

Voting Technology and the 2008 New Hampshire Primary¹

Michael C. Herron²

Walter R. Mebane, Jr.³

Jonathan N. Wand⁴

This Draft (4): January 21, 2008

Initial Draft: January 19, 2008

¹The authors thank Jasjeet Sekhon for computing resources and advice. Authors are listed alphabetically.

²Associate Professor of Government, Dartmouth College. 6108 Silsby Hall, Hanover NH 03755-3547. Phone: 603-646-2693. Email: Michael.Herron@dartmouth.edu.

³Professor of Political Science and Professor of Statistics, University of Michigan. 7735 Haven Hall, Ann Arbor, MI 48109-1045. Phone: 734-763-2220. Email: wmebane@umich.edu.

⁴Assistant Professor of Political Science, Stanford University, and Robert Wood Johnson Health Policy Scholar, University of Michigan. Encina Hall, 308 West, Stanford CA 94305. Email: wand@stanford.edu.

Executive Summary

We address concerns that the reported vote counts of candidates running in the 2008 New Hampshire Presidential Primaries were affected by the vote tabulating technologies used across New Hampshire.

- In the Democratic Primary, Hillary Clinton was more successful in New Hampshire wards that used Accuvote optical scan vote tabulating technology than was Barack Obama, receiving 4.3 more percentage points of the vote there (40.2% for Clinton versus 35.9% for Obama). In contrast, Clinton did worse than Obama in wards that counted paper ballots by hand, trailing by 6.1 percentage points (33.7% versus 39.8%).
- In the Republican Primary, Mitt Romney trailed John McCain by 3.6 points in Accuvote wards and by 15 points in wards that counted ballots by hand.
- In New Hampshire the choice of vote tabulating technology is made ward by ward, and electronic technology was used in wards that typically differ demographically and politically from wards that count ballots by hand. Wards that selected electronic tabulation are disproportionately from the southeast part of New Hampshire, and they tend to be more densely populated and more affluent. Accuvote and hand count wards have also typically produced divergent voting patterns in elections prior to the 2008 primary. It is plausible that most or all of the observed differences between vote tabulating technologies in the votes candidates received reflect such background differences and not anything inherent in the tabulation methods.
- Using a subset of New Hampshire wards that have similar demographic features and voting histories but differ in their vote tabulating technologies, we find no significant relationship between a ward's use of vote tabulating technology and the votes or vote shares received by most of the leading candidates who competed in the 2008 New Hampshire Presidential Primaries. Among Clinton, Edwards, Kucinich, Obama and Richardson in the Democratic primary and Giuliani, Huckabee, Paul, Romney and McCain in the Republican primary, we observe a significant difference only in the votes counted for Edwards, and that difference is small (a deficit of between 0.6 and 3.4 percent in the hand-counted votes).
- With respect to Hillary Clinton's surprise victory in the Democratic Primary and the differences across vote tabulating technologies in Clinton's and others' votes, our results are consistent with these differences being due entirely to the fact that New Hampshire wards that use Accuvote optical scan machines have voters with different political preferences than wards that use hand counted paper ballots.

1 Introduction

There have been widespread concerns that the reported vote counts of candidates running in the 2008 New Hampshire Presidential Primary were affected by the technologies used in New Hampshire to tabulate votes.¹ What is probably the most frequently discussed allegation asserts that a digital method of tabulating votes benefited New York Senator Hillary Clinton at the expense of her chief competitor, Illinois Senator Barack Obama, in the recent Democratic Primary. This allegation appears to be the motivating factor behind the statewide recount being pushed by Ohio Congressman Dennis J. Kucinich. As a candidate in the Democratic primary (according to pre-recount results Kucinich received 3,901 votes in the race), New Hampshire election law entitles Kucinich to request a recount as long as he funds it, and as of the writing of this paper the New Hampshire Secretary of State has announced plans to recount all the ballots cast in the Democratic and Republican Primaries.²

Assuming that the original paper ballots are available, a recount is the only comprehensive method for evaluating the accuracy of a given election's vote tabulations. However, comprehensive audits of elections remain rare even in the post-2000 presidential election period. Thus, we conduct here a statistical examination of reported election returns so as to evaluate whether voting technology in New Hampshire affected vote outcomes in the state's 2008 Presidential Primaries. Our approach to election forensics would normally be employed as a means of detecting election irregularities with any identified irregularity being the basis for requesting a recount. Since, as we will show shortly, we find no vote tabulating irregularities in the New Hampshire wards that we analyze, the forthcoming availability of the New Hampshire recount provides a useful benchmark against which to evaluate the performance of the election forensic methods we demonstrate.

The allegation that there were vote tabulating problems in the New Hampshire Democratic Primary is based on differences in candidate vote shares across voting technologies. These technologies are Accuvote optical scan voting (hereinafter "Accuvote") and hand-counted paper ballots

¹See, for instance, "Kucinich says he'll ask for a recount", *Manchester Union Leader*, January 11, 2008 and "Primary votes to be recounted," *Concord Monitor*, January 12, 2008.

²See <http://www.sos.nh.gov/recount%20press%20release.pdf> for details.

(hereinafter “PBHC”).³ No other technology is currently used in New Hampshire, and vote totals and percentages by tabulating technology for the top five Democratic and top five Republican candidates are listed in Table 1.⁴ As is evident in the table, Clinton was more successful in New Hampshire Accuvote wards than was Obama, receiving 4.3 more percentage points of the vote there (40.2% – 35.9%). In contrast, Clinton did worse than Obama in PBHC wards, trailing by 6.1 percentage points (33.7% – 39.8%).⁵

If one uses associations like these—where one candidate does better than another in areas with a given vote tabulating technology— as a metric for identifying vote tabulating problems, then it follows that the Republican Presidential Primary in New Hampshire suffered from such problems as well. While Arizona Senator John McCain performed better than former Massachusetts governor Mitt Romney in both Accuvote and PBHC wards, Romney trailed by 3.6 points in wards with Accuvote machines and by 15 points in PBHC wards. By this logic, Accuvote machines in the Republican Primary worked to the advantage of Romney.

