Available online at www.sciencedirect.com

ScienceDirect

Electoral Studies 26 (2007) 247–261

Electoral Studies

www.elsevier.com/locate/electstud

Assessing partisan bias in voting technology: The case of the 2004 New Hampshire recount

Michael C. Herron [a,*], Jonathan Wand [b]

[a] Department of Government, Dartmouth College, 6108 Silsby Hall, Hanover, NH 03755-3547, USA

[b] Department of Political Science, Stanford University. Encina Hall, 308 West. Stanford, CA 94305-6044, USA

Abstract

Following allegations that Accuvote optical scan machines used in New Hampshire during the 2004 presidential election produced unusually low vote totals for Democratic candidate John Kerry, third party candidate Ralph Nader requested and funded a hand recount of ballots cast in eleven New Hampshire precincts. Using statistical methods well-suited for identifying election irregularities, we find no evidence of systematic biases among New Hampshire's Accuvote machines. Nor do we find evidence of other technology-related tabulation problems in the state. Our findings explain why the New Hampshire presidential recount did not substantiate alleged Accuvote discrepancies, and indeed it recovered more votes for George W. Bush than it did for Kerry. More generally, our analysis demonstrates methods that can help avoid false allegations about vote fraud while enabling concerned citizens, election administrators, and researchers to find and remedy real election irregularities.

Keywords: Presidential election; Election irregularities; Recount; Partisan bias; Robust estimation; Matching; Diebold; Accuvote; Optech

1. Introduction

The first recount of the 2004 presidential race occurred in New Hampshire. At the behest of third party presidential candidate Ralph Nader, who paid approximately $14,000 to fund it, the presidential votes cast in eleven New Hampshire precincts were recounted by hand.[1]

The origin of the New Hampshire recount is not, as one might imagine, a dispute between the major party candidates who fought the 2004 presidential contest. According to pre-recount figures, Democratic presidential nominee John Kerry won New Hampshire's presidential race by 340,511 votes to Republican incumbent George W. Bush's 331,237.[2] Rather, the New Hampshire recount was instead the product of an alleged voting technology problem. In particular, six days after the 2004 election a study of New Hampshire vote returns (Briggs, 2004) claimed that some of the state's precincts that used *Accuvote*

* Corresponding author.

E-mail addresses: michael.c.herron@dartmouth.edu (M.C. Herron), wand@stanford.edu (J. Wand).

[1] We refer to all New Hampshire voting units as "precincts." Precincts sometimes consist of entire towns—e.g., Litchfield, a town in the southeastern portion of New Hampshire—but in other cases precincts are subunits of cites-e.g., Wards 6, 7, 8, and 9 of Manchester.

[2] New Hampshire vote returns and voting technology data were downloaded from the web site of the New Hampshire Secretary of State (http://www.sos.nh.gov).

0261-3794/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.electstud.2006.02.004

machines (a brand of optical scan voting technology manufactured by Diebold Election Systems) reported unusually low vote totals for John Kerry. Within the study Accuvote precincts were compared with other New Hampshire precincts that used either hand-counted paper ballots (PBHC) or the *Optech* brand of optical scan machines (manufactured by Election Systems and Software).[3]

The Briggs study appears to have been both a catalyst for the Nader-funded recount and for identifying precincts to be recounted. This underscores the gravity of the allegations in the study and the seriousness with which it was received. Similar allegations of biases penalizing Democratic totals were also prominent in the 2004 presidential tabulations of Florida and Ohio.

Contrary to the Briggs study, we show that Accuvote precincts in New Hampshire did not report unusually low Kerry returns. In fact, we find no evidence of any type of voting technology effects in the state's presidential or gubernatorial elections. This conclusion is consistent with results of the presidential recount, which failed to uncover significant discrepancies in vote tabulations. In an ironic twist, the recount recovered more votes for Bush than it did for Kerry, the opposite of what one would have expected if allegations of Accuvote problems were valid.

What appears to have led others to believe that Accuvote machines in New Hampshire were biased against Kerry is a failure to recognize that precincts in the state, like counties across the United States, have some discretion over what voting technologies they adopt. Consequently, voting technologies are not assigned randomly across New Hampshire just as they are not assigned randomly among counties (Knack and Kropf, 2002; Card and Moretti, 2005). We show that, once differences between precincts with Accuvote, Optech, and PBHC are accounted for, the allegations of bias in Accuvote tabulations are not supported by precinct voting returns.

2. The New Hampshire recount: a brief history

As a consequence of the 2000 presidential election and its controversial aftermath (e.g., Posner, 2001; Merzer, 2001; The Washington Post, 2001), the past several years have witnessed an increase in attention focused on voting technology, patterns of disfranchisement among different citizen groups, uncounted votes, and related problems in election administration. In light of such attention, it is not surprising that there have been numerous allegations of irregularities in the 2004 presidential election process. Many of these focus explicitly on voting technology. For instance, Hout et al. (2004) assert that electronic voting machines contributed to Bush's vote margin in Florida. Nonetheless, Sekhon (2004a) and Wand (2004) show that no such irregularities are evident in county-level election returns. Similarly, Card and Moretti' (2005) county-level analysis of touchscreen usage in the 2004 presidential election finds no evidence that voting technology itself affected presidential vote returns.

Allegations of Accuvote problems in New Hampshire were initiated in a study authored by Briggs (2004). Briggs examined New Hampshire precincts and compared the number of percentage points by which Gore beat Bush in 2000 with the number of points by which Kerry beat Bush in 2004. Briggs conjectured that turnout by precinct should be higher in 2004 than in 2000 and that the number of percentage points by which Kerry beat Bush in 2004 should be greater than the number of such points by which Gore beat Bush (or lesser than, in the case of Kerry losing to Bush and Gore losing to Bush).

An anomalous precinct, according to Briggs, had a much higher turnout rate in 2004 yet in the precinct Kerry did substantially worse than Gore, relatively speaking. For example, Briggs classified the town of Litchfield as being "out of trend" because its turnout rate in 2004 was 123% of the rate in 2000, the Gore-Bush difference was negative eight percentage points, and the Kerry-Bush difference in percentage points was negative fifteen, i.e., Kerry did relatively worse than Gore in Litchfield.[4] Briggs claims that Accuvote precincts are disproportionately represented among "out of trend" precincts, and this is the source of her contention that there were tabulating problems in these locales.

We begin our analysis by considering voting patterns grouped by type of voting technology, and subsequently we demonstrate why such a simple analysis can be a misleading basis for measuring the bias of a method of voting. Table 1 provides a summary of New Hampshire Democratic presidential and gubernatorial vote

[3] Electronic voting without paper records is not permitted in New Hampshire. For background on the Nader recount see "Losing by 335,000 in N.H., Nader Demands a Recount," Washington Post, November 10, 2004, "Nader Recount Goes Slowly," Valley News, November 19, 2004, and "Nader-requested recount continues," The Boston Globe, November 29, 2004.

[4] In her analysis Briggs compares some New Hampshire wards in 2004 to wards from 2000 that due to redistricting received new boundaries between the 2000 and 2004 general elections. This practice is not part of the analysis here.

Table 1
Distribution of Democratic vote shares across voting technologies and various general elections

	Accuvote	Optech	PBHC	Dispersion	Deviance	*p*-Dev	*p*-KS
President							
2000 (Gore)	0.503		0.484	44.3	132.4	0.1	0.001
		0.463	0.484	29.4	69.0	0.1	0.001
	0.503	0.463		88.9	331.6	0.1	0.001
2004 (Kerry)	0.510		0.519	71.1	37.6	0.5	0.001
		0.467	0.519	41.0	584.9	0.001	0.001
	0.510	0.467		151.5	506.4	0.1	0.001
Governor							
2002 (Fernald)	0.398		0.411	67.5	49.4	0.4	0.001
		0.347	0.411	39.9	593.7	0.001	0.001
	0.398	0.347		142.0	484.6	0.1	0.001
2004 (Lynch)	0.515		0.520	86.7	11.4	0.7	0.001
		0.468	0.520	41.3	582.1	0.001	0.001
	0.515	0.468		188.2	601.3	0.1	0.001
N towns	70	25	139				

Note: Dispersion is the estimated dispersion parameter from a quasi-binomial model; Deviance is the difference in deviances between two-mean and one-mean (pooled proportions) models; *p*-Dev is the probability that the Deviance is large enough to indicate the rejection of no difference between estimated proportions; and *p*-KS is the probability that weighted proportions for both technologies being compared have the same distribution based on bootstrapped Kolmogorov–Smirnov tests. The weighted proportion is $p_i/\sqrt{(\sigma^2 p_i(1-p_i)/n_i)}$, where in town n_i is the number of votes cast and p_i is the Democratic proportion of the two-party vote.

shares over 234 towns in the 2000, 2002, and 2004 general elections.[5] For simplicity, we examine significance levels of pairwise differences in average vote shares based on New Hampshire's use of three different voting machines. For instance, in comparing Accuvote towns and Optech towns (this is one of three possible pairwise comparisons), Table 1 shows that average Kerry vote share in 2004 was 0.510 in the former and 0.467 in the latter. Assuming an overdispersed binomial model, the observed difference between these two proportions is significant at the 0.1 level as shown in the *p*-Dev column. Moreover, the distribution of 2004 Kerry shares in Accuvote towns and the distribution of Kerry shares in Optech towns are, using a bootstrapped Kolmogorov–Smirnov test (Mebane and Sekhon, 2004a) significantly different as noted in the *p*-KS column. Comparing Accuvote versus PBHC and Optech versus PBHC yields similar conclusions.

Table 1's mildly significant difference in proportion *p* values in conjunction with its highly significant Kolmogorov–Smirnov *p* values show that town-level Democratic vote shares were distributed very differently across precincts with each type of voting technology. This statement holds for the 2000 and 2004 presidential races as well as the 2002 and 2004 gubernatorial contests.

Table 2 shows that a temporal difference-in-difference (DID) analysis of presidential and gubernatorial elections uncovers potential technology problems akin to the non-differenced results in Table 1. This is notable because time-invariant individual town effects might confound nominal Democratic vote shares in the cross-sectional analysis, and such effects will be eliminated by DID. For instance, Table 2 shows that the average 2004–2000 Democratic presidential vote share difference (i.e., Kerry share minus Gore share) was 0.149 in Accuvote towns and 0.215 in PBHC towns. And, the Accuvote and PBHC distributions of Kerry-Gore differences are significantly different at the 0.002 level. The 2004 gubernatorial race is particularly notable in Table 2: 2004–2002 changes in Democratic vote share (Lynch minus Fernald) appear highly correlated with voting technology, even more so than 2004–2000 changes in Democratic presidential vote shares.

It follows from Tables 1 and 2 that the 2004 presidential race was not appreciably different from recent races in New Hampshire with respect to potential voting technology effects. Indeed, on the basis of Table 1, one might be tempted to surmise that there were tabulating problems in New Hampshire in 2000 and in 2002 as well as in 2004.

[5] The 234 towns are those profiled by the New Hampshire Economic and Labor Market Information Bureau. The towns reported a positive number of presidential votes in 2004.

Table 2
Distribution of changes in Democratic vote share across voting technologies and various general elections

	Accuvote	Optech	PBHC	Dispersion	Deviance	*p*-Dev	*p*-KS
President							
2004–2000	0.149		0.215	0.080	0.197	0.116	0.002
		0.154	0.278	0.091	0.335	0.053	0.046
	0.105	0.823		0.082	0.009	0.740	0.219
Governor							
2004–2002	1.246		0.489	0.375	26.713	0.001	0.001
		1.352	0.695	0.338	9.152	0.001	0.001
	0.864	0.658		0.215	0.781	0.05	0.066

See Table 1 for definitions of Dispersion, Deviance, *p*-Dev, and *p*-KS. Weighted vote share changes are $\left(p_i - p_i'\right)\Big/\sqrt{(\sigma^2 p_i(1-p_i)/n_i)+\left(\sigma'^2 p_i'\left(1-p_i'\right)/n_i'\right)}$, where a prime indicates quantities from an earlier time period. See Goodman (1961).

These differences in vote shares and changes in vote shares across voting technologies would be compelling and seriously troubling if they were a function of a bias in a technology's vote tabulation. However, a causal interpretation could be attached to the differences between technologies in Tables 1 and 2 only if one could assume that voting technologies were randomly distributed across precincts in New Hampshire. This is unlikely to be the case. The seemingly ubiquitous technology effects in the state illustrate the difficulties of drawing causal inference from vote return data that are not the product of an experiment intended to evaluate the consequences of administrative practices, i.e., voting technology choices. That is, the results in Tables 1 and 2 could reflect town heterogeneity in matters than affect both the distribution of voter preferences *and* the probability of a town's using a particular voting technology.

Answering the question of whether there were technology-based irregularities in the New Hampshire presidential election of 2004 requires that one avoid confounding pre-election differences among precincts with differences in the means by which such precincts count votes. We thus carry out two statistical analyses of New Hampshire voting precincts, analyses that seek to take into account observable precinct heterogeneity. Section 3's regression analysis considers all precincts in New Hampshire, but it makes assumptions about the linearity of key relationships. Section 4's matching analysis considers only subsets of precincts that have common support on observables, and it makes fewer assumptions. Our two analyses are complementary in that they have alternative strengths and weaknesses. That results of these two complementary methods are similar strengthens our claim that there was not a bias in the tabulation of votes by a particular voting technology.

3. Regression analysis of precinct vote returns

We first examine the allegation of Accuvote problems in New Hampshire in 2004 with a series of precinct-level regressions where the dependent variable in our regressions is vote share for Kerry of the total Democratic and Republican vote. We estimate weighted least squares regressions as well as overdispersed grouped logistical regressions. With respect to the latter, the regressions are grouped since we observe votes aggregated by precinct as opposed to individual votes; they are logistic in that we use a logit link between precinct characteristics and vote shares; and they are overdispersed in that we allow for variance beyond nominal binomial variance that would be expected in an ordinary grouped logistic model.[6]

Because of the possibility that some New Hampshire precinct vote shares may have been produced by an irregular process, such as vote tabulating or other administrative problems, our grouped logistic regressions use an estimator robust to outliers. This estimate is also robust to errors in right hand side covariates, i.e., Gore vote share from 2000, which is important insofar as Tables 1 and 2 could be understood as implying that there were vote-tabulating problems in elections before 2004. The robust estimator was developed in Mebane and Sekhon (2004b), and we use software from Mebane and Sekhon (2004a). An earlier version of the regression model was used to uncover the Palm Beach County butterfly ballot effect in the 2000 presidential election (Wand et al., 2001).

