# The 2004 Florida Optical Voting Machine Controversy: A Causal Analysis Using Matching*

Jasjeet S. Sekhon †

11/14/2004 (23:48)

Preliminary and Incomplete, Comments Welcome

*This work is part of a joint project with Walter Mebane, Jr., Jonathan Wand and Michael Herron. I thank them for valuable comments, advice and data. The **R** package "Matching" was used to preform all of the analysis in this note. "Matching" is available at `HTTP://jsekhon.fas.harvard.edu/matching`. It is also available along with **R** at `HTTP://www.r-project.org/`.

†Associate Professor, Department of Government, Harvard University. 34 Kirkland Street, Cambridge, MA, 02138. `jasjeet_sekhon@harvard.edu`, `HTTP://jsekhon.fas.harvard.edu/`, Voice: 617.496.2426, Fax: 617.496.5149.

In the aftermath of the 2004 Presidential election, a significant number of activists and researchers have argued that the optical voting machines that are used in a majority of Florida counties caused John Kerry to receive fewer votes than "Direct Recording Electronic" (DRE) voting machines. Implicitly, these researchers wish to estimate the causal effect of using optical versus DRE voting machines. There are two difficulties which must be confronted when one attempts to estimate this causal effect. First, the type of voting machine that a county uses was not randomly assigned. Second, the 67 counties in Florida are extremely heterogeneous. For example, the portion of the population which is white (including Hispanic whites) ranges from 0.41 to 0.96,[1] and the proportion of the population which is registered Democrat ranges from 0.24 to 0.89.[2]

Both of these issues are common when one tries to make causal inferences from observational data. The issues give rise to the problem of confounding—i.e., the distributions of pre-treatment variables, such as party registration, past votes and demographics, differ greatly between those counties that use optical voting machines and the counties that use DREs. These baseline differences must be accounted for before any valid causal inference can be made. A common way to account for these variables is by comparing matched observations. Matching estimators do not make functional form assumptions and are hence appealing.[3] For the optical voting machine question, a matching estimator greatly reduces the imbalances in baseline variables and produces the most reliable statistical results we currently have. The estimated causal effect of optical voting machines on the Kerry vote is indistinguishable from zero. These results give **no** support to the conjecture that optical voting machines resulted in fewer Kerry votes than the DREs would have.

A simple multivariate matching model is used. This model attempts to match counties using the following variables: registration in 2004 (Republican, Democrat and Independent), the proportion of the population which is white, the proportion of the population which is

[1] 2003 Census Bureau estimates.

[2] 2004 registration data from XX.

[3] For an extended discussion of matching estimators and making causal inferences from observational data see Rosenbaum (2002).

black, the proportion of the population which is Hispanic and the population size. As is discussed later, this simple multivariate matching model achieves balance on a large number of additional variables including historical registration, votes and additional demographics. This is possible because the baseline variables are highly correlated.

Table 1 presents the results for this matching estimator. The estimated Average Treatment effect for the Treated (ATT) is 0.00540 with a standard error of 0.0211. This is clearly insignificant (p-value = 0.798). These results are obtaining using a regression bias adjustment. Footnote 4 lists the many baseline variables which are used to perform this bias adjustment. If bias adjustment is not used, the estimated ATT is −0.0104, and the standard error is 0.0216. Please see the "Methodological Details" section for further discussion.

Table 1: Average Treatment Effect for the Treated

| | |
|---|---|
| Estimate | 0.00540 |
| SE | 0.0211 |
| p-value | 0.798 |

The variances of the matched groups are also comparable. The variance of the optical counties is 0.262 and for the DREs counties it is 0.269. With an over-dispersed count model (McCullagh and Nelder 1989, 125), these two variances are not significantly different. Estimating the average treatment effect also results in an insignificant treatment effect, but balance is far worse than for ATT.

In order to achieve good balance on the baseline covariates, only 16 counties remain in the analysis: 8 optical and 8 DRE counties. Table 2 lists these counties. Table 3 lists the pre- and post-matching balance on 34 key baseline variables.

Table 3 makes clear the difficult inference problem this dataset creates. Before matching, the optical and DRE counties are radically different. Comparisons between counties which use optical voting machines and those which use DREs are hopelessly confounded. After matching, the counties are significantly more homogenous. Indeed, there is a surprising amount of balance given that our units of analysis are counties.

