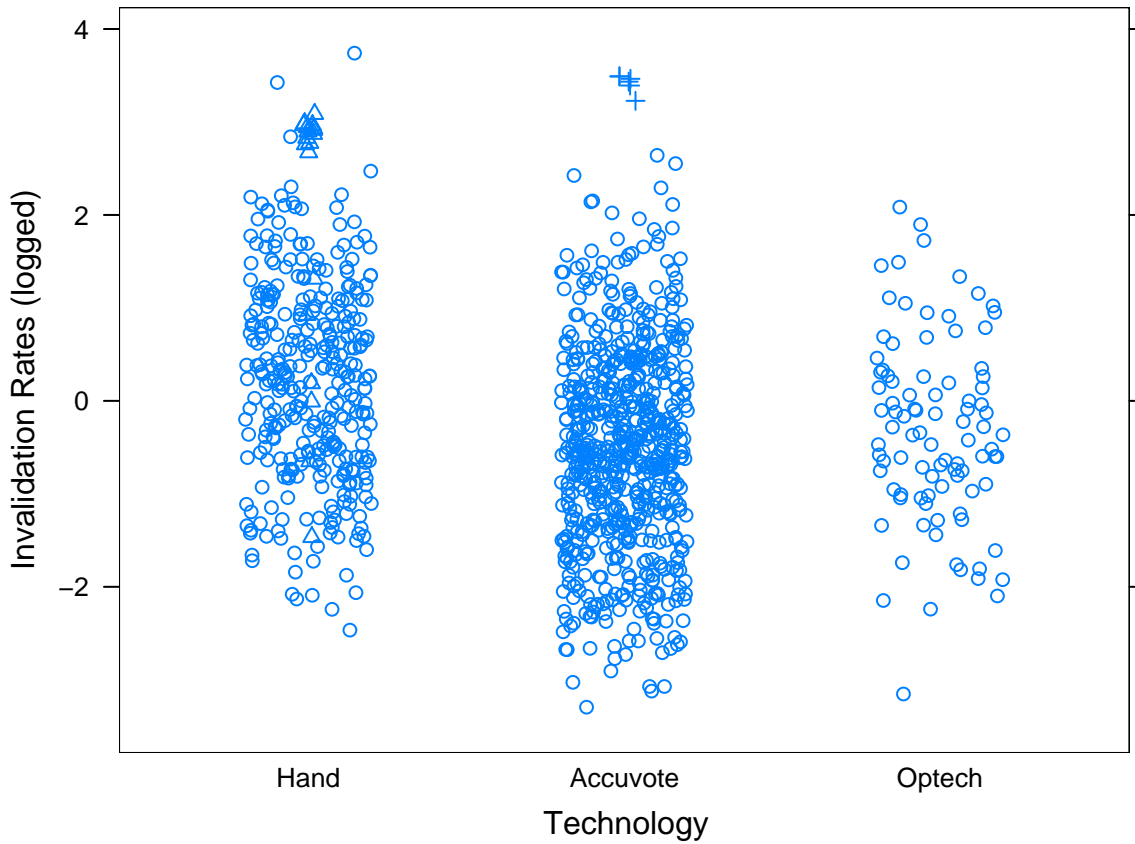


Figure 3: Invalidation rates by technology type for New Hampshire general and primary elections in 2000, 2002, and 2004. Bradford is indicated by triangles (Δ) and Dover Ward 3 is indicated by pluses (+). Groups are arranged from highest to lowest means (including outliers). While the inactivation rates for paper ballots are slightly higher than scanned ballots, the difference between Accuvote and Optech (ignoring outliers) is statistically indistinguishable. These data are summarized in Table 1.