The right hand side covariates appearing in our regression models are as follows. Gore (2000) is the

[6] See McCullagh and Nelder (1989) for a discussion of overdispersed models generated from clustering of similar individuals within units of observation.

fraction of presidential votes cast for Democrat Al Gore in the 2000 general election; Fernald (2002) is the vote share for Mark Fernald where Fernald was the losing Democratic candidate for New Hampshire governor in 2002; Shaheen (2002) is the vote share of Jeanne Shaheen, the Democratic U.S. Senate candidate in 2002 in New Hampshire (Shaheen lost to John Sununu); Democratic Checklist (2004) is the fraction of voters listed as Democrats of all voters on a precinct's checklist; Republican Checklist (2004) is the same but for Republican affiliation; Democratic Primary ballots is the fraction of ballots cast in the 2004 presidential primary that were Democratic; Per capita income (1999) is as described; Black (2000) is fraction of residents who describe themselves as single-race African-American; Accuvote is an indicator for a precinct's using Accuvote machines; and Optech is the same but for Optech voting machines. Our regression models also include nine county indicator variables that proxy for otherwise unmodeled county-wide effects.[7]

We include a variety of historical election returns as independent variables in our regressions, and this is because estimating voting technology effects requires controlling for underlying and heterogeneous precinct political tendencies. Thus, the Accuvote and Optech indicator variables in our regression models pick up machine voting effects on Kerry vote shares that are in addition to what Kerry would have received in a precinct based on its political predisposition. Furthermore, some of our regressions include income and race demographics insofar as there are known income and racial effects on political preferences.

Table 3 reports coefficient estimates for six regression models. Each model uses wards, i.e., subdivisions of towns, when possible. For Models 1 and 4, which compare 2004 election results with, among other things, 2002 election results, towns whose wards had constant boundaries are disaggregated by ward. However, wards in the town of Laconia received new boundaries for the 2004 election, so all of Laconia's wards are aggregated to the town level for these two models. Models 2 and 5 require constant boundaries between 2000 and 2004, and so in addition to Laconia the towns of Claremont, Concord, Dover, Keene, Nashua, and Rochester are aggregated by ward. The two demographics used in Models 3 and 6 are available only at the town level, so no wards can be used in estimating Model these two models.[8]

With respect to voting technology, all six models in Table 3 have qualitatively identical results: none of the twelve Accuvote and Optech estimates is significant at conventional confidence levels. Since we include wards when possible, Table 3's results are on the same level of analysis as the Briggs study of New Hampshire precincts. Unlike the latter, though, they control for precinct heterogeneity in a variety of ways. Doing so reveals what the recount itself illustrated: voting technology did not have an effect on presidential vote outcomes in New Hampshire.

One can see from the coefficient estimates in Table 3 that Kerry vote share in 2004 was positively associated with Gore vote share in 2000, with Fernald vote share in 2002, and with Shaheen vote share in 2002 as well. These results are intuitive and suggest that our right hand side covariates are picking up precinct political profiles.

Fig. 1 plots studentized residuals from Model 6 of Table 3 (comparable figures for Models 4 and 5 are qualitatively similar). These residuals, studentized so that they lie on a normal scale, capture Kerry vote shares that are unexplained beyond covariates.

The most salient feature of the figure is the lack of major differences between the densities of Accuvote, Optech, and PBHC studentized residuals. Had Accuvote precincts included lower Kerry vote shares even controlling for the precinct features noted in Table 3, then we would have seen evidence of this in a left-skewed density for Accuvote precincts.

Furthermore, had several Accuvote precincts been manipulated so as to produce unusually high Bush vote totals, we would have observed large and negative Kerry residuals in Fig. 1, just as Wand et al. (2001) found a large and positive Buchanan residual in Palm Beach County, Florida, in 2000. Indeed, the lack of anomalous residuals among Accuvote, Optech, and PBHC densities is evidence that the 2004 presidential election in New Hampshire was relatively clean.

As a consistency check on our presidential election tabulating results we estimate the six regression models of Table 3 with Lynch vote share as the dependent variable; recall that Lynch was the winning Democratic candidate for governor in 2004. It is

[7] Per capita income and African-American figures were culled from U.S. Census data. Although the Census Bureau publishes county total population projections between decennial censuses, it does not do this for New Hampshire towns.

[8] There are four New Hampshire towns, Dixville, Hale's Location, Millsfield, and Wentworth's Location, for which basic demographics are unavailable. These towns, all of which use PBHC, are not tracked by the Economic and Location Market Information Bureau.

Table 3
Regression estimates for Kerry vote share models

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.174*	0.0758*	0.0614*	−1.35*	−1.73*	−1.78*
	(0.0230)	(0.0280)	(0.0277)	(0.104)	(0.122)	(0.121)
Accuvote	−0.00206	−0.00632	−0.00646	0.00411	−0.196	−0.0158
	(0.00441)	(0.00434)	(0.00394)	(0.0186)	(0.0188)	(0.0162)
Optech	−0.00290	−0.00791	−0.00935	−0.0180	−0.0212	−0.0191
	(0.00543)	(0.00533)	(0.00492)	(0.0253)	(0.0262)	(0.0222)
Fernald (2002)	0.464*	0.447*	0.380*	1.93*	1.87*	1.60*
	(0.0375)	(0.0375)	(0.0380)	(0.158)	(0.168)	(0.173)
Shaheen (2002)	0.284*	0.109	0.134*	1.24*	0.493	0.636*
	(0.0512)	(0.0599)	(0.0608)	(0.232)	(0.253)	(0.251)
Democratic Checklist (2004)	0.193*	0.0905*	0.103*	0.678*	0.356*	0.421*
	(0.0303)	(0.0349)	(0.0372)	(0.131)	(0.149)	(0.141)
Republican Checklist (2004)	−0.135*	−0.0334	−0.0690	−0.569*	−0.179	−0.324*
	(0.0318)	(0.0354)	(0.0369)	(0.0143)	(0.155)	(0.153)
Democratic Primary ballots	−0.00381	−0.00357	0.00154	−0.0277	−0.242	−0.000435
	(0.00565)	(0.00581)	(0.00577)	(0.0219)	(0.0238)	(0.0245)
Gore (2000)		0.395*	0.407*		1.55*	1.49*
		(0.0708)	(0.0723)		(0.296)	(0.293)
Per capita income (1999)			0.0130*			0.0594*
			(0.00288)			(0.0107)
Black (2000)			0.160			1.19
			(0.343)			(1.48)
R^2	0.935	0.945	0.954			
σ^2	1.11	1.03	0.947	1.91	1.93	1.81
N	299	264	238	299	264	238

Note: models 1–3 are estimated by weighed least squares where a precinct's weight is the sum of its Democratic and Republican presidential vote; for these models σ^2 is residual variance. Models 4–6 are robust overdispersed grouped logistic regressions; σ^2 is the overdispersion parameter where values of σ^2 greater than one denote overdispersed data. For all models county indicator variables are suppressed, standard errors are in parentheses under estimates, and * denotes $p < 0.05$.

important to consider a set of Lynch models as well as Kerry models insofar as problems associated with voting technology, if they existed, would not necessarily be limited to presidential contests. Moreover, if the Accuvote machines used in New Hampshire were systematic biased in their tabulations of Kerry votes, there is no reason to think that they would be accurate for the Democratic Lynch, who won a close

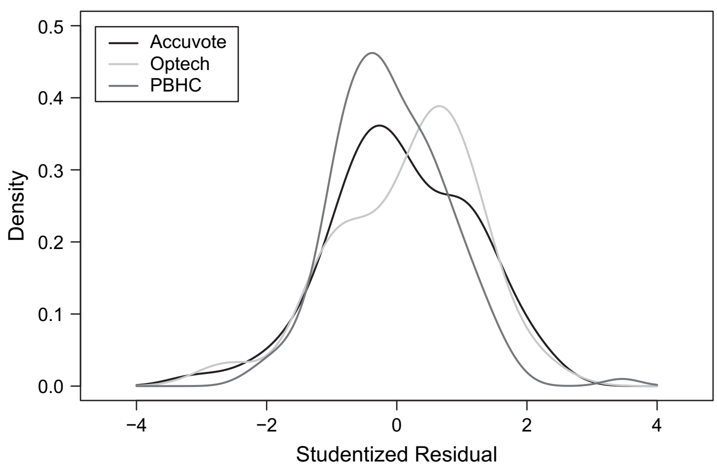

Fig. 1. Smoothed histograms of studentized residuals from Table 3, Model 6.

Table 4
Regression estimates for Lynch vote share models

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.196*	0.125*	0.117*	−1.27*	−1.45*	−1.44*
	(0.0264)	(0.0338)	(0.0360)	(0.106)	(0.124)	(0.128)
Accuvote	−0.000753	−0.00327	−0.00451	0.00385	−0.0135	−0.0167
	(0.00506)	(0.00524)	(0.00513)	(0.0192)	(0.0204)	(0.0184)
Optech	−0.00554	−0.00878	−0.0106	−0.03170	−0.0443	−0.0414
	(0.00623)	(−0.00644)	(0.00641)	(0.0260)	(0.0274)	(0.0241)
Fernald (2002)	0.678*	0.679*	0.630*	2.66*	2.64*	2.30*
	(0.0431)	(0.0454)	(0.0496)	(0.180)	(0.190)	(0.206)
Shaheen (2002)	0.141*	0.0103	0.022	0.762*	0.348	0.619
	(0.0588)	(0.0724)	(0.0791)	(0.0244)	(0.300)	(0.319)
Democratic Checklist (2004)	0.161*	0.0890*	0.115*	0.666*	0.391*	0.438*
	(0.0348)	(0.0422)	(0.0484)	(0.143)	(0.167)	(0.185)
Republican Checklist (2004)	−0.0279	0.0443	0.0162	−0.0878	0.0503	−0.0110
	(0.0365)	(0.0429)	(0.0480)	(0.147)	(0.162)	(0.166)
Democratic Primary ballots	−0.00413	−0.00225	0.00132	−0.0332	−0.0283	0.0138
	(0.00649)	(0.00703)	(0.00752)	(0.0266)	(0.0300)	(0.0304)
Gore (2000)		0.276*	0.302*		0.797*	0.739*
		(0.276)	(0.0939)		(0.342)	(0.362)
Per capita			0.00564			0.0107
income (1999)			(0.00376)			(0.0143)
Black (2000)			−0.227			−1.44
			(0.448)			(1.71)
R^2	0.926	0.932	0.954			
σ^2	1.45	1.51	1.59	2.11	2.16	2.05
N	299	264	238	299	264	238

Note: models 1–3 are estimated by weighed least squares where a precinct's weight is the sum of its Democratic and Republican presidential vote; for these models σ^2 is residual variance. Models 4–6 are robust overdispersed grouped logistic regressions; σ^2 is the overdispersion parameter where values of σ^2 greater than one denote overdispersed data. For all models county indicator variables are suppressed, standard errors are in parentheses under estimates, and * denotes $p < 0.05$.

race with Republican incumbent governor Craig Benson.

The results from the Lynch regression models can be found in Table 4, and they are qualitatively similar to those of the Kerry models. Note that, among other things, Gore fraction from 2000 is positively and significantly correlated with Lynch fraction in 2004, Fernald fraction in 2002 is correlated in this way, and Shaheen fraction in 2002 is significantly positively correlated as well. Most importantly, none of the Accuvote not Optech coefficients in the Lynch models is statistically significant at conventional confidence levels.

4. Matching analysis

Our regression results, like all regression results, are dependent on functional form assumptions such as linearity and additivity. Thus, we now conduct a complementary analysis of the 2004 New Hampshire presidential election using multivariate propensity score matching (Rosenbaum and Rubin, 1983; Diamond and Sekhon, 2005) which makes weaker assumptions and ensures that cases being compared have common support on observable characteristics. Multivariate propensity score matching provides a non-parametric method for drawing inference from a set of units, in our case New Hampshire precincts, that are comparable on baseline observables. Our matching analysis uses algorithms and code provided by Sekhon (2004b).

To illustrate the challenges posed by observational data and the use of matching to overcome these challenges, it is useful to consider first the advantages of hypothetically studying elections in an experimental context. If we were studying the effects of voting technology in New Hampshire and had control over the allocation of technologies, then prior to the 2000 general election we as experimenters would have randomly assigned some New Hampshire precincts Accuvote, some Optech, and the remainder PBHC.

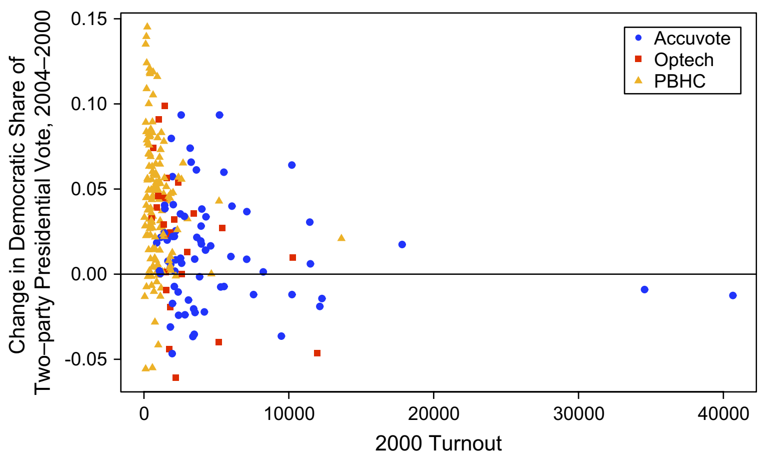

Fig. 2. Changes in presidential vote, 2004–2000, by precinct size and voting technology.

With random assignment of voting technology, the use of a particular technology would be expected to be a priori uncorrelated with future votes. Relatedly, groups of voters who used different technologies would also be expected to be *balanced* with respect to a large collection of variables. In other words, we would expect that the distribution of characteristics of precincts would be essentially the same for each type of precinct, type being defined by a precinct's use of voting technology. For example, in our hypothetical experiment we would expect that the fraction of Accuvote precincts that is large (or rural, or Democratic-leaning, and so forth) would be equivalent to the fraction of Optech and PBHC precincts that is large (or rural, etc.). If balance across precincts were achieved through random assignments, simple tests for differences in means, weighted by precinct sizes, could suffice to determine whether the Accuvote was responsible for vote tabulation problems.

Of course, a randomized experiment like this did not occur in New Hampshire. Instead, we have to rely on observational data, as is currently the case for most research on voting technology and voting administration (see Sinclair et al., 2000 for an exception).

Conclusions drawn from precinct-level observational data can suffer from bias if subject precincts are not balanced with respect to variables that influence electoral behavior. For example, the fraction of large Accuvote precincts in New Hampshire in 2004 may differ from the fraction of large Optech and large PBHC precincts. Large precincts may be more common in cities, and cities may be home to a disproportionate share of Democratic-leaning residents compared to smaller rural localities. This can induce a correlation between voting technology and Democratic vote share that is not a consequence of an irregularity in vote tabulation.

Indeed, the lack of comparability across precincts with different voting in New Hampshire is clear in Fig. 2, which plots the 2004–2000 change in the Democratic share of the two-party vote for president (Kerry minus Gore) at the precinct level by two-party turnout in 2000. The symbols in the figure indicate whether a given precinct used Accuvote in the 2004 general election. It is apparent from Fig. 2 that there are no small precincts that in 2004 used Accuvote (toward the left of Fig. 2 one finds triangles almost exclusively), and Accuvote machines are used by the very largest precincts (toward the right there are almost exclusively dots). Optech precincts, denoted with squares, are neither extremely small nor extremely large. Naively comparing precincts that are not comparable on baseline features, such comparing small New Hampshire PBHC precincts with medium-sized Accuvote precincts, may lead to biased estimates of the effect of Accuvote on vote share if precinct size is correlated with a precinct characteristic that affected the likelihood that a precinct's residents supported Kerry for president.

One could make this same point using variables other than precinct size. For instance, Fig. 3 contains a scatter plot of the same Kerry–Gore difference versus Gore vote share in 2000. The figure shows that there is a notable absence of extremely conservative (low Gore percentage) precincts in New Hampshire that use Accuvote machines. Toward the left of the figure one sees a plethora of triangles but almost no circles (and almost no squares).

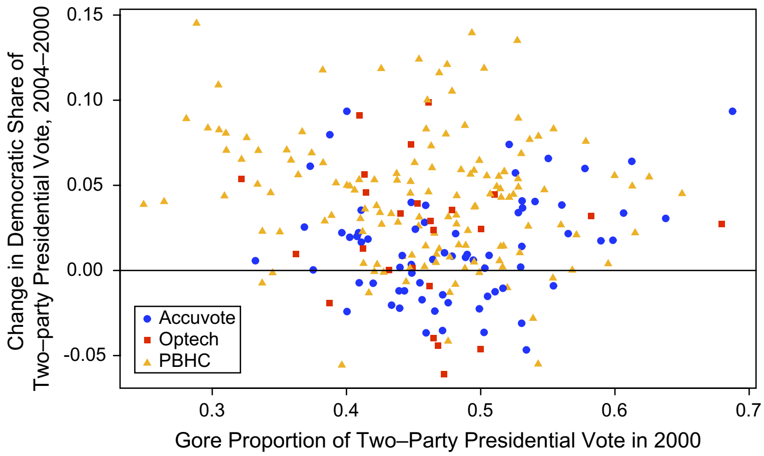

Fig. 3. Changes in presidential vote, 2004−2000, by Gore percent in 2000 and voting technology.

Finally, the non-uniform distribution of voting technology across New Hampshire is evident in Fig. 4, a map which describes Kerry vote share by town and town-level voting technology. Accuvote machines, and optical scan technology in general, are much more common in southern New Hampshire, an area that tends to be more Democratic and have higher population density than the rest of New Hampshire. Were Accuvote a randomly assigned treatment to New Hampshire precincts, we would expect to see some Accuvote machines in the northern tip of the state and more along the state's western border with Vermont.

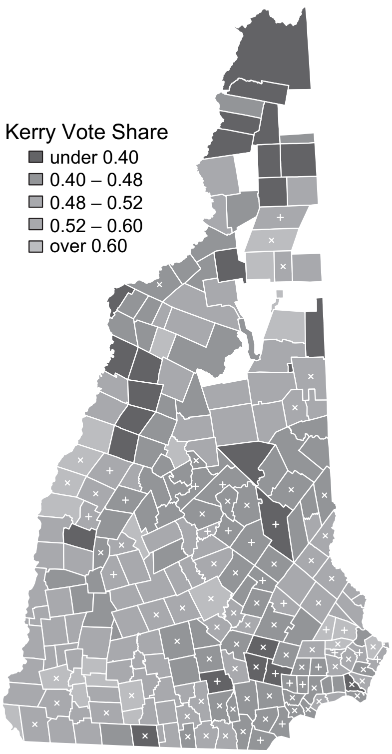

Fig. 4. Kerry vote share among 234 New Hampshire towns (×, Accuvote; +, Optech; no symbol, PBHC). White areas on the map reported zero presidential votes.

The two scatter plots and map together imply that one should be very cautious in drawing causal inferences about the effects of Accuvote (or Optech or PBHC) on vote tabulations without taking into account potentially confounding differences among types of voters who used these technologies. This suggests that we should treat with skepticism any inferences about the biases drawn from the differences and differences-in-differences shown in Tables 1 and 2, respectively.

Our matching analysis, which is a direct response to this skepticism, focuses on New Hampshire towns. This level of analysis takes advantage of demographic and economic characteristics that are available only at the town level. We aggregate vote returns to towns in places that are disaggregated into wards, e.g., the city of Manchester had twelve wards in 2004.[9] All towns with wards are unified in terms of having a single voting technology across all wards. In this section, then, the term "precinct" refers to a New Hampshire town.

[9] This requires aggregating in Laconia (six wards), Keene (four), Berlin (four), Lebanon (three), Manchester (twelve), Nashua (nine), Concord (ten), Franklin (three), Portsmouth (five), Dover (six), Rochester (six), Somersworth (five), and Claremont (three).

Our matching exercise consists of three separate matching analyses where each analysis considers two voting technologies from the set of three technologies used in New Hampshire. For example, our first analysis is Accuvote versus PBHC where Accuvote is considered a "treatment" to precincts. The criteria for matching in the Accuvote versus PBHC case is to choose for each Accuvote precinct a case from among the controls (i.e., from among PBHC precincts) that is closest in terms of (a) propensity to use Accuvote; (b) Gore's proportion of the two-party presidential vote in 2000 weighted by precinct size; and (c) Fernald's percentage of the two-party gubernatorial vote in 2002 weighted appropriately; (d) the average number of people voting in 2000 by number of wards in 2004; (e) precinct budget; and (f) precinct tax revenue. The propensity for a town to use Accuvote is a fitted probability from a logistic regression of Accuvote use on number of wards (minus one), number of 2000 voters divided by number of 2004 wards, and precinct budget. Distances between precincts are weighted according to the inverse of the variances of each matching variable, and a caliper excludes matches where the distance exceeds one standard deviation. Appendix A contains corresponding details on the Accuvote versus Optech analysis and the Optech versus PBHC analysis; it also describes a large collection of pre- and post-matching tests for balance.

Based on three sets of matched sets of precincts, Table 5 reports estimates of the effect of voting technology on presidential and gubernatorial vote shares in 2000, 2002, and 2004. The table is analogous to the earlier Table 1 except that Table 5 uses matched subsamples only.

The first three columns of Table 5 report vote shares. For instance, in the Accuvote versus PBHC test, Kerry vote share in matched Accuvote precincts was 0.476; in matched PBHC precincts it was 0.488. This difference in means is not significant at conventional confidence levels ($p \approx 0.5$), and based on a Kolmogorov–Smirnov test the distribution of Kerry vote share in Accuvote precincts is not statistically different than that in PBHC precincts ($p \approx 0.486$). In stark contrast to Table 1, differences that were statistically significant without consideration of balance become statistically insignificant once matched subsamples are used.[10]

[10] The dispersion parameters in Table 5 were not reestimated for matched samples. They are taken from the full sample estimates of Table 1.

Indeed, in all cases we observe a change from statistical significance to non-significance when comparing estimates in Table 1 and Table 5. For all three voting machine comparisons and for all races considered, there are in our matched subsamples no statistically significant voting machine effects at conventional confidence levels. The lack of significant results in matched subsamples is consistent with our regression results and with the claim that voting technology did not affect New Hampshire presidential vote shares in 2004.

Table 6 leads us to a similar conclusion. This table, analogous to the earlier Table 2, reports for matched subsamples voting machine effects on across-time changes in Democratic share of the two-party vote. As before, for all races considered and for all three voting machine comparisons, we cannot reject at conventional confidence levels the hypothesis that there were no voting machine-related tabulation biases in New Hampshire.

A visual depiction of our null results on voting technology effects can be see in Fig. 5, which plots the distribution of changes in 2004–2000 Democratic presidential support (Kerry minus Gore) and changes in 2004–2002 Democratic gubernatorial support (Lynch minus Fernald) by Accuvote and PBHC both before and after matching. The upper two panels in Fig. 5 reflect pre-matching vote returns, and it is clear from these panels that Accuvote precincts were associated with very large and positive changes in gubernatorial support from 2004–2002 compared to PBHC precincts. With respect to the 2004–2000 change in Democratic presidential support, one can see evidence of the opposite: Accuvote precincts had smaller changes. This, of course, motivated the 2004 presidential recount.

In contrast, the bottom two panels of Fig. 5 draw from post-matching data, and they depict Accuvote and PBHC densities that are very similar. As reported in Table 6, we cannot reject that the two changes in presidential support densities are different, and the same applies to the two changes in gubernatorial support densities. Although not shown, the other two matched comparisons (Accuvote versus Optech and Optech versus PBHC) produce very similar figures.

5. Aftermath of the recount

The New Hampshire presidential recount covered eleven precincts and concluded on November 30, 2004. Pre-recount and post-recount vote totals for

Table 5
Distribution in matched subsamples of Democratic share of two-party vote; number of matches: 29 (Accuvote versus PBHC), 18 (Optech versus PBHC), and 24 (Optech versus Accuvote)

	Machine						
	Accuvote	Optech	PBHC	Dispersion	Deviance	*p*-Dev	*p*-KS
President							
2000 (Gore)	0.476		0.488	44.3	21.1	0.5	0.486
2000 (Gore)		0.457	0.472	29.4	0.5	0.97	0.432
2000 (Gore)	0.482	0.484		88.9	0.27	0.96	0.983
2004 (Kerry)	0.509		0.509	71.1	0.1	1.0	0.708
2004 (Kerry)		0.440	0.439	41.0	14.1	0.56	0.429
2004 (Kerry)	0.482	0.499		151.5	43.3	0.6	0.987
Governor							
2002 (Fernald)	0.408		0.385	67.5	58.6	0.4	0.732
2002 (Fernald)		0.348	0.354	39.9	1.6	0.84	0.225
2002 (Fernald)	0.369	0.389		142.0	39.1	0.6	0.983
2004 (Lynch)	0.518		0.502	86.7	43.9	0.5	0.911
2004 (Lynch)		0.466	0.486	41.3	25.4	0.43	0.946
2004 (Lynch)	0.483	0.496		188.2	26.3	0.7	0.844

See Table 1 for definitions of Dispersion, Deviance, *p*-Dev, and *p*-KS.

Bush, Kerry, and Nader from the eleven are listed for the eleven in Table 7.

The three key columns in Table 7 are those labeled "Change," and it is fair to say that the recount's total vote changes for Bush (87) and Kerry (72) are negligible. The net Bush–Kerry change is 15 votes among 49,871 post-recount votes for Bush and Kerry. If, as alleged before the recount, there were Accuvote tabulating problems in New Hampshire that disproportionately and negatively affected Kerry, then we should have observed large net gains for Kerry in Litchfield, Manchester Wards 6, 7, 8, and 9, and in Somersworth Ward 4. Instead, Table 7 shows that Kerry gained 22 votes in recounted Accuvote precincts and 40 in Optech precincts. Moreover, and somewhat ironically, Bush recovered more votes than did the Kerry from the recount. The sizes of the shifts in vote totals due to the recount are small.

The results of a historical study of recounts help to further put the observed magnitude of the recount vote shifts into perspective. Ansolabehere and Reeves (2004) show that optical scan recounts in New Hampshire change on average 0.55% of the vote share for candidate. This figure is based on recounts across many different types of races, and vote share is defined as total votes for a candidate divided by ballots cast. Overall percentage changes for the 2004 presidential recount are 0.17% for Bush, 0.12% for Kerry, and 0.014% for Nader. Simple calculations show that corresponding percentage changes by precinct are slightly higher, since negative changes cancel out when computing the recount-wide adjustment, but not appreciably so. Thus, vote total changes produced by the 2004 recount are consistent with historical averages and if anything are less dramatic than the sorts of changes that have occurred in previous elections.

6. Conclusion

In events reminiscent of what transpired in New Hampshire, veteran political activist Jesse Jackson pointed out in December, 2004 that Ohio presidential election "challengers noticed Bush generally received more votes in counties that use optical-scan voting machines and [Jackson] questioned whether the machines were calibrated to record votes for Bush."[11] This is a serious charge and one that speaks to fundamental concerns about the legitimacy of presidential election outcomes. Nonetheless, assessing whether charges like Jackson's are valid requires a recognition that precincts and other voting aggregates like counties often differ with one another and that, most importantly, voting technology is not randomly assigned among them.

While studies of voting technology problems are almost inherently observational in nature as opposed to being experimental, the combination of techniques we apply here to the New Hampshire recount can be used fruitfully in any number of venues to identify election irregularities and investigate specific allegations of vote

[11] See "Jackson helps challenge ballot totals in Ohio," Chicago Sun-Times, December 14, 2004.

Table 6
Distribution in matched samples of changes in Democratic share of two-party vote; number of matches: 29 (Accuvote versus PBHC), 18 (Optech versus PBHC), and 24 (Optech versus Accuvote)

	Machine						
	Accuvote	Optech	PBHC	Dispersion	Deviance	*p*-Dev	*p*-KS
President							
2004–2000	0.234		0.147	0.080	0.168	0.148	0.095
		0.439	0.324	0.091	0.056	0.433	0.703
	0.138	0.956		0.082	0.015	0.665	0.062
Governor							
2004–2002	0.795		0.785	0.375	0.002	0.947	0.726
		1.316	1.405	0.338	0.049	0.703	0.21
	0.771	0.790		0.215	0.003	0.914	0.205

See Table 1 for definitions of Dispersion, Deviance, *p*-Dev, and *p*-KS.

fraud. Indeed, we believe adamantly that elections need to be subject to open and rigorous auditing far beyond what is currently done outside of rare cases (Mebane et al., 2004). In democratic elections, a primary goal should be that the preferences of all voters are accurately recorded and tabulated. Everything from the kind of voting machines used to ballot formats to the way poll workers respond to voters' questions have been shown to affect the rate at which voters make mistakes or the rate at which other tabulation errors occur (e.g., Brady et al., 2001; Wand et al., 2001; Tomz and van Houweling, 2003). Nonetheless, if one asks for a precise statement about the consequences of a specific election administration practice, or if the question at hand is which combination of election administration procedures is optimal, it quickly becomes apparent that academics and election administrators know remarkably little. To ensure the integrity of elections and to correct problems *before* they play a pivotal role in an election outcome, vigilant testing and auditing is required.

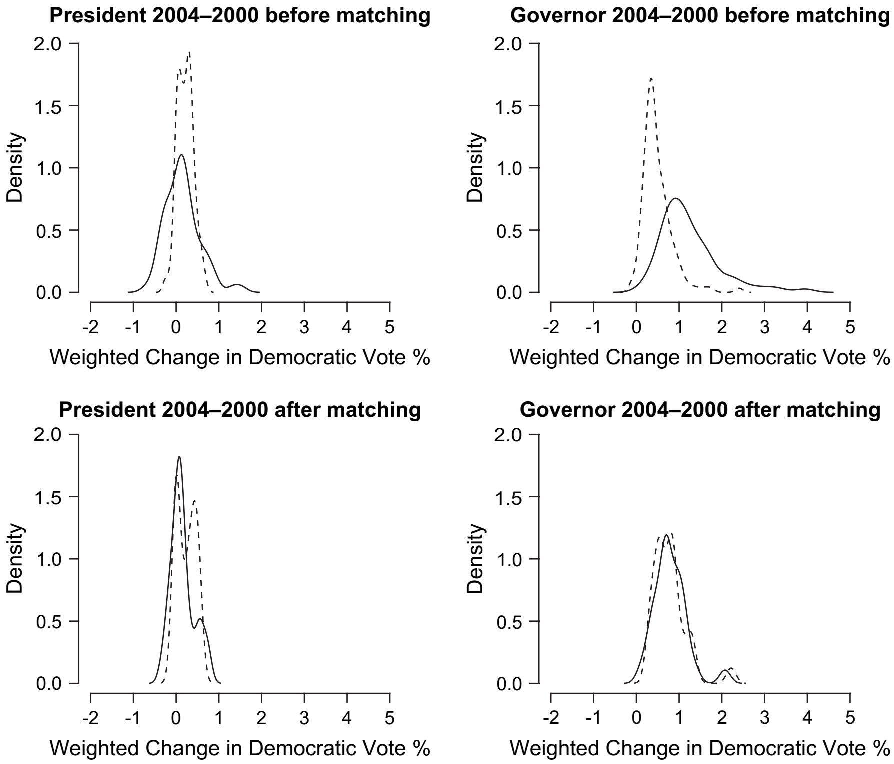

Fig. 5. Density of changes in Democratic share of two-party vote, Accuvote versus PBHC. Note: Solid lines denote Accuvote precincts and dashed lines, PBHC precincts.

Table 7
Recount results for eleven New Hampshire precincts

Precinct	Technology	Ballots	Bush			Kerry			Nader		
			Original	Recount	Change	Original	Recount	Change	Original	Recount	Change
Danville	Optech	2319	1261	1264	3	929	929	0	21	21	0
Litchfield	Accuvote	4170	2386	2389	3	1747	1750	3	16	17	1
Manchester 6	Accuvote	4270	2263	2265	2	1954	1956	2	21	21	0
Manchester 7	Accuvote	3656	1795	1795	0	1788	1791	3	30	29	−1
Manchester 8	Accuvote	4646	2613	2616	3	1983	1984	1	22	20	−2
Manchester 9	Accuvote	3996	2022	2032	10	1912	1919	7	21	22	−1
Pelham	Optech	6565	3725	3748	23	2755	2770	15	35	36	1
Newton	Accuvote	2302	1160	1160	0	1103	1102	−1	30	30	0
Salem	Optech	14535	7797	7821	24	6472	6490	18	91	97	6
Sandown	Optech	2866	1655	1670	15	1158	1165	7	16	17	1
Somersworth 4	Accuvote	1335	517	521	4	727	734	7	9	9	0
Total		50660	27194	27281	87	22528	22590	62	312	319	7

Acknowledgments

The authors thank John Carey, Linda Fowler, and two anonymous referees for helpful comments on earlier drafts.

Appendix A. Matching details

The criteria for determining whether treatment and control sets of precincts have been matched is based on whether both the means and the distributions of observable pre-treatment variables are indistinguishable between the two groups being tested, e.g., Accuvote towns and Optech towns. Tests for balance were performed on a large collection of observables, some of which are summarized in Table A1.[12]

The Accuvote versus PBHC matching analysis is described in the body of the paper. For Optech versus PBHC, matches are based on propensity to use Optech 2000 Gore percentage weighted appropriately, and number of 2000 voters divided by number of 2004 wards, with a three-quarters standard deviation caliper. The propensity score is a fitted probability from a logistic regression of Optech on number of wards (minus one) and number of 2000 voters divided by number of 2004 wards. For Optech versus Accuvote, we could not find a propensity score that produced matches; instead, matches were based on 2000 Gore percentage weighted appropriately and number of votes in 2000 divided by number of wards in 2004, with a one-half standard deviation caliper.

The top section of Table A1 reports tests for balance in the Accuvote versus PBHC test. "Before matching" results are based on the full set of Accuvote and PBHC precincts, and "After Matching" refers to matched subsamples. The table reports means for various observables, p values for difference in means tests, and p values for bootstrapped, univariate Kolmogorov–Smirnov tests. The middle section of the table pertains to Accuvote versus Optech and the bottom to Optech versus PBHC.

The majority of difference in means tests and Kolmogorov–Smirnov tests are statistically significant before matching. For instance, compared to PBHC precincts Accuvote precincts had more wards (1.73 versus 1.09), larger budgets (15.5 million versus 2.57 million), more black residents, more Hispanic residents, and fewer white residents. All of these differences are significant at conventional confidence levels. This result and others like them show that Accuvote and PBHC precincts are not similar on key observables and as such naive comparisons between them may confound real and spurious effects of technology. The low p values in the bottom two thirds of Table A1 imply that this would be true for an Accuvote versus Optech comparison as well as an Optech versus PBHC comparison.

The rightmost four columns of Table A1 report results for matched subsamples. In contrast to the unmatched samples, for each of the variables considered the three groups of matched subsamples are indistinguishable in their means and distributions. For instance, matched PBHC and matched Accuvote precincts have similar numbers of wards (1.21 versus 1.24), and the difference between these two averages is not significant at a conventional confidence level ($p \approx 0.318$).

[12] Source and definitions for town data on budgets and services is available from the New Hampshire Economic and Labor Market Information Bureau at http://www.nhes.state.nh.us/elmi/textonly/communpro.htm.

Table A1
Comparison of precinct characteristics before and after matching

Variable	Before matching				After matching			
	Mean 1	Mean 2	*p-t*	*p*-KS	Mean 1	Mean 2	*p-t*	*p*-KS
Propensity	0.760	0.121	<0.001	<0.001	0.586	0.567	0.101	0.992
Wards	1.73	1.09	0.0185	0.003	1.21	1.24	0.318	0.928
Per capita budget	1.11	1.14	0.837	0.324	1.05	1.19	0.408	0.176
Budget	15.5	2.57	0.000486	<0.001	6.60	7.55	0.18206	0.093
Tax ratio	77.9	78.5	0.822	0.235	73.3	73.0	0.919	0.456
Tax	24.6	23.8	0.490	0.644	25.3	26.4	0.101	0.528
Commercial	15.3	8.04	<0.001	<0.001	12.9	13.5	0.781	0.297
Per capita income	24858	22101	0.00147	0.001	24196	22471	0.102	0.11
Black	0.00418	0.00202	0.000119	<0.001	0.00309	0.00307	0.970	0.903
Hispanic	0.0101	0.00608	0.000391	<0.001	0.00786	0.00682	0.191	0.097
White	0.965	0.975	0.00109	0.005	0.972	0.974	0.368	0.326
Population	11599	2020	<0.001	<0.001	5090	5300	0.398	0.99
Propensity	0.333	0.238	0.00168	<0.001	0.341	0.340	0.824	0.983
Wards	1	1.73	0.00638	0.04	1	1	1	1
Per capita budget	0.983	1.12	0.303	0.706	0.964	1.19	0.150	0.393
Budget	5.87	15.5	0.0131	0.016	5.035	5.73	0.231	0.572
Tax ratio	75.3	77.9	0.518	0.17	75.8	74.1	0.758	0.632
Tax	23.9	24.6	0.672	0.589	24.1	25.1	0.694	0.608
Commercial	11.0	15.3	0.105	0.03	10.8	12.7	0.526	0.369
Per capita Income	24652	24858	0.874	0.899	24107	24206	0.947	0.628
Black	0.00285	0.00418	0.0327	0.276	0.00284	0.00272	0.851	0.381
Hispanic	0.00806	0.0101	0.119	0.123	0.00801	0.00770	0.768	0.386
White	0.970	0.965	0.205	0.588	0.970	0.970	0.935	0.196
Population	5723	11599	0.0136	<0.001	5200	5264	0.849	0.397
Propensity	0.427	0.103	<0.001	<0.001	0.251	0.249	0.892	0.698
Wards	1	1.094	0.0631	0.206	1	1	1	1
Per capita budget	0.982	1.14	0.276	0.311	0.949	1.13	0.263	0.108
Budget	5.87	2.57	0.0503	<0.001	3.08	3.54	0.470	0.705
Tax ratio	75.3	78.5	0.398	0.118	74.5	68.9	0.324	0.21
Tax	23.9	23.8	0.910	0.544	25.1	28.1	0.191	0.415
Commercial	11.0	8.04	0.228	0.189	10.0	11.3	0.728	0.705
Per capita income	24652	22101	0.0415	<0.001	23584	22094	0.202	0.208
Black	0.00285	0.00202	0.0494	<0.001	0.00243	0.00203	0.457	0.101
Hispanic	0.00806	0.00608	0.0343	<0.001	0.00709	0.00618	0.346	0.682
White	0.970	0.975	0.0753	0.028	0.974	0.976	0.419	0.711
Population	5723	2018	0.00601	<0.001	3213	3172	0.786	0.941

Note: the top one-third compares Accuvote (Mean 1) versus PBHC (Mean 2); the middle third is Accuvote (Mean 1) versus Optech (Mean 2); and the final third is Optech (Mean 1) versus PBHC (Mean 2). The two columns labeled "*p-t*" report *p* values for difference of mean *t*-tests, and the "*p*-KS" columns refer to *p* values from Kolmogorov–Smirnov tests. Budget is in millions of dollars.

Results from additional balance tests on political variables are available from the authors. All tests imply that our treatment and control groups are indeed balanced.

References

Ansolabehere, S., Reeves, A., 2004. Using Recounts to Measure the Accuracy of Vote Tabulations: Evidence from New Hampshire Elections 1946–2002. Working paper, Massachusetts Institute of Technology.

Brady, H.E., Buchler, J., Jarvis, M., McNulty, J., 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.

Briggs, I., 2004. New Hampshire Data. Available at http://www.invisibleida.com/New_Hampshire.htm.

Card, D., Moretti, E., 2005. Does Voting Technology Affect Election Outcomes? Touch-Screen Voting and the 2004 Presidential Election. NBER Working Paper 11309.

Diamond, A., Sekhon, J.S., 2005. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. Working Paper, University of California at Berkeley.

Goodman, L.A., 1961. Modifications of the Dorn-Stouffer-Tibbitts method for 'Testing the significance of comparisons in sociological Data'. The American Journal of Sociology 66 (4), 355–363.

Hout, M., Mangels, L., Carlson, J., Best, R., 2004. The Effect of Electronic Voting Machines on Change in Support for Bush in the 2004 Florida Elections. Unpublished working paper, University of California, Berkeley.

Knack, S., Kropf, M., 2002. Who uses inferior voting technology? PS: Political Science and Politics 35 (3), 541—548.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, Second ed. Chapman and Hall, Cambridge.

Mebane Jr., W.R., Sekhon, J.S., 2004a. MultinomRob: Robust Estimation Software. R library. http://sekhon.berkeley.edu/robust/.

Mebane Jr., W.R., Sekhon, J.S., 2004b. Robust estimation and outlier detection for overdispersed multinomial models of count data. American Journal of Political Science 48 (2), 391—410.

Mebane Jr., W.R., Sekhon, J.S., Wand, J., 2004. Detecting and Correcting Election Irregularities. Working paper, http://wand.stanford.edu/research/detecting.pdf.

Merzer, M., 2001. The Miami Herald Report: Democracy Held Hostage. St. Martin's Press, New York.

Posner, R.A., 2001. Breaking the Deadlock: The 2000 Election, the Constitution, and the Courts. Princeton University Press, Princeton, NJ.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41—55.

Sekhon, J.S., 2004a. The 2004 Florida Optical Voting Machine Controversy: A Causal Analysis Using Matching. Unpublished working paper, Harvard University.

Sekhon, J.S., 2004b. Matching: Multivariate and Propensity Score Matching Software. R library. http://sekhon.berkeley.edu/matching/.

Sinclair, R.C., Mark, M.M., Moore, S.E., Lavis, C.A., Soldat, A.S., 2000. An electoral butterfly effect. Nature 408, 665—666.

The Washington Post, 2001. Deadlock: The Inside Story of America's Closest Election. PublicAffairs, New York.

Tomz, M., van Houweling, R.P., 2003. How does voting equipment affect the racial gap in voided ballots? American Journal of Political Science 47 (1), 46—60.

Wand, J., 2004. Evaluating the Impact of Voting Technology on the Tabulation of Voter Preferences: the 2004 presidential election in Florida. Working paper, http://wand.stanford.edu/elections/us/FL2004.

Wand, J.N., Shotts, K.W., Sekhon, J.S., Mebane Jr., W.R., Herron, M.C., Brady, H.E., 2001. The butterfly did it: the aberrant vote for Buchanan in Palm Beach County, Florida. American Political Science Review 95 (4), 793—810.