The matching estimator is able to greatly reduce the risk of confounding by observed variables, but given the limitations of this data, it is not able to eliminate it entirely. Many counties must be dropped from the analysis because it is impossible to find good matches for them. To make further progress, precinct level data must be used. At that level of resolution, it is plausible that matches could be found that would remove a greater proportion of the bias induced by the baseline variables. One may be able to make inference for regions of the state where the current analysis is unable to say anything. For example, with the current data, inferences about Palm Beach County or Miami-Dade are hopelessly confounded.

In conclusion, there is no support in this data for the contention that optical voting machines had a significant causal effect on the Kerry vote.

Table 2: Matched Counties

| Optical Counties | | DRE Counties | |
|---|---|---|---|
| | Kerry prop* | | Kerry prop |
| Bay | .281 | Sumter | .364 |
| Citrus | .421 | Pasco | .444 |
| Hernando | .462 | Pasco | .444 |
| Manatee | .427 | Lee | .391 |
| Marion | .410 | Sumter | .364 |
| Orange | .498 | Hillsborough | .462 |
| St. Johns | .306 | Martin | .417 |
| Walton | .259 | Nassau | .262 |

* Kerry vote proportion.

# 1 Methodological Details

The **R** package "Matching" was used to preform all of the analysis in this note. "Matching" is available at `HTTP://jsekhon.fas.harvard.edu/matching`. It is also available along with **R** at `HTTP://www.r-project.org/`.

A caliper of 0.25 standard deviations was used to perform the matching. And bias

Table 3: Balance Tests

| Variable | Before Matching: Mean Optical | Mean DRE | t-test | ks test | After Matching: Mean Optical | Mean DRE | t-test | ks test |
|---|---|---|---|---|---|---|---|---|
| Population | 143093 | 638550 | 0.015 | 0.000 | 268287 | 332241 | 0.300 | 0.561 |
| Dem reg ’04 | 0.551 | 0.361 | 0.000 | 0.000 | 0.369 | 0.365 | 0.556 | 0.534 |
| Rep reg ’04 | 0.314 | 0.430 | 0.000 | 0.001 | 0.441 | 0.439 | 0.723 | 0.91 |
| Ind reg ’04 | 0.113 | 0.179 | 0.000 | 0.000 | 0.162 | 0.167 | 0.382 | 0.914 |
| Ref reg ’04 | $3.75\times10^{-4}$ | $4.25\times10^{-4}$ | 0.363 | 0.359 | $4.06\times10^{-4}$ | $4.38\times10^{-4}$ | 0.126 | 0.698 |
| Green reg ’04 | $4.59\times10^{-4}$ | $5.83\times10^{-4}$ | 0.212 | 0.011 | $6.49\times10^{-4}$ | $4.93\times10^{-4}$ | 0.231 | 0.537 |
| Dem white reg ’04 | 0.418 | 0.266 | 0.000 | 0.000 | 0.289 | 0.295 | 0.608 | 0.914 |
| Dem black reg ’04 | 0.103 | 0.0589 | 0.017 | 0.005 | 0.0506 | 0.0438 | 0.184 | 0.919 |
| Rep white reg ’04 | 0.289 | 0.386 | 0.010 | 0.001 | 0.412 | 0.414 | 0.786 | 0.909 |
| Rep black reg ’04 | $4.80\times10^{-3}$ | $4.32\times10^{-3}$ | 0.504 | 0.335 | $5.11\times10^{-3}$ | $3.36\times10^{-3}$ | 0.148 | 0.55 |
| Prop white | 0.600 | 0.634 | 0.454 | 0.001 | 0.670 | 0.687 | 0.369 | 0.2 |
| Prop black | 0.153 | 0.098 | 0.015 | 0.006 | 0.086 | 0.0839 | 0.803 | 0.928 |
| Prop Hispanic | 0.0853 | 0.138 | 0.203 | 0.003 | 0.0696 | 0.0905 | 0.0541 | 0.055 |
| Foreign born | 0.0589 | 0.131 | 0.039 | 0.000 | 0.062 | 0.0706 | 0.228 | 0.061 |
| Adults in poverty | 9922 | 46507 | 0.051 | 0.000 | 17529 | 21291 | 0.370 | 0.496 |
| Turnout ’00 | 0.669 | 0.718 | 0.000 | 0.000 | 0.681 | 0.714 | 0.0811 | 0.015 |
| Tot Votes ’00 | 51436 | 219063 | 0.007 | 0.000 | 94199 | 120056 | 0.255 | 0.508 |
| Dem vote ’00 | 0.417 | 0.458 | 0.163 | 0.135 | 0.410 | 0.429 | 0.374 | 0.914 |
| Rep vote ’00 | 0.558 | 0.518 | 0.178 | 0.123 | 0.563 | 0.547 | 0.418 | 0.926 |
| Nader vote’00 | 0.015 | 0.0176 | 0.130 | 0.036 | 0.0187 | 0.0180 | 0.558 | 0.546 |
| Buchanan vote ’00 | 0.00525 | 0.00285 | 0.000 | 0.000 | 0.00406 | 0.00348 | 0.213 | 0.931 |
| Dem reg ’00 | 0.606 | 0.381 | 0.000 | 0.000 | 0.415 | 0.408 | 0.605 | 0.923 |
| Rep reg ’00 | 0.284 | 0.444 | 0.000 | 0.000 | 0.419 | 0.424 | 0.692 | 0.518 |
| Ref reg ’00 | $5.16\times10^{-4}$ | $5.90\times10^{-4}$ | 0.285 | 0.068 | $6.33\times10^{-4}$ | $6.01\times10^{-4}$ | 0.716 | 0.928 |
| Green reg ’00 | $2.24\times10^{-4}$ | $2.82\times10^{-4}$ | 0.306 | 0.00 | $2.70\times10^{-4}$ | $2.16\times10^{-4}$ | 0.260 | 0.53 |
| Dem white reg ’00 | 0.479 | 0.297 | 0.00 | 0.001 | 0.342 | 0.342 | 0.986 | 0.92 |
| Dem black reg ’00 | 0.104 | 0.0553 | 0.006 | 0.035 | 0.0516 | 0.0448 | 0.184 | 0.908 |
| Rep white reg ’00 | 0.266 | 0.405 | 0.000 | 0.000 | 0.398 | 0.404 | 0.580 | 0.529 |
| Rep black reg ’00 | $4.46\times10^{-3}$ | $4.12\times10^{-3}$ | 0.635 | 0.921 | $3.91\times10^{-3}$ | $3.24\times10^{-3}$ | 0.203 | 0.914 |