The difference between pre-election polls and the Democratic Primary outcome (Clinton over Obama) have perhaps contributed to skepticism of the accuracy of vote tabulations in this contest. Pre-election polls were consistent in their view that Obama, who won the previously held Iowa Caucuses, would beat Clinton by somewhere between five and 13 percentage points. Depending on which pre-election poll one consults, there was perhaps a 10-12 point swing between Clinton’s pre-election deficit to Obama compared to her observed electoral margin over the Illinois Senator.⁶ In contrast, the victory in the Republican Primary of John McCain was predicted by many, if not

³For the 2004 Primary and General elections there were two different optical scan technologies used in New Hampshire: Accuvote and Optech. Optech is no longer used in the state. New Hampshire’s Accuvote machines are manufactured by Premier Election Systems which was previously known as Diebold Election Systems.

⁴The 2008 New Hampshire Democratic and Republican Presidential Primaries were held on January 8, 2008. Both primaries were contested by 21 candidates each; that both featured the same number of candidates is coincidental. Pre-recount vote totals, which remain unofficial, were retrieved from the New Hampshire Secretary of State’s website at <http://www.sos.nh.gov/presprim2008/index.htm> at 14:26 on January 10, 2008. Voting technology information was downloaded from <http://www.sos.nh.gov/voting%20machines2006.htm> at 17:12 on January 9, 2008.

⁵A New Hampshire ward is either a town, e.g., Hanover, or a subdivision of a town, e.g., the first ward in Manchester. Most towns in New Hampshire are not broken down by ward, but the more populous ones are.

⁶See http://www.realclearpolitics.com/epolls/2008/president/nh/new_hampshire_democratic_primary-194.html for details.

Table 1: Top Candidate Vote Totals and Percentages

	Total		Accuvote		PBHC	
	Votes	Percent	Votes	Percent	Votes	Percent
<i>Democratic Primary</i>						
Clinton	112,606	39.4	97,388	40.2	15,218	33.7
Obama	105,004	36.8	87,066	35.9	17,938	39.8
Edwards	48,818	17.1	40,871	16.9	7,947	17.6
Richardson	13,239	4.6	10,652	4.4	2,587	5.7
Kucinich	3,901	1.4	3,063	1.3	838	1.9
<i>Republican Primary</i>						
McCain	88,570	37.7	73,684	36.5	14,886	39.9
Romney	75,546	32.2	66,246	32.9	9,300	24.9
Huckabee	26,859	11.4	21,964	10.9	4,895	13.1
Giuliani	20,439	8.7	17,375	8.6	3,064	8.2
Paul	18,307	7.8	14,875	7.4	3,432	9.2

Notes: Candidates are listed in order of total votes. Edwards refers to former North Carolina Senator John Edwards; Richardson to New Mexico Governor Bill Richardson; Romney to former Massachusetts governor Mitt Romney; Huckabee to Arkansas Governor Mike Huckabee; Guiliani to former New York City mayor Rudolph Guiliani; and Paul to Texas Congressman Ron Paul. Vote shares are rounded and do not sum to one because of candidates not listed.

all, opinion polls that circulated immediately prior to the election.⁷

With this setup mind, we investigate the claim that voting technology differentially aided Clinton and more generally we consider whether the New Hampshire Democratic and Republican Primaries were fair in the sense of having accurate vote tabulations. Simply put, we seek to understand whether a facet of election administration—ward choice of voting technology—can explain why a given candidate received either many or few votes. In an ideal world, all such administrative choices would have no effect on election outcomes or at least not systematically bias the results in favor of any particular candidate. Thus, to the extent that we can explain differences between Clinton’s and Obama’s vote counts (and similarly, between McCain’s and Romney’s vote counts) without recourse to voting technology, the more that New Hampshire voters can trust that the elec-

⁷See http://www.realclearpolitics.com/epolls/2008/president/nh/new_hampshire_republican_primary-193.html for details.

tion outcomes in the recent 2008 Presidential Primaries were not artifacts produced by a particular method of tabulating votes.

Our objective is similar to prior efforts aimed at evaluating the effect of voting technology and other election administration matters on vote outcomes including studies of the 2004 New Hampshire Democratic Presidential Primary (Wand, 2004) and the 2004 Presidential Election in New Hampshire (Herron and Wand, 2007). It is also similar to analyses of elections across the United States (e.g., Brady et al., 2001; Wand et al., 2001; Ansolabehere, 2002; Tomz and van Houweling, 2003; Mebane, 2004). The question of whether the administration of an election is without error is one that should be asked after every election, but the question has particular salience in the 2008 New Hampshire Democratic primary because of Clinton’s surprise victory. All elections should be subject to audits, and particular attention should be paid to elections that surprise most political observers.

2 Background on New Hampshire Wards

This section sets the stage for statistical matching results that follow shortly, and here we make two points. First, we present several illustrations of the fact that voting technology in New Hampshire is not distributed across wards independently of ward-level political characteristics. That is, we provide several examples of the fact that the type of voting technology used in a New Hampshire ward is related to ward features. And second, we show that Clinton vote share varied systematically by ward characteristics.

Note that wards are the smallest New Hampshire voting units and that each ward chooses its own voting technology. There are ten New Hampshire counties, and no county is uniform in its use of voting technology.