	<i>N</i>	Weighted Average	Median	Standard Error	Maximum ^a
Hand-Counted	454	2.10	0.86	0.15	42.11
Hand-Counted (without Bradford)	435	1.83	0.81	0.12	42.11
Machine Counted	903	0.93	0.50	0.10	32.80
Machine Counted (without Dover Ward 3)	897	0.67	0.50	0.04	14.03
ACCUVOTE	787	0.96	0.50	0.03	32.80
ACCUVOTE (without Dover Ward 3)	781	0.67	0.49	0.04	14.03
OPTECH	116	0.69	0.53	0.09	8.04

Table 1: **Invalidation Rates of Hand-Counted and Scan Ballots**, New Hampshire 2000, 2002, and 2004 General and Primary Elections

^aThe minimum values is 0 in all cases.

nity. Smaller communities are more likely to have a large discrepancy in their tabulations and they are more likely to use paper ballots. We can untangle these effects by controlling for the office at stake and the size of the vote cast. We performed two regression analyses to estimate the effect of *Scan Ballot*, which equals 1 if the town used optically scanned ballots and 0 if the town used hand-counted paper ballots. We control for total votes cast in a town and the initial count for a candidate. We also control for the office at stake, as some elections spanned several towns. This allows us to hold constant the candidates on the ballot. Table 2 presents two analyses, one of which removes the effect of office (see the column labeled Fixed Effects) and one does not (see the column labeled Least Squares). We present both as a robustness check.

The key coefficient of interest is the effect of *Scan Ballot*. The coefficient in both specifications is approximately -.8, which means that the percentage difference between the initial count and the recount is approximately eight-tenths of one-percentage point lower in towns

	Least Squares	Fixed Effects for Office
	b (S.E.)	b (S.E.)
Constant	1.84 (0.11)	1.48 (0.18)
Scan Ballot	-0.78 (0.13)	-0.77 (0.13)
Town Vote (10,000s)	-0.03 (0.06)	-0.06 (0.06)
Initial Candidate Vote (10,000s)	-0.13 (0.04)	0.26 (0.16)
R^2	0.04	0.04
N	1331	1331

Table 2: **Estimated Effect of Machine v. Hand Count** on Percent Change in Votes Counted, 2000 – 2004

using scanners than in towns using hand-counted paper, holding constant the initial count, the total vote, and the office sought. These coefficients are slightly smaller than the observed difference between the mean for hand-counted and machine counted (which is $1.7 - .8 = .9$).

To see the practical effect of these estimates consider an election with 10,000 votes where the candidate in question received exactly half of the votes.¹³ The predicted discrepancy between counts is 1.55 percent if the town tabulates by hand (i.e., $1.48 - .06 \times 1 + .26 \times .5 = 1.55$) and 0.78 percent if the town uses an optical scanner (i.e., $1.48 - .77 - .06 \times 1 + .26 \times .5 = .78$).

Studying recounts allows us to observe the *true* underlying distribution of the votes, which we do not observe when votes are counted only once. Despite having a dramatic impact on the likelihood of a reversal, it is the discrepancy between the initial count and the distribution of the votes ($p_i - \rho_i$), that researchers know the least about. Election recounts

¹³We caution against making out of sample predictions. The town with the largest number of votes cast is Keene, which recorded approximately 45,000 votes in a 2002 race. The effect of town vote likely tends to an asymptote and the linear specification used cannot capture the predicted discrepancy for larger communities.

themselves may be controversial, influenced by who won the initial count and who did the recounting.

In our statistical model, we consider the relationship between the observed proportion of candidate 1's (or the cusp winners's) vote in the initial count (p_i) and the recount (ρ_i). We now turn to an analysis of the difference between these two counts of the same votes. Although random error may cause the two counts to not be identical, we test for more devious systematic factors that may cause the initial count to vary from the recount. The dependant variable here is the difference in vote share between the initial count and the recount for each candidate in general election contests in 2000, 2002, and 2004. Using the notation from earlier in the paper we can express the dependent variable as $p_i - \rho_i$. Independent variables include the partisanship of the district, measured as the district's percent of the two party vote cast for Bush in 2000 and 2004 (if the state candidate is a Republican) or Gore and Kerry in 2000 and 2004 respectively (if the state candidate is a Democrat). We also include an indicator variable for if the candidate is Republican to gauge any systematic partisan differences in the vote count. Although we treat the recount as the true count, it may be the case that bias and corruption occurs when the votes are recounted. To account for the possibility that the recount might favor the cusp candidates, we include an indicator variable of the cusp winner (the winning candidate with the lowest vote total in the initial count) and the cusp loser (the losing candidate with the highest number of votes in the initial count) for each contest.¹⁴

Table 3 presents a least squares regression described above. The results in Table 3 provide no evidence for systematic political bias in recounts for the 2000, 2002, and 2004 general elections in New Hampshire. In addition to the size of the coefficients being extremely small,

¹⁴For instance, consider the following race for two legislative seats:

Candidate	Votes
A	50
B	45
C	44
D	30

In this two seat example, candidate B is the cusp winner and candidate C is the cusp loser.

the standard errors are relatively large. Interestingly, the cusp candidates, both winner and loser, do appear to have a slight, positive differences between the initial count and recount than non-cusp candidates. Cusp candidates tend to receive about two tenths of a percentage point more of the initial vote share than in the recount. One possibility is that the vote counters perceive the closeness of the races and attempt to affect the outcome; although it is equally possible that the recounted votes of the cusp candidates are put under closer scrutiny and are more likely to be challenged and excluded from the recounted votes.

	b (S. E.)
Constant	0.0004 (0.0006)
Candidate is Republican	0.0002 (0.0008)
District Partisanship	0.00010 (0.00007)
Cusp Loser	0.0019 (0.0009)
Cusp Winner	0.0020 (0.0008)
R^2	0.07
N	157

Table 3: **Biases in election recounts** based on political party of the candidate, their success in the initial count and the partisanship of the district, 2000, 2002, and 2004 General elections only.

3 Conclusions

The theoretical analysis offered here cuts to a fundamental democratic issue, the legitimacy of elections. Recounts are rare, especially considering the magnitude of the problems with election administration that have been uncovered in recent years. How is it that even after the problems unearthed in Florida and other states that we still readily accept elections

decided by only small margins? The reason, we think, has to do with the laws of probability at work. In a large electorate, a recount is highly unlikely to alter the results of any election except for the slightest of margins. This result is little more than the law of large numbers at work in elections, a law whose operation has been noted at least since Condorcet and the Enlightenment

That insight coupled with the data mustered here speak immediately to four practical questions.

First, how confident should we be in vote counts? The tabulation invalidation rate was low, especially for optical scanning. However, it was not trivial. In a US House election with 250,000 votes, the invalidation rate of .007 for scanners amounts to 1750 votes. If the race is close, say split 50-50, then the tabulation might affect the margin between the candidates by as much as 1.4 percentage point of the vote (i.e., if .7 percent of the votes for candidate A went to candidate B and candidate B lost none of her votes).

Second, is the fear of hacking a legitimate concern? The data here are inconsistent with assertions of vote stealing through the tabulation software of optical scan equipment. The discrepancies with scanners were consistently small and showed no partisan tilt. While security concerns over hacking may be relevant, there is no evidence that hacking has been a factor in New Hampshire elections. The methods here offer an easy way to check for tampering of optical scan software.

Third, What do tabulation errors tell us about the relative quality of voting technology? Optical scanning performed better than hand counted paper ballots. Looking at recounts from 2000, 2002, and 2004, controlling for total vote, initial vote, and office sought, optically scanned paper produced a lower discrepancy between the initial count and the recount compared to hand-counted paper.

Fourth, when should we have a recount? Most states have no set standard for an automatic recount; the courts or state election officer decides whether a recount is appropriate. Some states, however, have provisions for automatic recounts – typically, if the election is

closer than one-half of one percent of ballots cast (Bauer and Rylander, 2002). This standard appears inadequate in three respects. First, in very large electorates it may be excessive. Second, in very small (local) elections it may be too restrictive. Third, it ought ultimately depend on the potential for biases. However, as a general rule, an automatic recount standard, such as .5 percent provides at least some guide as to when recounts should be undertaken. Refinement of such a standard requires further analysis of election recounts, especially to determine the generality of the results offered here. At the very least, though, the simple probability model presented in this paper identifies the key quantities that must be studied to further understanding of the legitimacy of vote counts in American elections.

A Derivation

The probability of a reversal is based on two elements: the total number of votes cast (N) and the discrepancy between candidate i 's proportion of the vote in the initial count (p_1) and the proportion of the vote in the recount (ρ_i).

Like drawing balls of two colors from an urn, vote counts between two candidates can be modeled using the binomial distribution. Because we consider two candidates, the proportion of the vote candidate 1 receives in the initial count is p_1 and the proportion of candidate 2's vote is $1 - p_1$. Studying recounts allow us to observe the *true* proportion of that vote that candidate i receives, which we represent as ρ_i .

Using the binomial distribution, we can calculate the probability of candidate 1 receiving n_1 votes (where $n_1 = N \times p_1$) as

$$P(Y = n_1 | \rho_1) = \binom{N}{n_1} \rho_1^{n_1} (1 - \rho_1)^{N - n_1}. \quad (4)$$

By taking the conditional distribution function (cdf) of the binomial distribution, we can calculate the probability of a reversal as

$$1 - F(Y = n_1) = P(Y < n_1), \quad (5)$$

where $F(\cdot)$ is the conditional distribution function for the binomial distribution.

We can also use the normal approximation of the binomial distribution to approximate the probability of a reversal where

$$P(Y = p_1) \sim N(0, 1),$$

$$P(Y = p_1 | \rho_1) = \frac{p_1 - \rho_1}{\sqrt{\frac{\rho_1(1 - \rho_1)}{N}}}, \quad (6)$$

$$P(Y = p_1 | \rho_1 = .5) = \frac{2p_1 - 1}{\sqrt{\frac{1}{N}}}, \quad (7)$$

and

$$P(\text{Reversal}) = 1 - F\left(\frac{(p_1 - \rho_1)}{\sqrt{\frac{\rho_1(1-\rho_1)}{N}}}\right) \quad (8)$$

If N is large enough, we can use the χ^2 distribution to provide a computationally different but substantively similar statistic based on how similar the initial count is to the recount. This also provides as a measure of when a reversal is likely. Where

$$\chi_1^2 = \frac{(Np_1 - N\rho_1)^2}{N\rho_1} + \frac{[N(1 - p_1) - N(1 - \rho_1)]^2}{N(1 - \rho_1)}. \quad (9)$$

References

- Ansolabehere, Stephen and Charles Stewart III. 2005. "Residual Votes Attributed to Technology." *Journal of Politics* 67(2):365–389.
- Ansolabehere, Stephen and Thomas R. Belin. 1993. "Poll Faulting." *Chance* 6(2):22–27.
- Bauer, Robert F. and Jason C. Rylander. 2002. "Task Force on Constitutional and Federal Election Law, Recounts and Contents" in *To Assure Pride and Confidence in the Electoral Process: A Report of the National Commission on Federal Election Reform*. Washington, DC: Brookings.
- Caltech / M.I.T. Voting Technology Project. 2001. *Voting: What Is, What Could Be*. Pasadena and Cambridge: Caltech and M.I.T.
- Herron, Michael and Jonathan Wand. 2007. "Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount." *Electoral Studies* 26(2):247–261.
- Knack, Stephen and Martha Kropf. 2003. "Voided Ballots in the 1996 Presidential Election: A County-Level Analysis." *Journal of Politics* 65(3):881–898.
- Kohno, Tadayoshi, Adam Stubblefield, Aviel D. Rubin and Dan S. Wallach. 2004. "Analysis of an Electronic Voting System." *IEEE Symposium on Security and Privacy*.
- Niemi, Richard G. and Paul S. Herrnson. 2003. "Beyond the Butterfly: The Complexity of U.S. Ballots." *Perspectives on Politics* 1(2):317–326.
- Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Seekhon, Walter Webane Jr., Michael Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95(4):793–810.