Note: The after matching t-test is a paired t-test while the pre-matching t-test is the usual two sample t-test. The “ks test” is a bootstrapped Kolmogorov-Smirnov test. This test provides correct coverage even when the distributions being compared are not entirely continuous. 1,000 bootstrap simulations were run.

Table 4: Balance Tests (continued)

| | Before Matching | | | | After Matching | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Mean Optical | Mean DRE | t-test | ks test | Mean Optical | Mean DRE | t-test | ks test |
| Turnout '96 | 0.647 | 0.704 | 0.000 | 0.000 | 0.683 | 0.706 | 0.056 | 0.207 |
| Tot Votes '96 | 45548 | 195538 | 0.007 | 0.000 | 80151 | 105953 | 0.234 | 0.568 |
| Dem vote '96 | 0.426 | 0.452 | 0.331 | 0.394 | 0.407 | 0.437 | 0.099 | 0.555 |
| Rep vote '96 | 0.445 | 0.444 | 0.983 | 0.654 | 0.471 | 0.441 | 0.237 | 0.537 |
| Perot vote '96 | 0.124 | 0.100 | 0.006 | 0.010 | 0.117 | 0.118 | 0.954 | 0.929 |
| Prop adults in pov | 0.136 | 0.094 | 0.000 | 0.000 | 0.104 | 0.0975 | 0.324 | 0.913 |
| Prop highly educated* | 0.153 | 0.218 | 0.000 | 0.000 | 0.192 | 0.178 | 0.218 | 0.537 |
| Prop lowly educated* | 0.0828 | 0.0603 | 0.022 | 0.000 | 0.0545 | 0.0569 | 0.486 | 0.208 |

Note: The after matching t-test is a paired t-test while the pre-matching t-test is the usual two sample t-test. The "ks test" is a bootstrapped Kolmogorov-Smirnov test. This test provides correct coverage even when the distributions being compared are not entirely continuous. 1,000 bootstrap simulations were run.