2.1 Voting Technology across Wards

A very simple point about the distribution of voting technology across New Hampshire wards is that, perhaps not surprisingly, counties with small numbers of voters tend to use PBHC and counties with cities, Accuvote. For example, the northernmost county in New Hampshire, Coos County, contains 46 towns that contribute votes to statewide elections. Of these, 40 use PBHC and only six use Accuvote. In contrast, Hillsborough County, which contains Manchester, New Hampshire's most populous city, has 40 wards of which 31 use Accuvote. The point here is that ward size is correlated with ward voting technology.

Consider Figure 1(a); Figures 1(b)–1(d) are qualitatively similar. Figure 1(a) displays smoothed densities based on ward-level Howard Dean vote shares from the 2004 Democratic Presidential Primary. The solid line density depicts Dean vote share among Accuvote wards and the dashed line density, Dean vote share among PBHC wards. What is notable is that these densities differ: for example, the modal Accuvote ward contained on the order of 20% Dean supporters while the modal PBHC ward contained on the order of 30% to 40% Dean supporters. Explaining patterns of Dean vote share from 2004 is beyond the scope of this analysis and in fact what ultimately motivated Dean voters to support Dean is not germane to the exercise presented here. The point we want to make is that Dean received disproportionately more support in PBHC wards than in Accuvote wards, and in contrast Figure 1(b) shows that John Kerry received disproportionately more support in the 2004 Democratic Presidential Primary among Accuvote wards than among PBHC wards.

Continuing with this logic, Figure 1(c) highlights differences by voting technology in the fraction of Democratic votes cast in the 2006 Republican Gubernatorial Primary. These votes are write-ins and are cast by individuals who were given Republican ballots yet selected a Democratic candidate.⁸ Whether this behavior reflects moderate policy preferences, confused voters, voters who forgot to change party affiliations before a primary, ward administration deficiencies, or some other factor is not known. What we can say, though, is that whatever drives this sort of crossover

⁸Such votes were not added by the New Hampshire Secretary of State to the candidates' vote totals in the Democratic Primary.

voting varies by type of ward.

Finally, Figure 1(d) highlights the correlation between ward size and voting technology that we commented on earlier when discussed Coos and Hillsborough Counties. The figure uses the total number of votes in the 2006 gubernatorial election as a measure of ward size. It is clear from the two densities in the figure that PBHC is found only in relatively small wards but that some small wards use Accuvote.

We note that Figure 1 contains only four variables that are correlated with voting technology. Many more such variables exist, and we draw on the four from the figure plus others in our subsequent matching analysis.

2.2 Clinton Vote Share across Wards

We now turn to Figure 2, which presents four scatterplots analogous to the four densities highlighted above. Each panel in the figure plots Clinton vote share among all candidates against a ward-level variable. For instance, Figure 2(a) (which parallels Figure 1(a)) plots Clinton vote share from 2008 against Dean vote share from the 2004 Democratic Presidential Primary. The correlation between these variables is -0.539 : wards with many Dean voters in 2004 had fewer Clinton voters in 2008.

In contrast, the correlation between 2008 Clinton vote share and Kerry vote share from the 2004 Presidential Primary is 0.579 ; these two vote shares are depicted in Figure 2(b). We do not observe a notable bivariate correlation between Clinton vote share and the Democratic vote in the 2006 Republican Primary (correlation is -0.119), but we do observe a correlation of 0.239 between Clinton vote share and the size of the governor vote in the 2006 general election. Namely, larger wards were more pro-Clinton.

Why precisely wards that were pro-Dean in 2004 were anti-Clinton in 2008 is beyond our scope. One could conjecture that Dean voters were anti-establishment and that Clinton is seen, among contending Democrats, as a rather pro-establishment figure. Regardless of whether this conjecture is true, the combination of Figures 1(a) and 2(a) suggests a plausible explanation for

a correlation between Clinton vote share and voting technology that has nothing to do with vote tabulating problems.

3 Estimates of Vote Tabulating Technology Effects

We have now illustrated that, as of January, 2008, the distribution of voting technology across wards in New Hampshire depends on ward characteristics. Any analysis of the effect of voting technology in New Hampshire on ward-level election outcomes will be confounded by features of wards that are correlated with both voting technology and ward political profile. To estimate how much of the Accuvote-PBHC difference in the candidates' votes is due to something inherent to technology, we need a way to purge the comparison of the multitude of differences between wards that used each technology.

To do this we use genetic matching (Sekhon, 2008) to find a set of matched pairs of wards. Each pair contains one ward that used Accuvote technology and one that used PBHC, and the two wards are similar with respect to a large set of observable characteristics whose values were determined before the 2008 primaries. A set of matched wards is deemed to be well-balanced and suitable for making comparisons if each observable characteristic has the same distribution in the two types of wards. Borrowing the language of experiments, we look at PBHC as the “treatment” and Accuvote as the “control” technology, and we investigate the effect of having PBHC technology on the wards that had PBHC technology. This is formally referred as estimating the Average Treatment Effect on the Treated (ATT).⁹ Because vote tabulating technology is not randomly assigned to wards, we do not expect and indeed do not observe that there are good matches among the wards that used Accuvote for all of the wards that used PBHC. In fact, of the 110 PBHC wards in our analysis, only 24 have a sufficiently good match among the wards that used Accuvote.¹⁰ In a strict sense, our

⁹Ideally we would also like to estimate the effect of having Accuvote technology on the wards that had Accuvote technology, which would allow us to draw more comprehensive conclusions about the effects of the technology. Unfortunately there is insufficient overlap (Abadie and Imbens, 2006, 237-238) between the two types of wards to support such an analysis. Across New Hampshire more wards used Accuvote technology, and their sizes and urban or suburban characters make too many of them substantially different from the wards that used PBHC.

¹⁰In all, 148 of New Hampshire's 323 wards used PBHC, but of the PBHC wards 18 recorded zero votes in the 2008 primary, 19 recorded zero votes in the 2006 primary, 23 were missing 2006 average wage data, and 26 were missing

inferences about the effects of vote tabulating technology apply only to this subset of the wards. But these are the only wards about which we can have any confidence that other characteristics we are able to observe—and which we know are both imbalanced across technologies and related to the distribution of votes—are not responsible for any differences that may appear.

The matching method we use is one-to-one matching with replacement.¹¹ Table 2 shows the definition of the treatment indicator variable ($vs10 = 1$ for wards using PBHC, 0 for wards using Accuvote), along with the set of variables used to estimate the probability that each ward uses PBHC, conditional on those characteristics. This probability is known as the estimated propensity to receive the treatment, given the values of the conditioning variables. The conditioning variables are a collection of vote proportions, vote counts and functions of proportions and counts from the 2004 New Hampshire Presidential Primary and 2006 New Hampshire Gubernatorial Primary. The set of conditioning variables also includes a variable that measures one aspect of the economic profile of each ward (the percent of persons in poverty as of 2000). We use a simple logistic regression model to compute the estimated propensities. The exact formulation for this model using the `glm` function of **R** (R Development Core Team, 2005) appears at the end of Table 2. We match on the estimated propensity score along with a subset of the variables used to compute the estimated propensity score.¹² We set a caliper of one standard deviation for each of the matched variables, which means that we drop all matches that have more than one standard deviation of

poverty rate data. Excluding the wards with zero votes or missing data leaves 110 PBHC wards. Only one of the 175 wards that used Accuvote was excluded for these reasons (one ward is missing 2006 average wage data).

¹¹ The relevant calls to functions in the `Matching` (Sekhon, 2008) package have arguments as indicated in the following, where Y is the vector of outcomes of interest (e.g., vote counts or vote shares for a candidate), X is a matrix of characteristics to be matched on, and BM is a matrix of the characteristics on which the genetic matching algorithm attempts to achieve balance.

```
genout <-  
GenMatch(Tr=Tr,X=X, BalanceMatrix=BM, estimand="ATT", M=1,  
  pop.size=10000, max.generations=101, wait.generations=100,  
  caliper=rep(1,ncol(X)))  
Match(Y=Y, Tr=Tr, X=X, estimand="ATT", Weight.matrix=genout,  
  caliper=rep(1,ncol(X)))
```

¹²The variables used for matching are the estimated propensity score plus all the covariates used to estimate the propensity score except `fdp06R` and `f06D - f06R`. These are the variables included in the matrix X in the invocation of `GenMatch`.

discrepancy for any of these matching variables. Table 3 lists the variables on which the matching algorithm seeks balance. This set includes some additional measures of characteristics of the 2004 and 2006 primaries, some additional measures of each ward's income profile and a measure of the number of people living in each ward as of 2002.

Table 4 reports the mean values of several variables in the wards that used PBHC and the wards that used Accuvote, before and after matching.¹³ Even though we are not concerned with any kind of hypothesis testing at this stage, the p -value from a t -test of the formal hypothesis of no difference between these means provides a useful summary for how well both the mean and variance of the distributions in the respective subsets of wards correspond. Large values for these p -values do not guarantee that the achieved level of balance is sufficient for reliable inferences about the effects of the treatment (Diamond and Sekhon, 2005), but nonetheless it is noteworthy and reassuring that none of the p -values in the matched subset of data are small.

In addition to the estimated average effects that using PBHC tabulation had on the votes recorded in wards that used PBHC tabulation, we report three estimators for the uncertainty in those estimates. We apply each estimator both to the raw vote counts recorded for several candidates and to the proportion of the total votes cast in the appropriate primary for each of the candidates. The candidates we focus on are Clinton, Edwards, Kucinich, Obama and Richardson in the Democratic primary and Giuliani, Huckabee, Paul, Romney and McCain in the Republican primary. Hill and Reiter (2006) discuss the rationale for and performance of the estimation uncertainty estimators we use. Hill and Reiter (2006) directly study one estimator we use, namely the Hodges-Lehmann aligned rank test. This estimator gives directly for each outcome variable a 95% confidence interval for the average treatment effect measured on the scale of the outcome variable.

Our other two estimators are variations of the weighted least squares estimators Hill and Reiter (2006) discuss. We apply the weights they define to generalized linear models that are suitable for either counts or proportions.¹⁴ For the vote counts this is an overdispersed Poisson model, and for

¹³These statistics are computed using the `MatchBalance` function in the `Matching` (Sekhon, 2008) package.

¹⁴To define the weights, for sample size N , let c_j be the number of times observation $j = 1, \dots, N$ is included in the matched sample. Let n_m denote the number of distinct observations in the matched sample. The weight for each observation is $w_j = n_m c_j / \sum_{j=1}^N c_j$ (Hill and Reiter, 2006, 2232).

the vote proportions this is an overdispersed binomial model.¹⁵ One version of these estimators incorporates a regression adjustment for bias reduction: the variables we matched on are included as regressors along with the treatment indicator variable (`vs10`) and a constant (for the intercept) in either a weighted overdispersed Poisson model (for the vote counts) or a weighted overdispersed binomial model (for the proportions) for the outcome variable in the matched data set. The estimated average treatment effect is represented as the coefficient of the treatment indicator variable in the regression. If the average treatment effect is zero, then the coefficient is zero, but it is not straightforward to translate a nonzero coefficient into an expression of the treatment effect on the original scale of the data. We will be content to determine whether a symmetric 95% confidence interval contains zero.¹⁶ The other version of these estimators does not feature any bias adjustment. For this estimator the treatment indicator variable and a constant are the only variables included in the regressions. For both of these estimators we use the sandwich estimator (Huber, 1967) to obtain the coefficients' standard errors.

Table 5 shows that all of estimates for the effect that using PBHC tabulation had on votes recorded in wards that used PBHC tabulation are small, and for all but one of the effects the estimated confidence intervals suggest the effects do not differ significantly from zero. The column labeled ATT in the table reports the point estimates for the average treatment effects. For the vote counts these estimates are on the scale of votes, and for the vote proportions they are on the scale of the proportions. The estimates for vote counts range from an average 30.5 vote deficit for Clinton due to having PBHC tabulation to an average 4.8 vote gain for Richardson. These small counts correspond to proportional effects ranging from a 2.1 percent deficit for Edwards due to having PBHC tabulation up to a 2.5 percent gain for McCain. The Hodges-Lehmann 95% confidence interval estimates suggest, however, that among the ten candidates only the effects for Edwards differ significantly from zero. The upper bounds of the Hodges-Lehmann intervals are less than zero both for Edwards's vote counts and for his vote proportions. For the other

¹⁵In **R** (R Development Core Team, 2005), these models are estimated using the `glm` functions with respectively the "quasipoisson" and "quasibinomial" family arguments.

¹⁶Using \hat{b} to denote the coefficient estimate of interest and $SE(\hat{b})$ to denote its estimated standard error, we compute the 95% confidence interval using $\hat{b} \pm 1.96 \cdot SE(\hat{b})$.

candidates the Hodges-Lehmann intervals include zero. The 95% confidence intervals for the coefficient of the treatment indicator variable in the weighted regression models for vote counts mostly confirm these inferences. All of the intervals for the bias-adjusted models include zero; this confirms the Hodges-Lehmann results for all the candidates except Edwards. The 95% confidence intervals for the vote count models without bias adjustment show significant negative effects for both Edwards and Huckabee. The intervals for the coefficient of the treatment indicator variable in the weighted regression models for vote proportions show significant negative effects for Edwards and significant positive effects for both Kucinich and Richardson. In view of the small vote shares both of the latter candidates received, it is likely that the vote proportion results for them stem from differences not in votes for them but in the aggregate of votes for all the other candidates.

4 Conclusion

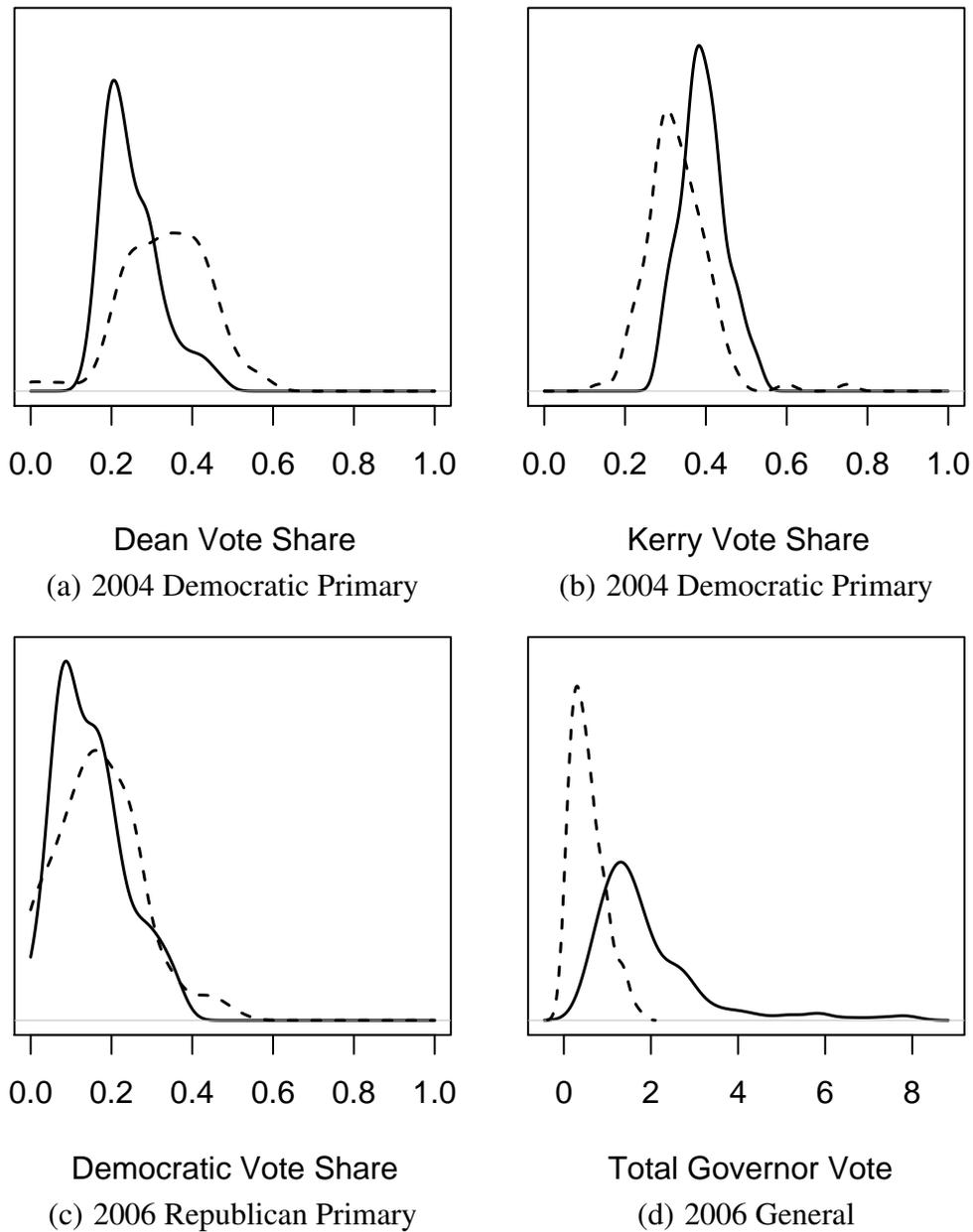
We find no significant relationship between a ward's use of vote tabulating technology and the votes or vote shares received by most of the leading candidates who competed in the 2008 New Hampshire Presidential Primaries. Among Clinton, Edwards, Kucinich, Obama and Richardson in the Democratic primary and Giuliani, Huckabee, Paul, Romney and McCain in the Republican primary, we observe a significant average effect of using PBHC technology on the wards that used PBHC technology only in the votes counted for Edwards, and that difference is small. The effects for Edwards also do not appear to be significant when a regression-based bias adjustment is applied.

The particular set of variables used for the matching analysis in this study does not exhaust the range of observable ward attributes. It is possible that another set of matching variables and matched pairs of wards would produce even better balance among observables across technologies than we have found. It is also possible that some observables we have not examined in this study remain imbalanced, contributing to bias in our estimates of the average treatment effects. The observable features we have examined, however, include variables that measure many aspects of

the preceding primary elections in the state, as well as many demographic features of wards in the state.

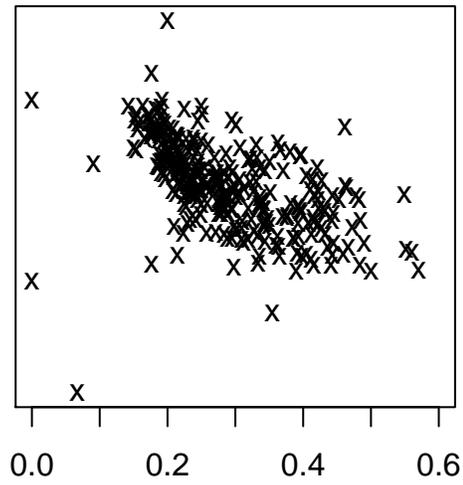
The biggest limitation of this study is that the matching exercise produced only 24 matched wards. For most of the remaining hundreds of wards it is not possible to obtain a direct estimate of the effects of vote tabulating technology that is not confounded with known extraneous factors. There are other approaches one might use for an analysis that would use many more or perhaps all of New Hampshire's wards, but such methods depend on strong and unverifiable assumptions about features of abstract analytical models. No analysis is free from assumptions, but the matching approach we use has the virtue of remaining very close to the observable data. The key assumption we need is that the inherent relationship between the two vote tabulating technologies is roughly the same throughout New Hampshire. In that case, the average treatment effect we estimate in the subset of matched wards is telling us about what is true throughout the state.

If one suspects that either the 2008 Democratic Primary or the 2008 Republican Primary election in New Hampshire was affected by irregularities, then naturally one may believe that such irregularities are haphazard or perhaps even artfully disguised. This study is not intended to address such suspicions. But with respect to Hillary Clinton's surprise victory in the Democratic Primary and the notable differences across vote tabulating technologies in Clinton's and others' levels of support, our results are consistent with these differences being due entirely to the fact that New Hampshire wards that use Accuvote optical scan machines typically have voters with different political preferences than wards that use hand counted paper ballots.

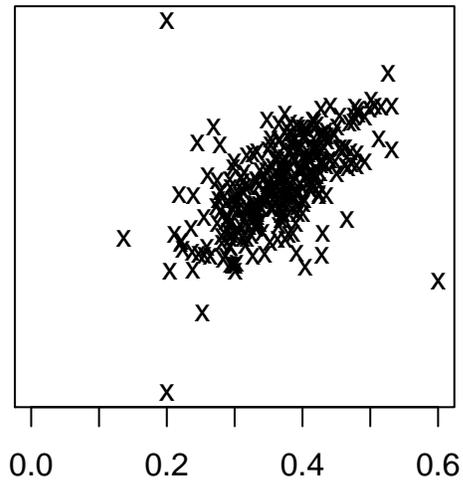


Note: Each figure displays two smoothed densities based on ward-level variables. Solid line densities are for Accuvote wards and dashed line densities for PBHC wards. The 2004 Primary refers to the 2004 Democratic Presidential Primary, and the 2006 Republican Primary refers to the 2006 Gubernatorial Primary. Finally, Total Governor Vote in the 2006 General Election is in thousands of voters.

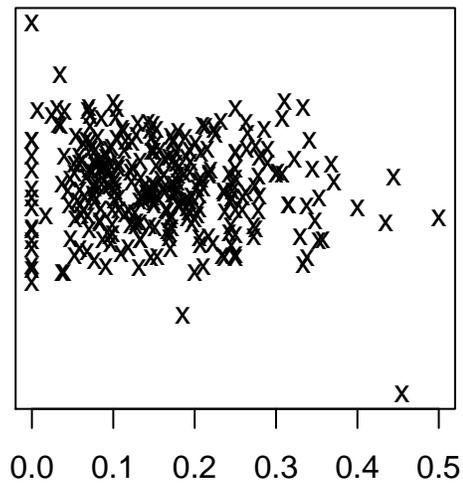
Figure 1: Variance Across New Hampshire Wards by Voting Technology



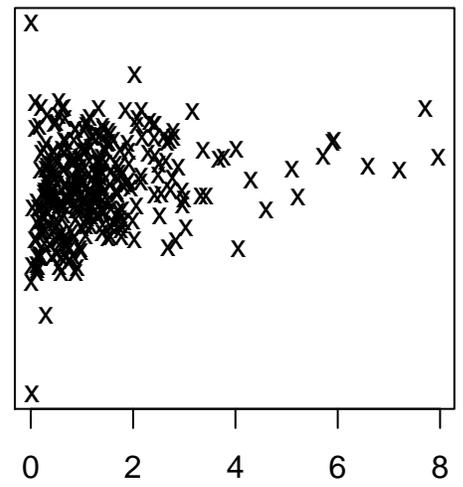
Dean Vote Share
(a) 2004 Democratic Primary



Kerry Vote Share
(b) 2004 Democratic Primary



Democratic Vote Share
(c) 2006 Republican Primary



Total Governor Vote
(d) 2006 General

Note: Plot labels are identical to those in Figure 1. Each “x” in the figure denotes a New Hampshire ward, and each panel plots Clinton share of the total Democratic primary vote against another variable.

Figure 2: Variance in Clinton Vote Share

Table 2: Treatment and Propensity Variables

Treatment Variable

vs10 vote tabulating technology: 1 if PBHC, 0 if Accuvote

“Propensity” Variables

wpkerry04pd standardized Kerry proportion in 2004 Democratic presidential primary
wdpgov06T standardized difference between proportion voting Democratic in the 2006 gubernatorial primary and Kerry’s proportion in 2004 Democratic presidential primary
Pdean04p Dean proportion in 2004 Democratic presidential primary
Pclark04p Clark proportion in 2004 Democratic presidential primary
Plieberman04p Lieberman proportion in 2004 Democratic presidential primary
Pkerry04p Kerry proportion in 2004 Democratic presidential primary
pforRinD04 proportion voting for a Republican candidate in the 2004 Democratic presidential primary
pforDinR04 proportion voting for a Democratic candidate in the 2004 Republican presidential primary
frp06D proportion voting for a Democratic candidate in the 2006 Republican gubernatorial primary
fdp06R proportion voting for a Republican candidate in the 2006 Democratic gubernatorial primary
f06D - f06R difference between proportion of all voters voting in the Democratic 2006 gubernatorial primary and proportion voting in the Republican 2006 gubernatorial primary
povrate percent of persons in poverty: persons are classified as being below the poverty level by comparing their total 1999 income to an income threshold. Refer to <http://www.census.gov> for more detailed information. Source: 2000 Census SF-3

Note: The propensity variable `pfit` is computed as follows.

```
z <- glm( vs10 ~ wpkerry04pd + wdpgov06T + povrate +
          Pdean04p + Pclark04p + Plieberman04p + Pkerry04p +
          pforRinD04 + pforDinR04 + frp06D + fdp06R + I(f06D - f06R) ,
          family="quasibinomial")
pfit <- as.real(fitted(z))
```

Table 3: Variables Used to Measure Balance When Matching

wpkerry04pd wdpgov06T	standardized Kerry proportion in 2004 Democratic presidential primary standardized difference between proportion voting Democratic in the 2006 gubernatorial primary and Kerry's proportion in 2004 Democratic presidential primary
Pdean04p	Dean proportion in 2004 Democratic presidential primary
Pclark04p	Clark proportion in 2004 Democratic presidential primary
Plieberman04p	Lieberman proportion in 2004 Democratic presidential primary
Pkerry04p	Kerry proportion in 2004 Democratic presidential primary
pforRinD04	proportion voting for a Republican candidate in the 2004 Democratic presidential primary
pforDinR04	proportion voting for a Democratic candidate in the 2004 Republican presidential primary
povrate	percent of persons in poverty: persons are classified as being below the poverty level by comparing their total 1999 income to an income threshold. Refer to http://www.census.gov for more detailed information. Source: 2000 Census SF-3
pop2002e	estimated population in 2002. Source: NH Office of Energy and Planning
pfim	median family income: total income received in 1999 by all family members 15 years of age and older. Source: 2000 Census SF-3
pci	per capita income in 1999: Source: 2000 Census SF-3
pballots.abs06	proportion absentee ballots in the 2006 gubernatorial primary elections
pchecklist.undl06	proportion of voters with partisanship undeclared in 2006 gubernatorial primary election
pchecklist.dem06	proportion of voters with Democratic partisanship in 2006 gubernatorial primary election
gov06T	number of votes recorded in 2006 Democratic and Republican primaries
dpgov06T	number of votes recorded in 2006 Democratic gubernatorial primary
rpgov06T	number of votes recorded in 2006 Republican gubernatorial primary
f06D - f06R	difference between proportion of all voters voting in the Democratic 2006 gubernatorial primary and proportion voting in the Republican 2006 gubernatorial primary
fp06D	proportion voting for a Democratic candidate in either Democratic or Republican 2006 gubernatorial primary
frp06D	proportion voting for a Democratic candidate in the 2006 Republican gubernatorial primary
fdp06R	proportion voting for a Republican candidate in the 2006 Democratic gubernatorial primary
T2004D	number of votes recorded in 2004 Democratic presidential primary

Note: These are the variables included in the matrix `BM` in the invocation of `GenMatch`.

Table 4: Balance Tests Before and After Matching

variable	Before Matching			After Matching		
	mean		<i>t</i> -test	mean		<i>t</i> -test
	PBHC	Accuvote	p-value	PBHC	Accuvote	p-value
pfrit	0.84874	0.095623	< 0.00001	0.54114	0.47607	0.13273
wpkerry04pd	3.7192	8.1903	< 0.00001	5.2914	5.567	0.18664
wdpgov06T	3.7885	5.6025	0.0019328	4.6297	4.6594	0.9542
povrate	0.064598	0.068807	0.31696	0.055721	0.053367	0.53088
Pdean04p	0.34102	0.25129	< 0.00001	0.31170	0.31363	0.82156
Pclark04p	0.12407	0.12478	0.87565	0.12037	0.12030	0.98418
Plieberman04p	0.058903	0.086213	< 0.00001	0.07099	0.073491	0.4497
Pkerry04p	0.32683	0.39369	< 0.00001	0.34959	0.35209	0.69207
pforRinD04	0.00058559	0.0013068	0.0011216	0.00050928	0.00051322	0.97256
pforDinR04	0.10735	0.12720	0.02611	0.10433	0.10146	0.7454
pop2002e	1584.4	6248.1	< 0.00001	2693.1	2984.9	0.18405
pfim	51807	57333	0.000042	55403	55680	0.89085
pci	21961	23367	0.010485	23231	23193	0.9529
pballots.abs06	0.06533	0.055516	0.026040	0.056382	0.056036	0.96201
pchecklist.undl06	0.50046	0.43456	< 0.00001	0.48623	0.47197	0.55152
pchecklist.dem06	0.20656	0.27398	< 0.00001	0.21323	0.23334	0.16148
gov06T	597.31	1910.2	< 0.00001	986.5	1055.5	0.26347
dpgov06T	67.273	204.24	< 0.00001	105.08	109	0.68952
rpgov06T	56.845	164.28	< 0.00001	89.125	94.958	0.61281
f06D	0.74752	0.75467	0.35224	0.75259	0.75254	0.99533
fp06D	0.59197	0.62149	0.054287	0.61254	0.6089	0.8105
fdp06R	0.0021841	0.00091978	0.36341	0.00090086	0.0012821	0.71886
frp06D	0.17569	0.14901	0.023768	0.16112	0.15386	0.43305
T2004D	303.83	1056.0	< 0.00001	511	548.92	0.29725

Note: Before matching there are 110 PBHC and 174 Accuvote observations. After matching there are 24 matched pairs. Before matching we use the two-sample *t*-test, and after matching we use the paired *t*-test.

Table 5: Estimated Average Treatment Effects

Vote Counts		Weighted Regressions					
variable	ATT	Hodges-Lehmann		Bias-adjusted		Not adjusted	
		lower	upper	lower	upper	lower	upper
Clinton	-30.5	-72.5	19.5	-0.361	0.053	-0.301	0.070
Obama	-19.8	-59.5	16.5	-0.279	0.122	-0.257	0.083
Edwards	-24.9	-43.0	-3.5	-0.450	0.021	-0.398	-0.013
Kucinich	1.2	-3.5	5.0	-0.174	0.682	-0.168	0.538
Richardson	4.8	-5.5	14.5	-0.142	0.334	-0.044	0.383
Giuliani	-3.9	-13.5	6.0	-0.326	0.163	-0.334	0.050
Huckabee	-14.5	-31.5	4.0	-0.520	0.088	-0.474	-0.012
Paul	-7.4	-19.5	5.0	-0.459	0.107	-0.447	0.049
Romney	-14.4	-36.5	8.0	-0.326	0.123	-0.318	0.028
McCain	-12.9	-47.0	23.0	-0.267	0.121	-0.270	0.054

Vote Proportions		Weighted Regressions					
variable	ATT	Hodges-Lehmann		Bias-adjusted		Not adjusted	
		lower	upper	lower	upper	lower	upper
Clinton	-0.0125	-0.0400	0.0183	-0.185	0.064	-0.155	0.086
Obama	0.0149	-0.0083	0.0422	-0.068	0.187	-0.079	0.140
Edwards	-0.0210	-0.0330	-0.0063	-0.230	-0.013	-0.218	-0.057
Kucinich	0.0022	-0.0026	0.0066	0.032	0.718	0.014	0.583
Richardson	0.0156	0.0024	0.0224	0.051	0.394	0.160	0.499
Giuliani	0.0046	-0.0078	0.0159	-0.111	0.186	-0.106	0.134
Huckabee	-0.0135	-0.0334	0.0106	-0.299	0.070	-0.266	0.040
Paul	-0.0102	-0.0263	0.0096	-0.249	0.118	-0.211	0.099
Romney	-0.0021	-0.0306	0.0275	-0.140	0.181	-0.116	0.137
McCain	0.0251	-0.0063	0.0542	-0.036	0.177	-0.014	0.161

Note: The weighted regression is a overdispersed Poisson model for the vote counts and an overdispersed binomial model for the vote proportions. ATT point estimates and Hodges-Lehmann confidence intervals are in vote count or vote proportion units. Weighted regression confidence intervals are for the coefficient of the treatment indicator variable.

References

- Abadie, Alberto and Guido W. Imbens. 2006. “Large Sample Properties of Matching Estimators for Average Treatment Effects.” *Econometrica* 74(1):235–267.
- Ansolabehere, Stephen. 2002. “Voting Machines, Race, and Equal Protection.” *Election Law Journal* 1(1):61–70.
- Brady, Henry E., Justin Buchler, Matt Jarvis and John McNulty. 2001. “Counting All the Votes: The Performance of Voting Technology in the United States.” Report published by the Survey Research Center and Institute of Governmental Studies, University of California, Berkeley.
- Diamond, Alexis and Jasjeet S. Sekhon. 2005. “Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies.” Working paper.
- Herron, Michael C. and Jonathan N. Wand. 2007. “Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount.” *Electoral Studies* 26(2):247–261.
- Hill, Jennifer and Jerome P. Reiter. 2006. “Interval Estimation for Treatment Effects Using Propensity Score Matching.” *Statistics in Medicine* 25(13):2230–2256.
- Huber, Peter J. 1967. The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. Lucien M. Le Cam and Jerzy Neyman. Vol. 1 Berkeley, CA: University of California Press pp. 221–233.
- Mebane, Jr., Walter R. 2004. “The Wrong Man is President! Overvotes in the 2000 Presidential Election in Florida.” *Perspectives on Politics* 2(3):525–535.
- R Development Core Team. 2005. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- URL:** <http://www.R-project.org>

- Sekhon, Jasjeet S. 2008. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." Forthcoming, *Journal of Statistical Software*.
- Tomz, Michael and Robert P. van Houweling. 2003. "How Does Voting Equipment Affect the Racial Gap in Voided Ballots?" *American Journal of Political Science* 47(1):46–60.
- Wand, Jonathan. 2004. Evaluating Voting Technologies: 2004 New Hampshire Democratic Primary. Technical report Stanford University.
URL: <http://wand.stanford.edu/elections/us/NH/nh2004primary.pdf>
- Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. 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.