* The proportion highly educated is the proportion of adults with bachelor, graduate or professional degrees. And the proportion lowly educated is the proportion of adults with less than a grade 9 education.

corrected estimates are reported. Bias correction is performed using linear regression on the matched data.[4] The standard errors are based on the matching standard errors derived by Abadie and Imbens (2004). Please see the matching software for details about how the software implements these features. Note that the reported results make no correction for the varying size of the counties, but such a correction does not significantly alter the results.

### 1.1 Methodological Background on Matching Estimators

Underlying conditional probability is a notion of counterfactual inference.[5] A randomized experiment follows counterfactual reasoning. For example, we could either contact Jane to prompt her to vote as part of a turnout study or we could not contact her. But we cannot observe what would happen if we both contacted Jane and if we did not contact Jane—i.e., we cannot observe Jane's behavior both with and without the treatment. If we contact Jane, in order to determine what effect this treatment had on Jane's behavior (i.e, whether she voted or not), we still have to obtain some estimate of the counterfactual in which we did not contact Jane. We could, for example, seek to compare Jane's behavior with someone exactly like Jane whom we did not contact. In a randomized experiment we obtain a group of people (the larger the better) and we assign treatment to a randomly chosen subset (to contact) and we assign the remainder to the control group (not to be contacted). We then observe the difference in turnout rates between the two groups and we attribute any differences to our treatment.

In principle the process of random assignment results in the observed and unobserved

[4]In addition to all of the variables used for matching, the bias adjustment is based on the following variables: 2003 Hispanic, Democrat, Republican and Nader vote proportions in 2000 and 1996, Perot vote proportion in 1996, Buchanan vote proportion in 2000, turnout proportion in 2000 and 1996, Democrat, Republican, Green and Reform registration proportions in 2004 and 2000, proportion of all registrants who are white Democrats in 2004 and 2000, proportions of all registrants who are black Democrats in 2004 and 2000, proportion of all registrants who are white Republicans in 2004 and 2000, proportions of all registrants who are black Republicans in 2004 and 2000, the proportion of the population which is foreign born (2000 census), adult population in poverty (2000 census) and the proportion of the population who is "Cuban" (2000 census).

[5]Holland 1986; Rubin 1990, 1978, 1974; Splawa-Neyman 1990 [1923].

baselines variables of the two groups being balanced.[6] With observational data some other correction must be made.

Let $Y_{i1}$ denote the change in voter $i$'s vote intention from baseline when voter $i$ changes information state (i.e., is in the treatment regime), and let $Y_{i0}$ denote the change in vote choice when voter $i$ does not change information state (i.e., is in the control regime). The causal inference problem is a missing data problem because both $Y_{i1}$ and $Y_{i0}$ cannot be observed—a given voter cannot both change and not change information state at the same time. Let $T_i$ be a treatment indicator: 1 when $i$ is in the treatment regime and 0 otherwise. The observed outcome for observation $i$ is then $Y_i = T_i Y_{i1} + (1 - T_i) Y_{i0}$. The treatment effect for observation $i$ is defined by $\tau_i = Y_{i1} - Y_{i0}$.[7]

In principle, if assignment to treatment is randomized, the inference problem is straightforward because the two groups are by construction drawn from the same population. The observed and unobserved baseline variables of the two groups are balanced; treatment assignment is independent of all baseline variables. This occurs with arbitrarily high probability as the sample size grows. With the independence assumption, the missing data problem is simple to resolve because treatment assignment is independent of $Y_0$ and $Y_1$—following Dawid's (1979) notation, $\{Y_{i0}, Y_{i1} \perp T_i\}$. Hence, for $j = 0, 1$

$$E(Y_{ij}|T_i = 1) = E(Y_{ij}|T_i = 0) = E(Y_i|T_i = j)$$

Therefore the average treatment effect (ATE) can be estimated by:

$$\begin{aligned} \tau &= E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 0) \\ &= E(Y_i|T_i = 1) - E(Y_i|T_i = 0) \end{aligned} \tag{1}$$

It is possible to estimate Equation 1 because observations in the treatment and control

[6]This occurs with arbitrarily high probability as the sample size grows. Balance implies that treatment assignment is independent of baseline covariates.

[7]For comparability with the literature, the notation here is the same as in Dehejia and Wahba (1999).

groups are exchangeable. In the simplest experimental setup, individuals in both groups are equally likely to receive the treatment, and hence assignment to treatment will not be associated with anything which may also affect one's propensity to vote in a particular fashion. Even in an experimental setup, much can go wrong which requires statistical correction (e.g., Barnard, Frangakis, Hill, and Rubin 2003). In an observational setting, unless something special is done, the treatment and non-treatment groups are almost never balanced.

With observational data, the treatment and control groups are not drawn from the same population. Thus, a common quantity of interest, and the one I use, is the average treatment effect for the treated (ATT):

$$\tau|(T=1) \quad = \quad E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1). \tag{2}$$

Unfortunately, Equation 2 cannot be directly estimated because $Y_{i0}$ is not observed for the treated. Progress can be made if we assume that the selection process is the result of only observable covariates denoted by $X$. Following Rosenbaum and Rubin (1983), one can assume that once one conditions on $X$, treatment assignment is unconfounded ($\{Y_0, Y_1 \perp T\} \,|X$) and that there is overlap: $\delta < Pr(T = 1|X) < 1-\delta$, for some $\delta > 0$. Unconfoundedness and overlap together define strong ignorability. These conditions are required to identify ATE. Heckman, Ichimura, Smith, and Todd (1998) demonstrate that to identify ATT, the unconfoundedness assumption can be weakened to mean independence: $E\left[Y(w)|T, X\right] = E\left[Y(w)|X\right]$, for $w = 0, 1$.[8] The overlap assumption for ATT only requires that the support of $X$ for the treated is a subset of the support of $X$ for the control observations.

Then, following Rubin (1974, 1977) we obtain

$$E(Y_{ij}|X_i, T_i = 1) = E(Y_{ij}|X_i, T_i = 0) = E(Y_{ij}|X_i, T_i = j). \tag{3}$$

Therefore, conditioning on the observed covariates, $X_i$, the treatment and control groups

[8] Also see Abadie and Imbens (2004).

have been balanced. The average treatment effect for the treated can then be estimated by

$$\tau|(T=1) \quad = \quad E\left\{E(Y_i|X_i, T_i = 1) - E(Y_i|X_i, T_i = 0) \mid T_i = 1\right\}, \tag{4}$$

where the outer expectation is taken over the distribution of $X_i|(T_i = 1)$ which is the distribution of baseline variables in the treated group.

The most straightforward and nonparametric way to condition on $X$ is to exactly match on the covariates. This is an old approach going back to at least Fechner (1966 [1860]), the father of psychophysics. This approach of course fails in finite samples if the dimensionality of $X$ is large. So nearest neighbor matching is often used, as it is in this case. The weights for each variable are given by the inverse variances. For additional details about the implementation, please see the "Matching" software package: `HTTP://jsekhon.fas.harvard.edu/matching`.

## References


Abadie, Alberto and Guido Imbens. 2004. "Large Sample Properties of Matching Estimators for Average Treatment Effects." Working Paper.

Barnard, John, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. Rubin. 2003. "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City." *Journal of the American Statistical Association* 98 (462): 299–323.

Dawid, A. Phillip. 1979. "Conditional Independence in Statistical Theory." *Journal of the Royal Statistical Society, Series B* 41 (1): 1–31.

Dehejia, Rajeev and Sadek Wahba. 1999. "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448): 1053–1062.

Fechner, Gustav Theodor. 1966 [1860]. *Elements of psychophysics, Vol 1.*. New York: Rinehart & Winston. Translated by Helmut E. Adler and edited by D.H. Howes and E.G. Boring.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–1098.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–960.

McCullagh, Peter and John A. Nelder. 1989. *Generalized Linear Models*. New York: Chapman & Hall 2d edition.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag 2nd edition.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.

Rubin, Donald B. 1977. "Assignment to a Treatment Group on the Basis of a Covariate." *Journal of Educational Statistics* 2: 1–26.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics* 6 (1): 34–58.

Rubin, Donald B. 1990. "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5 (4): 472–480.

Splawa-Neyman, Jerzy. 1990 [1923]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science* 5 (4): 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed.