Essays on Causal Inference and Political Representation

Thesis by

Delia Bailey

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2007
(Defended May 10, 2007)

Acknowledgements

The members of my dissertation committee, Jonathan Katz, Mike Alvarez, Bob Sherman, and Gary Lorden, have greatly shaped the course of my research and I thank them for the time and energy that they have dedicated to my academic and professional development. The advice and insight of Jonathan Katz and Mike Alvarez has been invaluable. Together, they taught me most of what I know about political science, and everything I know about academia. They have both mentored me on aspects of life both personal and professional, and are more than worthy of the moniker "advisor." I thank Jonathan for his (at times brutal) honesty and insistence on excellence; I thank Mike for his persistence, sense of humor, and unending patience. Most of all, I thank them for their faith in me and for being my friends. Bob Sherman started my fascination with identification and causal inference, and for that I am eternally grateful. Bob's example has taught me the value of clarity in exposition. He also deserves many thanks for impressing upon me the importance of stating all assumptions, both implicit and explicit, and for always asking the tough questions. Gary Lorden helped me to clarify the presentation of my analyses, and offered insight on my research agenda from a prospective outside social science.

This research has benefited greatly from discussants and participants at the Annual Meetings of the Midwest Political Science Association and the Summer Meetings of the Society for Political Methodology. I am grateful to Thad Hall, Gary King, and Jonathan Nagler for extensive discussions on this research. The financial support of the Division of Humanities and Social Sciences at Caltech, the Caltech/MIT Voting Technology Project, the Institute for Quantitative Social Sciences at Harvard University, and the Los Angeles chapter of ARCS is gratefully acknowledged. I thank

Steve Ansolabhere and Charles Stewart, Gary King, and Gary Jacobson for sharing their data with me. Much of this research would not have happened without their support.

I thank my classmates, other HSS students, and the many political science graduate students elsewhere who have given me perspective on my research and provided a source of encouragement. Betsy Sinclair has been instrumental to this project: conversations with her have influenced the direction of my research, and they have also kept me sane. I thank Betsy for challenging me with her endless questions and for supporting all my many neuroses. She will always be my favorite surfing political methodologist. I thank Sarah Sled for her emotional support and for her indulgence of my various and sundry misanthropic tendancies and for sharing the same indulgent priorities. Joel Grus deserves a special thank you for carrying more than his weight on our first-year problem sets, and for making me laugh — when all I really wanted to do was to cry. Finally, I thank Laurel Auchampaugh for her support and assistance. I literally could not have done this without her.

I thank my mother for always believing I was the smartest kid in the room and supporting me accordingly. I thank her for her unconditional love and for teaching me independence, even when it meant that I would one day grow up and leave. I thank my father for teaching me to hold my own among strong-willed men, for financial and emotional support of my educational pursuits and for giving my perspective a much-needed challenge from time-to-time. I thank my husband, John, for his love and support. He kept me fed and in clean clothes for the past five years, and he truly is my best friend. This work is dedicated to him.

Abstract

I present three political science examples of observational studies where modern causal inferences techniques are used to improve upon previous estimates. Difference-in-differences, fixed effects estimators, and a propensity score matching model are used to demonstrate model dependence in previous studies of the impact of voting technology on residual vote rates. Measuring the incumbency advantage serves as an example of when the assumptions of matching methods fail, and given the data, a linear model is most appropriate. The impact of voter identification on turnout is properly modeled in two ways: first, a multilevel logistic regression is used to appropriately model how state and individual covariates, and their interactions, affect the decision to participate; second, a Bayesian shrinkage estimator is used to properly model the ordinal nature of the voter identification treatment variable. In each essay, the benefit of using causal inference techniques to more efficiently estimate quantities of interest in questions of political representation and policy outcomes is demonstrated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis contains three essays that were written independently, but that contain overlapping themes and ideas. As indicated by the title, all three essays concern questions of political representation and the causal inference techniques that can be used to measure these quantities of interest. In the first essay, difference-in-differences, fixed effects estimators, and a propensity score matching model are used to demonstrate model dependence in previous studies of the impact of voting technology on residual vote rates. In the second essay, measuring the incumbency advantage serves as an example of when the assumptions of matching methods fail, and given the data, a linear model is most appropriate. In the final chapter, the impact of voter identification on turnout is properly modeled in two ways: first, a multilevel logistic regression is used to appropriately model how state and individual covariates, and their interactions, affect the decision to participate; second, a Bayesian shrinkage estimator is used to properly model the ordinal nature of the voter identification treatment variable. In each essay, the benefit of using causal inference techniques to more efficiently estimate quantities of interest in questions of political representation and policy outcomes is demonstrated.

The first essay asks whether the method used to cast and count ballots affects the quality of preference recording in the voting booth. Quality is measured with the residual vote rate, which is calculated using county level data for the presidential elections in 1988–2004. Difference-in-differences, fixed effects models and propensity score matching methods are used to isolate attributable effects to technology. Punch cards consistent perform the worst, but the ranking of other technologies is model dependent. The magnitude of the effects varies across estimation methods as well.

The second essay notes that the problem of measuring the incumbency advantage is really a missing data problem. Given this, the essay asks whether matching methods can be utilized to avoid linear model dependency. The data used are election returns, incumbency status, and party identification from the 1898–2002 U.S. congressional elections. The model of Gelman-King (1990) is extended to include more information about previous vote shares in each district, and then a propensity score matching model is used to try and isolate the average gain in vote share to incumbents. The results show that the classical linear model produces the most reliable estimates of the incumbency advantage. In addition, the essay demonstrates that if the propensity score used for matching is not a good estimate of the true propensity score, then matching results are essentially based on random samples of the data and are not reliable.

The third essay concerns the impact of voter identification on turnout, particularly in subpopulations such as the elderly, the lower educated, and racial minorities. Voter identification requirements are measured at the state level and are ordinal. The data utilized are individual responses to the Current Population Survey Voter Supplement in 2000 and 2004. Two models are estimated. First, a multilevel logistic regression with interactions attempts to uncover the impact of voter identification on subpopulations. Second, a Bayesian shrinkage estimator is used to properly model the ordinal nature of the voter identification variable and to suggest that conventional constrained models are insufficient. The results show that conditional on registration, voter identification requirements have little to no effect on voter turnout, even within important subpopulations. In addition, modeling choice of the ordinal variable matters, as does proper modeling of the state and nationwide trends in turnout.

All three essays relate to consequences for political representation. The first and second essays concern consequences of representation at the ballot box—as the quality of preference recording in casting ballots can affect the choice of elector, and an incumbency advantage may insure an elector with preferences that are not representative of his/her consituents is chosen. Both the first and third essay address the principle of "one person, one vote"—as subpopulations may be disenfranchised by unequal ability to operate technologies, non-uniform enforcement of voter identification laws, or a heftier burden of the tax of acquiring identification.

Unlike much of the data in other sciences, experiments are rare in political science and field experiments are only seldom implementable. As observational data is most often available, and in addition political science data is often messy and sparse, inference can be tricky. There continues to be a gap in the literature between the theoretical properties of causal inference techniques and the practical applications of them. It is shown clearly that regression adjustment and matching methods together reduce bias, and reducing heterogeneity leads to more efficient estimates. But in practice, unobservable covariates often exist and finding the "true" propensity score can be a difficult task. In addition, often the reduction of heterogeneity leads to a very small sample and reduced direction of inference. In addition, political scientists are faced with a tradeoff between answering important policy questions and choosing models that fit the data best. Given all this, it is still arguable that causal inference techniques should be used whenever possible. It is extremely important, however, that researchers are aware of all the assumptions—explicit and implicit—made by the theoretical models they employ and that they thoroughly evaluate the reasonable nature of the assumptions for their practical problem.

Chapter 2

Model dependency and measuring the effect of voting technology on residual votes

After the 2000 election, political scientists became increasingly interested in measuring the extent to which different voting technologies impact residual vote rates in the United States. The question in which I am interested is how robust these measurements are to the choice of specification and estimation technique. Most of the previous research has conducted multivariate regression analysis on cross-sectional data, with the exception of a few regionally concentrated panel studies.[1] Brady et al. (2001) evaluate performance of technologies in U.S. counties in the 2000 presidential election. Using data from the 1996 election, Knack and Kropf (2003) find a positive relationship between voided ballots and the percentage of African Americans in the county, specifically in counties with voting equipment that allow overvotes. Kimball et al. (2004) utilize a generalized least squares approach to estimate the number of unrecorded votes in the 2000 election. Ansolabehere and Stewart (2005) advance the methodology considerably by estimating a fixed effects model on a pooled time-series data set, consisting of data from the 1988–2000 presidential, gubernatorial, and senatorial election returns in U.S. counties.

This analysis begins with a replication of Ansolabehere and Stewart (2005) with an additional panel of data. To investigate the degree of model dependence in their results, the analysis is replicated using several different causal estimators. Specifically, using data from the 1988–2004 presidential elections, the effect of voting technology on residual vote rates is analyzed via several econometric estimators. A difference-in-differences estimator is used estimate the effect on an average county of switching from punch cards to optically scanned ballots for each election cycle. Fixed effects regression models provide a generalization of the difference-in-differences approach, estimating the effect of changing technology on residual vote rates within counties over time, for each type of voting equipment currently in use in the U.S. Both difference-in-differences and fixed effects models attempt to isolate the effect of a technology change on residual votes by controlling for confounding factors that are unobservable and are fixed, or at least slowly changing over time. In contrast, the propensity score matching method applied here generates a balanced data set by conditioning on observable confounders; several estimators are then applied to this data set—a simple differences estimator and a parametric regression.

The pattern of the results is not robust to the different methods. Applying the parametric estimators to the raw data indicates that paper ballots and lever machines produce the lowest rates of residual votes, followed by optically scanned ballots, direct recording electronic machines, and punch cards. After producing matched samples and repeating the analysis, electronic machines, optical scanners and paper ballots are proven the superior technologies (followed by lever machines). Punch cards are universally represented as the poorest choice in terms of residual votes.

The remainder of the chapter is organized as follows: Section 2.1 defines the different technologies and discusses various ways residual votes can occur. Section 2.2 addresses the problem of estimating treatment effects. Sections 2.3 and 2.4 describe the data and methods used. I report results of the analyses in Section 2.5. Section 2.6 further explores the results by looking more in depth at 2004 election data and provides directions for future research.

2.1 Voting technology and residual votes

The residual vote rate is defined as the fraction of total ballots cast for which no vote for president was counted (Caltech/MIT Voting Technology Project, 2001). Residual votes can occur when a vote is cast for more than one candidate in a single race, when a single vote is marked in a way that is uncountable, or when the ballot is left blank. In the voting literature, other terms used to refer the difference between ballots cast and votes counted are “over votes,” “under votes,” “spoiled ballots,” “drop off,” “roll off,” “voter fatigue,” or the “error rate” (Caltech/MIT Voting Technology Project, 2001). In this paper, the term residual vote is used because it encompasses each of these cases—error on the part of the voter, mechanical or technological failure of the voting equipment, and abstention.

Abstention is the most obvious way that the number of ballots cast might differ from the number of votes counted for president, but there are other factors that might affect the variance in residual vote rates across counties and election years.[2] The primary focus of this paper is on how to measure the causal effect of different voting technologies on the rate of residual votes in an average county. There are five general types of technologies in use in the United States represented in the data: hand-counted paper ballots, mechanical lever machines, punch cards, optically-scanned paper ballots, and direct recording electronic machines (DREs).[3]

The technologies used to record a voter’s preferences may affect the residual vote because of mechanical (or other) failures. Paper ballots generally only “fail” when a precinct runs out of ballots. Of course, human counting errors can also cause paper ballots to “fail.” All other machine types can break down, which presents a serious problem if the break down is not caught. Optically-scanned ballots can be treated as paper ballots, if the officials are alerted to the malfunction in the scanner. The counters, whether external and mechanical, or internal and electronic, on lever and DRE machines may malfunction without being caught—leaving no way to recover lost ballots. Punch cards are now famous for their failures—the “pregnant and hanging chads” of Palm Beach County, FL, in the 2000 election are examples of failures of the punch card system to perform mechanically (Ansolabehere and Stewart, 2005).

There are, however, many county-specific factors that affect the ability of a voter to cast a vote and have it counted that are independent of the voting technology. If not controlled for properly, these may confound the estimated effect different technologies have on residual votes.[4] Voter-specific characteristics, such as literacy and English-language proficiency, may affect a voter's ability to complete a ballot, as might physical impairments such as arthritis or poor eyesight. The county's size in terms of population and wealth impact finances available for the adminstration of elections, and in turn, affect the level of quality of trained poll workers available to assist voters on election day, and the number of qualified workers on hand to count ballots at the end of the day. The presence of a particularly salient issue or prominent race on the ballot may bring voters to the polls that might not usually vote, or a county may have a higher than average number of young people participating in their first election—both could affect the rate of residual votes in a particular county or election year. Finally, the introduction of a new technology may affect residual votes, although it is not immediately clear in what direction. Voters may be confused with the new machinery and therefore make more mistakes, or election officials may anticipate these problems and increase educational efforts both before and during the election, countering the effect and possibly lowering the residual vote rate.

2.2 Estimating treatment effects

The literature on the effect of voting technology on residual votes is filled with interesting counterfactual questions. For example, Wand et al. (2001) find that in the 2000 presidential election, more than 2,000 Democratic voters in Palm Beach County voted for Pat Buchanan by mistake because of the use of the "butterfly ballot." Ansolabehere and Stewart (2005) claim that if all jurisdictions in the United States that used punch cards in the 2000 presidential elections had instead used optical scanners, approximately 500,000 more votes would have been counted in presidential election returns nationwide. Fundamentally, these research questions are concerned with issues of cause and effect—in an average county, what percentage change in residual votes can be expected if a change is made from voting technology X to voting technology Y? But evaluating the impact of a policy change on individual (or county-level) behavior is extremely difficult, as evidenced by the following simple example (adopted from Duflo, 2002).

Suppose we are in a simpler situation where there are only two voting technologies available, punch cards and optical scan machines. And suppose that it is not possible to have a mixed technology county. Let Y_i^{OS} represent the residual vote rate in a given county i if the county uses optical scanners, and Y_i^{P} represent the residual vote rate in the same county i if the county uses punch cards. The quantity of interest is the difference $Y_i^{\mathrm{OS}} - Y_i^{\mathrm{P}}$, the effect of using optical scanners relative to using punch cards in county i. But the inherent problem is that we will never have a county i with all ballots counted by optical scan machines and with all ballots cast on punch cards simultaneously. We can only hope to infer the expected treatment effect, $E[Y_i^{\mathrm{OS}} - Y_i^{\mathrm{P}}]$.

Now, imagine that we have collected data on a large number of counties in the United States. Some of these counties use optical scanners, while the others use punch cards. We can calculate the average residual vote in counties with optical scanners and the average residual vote in counties using punch cards, and then take the difference between the two averages. This can be represented as:

Difference $= E[Y_i^{\mathrm{OS}}|\text{county uses optical scan}] - E[Y_i^{\mathrm{P}}|\text{county uses punch cards}] = E[Y_i^{\mathrm{OS}}|\mathrm{OS}] - E[Y_i^{\mathrm{P}}|\mathrm{P}]$.

But this is potentially a biased estimate of the expected treatment effect. If Y_i^{P} differs systematically between counties in group OS and counties in group P, then Y_i^{P} is estimated incorrectly for the treated group (OS), because we only observe Y_i^{P} for the control group (P).

Moving from this example to the real world, we have not two technologies, but five. And it is indeed possible to have mixed technology counties. How are we to learn the average effect of using another technology, such as paper ballots, relative to punch cards in U.S. counties? Ideally, we would be in a situation where we could conduct an experiment, controlling the assignment of treatment to subjects and thereby ensuring that subjects who receive different treatments are comparable (Rosenbaum, 2002). Because laboratory experiments are often not feasible in political science, Green and Gerber (2002) argue field experiments should be employed, when possible, to answer questions of causality. However, in our particular example, even field experiments are not feasible. Even if we could convince a sampling of counties to allow us to randomly assign which voting technology they will use in the next presidential election, voting equipment is extremely expensive and the sheer cost of implementation would be enough to prohibit an experiment. Thus, we find ourselves in the world of observational studies (Cochran, 1965). As in any observational study, modeling assumptions must be made in order to identify causal effects. As policy decisions are made based on the outcome of such studies, it is often useful to examine the assumptions made and compare the outcome under each set of assumptions.

2.3 Data

Because the decision of which voting technology to use in elections is generally made at the county level, the unit of analysis is a (county, year) pair. To calculate the residual vote rate in U.S. elections, I obtained data that recorded the total number of ballots and the number of presidential votes cast in the 1988–2004 presidential elections in each county in the sample.[5] Also noted is whether another prominent race is on the ballot in that observation, such as governor or U.S. senator. Data from 1988 to 1996 were obtained from Election Data Services (EDS), for 2000 and 2004 from local election officials, and additional 2004 data from the *Atlas of U.S. Presidential Elections* (Leip 2004).

Data on the voting equipment used in each of the counties was assembled for 1988 to 1996 from EDS, for 2000 from local election officials, and 2004 data from both EDS and the Verified Voting Organization. The focus of this paper is on the five general types of technologies, without making distinctions within the types. This is relaxed in Section 2.6. There are some counties without a uniform voting technology. These are referred to in the paper as “mixed” technology counties. Such counties occur most often in the New England states, where the municipal governments administer elections. Over the course of this sample, counties increased their usage of optical scanners and DREs, decreasing their use of the older machine types. Figure 2.1 plots the distribution of voting technology types across counties and across the percent of the voting population covered by each technology type in 1988–2004.

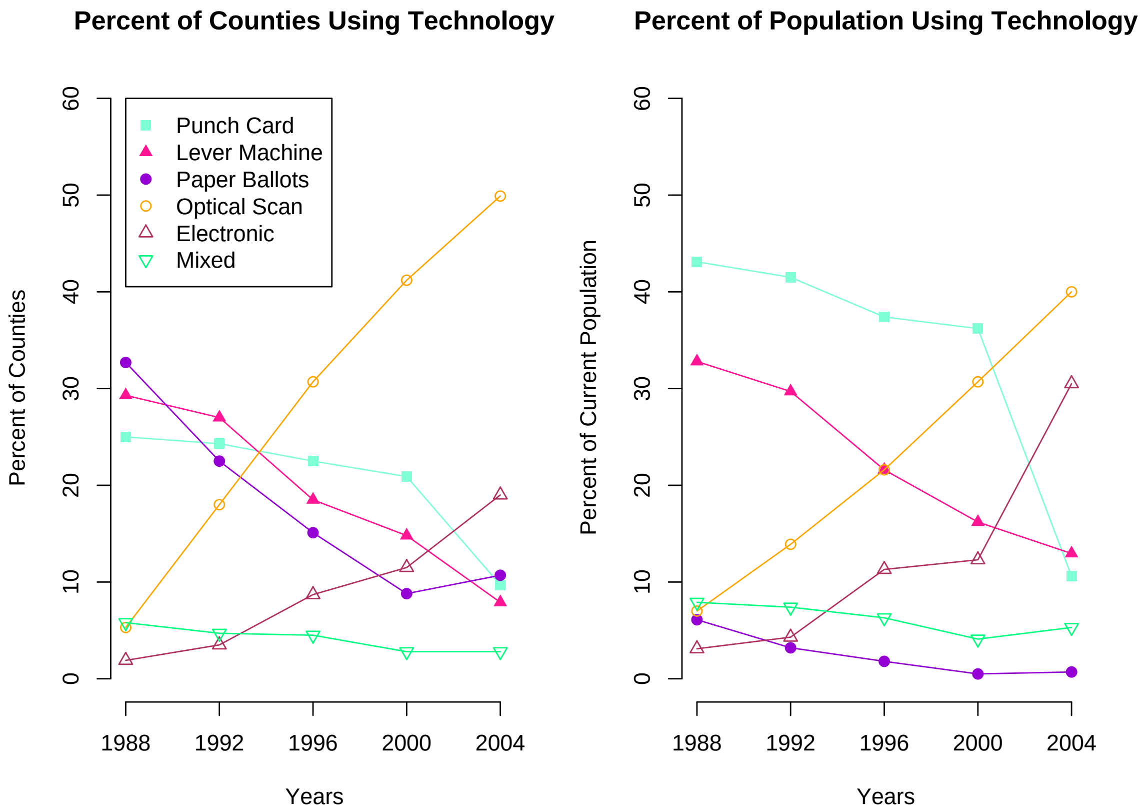

Figure 2.1: Usage of Voting Technologies in the 1988–2004 Presidential Elections

A map depicting the distribution of machine types across the United States in the 2000 elections can be found in Figure 2.2, while Figure 2.3 presents the distribution of machine types in 2004. Paper ballots are most used in the Midwestern states; New York and Louisiana are the main states still using lever machines. The Southeast and Western United States show a preference for the electronic machines, with punch cards interspersed throughout the regions. Optical scanners are the most widely used technology, covering 40% of the population. Some states, such as Arizona, Georgia, Maryland and New York use one technology only, whereas others, such as Arkansas, Colorado, North Carolina and West Virginia have no single dominant technology. Comparing the two maps reveals the large scale changes in Georgia and Nevada after the 2000 elections.

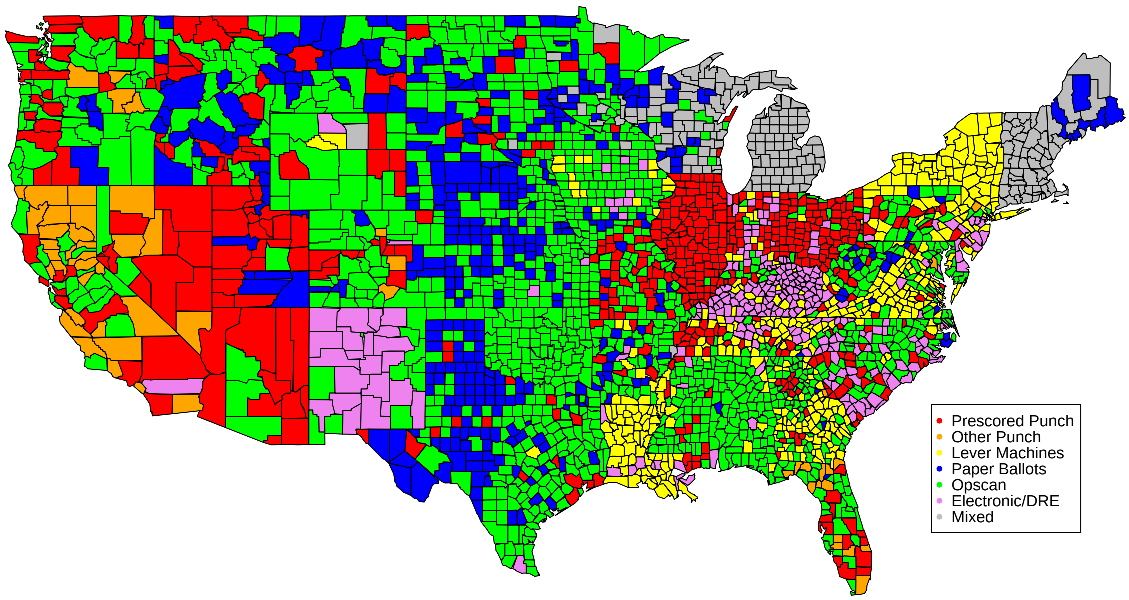

Figure 2.2: Voting Technology Usage in the 2000 Presidential Election

In addition to election returns and voting technology data, I obtained estimates of county population by race and age and median income for each year in the sample from the U.S. Census Bureau. Income was inflated to represent 2000 dollars using a multiplier from the Bureau of Economic Analysis.

Data was acquired for approximately one-half of the 3,155 counties in the United States over five presidential elections, 1988, 1992, 1996, 2000, and 2004. Over the course of the sample, several states were excluded in their entirety because they do not require counties to report turnout separately from the number of votes cast for

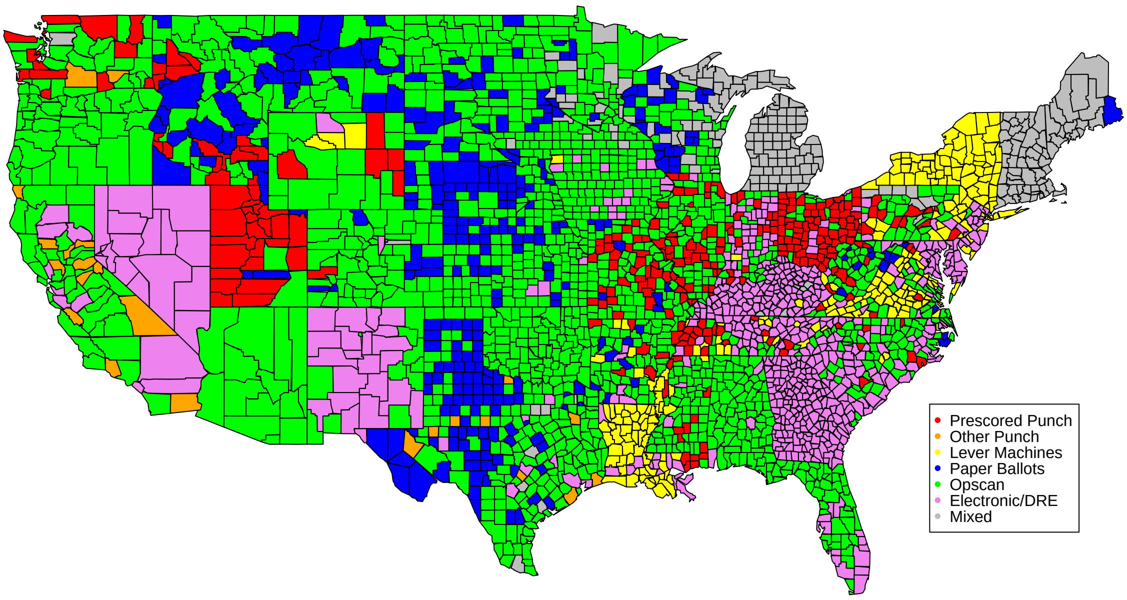

Figure 2.3: Voting Technology Usage in the 2004 Presidential Election

president. This is of great concern if states that do not report total ballots cast differ systematically in their relationship between voting technologies and residual vote rates. Lacking a theoretical model of how the relative performance of voting equipment to residual votes would differ in these states that do not require the reporting of turnout, this concern cannot be directly addressed.

States with mixed-technology counties are excluded from the sample. Massachusetts, New Hampshire, and Vermont administer elections at the town level for many of the years in our sample and Alaska administers elections at the State House district level. These were excluded to maintain a constant unit of analysis. Finally, some observations were selectively excluded from the data due to strong suspicions of typographical errors made when election returns were recorded.[6] The total cases included in the sample can be found in Table A.2 in the Appendix.

Residual vote rates in U.S. counties, averaged over the entire time frame, range from 1.9% (lever machines) to 2.9% (punch cards). When viewed as a percent of all ballots cast nationwide, the lowest average residual vote is 1.5% (optical scanners) and the highest is once again punch cards, with 2.5%. Residual vote rates have decreased substantially over time for both optical scanners and DREs. Punch cards perform worse than any other technology in 2000 and 2004, while DREs present higher residual vote rates than other technology types from 1988–1996, regardless of whether the unit of analysis is U.S. counties or U.S. voters.

2.4 Methods

When a county changes its voting technology, a natural experiment occurs. One way to exploit this natural experiment is to simply estimate the difference in mean residual vote rates before and after the change in technology. The problem with this approach is that is impossible to distinguish changes in residual votes due to the technological switch and changes due to other factors. These changes may be due to demographic changes, a particularly competitive election, or any number of observable or unobservable factors.

2.4.1 Difference-in-differences

The idea behind difference-in-differences is that we can improve on the method of simple differences in means by subtracting out the differences in means of a control group. In the context of this data, there is one (treatment, control) pair of technologies for which there is sufficient N over all time periods that it makes sense to produce difference-in-differences estimates: optically scanned ballots and punch cards. The average change in residual vote rates when a county changes from punch cards to optical scan machines is estimated for each of four time periods: 1988–1992, 1992–1996, 1996–2000, and 2000–2004. Operationally, this is done by running least squares regression on the following equation:

$$\log(\mathcal{F}(Y_{it})) = \alpha + \beta \cdot 1\{p = 1\} + \gamma \cdot 1\{i \in \mathrm{OS}\} + \eta \cdot 1\{p = 1\} \cdot 1\{i \in \mathrm{OS}\} + \varepsilon_{it} \quad (2.1)$$

where Y_{it} represents the residual vote rate in county i at time t; $\mathcal{F}(\cdot)$ is a function used to transform the dependent variable, to be discussed in detail below; α is a constant; $1\{p = 1\}$ is a dummy variable equal to unity if the observation is in the latter half of the period (i.e., for the 1988–1992 period, p=1 in 1992); and $1\{i \in \text{OS}\}$ is a dummy variable with a value equal to unity indicating that the observation belongs to the treatment group (counties that switch to optical scanners). The OLS estimate of η is numerically identical to the difference-in-differences estimate, DD.

The distribution of residual vote rates is skewed to the right and a transformation is necessary to maintain the normality assumption in the least squares specification. A commonly used transformation for variables with a skewed distribution is the log transformation. However, it is also the case that the distribution of residual vote rates, Y_{it}, has a mass at zero, which is problematic for the log transformation. To avoid dropping all of the zero residual vote observations, the following transformation was utilized: $\mathcal{F} = 0.005 + 0.99 * Y_{it}$ (Fox, 1997, 59–81). The transformation function $\mathcal{F}$ maps residual votes from the $[0, 1]$ interval into the $[0.005, 0.995]$ interval. To sensibly interpret the estimated coefficients, β_i, we must "back-transform" the estimates from the log transformation, using the formula: $\tau(\hat{\beta}_i) = 100[\exp(\hat{\beta}_i - \frac{\hat{Var}(\hat{\beta}_i)}{2}) - 1]$ (Halvorsen and Palmquist, 1980, and Kennedy, 1981). Standard errors are calclulated using an approximate variance formula, $\hat{Var}(\tau(\hat{\beta}_i)) = 100^2 \exp(2\hat{\beta}_i)[\exp(-\hat{Var}(\hat{\beta}_i)) - \exp(-2\hat{Var}(\hat{\beta}_i))]$ (van Garderen and Shat, 2002). Additionally, all observations are weighted by turnout, so the interpretation of the dependent variable is relative to the total number of votes cast.

2.4.2 Fixed effects models

Fixed effects regression generalizes the difference-in-differences approach to include more than two time periods and more than one treatment group.[7] The average change in residual vote rates that occurs when a county changes voting technology is estimated with several parametric specifications. All of the specifications are variations on the following equation:

$$\log(\mathcal{F}(Y_{it})) = \alpha_i + \gamma_t + T_{it}^j \lambda_j + X_{it}\beta + \varepsilon_{it} \tag{2.2}$$

where Y_{it} is the residual vote rate of county i in year t; $\mathcal{F}(\cdot)$ is a transformation of the dependent variable, discussed above; α_i are state *or* county fixed effects, depending on the particular specification; γ_t are year fixed effects; T_{it}^j are binary variables equal to unity if county i uses voting technology j in year t; and X_{it} is a vector of observation specific attributes acting as controls. Variables that appear in X_{it} are: log of turnout, an indicator variable denoting whether there was a change in technology since the last presidential election, an indicator variable denoting whether there is a concurrent gubernatorial or senatorial election on the ballot, racial breakdown of the population in percentage terms, percent of the population aged 18–24, percent of the population 65 and older, median income and median income, squared.[8] In all estimation procedures, punch cards are treated as the "control" group. Again, all observations are weighted by turnout.

In the first model, α_i are state fixed effects. Consequently, a larger number of county-specific control variables are included in X_{it} for this model. In addition to state and year fixed effects, the first model includes all possible variables in the vector X_{it}. In the next model, α_i are county fixed effects. The control variables are an indicator for the presence of another prominent race on the ballot, an indicator variable denoting whether the county experienced a shift in technology, and the log of turnout. All models are estimated by fixed effects regression on an unbalanced panel, in which the unit of analysis is a (county, year) pair.[9]

2.4.3 Propensity score matching model

Rather than controlling for unobservable variables that are fixed across groups or time, the propensity score matching methods developed in Rosenbaum and Rubin (1983) attempt to overcome the problem of discerning treatment effects in observational studies by explicitly conditioning on observables. The problem in identifying treatment effects is essentially a missing data problem—the treatment group is observed and the outcome conditional on treatment assignment is observed, but the counterfactual *is not* observed.

The problem with comparing average effects in observational studies is that typically treated units differ systematically from control units. Rosenbaum and Rubin (1983) define treatment assignment to be *strongly ignorable* if we can find a vector of covariates, X, such that

$$Y^1, Y^0 \perp T \mid X, \quad 0 < pr(T = 1 \mid X) < 1.$$

That is, the outcomes under the treatment and the control Y^1 and Y^0 are independent of the treatment assignment, T, conditional on observable covariates, X and that there is overlap in the treatment probability. Intuitively, this says that conditional on observables, the treatment assignment is random and that there is some non-zero probability of each subject receiving the treatment or control.

Typically X is multidimensional and often contains continuous variables, making exact matching highly impractical. However, a result due to Rosenbaum and Rubin (1983) demonstrates that it is enough to condition on the propensity score, $p(X) = pr(T = 1|X)$. The true propensity score is not known, but is estimated via logistic regression of T_{it} on a constant term and X_{it}, without regard to the dependent variable Y_{it}. In the context of this particular data, T^0_{it} is punch card machines, whereas the treatment group is one of the other equipment types, considered one at a time.[10] The vector X_{it} differs depending on the treatment in question, but generally consists of the same covariates used as controls in the fixed effects regression.

After estimating the propensity score, an algorithm for matching is needed. A simple way to generate matched pairs is the so-called "nearest available" or "nearest neighbor" matching,[11] in which each observation in the treatment group is paired with the observation in the control group with the propensity score that is closest to it, typically in terms of absolute value (Rosenbaum and Rubin, 1985; see Deheija and Wahba, 2002 for a detailed application).

After matching, without consideration to the dependent variable, the propensity score model is adjusted and the matching algorithm is repeated as many times as necessary to acheive balance. In this particular context, the propensity score model generally contained the covariates from the fixed effects regression analysis. Higher-order terms and interactions are included when they increase balance. Additionally, binary variables or factor variables, such as year, are matched on exactly if that increases balance. For this analysis, balance is evaluated by comparing differences in means and qq-plots across covariates, for the treatment and control groups, before and after matching.[12]

Once a balanced sample is achieved, the average treatment effect is estimated by taking the difference of the average of the transformed residual vote rate, weighting control units by the number of times they appear in the matched sample. Observations are also weighted by turnout, to facilitate comparision with the other methods. Standard errors of the estimate are calculated by summing the weighted matched sample variances for the treated and control groups, and then taking the square root.[13] Additionally, the fixed effects regressions are re-estimated on the balanced sample.

2.5 Empirical results

Figures 2.4– 2.7 compare the estimated treatment effects presented in the previous section, by treatment type.[14] For paper ballots, most of the estimates are fairly similar, with the exception of county and year fixed effects, which produce the largest negative estimate. When this same estimator is applied to a matched data set, the point estimate decreases, although still within the same range. Lever machines also produced fairly similar results across the estimators, particularly given the large uncertainty around the final matched data estimate. A varied picture emerges however, when looking at the estimates for counties switching to optical scanners or electronic machines. One reason for this variability, is potentially the heterogeneity of machine types within this category, not because of model dependence. To address this concern, data from the 2004 election is employed in Section 2.6. First, each of the estimates are addressed in turn.

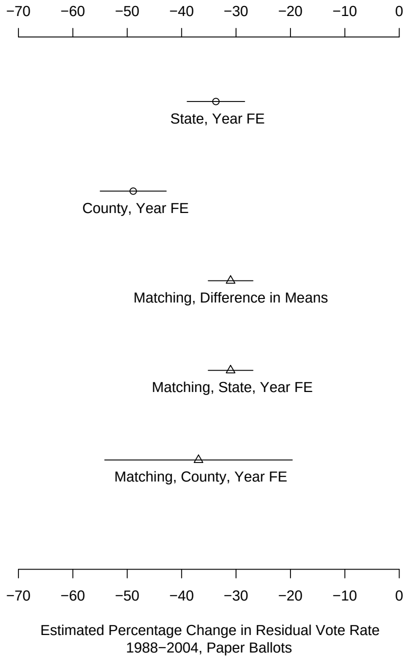

Figure 2.4: Estimated Treatment Effect for an Average County Switching from Punch Cards to Paper Ballots

2.5.1 Difference-in-differences estimates

The estimates for the 1988–1992 period are not significant. This is very likely due to the small number of observations in that time period. In each of the remaining three periods, counties switching to optical scan machines from punch cards experienced an average drop in residual vote rates, relative to those counties using punch cards in both elections. This decline ranges from 24% for those switching from 2000 to 2004, to 43% for counties making the switch between 1996 and 2000. Counties in the 1992–1996 time period experienced a decline of 28% in their residual vote rates.[15]

2.5.2 Fixed effects estimates

The first model includes state and year fixed effects as well as county-specific control variables. Paper ballots are the "best" technology in terms of reduction in residual votes—they produce a 34% reduction in the rate of residual voting, relative to punch

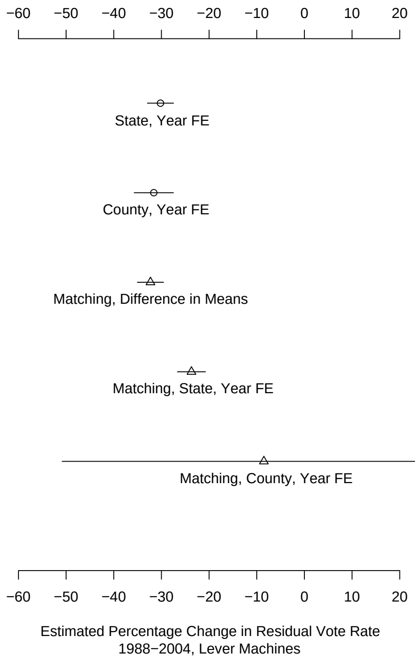

Figure 2.5: Estimated Treatment Effect for an Average County Switching from Punch Cards to Lever Machines

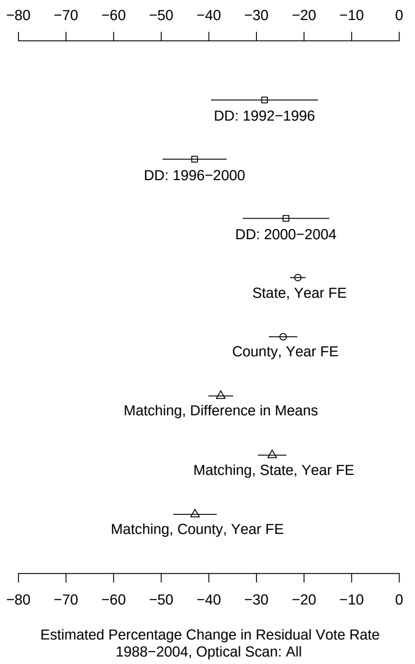

Figure 2.6: Estimated Treatment Effect for an Average County Switching from Punch Cards to Optical Scanners

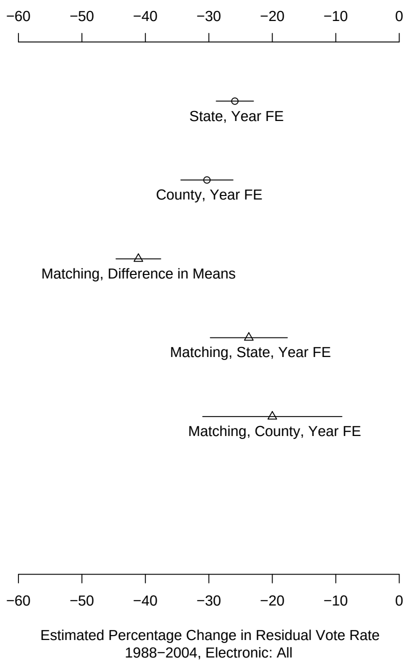

Figure 2.7: Estimated Treatment Effect for an Average County Switching from Punch Cards to DREs

cards. Lever machines are a close second, with a rate of residual voting 30% lower than punch card machines. Electronic machines and optically scanned ballots produce smaller improvements over punch cards, but improvements nonetheless. Electronic machines produced 26% lower rates of residual voting than punch cards. Counties switching to optical scanners experienced an average decline of 21% in residual votes over punch card counties. This is a smaller estimate than each of the three difference-in-difference estimates discussed previous.

The percent of the population aged 18–24 and median income are negatively related to residual vote rates, while percent of the population that is minority and percent of the population 65 and older are positively related to residual votes. The results indicate a negative relationship between shifts in technology and residual votes. This relationship could be due to counties taking extra precautions to educate voters during years when a shift in technology occurs. Most of the coefficients on the control variables have the expected sign. The positive relationship between other prominent offices on the ballot and residual votes and the negative relationship between young voters and residual votes, however, are not as expected. These results may be due to the omission of other confounding factors at the county-level.

The second model introduces county-level fixed effects, in order to control for the many unobservable county-level characteristics that remain relatively constant over time. After controlling for other confounding variables such as contemporaneous gubernatorial and senatorial races, shifts in technology, and the year of the election, averaging over the changes in residual vote rates as counties change technology provides a better estimate of the effect of each particular technology type.

This specification produces the same ordering of the equipment types, in terms reduction in residual vote rates under punch card machines, than the previous model. Counties using paper ballots generated 49% lower residual vote rates than counties using punch cards, whereas lever and electronic machines produced 32% and 30% lower rates of residual voting, respectively, than punch cards. Optical scanners are still the closest to punch cards, but the magnitude is larger than in the previous regression—24% rather than 21% lower rates of residual voting.

In sum, the fixed effects models overwhelmingly indicate that punch cards are the worst technology in terms of rates of residual voting. Paper ballots and lever machines produce the lowest rates of residual voting. Although electronic machines and optically scanned ballots do not reduce residual votes at the level estimated for paper ballots, they certainly fare much better than punch cards.

2.5.3 Propensity score matching estimates

The final estimation method considered is propensity score matching. Taking simple differences in mean residual vote rates across the matched samples results in a distinctly different pattern than the fixed effects estimates. Here, electronic machines and optical scanners fare the best, causing a 41% and 38% reduction in residual votes for counties that switch from punch cards, respectively. Paper ballots and lever machines still represent marked improvements over punch cards, with 31% and 32%

reductions, respectively.

As it is unlikely that we acheive uniform improvement in all observables in the matching procedure, the minor differences that remain are adjusted by running a parametric analysis on the parametric data (Ho et al., 2006). Additionally, by including fixed effects in the parametric analysis, unobserved fixed confounding variables are controlled for as well. Again, it is useful to note that the matching procedure discards observations in one group that are "far away" from the observations in the opposite group, resulting in a matched data set that looks similar in observed characteristics, and therefore relies less heavily on linearity assumptions when calculating counterfactuals.

Applying the first fixed effects model to the data, using the same covariates as in the matching procedure, as well as state and year fixed effects, once again yields a new pattern. Here, paper ballots are the stars, with a 31% reduction in residual vote rates, while optical scanners are a close second with a 28% reduction in rates. Electronic machines produce an estimated 24% reduction in residual vote rates, when counties switch from punch cards. The second equation continues to champion paper ballots, optical scanners, and electronic machines, however lever machines no longer are distinguishable from punch cards in their effect on residual votes.

2.6 An extension and future research

After research on the 2000 election debacle emerged, better data-collection practices have been advocated. One of the results of this advocacy is the availability of specific manufacturer or model types, for much of the data. This information allows the separation of counties using optical scanners into two types—those who count their optical scan ballots at a central location, away from the voter, and those who count their ballots in the precinct, allowing voters the opportunity to resubmit voided ballots. For electronic machines, we can again distinguish two types of counties—those who record the votes mechanically, similarly to a lever machine, and those who record the votes electronically, on the newer ATM-style touchscreen machines. One additional piece of data available for 2004 is the Election Assistance Commission's *Election Day Survey* (2005). From this survey, two additional covariates are added to the analysis: First, the reported average number of poll workers in a precinct, and second, the reported average number of polling places in a precinct. These both may contribute to educational and organization factors that influence residual votes, and are certaintly correlated with technology.

The analysis of the 2004 data begins with a linear model, which has the transformed residual vote rate as the dependent variable, and the same covariates as in the first fixed effects regression above. Additionally, the two EAC variables were included, as well as state fixed effects. Then a matched sample was obtained for each of the four treatment groups by repeating the matching algorithm with different propensity models until balance was as good as possible. In each case, the treated and control groups in the matched sample were more alike in their distribution of covariates than these groups in the raw data. Because uniform improvement in all covariates to the degree of 100% bias reduction is unlikely, the fixed effects estimator was applied to the matched samples.[16] Unfortunately, there were relatively few cases of counties using touchscreen technology that were sufficiently similar to punch card counties to create a sizable matched sample. As a result, the parametric estimation was not performed on this group.

Figures 2.8–2.11 compare the estimates for this analysis. Counties switching from punch cards to centrally-counted optical scanners do not gain a distinguishable improvement in residual vote rates, by any of the estimates. Precinct-counted optical scanners do quite well by the regression estimate and the simple differences estimate on the matched sample. Once controlling for unobserved factors on the matched sample, however, these benefits are no longer apparent. For both of the electronic machine types, the pattern is varied. Matching produces estimates of large reductions in residual votes when changing to electronic machines, but regression estimates do not pick up this distinction.

Although it appears that model dependency is driving some of the results, there is a final consideration for the electronic machines. Rather than focusing on the tallying

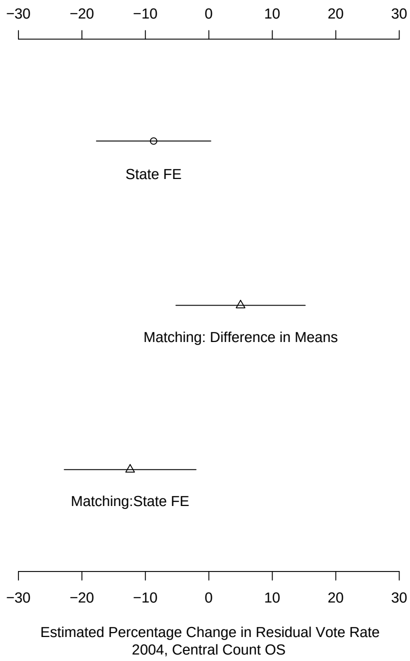

Figure 2.8: Estimated Treatment Effect for an Average County Switching from Punch Cards to Central Count Opscan in 2004

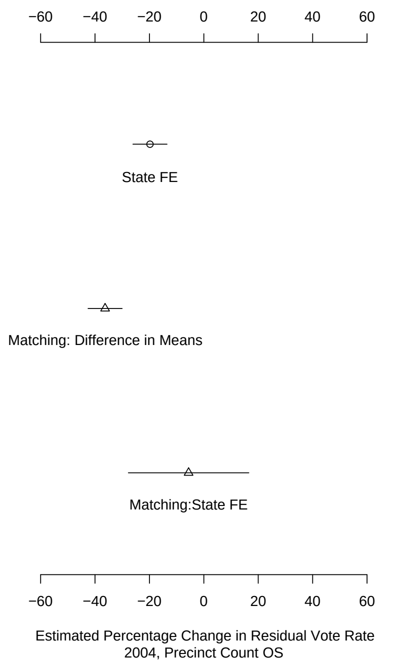

Figure 2.9: Estimated Treatment Effect for an Average County Switching from Punch Cards to Precinct Count Opscan in 2004

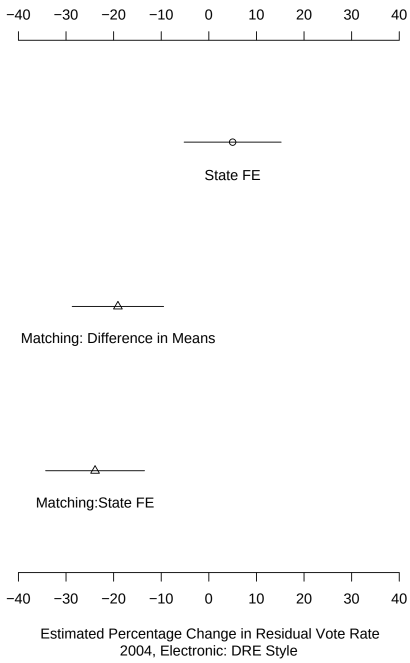

Figure 2.10: Estimated Treatment Effect for an Average County Switching from Punch Cards to Electronic DRE

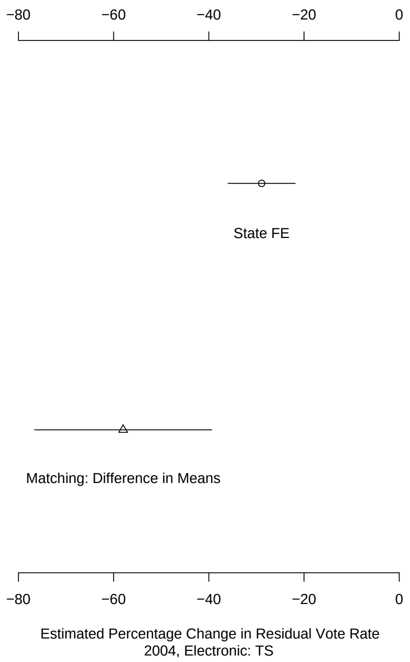

Figure 2.11: Estimated Treatment Effect for an Average County Switching from Punch Cards to Electronic Touchscreen

of votes—mechanical or electronic—perhaps the machines should be sorted according to whether there are prompts or mechanisms that prevent overvoting or undervoting of races. For instance, some touchscreen machines display a warning message if a voter attempts to leave a race blank, whereas other touchscreen machines have no such message. This may be the right place to focus, and is an interesting focus for future research. Another suggestion for future research would be to incorporate data from contests for offices other than president, such as U.S. senator and gubernatorial races. This would provide another four panels to the data, if all midyear elections between 1988 and 2004 were added, and might help clear up the picture. Although the results do appear to be somewhat dependent on the linearity assumptions of the regression models, there is the comforting fact that punch cards nonetheless perform poorly across the board, and the current push toward precinct-counted optically scanned ballots and electronic machines appears to be in the right direction.

2.7 Notes

[1]See Ansolabehere and Stewart 2005 for a detailed discussion of previous literature on the performance of voting technologies.

[2]Although historically some citizens have faced challenges to voting, such as difficulties with registration or polling place accessibility, these subjects are beyond the scope of this paper.

[3]For a detailed account of particular machine types and their features, please see the report at the Center for American Politics and Citizenship (2003).

[4]The different variables that influence a voter's ability to indicate his intended preference in an election, both in and out of the voting booth, are interesting questions in their own right and have been the subject of recent research. See especially Caltech/MIT Voting Technology Project, 2001; Tomz and Van Houweling, 2003; and Highton, 2004.

[5]This data was kindly shared within the VTP by Steve Ansolabehere and Charles Stewart.

[6]These cases represent less than 2% of the total sample. Some examples are observations where the residual vote rate is negative, or when it is much higher than is generally believed to be plausible, such as a residual vote rate of 50%. These cases have been eliminated until the election returns can be confirmed.

[7]See Greene, 2003, 287–293, for a detailed discussion of fixed effects models.

[8]Demographic variables are linearly interpolated from the U.S. Decentennial Census and median income comes from the Bureau of Economic Analysis.

[9]These regressions were estimated using the STATA command `areg`. In addition to the specifications presented above, I tested for the effect of outliers by re-estimating the models after dropping outliers from the data. Outliers are defined to be those

observations with unusually low (less than 0.1%) or unusually high (greater than 10%) levels of residual vote rates. The results on the restricted sample do not change the coefficients on the technology indicators significantly and therefore are not presented here.

[10]Although a generalized propensity score model (as suggested in Imbens (2000) and further developed in Imai and Van Dyk (2004))—such as estimating a multinomial propensity score for the treatment, $T_{it}^{j}, \quad j = 0, 1, \ldots, 4$—prevents the loss of information that occurs when considering each pair, (T_{it}^{1}, T_{it}^{0}), separately, the full implications of the model in this context are not immediately clear. Additionally, it is not obvious that the computational difficulties will outweigh the benefit. This is left for future research.

[11]Operationally, the matching algorithm was implemented with the package "Matching" written for the R programming language and statistical software by Sekhon (2006).

[12]The results from the final matching stages are presented in the Appendix.

[13]This estimate is too small, because it does not take into account uncertainty involved in the matching procedure or uncertainty from estimating the propensity score. See Abadie and Imbens (2004) for an alternative approach.

[14]Full estimation results are presented in the Appendix.

[15]To test whether partial or total pooling is appropriate for this data, dummy variables for each period were interacted with the independent variables in the estimated equation, detailed in Section 5.1. The intercepts of the equations corresponding to each time period differ significantly. The unobservable, fixed temporal effects can be captured by introducing year fixed effects into the model, the starting point for the specifications in the next section.

[16]The results of the final matched samples are available from the author.

Chapter 3

Incumbency advantage as an illustration of obtaining reliable causal estimates with classical linear models

Incumbency advantage is a much-studied phenomenon in political science.[17] Studies have tackled the problem of measuring incumbency advantage in various ways, ranging from the measures of sophomore surge and retirement slump; the use of term limits, redistricting, and deaths in office to avoid selection bias; and measurement models in the form of both classical regression and hierarchical Bayesian regression. At present, I know of no other study that approaches the problem of measuring incumbency advantage using matching methods. As many proponents of matching methods will argue, matching is generally preferable to classical linear regression because the assumption of independence conditional on observing all relevant covariates is explicit in both models, but matching methods do not suffer from model dependency, as classical linear models can.

Incumbency advantage is an interesting example of causal inference in political science because the treatment—incumbency status—is relatively independent of other predictors in the model from the start. One would think that in this scenario, matching would produce the same estimates as the linear model but with increased uncertainty, as a consequence of the reduced sample size. However, the results in this paper show that this is not always the case. If the proxy for treatment assignment is a bad predictor, such as when the researcher chooses the propensity score model based on predictors that are correlated with the outcome—typically control variables in a linear regression—but without regard to the fact that they are uncorrelated with the treatment, the matching procedure will be essentually an ad hoc sample of the original data. If the researcher has a bad draw—even if the matched sample appears to be "balanced"—the estimates from the matched sample will not only be biased, they will be excessively precise, leading to confidence about incorrect estimates. As the process is subject to uncertainty, researchers considering matching in favor of linear regression should consider the likelihood of model dependency and proceed with caution if the best propensity score model is not likely to be close to the "true" propensity score model. This point seems straightforward, but it is important to emphasize the potential error in poor modeling choices that may arise from choosing estimators by the current fashion rather than by careful analysis.

The rest of the chapter proceeds as follows: Section 3.1 formulates incumbency advantage as a causal inference problem and discusses in further detail previous attempts to measure it. Section 3.2 describes the data utilized and presents the model. Sections 3.3 and 3.4 contrast linear regression estimates and average treatment effect calculations after matching. Section 3.5 concludes.

3.1 Formulating incumbency advantage as a causal inference problem

The incumbency advantage, informally, is the benefit in vote share that an incumbent legislator enjoys over a non-incumbent challenger. Isolating this quantity is difficult for several reasons. Essentially, the problem is missing data: we want to estimate the difference between the proportion of the vote share that the incumbent legislator receives when running for re-election (which is unobserved in an open seat race) and the proportion of the vote share the incumbent party wins if the incumbent legislator does not run (which is unobserved in an election with an incumbent candidate). If incumbency were randomly assigned, we could make assumptions regarding the exchangability of the two scenarios and achieve reasonable estimates for the unobserved counterfactuals. But, in fact, incumbency is not randomly assigned across districts. We cannot perform an experiment randomly distributing incumbent status across candidates in a controlled election setting. Thus we find ourselves squarely in the realm of observational studies.

One problem we might face is selection bias. Perhaps incumbent legislators are systematically higher-quality candidates than their challengers. Or it may be the case that incumbent legislators only enter into a race if they expect to win it. If either of these are the case, we do not observe two critical instances for measurement: first, the low-quality incumbent (of such a low quality that he is not elected in the first place), and second, the incumbent legislator who knows he will lose re-election.

Previous research has suggested that term limits provide a natural experiment from which to study incumbency advantage, as it avoids the above problem of strategic retirement (Ansolabere and Snyder 2002). This approach suffers from selection bias as well. To see that this is the case, let the terminal election cycle when term limits force the incumbent candidate out of office be in time T. One way to measure the incumbency advantage, then, is to compare the incumbent candidate's vote share in election $T-1$ (his last election) and the vote share of the incumbent's party in election T (the election for the open seat). At first glance this may seem sensible, but consider that challengers face a decision as to when to enter the race. High-quality challengers may systematically avoid running against an incumbent in time $T-1$ and defer their candidacy to the open seat race in time T. And then the measure of incumbency advantage would be biased.

Two other popular measures of incumbency advantage are sophomore surge and retirement slump. Sophomore surge is defined to be the average vote gain for freshman winners in election 1 who run again in election 2. Retirement slump is defined to be the average vote loss for parties who candidates won in election 1 and did not run in election 2. But, as proved in King and Gelman (1990), both of these measures are biased. Sophomore surge underestimates the incumbency advantage, while retirement slump overestimates it. Further, while the average of these two quantities, the "slurge," is a better estimate of the incumbency advantage, it is still biased. The paper goes on to define a linear regression model that uncovers an unbiased measure incumbency advantage. This model is a starting point for this analysis and is discussed in detail in Section 3.2 below.

3.2 Data and methods

The data available on congressional elections in the United States is quite extensive. The unit of analysis for this data is a (district, year) pair and the data span all congressional districts in the United States from 1898–2002. The data include election returns for both major parties, a variable denoting whether an incumbent ran for reelection (and if so, his party), and a variable denoting whether redistricting occured immediately prior to that election cycle.[18] Years in which redistricting occured are eliminated from the sample.

A useful starting point in modeling the data is the unbiased estimator from Gelman and King (1990)

$$v_{jt} = \beta_{0t} + \beta_{1t} v_{j,t-1} + \beta_{2t} P_{jt} + \Phi_t I_{jt} + \nu_{jt} \tag{3.1}$$

where:

v_{jt} is the Democratic share of the two-party vote in district j at election t;

$v_{j,t-1}$ is the Democratic share of the two-party vote in district j at election $t-1$;

P_{jt} equals 1 if the Democrats are the incumbent party (i.e., a Democratic candidate won at election $t-1$) and -1 if the Republicans are the incumbent party in district j at election t;

I_{jt} equals 1 if there is a Democratic incumbent, 0 if there are no incumbents and -1 if there is a Republican incumbent seeking reelection in district j at time t (note that this definition ignores the possibility of two or more incumbents in a given district since redistricting years are excluded from the analysis);

and ν_{jt} is an error term.

This formulation compares vote shares in incumbent districts to vote shares in districts with open seats, controlling for the effect of the incumbent *party* and the vote share in the previous election. It is useful to reparameterize the coding such that the treatment is binary. Let the treatment be the decision of candidate to run for reelection. Then,

$$I_{jt} = T_{jt}P_{jt} \tag{3.2}$$

where:

T_{jt} equals 1 if the party runs a new candidate in district j at time t, and 0 if the incumbent legislator is running for reelection in district j at time t,

and Equation 3.1 above becomes:

$$v_{jt} = \beta_{0t} + \beta_{1t}v_{j,t-1} + \beta_{2t}P_{jt} + \Phi_t T_{jt}P_{jt} + \nu_{jt}. \tag{3.3}$$

Gelman and Huang (2007) present a hierarchical model of the incumbency advantage that allows the advantage to vary between incumbents. As I am not interested in the variation in incumbency advantage, but rather in a precise estimate of the average effect of incumbency across districts, the classical model is employed here. An extremely important, though under-emphasized, feature of the Gelman-Huang model is the inclusion of a lagged value of the incumbent party. That is, for election t, the term $P_{j,t-1}$ is included in the model. Including this term in the model maintains the structure of estimating incumbency advantage for pairs of elections—comparing districts with incumbents to districts with open seats—but results in significantly more precise estimates. Thus, the full model used for the classical linear regression here, and as a starting point for the propensity score model below, is:

$$v_{jt} = \beta_{0t} + \beta_{1t}v_{j,t-1} + \beta_{2t}P_{j,t-1} + \Phi_t T_{jt}P_{jt} + \nu_{jt}. \tag{3.4}$$

The intuition behind this result is that as more information is added to the model, the estimator gains efficiency. To see this, first note that one of the best-known predictors of current outcomes is generally the previous outcome(s). To this end, the Gelman-King structure models vote share, v_{jt}, as a function of incumbency, I_{jt}, incumbent party, P_{jt}, and previous vote share, $v_{j,t-1}$. Next, note that the term P_{jt} adds little to no new information to the model, as in virtually all cases, $P_{jt} = 1 \iff v_{j,t-1} > 0.5$. But, including $P_{j,t-1}$ in the model does give new information—information about $v_{i,t-2}$ – as $P_{i,t-1} = 1 \iff v_{i,t-2} > 0.5$ generally. The structure of the model as estimating the change between election t and election $t-1$ is maintained, but information about the twice-lagged vote share is implied.

3.3 Classical linear regression approach

The estimates from Equations 3.4 and 3.3 can be seen in Figure 3.1 below. The left graph entitled "Model 1" refers to Equation 3.4, while the right graph refers to Equation 3.3. The y-axis on both graphs ranges from -0.05 to 0.15 and represents the average gain in democratic vote share for an incumbent. The x-axis ranges from 1898–2002 representing the election years in the sample. Each point on the graph is an estimate of $\hat{\Phi}_t$ and the vertical bars through each point represent the 95% confidence interval around the estimate. Note that the years ending in "2" do not have estimates. This is because redistricting occurs after each decentennial census and the election immediately following redistricting is dropped from the sample.

Looking first at the graph on the right, the estimates replicate Gelman and King (1990) with the addition of recent election years. As they point out, consistent with congressional elections literature, the incumbency advantage has been increasing over time and—contrasting with estimates of the "slurge"—there is positive benefit to incumbency as far back as 1900. The graph on the left is the new specification that is a hybrid of the Gelman-King and the Gelman-Huang estimators. The general pattern remains the same as in the original model. Incumbency advantage has increased over time and there are positive benefits to incumbency estimated as far back as 1900. In contrast to the original model, however, the measure is much less variable. This

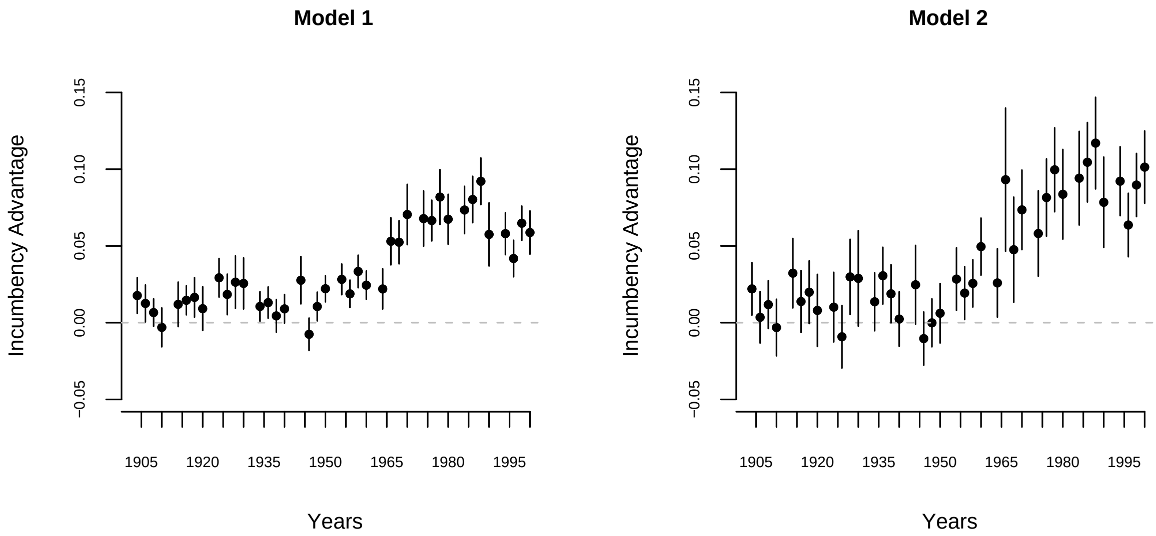

Figure 3.1: Estimates and 95% Confidence Intervals of Incumbency Advantage, Raw Data

is because the model is more reliable and statistically efficient as a result of the additional information provided by the $P_{j,t-1}$ term.

3.4 Propensity score matching approach

The propensity score matching methods developed in Rosenbaum and Rubin (1983) attempt to overcome the problem of discerning treatment effects in observational studies by explicitly conditioning on observables. The problem in identifying treatment effects is essentially a missing data problem—the treatment group is observed and the outcome conditional on treatment assignment is observed, but the counterfactual *is not* observed. In this case, we observe incumbency status, and we observe the democratic vote share in each district j in election t *conditional on incumbency status.*

Advocates of matching argue that we should prefer these semi-parametric methods since there is clear standard for choosing an optimal model—it is the one that balances the covariates, X, and no functional form is implied for the relationship between the treatment and the outcome (Sekhon, 2004). But a key assumption in matching, as in classical linear regression, is that there are no variables left unobserved that are correlated with treatment, T_{jt}, and the outcome, v_{jt}. As exact matching is not practical in many scenarios—X is multidimensional and contains continuous covariates—practitioners generally rely on a result due to Rosenbaum and Rubin (1983), which demonstrates that it is enough to condition on the propensity score, $p(X) = pr(T = 1|X)$. As the true propensity score is not known, it must be estimated. The generally acceptable place to start is with a logistic (or probit) regression of the treatment on the covariates in the linear model. Of course, as the propensity score is not known, only estimated, we could very well be conditioning on the wrong scalar! Ho et al. (2006) point out what they call the "propensity score tautology":

> The estimated propensity score is a balancing score when we have a consistent estimate of the true propensity score. We know we have a consistent estimate of the propensity score when matching on the propensity score balances the raw covariates. Of course, once we have balance on the covariates, we are done and do not need to look back. That is, it works when it works, and when it does not work, it does not work.

In the case of incumbency advantage, the seemingly obvious place to start is a model of incumbency as a function of the prior outcome—vote share in election $t-1$—and additional covariates such as the incumbent party.

$$T_{jt} = \Phi(\beta_0 + \beta_1 v_{j,t-1} + \beta_2 P_{jt} + \beta_3 P_{j,t-1}) \tag{3.5}$$

where the variables are as above and Φ is the normal cdf.

Previous researchers have noted that the lack of correlation between incumbency and previous vote status in the district make the problem less susceptible to model dependency (Gelman and King, 1990; Gelman and Huang, 2007). And indeed this is the case for most years in the data.

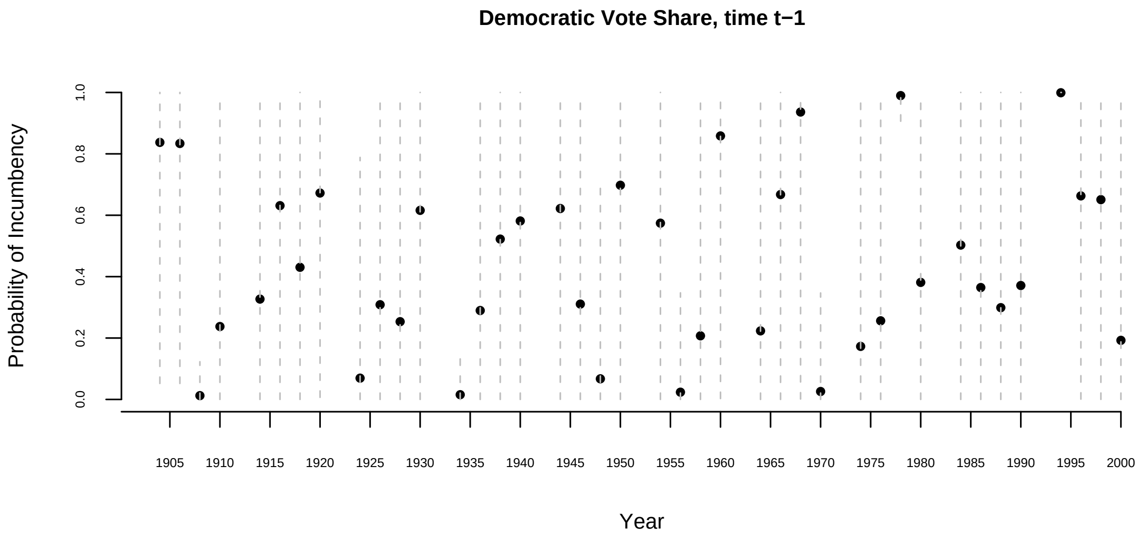

Figure 3.2: Estimates and 95% Confidence Intervals of Marginal Effect of Vote Share on Incumbency Advantage

Figure 3.2 presents the point estimates representing the average marginal effect of vote share in the previous election on the probability of an incumbent running for election at time t across all districts j based on Equation 3.4. That is, the point estimates plot $\hat{\beta}_1$ and the dashed lines represent the 95% confidence intervals across the years in the sample. For the majority of the years, the confidence interval spans the unit interval indicating that there is virtually no predictive power of $v_{j,t-1}$ on T_{jt}. One year that stands out in the graph is 1994. This is an interesting case politically, as 1994 was the first time the Republican party took control of the House of Representatives in 40 years. Many changes in the electoral environment have been hypothesized to effect the Republican victory, but it is certain that a larger than normal number of districts that would otherwise have held incumbents no longer did. That election is systematically distinct from the other years and, as such, we would expect the results from that year to be a bit off.

Nonetheless, suppose that a researcher did not consult the uncertainty estimates of the propensity score estimate, but rather turned to the question of balancing the data. As the treatment is uncorrelated with the covariates in question, the two samples would appear to be balanced for virtually any sample drawn from them, regardless of the propensity score.

To consider this scenario, two matched samples are drawn. One is based on a propensity score, as in Equation 3.4, and another includes the lagged-value of the treatment, $T_{j,t-1}$, in the equation as well. A balanced sample is achieved based on any traditionally accepted metric: minimizing the maximum distance between qq-plots of the two samples, minimizing the average distance between qq-plots of the two samples, and minimizing the median distance between the two qq-plots.[19] Once a balanced sample is achieved, the average treatment effect is estimated by repeating the estimation procedure in Section 3.3 above.

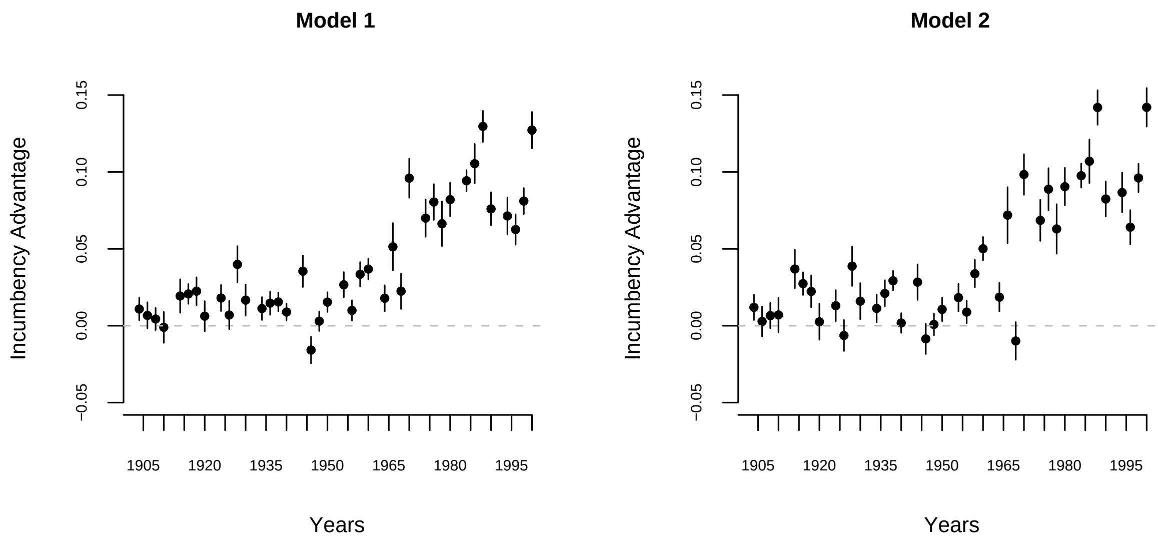

Figure 3.3: Estimates and 95% Confidence Intervals of Incumbency Advantage, Matched Data — Specification 1

The results from estimating incumbency advantage using Equations 3.4 and 3.3 on matched samples obtained from the two propensity score specifications can be seen in Figure 3.3 and Figure 3.4 above. In each figure, the left graph entitled "Model 1" refers to Equation 3.4, while the right graph refers to Equation 3.3. The y-axis on both graphs ranges from -0.05 to 0.15 and represents the average gain in Democratic vote share for an incumbent. The x-axis ranges from 1898–2002 representing the election

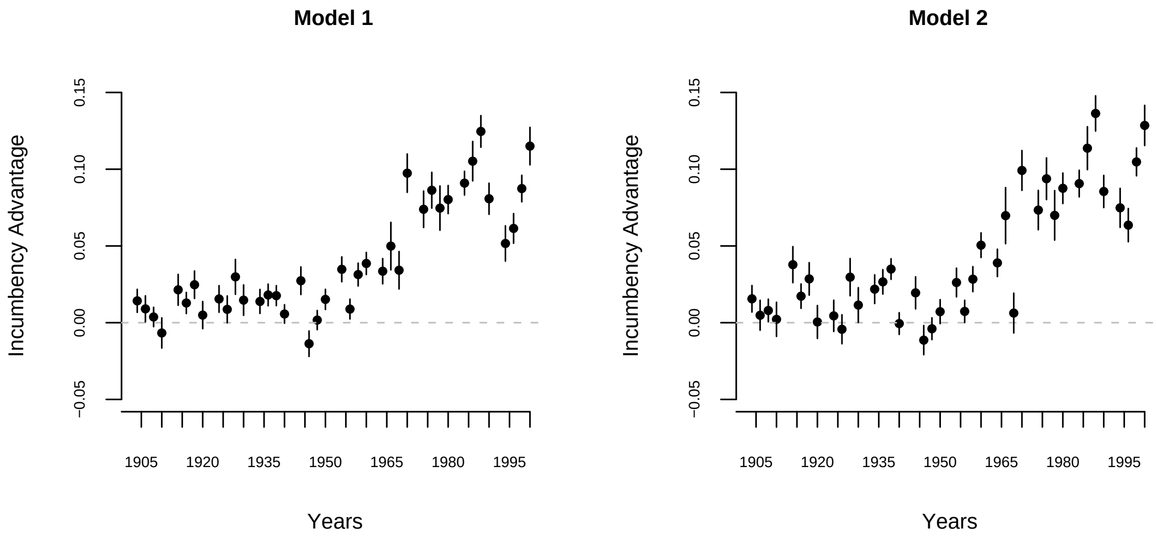

Figure 3.4: Estimates and 95% Confidence Intervals of Incumbency Advantage, Matched Data — Specification 2

years in the sample. Each point on the graph is an estimate of $\hat{\Phi}_t$ and the vertical bars through each point represent the 95% confidence interval around the estimate. Note that the years ending in "2" do not have estimates. This is because redistricting occurs after each decentennial census and the election immediately following redistricting is dropped from the sample.

There are two major points to take away from these graphs. First, although there is no reason based on balance tests to choose one propensity score model over the other, the results from each specification vary greatly. That is, how the sample is drawn affects the estimates, even if both samples appear to be "balanced." Second, comparing these figures to Figure 3.1 above, the estimates on the matched samples are more variable across time, but generally more precise within an election year. The estimates change not only between the raw data and the matched sample, but across the matched samples as well. The estimates in Figures 3.3 and 3.4 have smaller standard errors in general than the estimates in Figure 3.1, but jump around a great deal. This is a concern because we would expect that changes in incumbency advantage over time would occur smoothly, with the exception of a few shocks to the system, as we see in Model 1 in Figure 3.1. Another great concern is the difference in estimates across the two matched samples. Another advantage to the parametric model applied to the raw data is the replicability of the results.

3.5 Discussion

So, what's going on here? Do we have a case of model dependency? Is it a case of poor specification? Our initial hypothesis is that matching shouldn't help much, but it shouldn't hurt either. All the sensible candidates for confounders aren't in fact confounders. They aren't correlated with the treatment, so we'd expect standard errors to increase as a result of a decreased sample size, but that's about it. In fact, we are in a situation where the practical application of matching is at odds with the theoretical properties behind matching. Applied researchers turn to matching to simulate a quasi-experimental setting. The idea is to make sure the treatment and control groups look alike, so that any heterogeneity that might confound the causal effect of the treatment on the outcome is eliminated. But the problem here is that we're selecting matched pairs that are alike, but they might be alike in the wrong way. It is a case where we don't know the propensity score in truth and we probably can't know it with the data in hand—so matching might actually do worse. Since the treatment isn't correlated with these covariates, a propensity score based on them is the wrong metric on which to base a sample, even though they are the right covariates in the regression and even though traditional balance metrics pass.

3.6 Notes

[17]See e.g., Erikson, 1971; Alford and Hibbing, 1981; Gelman and King, 1990; Cox and Katz, 1996; Ansolabehere, Snyder and Stewart, 2000; and Gelman and Huang, 2004.

[18]The data for the years 1898–1992 are from ICPSR data set 6311 "Elections to the United States House of Representatives, 1898–1992," Gary King (Principal Investigator). The data for years 1994–2002 come from Gary Jacobson. Professor Jacobson also supplied data spanning 1946–1992 which was cross-checked against the ICPSR data.

[19]These results are available from the author.

Chapter 4

Measuring the impact of voter identification laws on turnout as an example of causal inference with ordinal treatment variables

Since the 2000 presidential election, much debate has centered on election reform—either through changing the technology used to cast and count ballots, expanding ballot access to include absentee and early voting, introducing statewide uniformity into the registration databases, or requiring identification at the polls. One of the many reforms included in the Help America Vote Act (2002) was to require that all new registrants show proof of identification, either with their application or the first time they visit the polls. Since the passage of HAVA, many states have pushed for additional voter identification requirements, in particular, requirements that all voters show identification before they are allowed to participate in any election. As of just prior to the 2006 general election, roughly half of the states were requiring some form of voter identification from all voters (Electionline.org, 2006).

There is little research on the effect that voter identification requirements, of any form, have on voter participation. Thus, while both HAVA and nearly half of the states have implemented voter identification requirements by the 2006 general election, there is little understanding about whether these requirements reduce voter participation, and whether they reduce the participation rates of certain classes of voters. This is a difficult problem. Identification requirements are not randomly distributed across states and implementation may vary across states, within an identification requirement. The data is sparse—there are only 50 states and two years in our data—and the only implementation of photo identification requirements are outside the range of the current data.[20] Finally, the treatment variable—the level of identification required—is ordinal. This chapter approaches the problem in two ways. First, a novel methodology is used to study the effects of voter identification requirements on the likelihood that voters participated in these two presidential elections, a multilevel binary logit model which allows us to appropriately model how covariates from both the individual and state level, and their interaction, affect the decision to participate (Gelman and Hill, 2006). Second, I suggest a model for the proper handling of the ordinal nature of the voter identification variable. Each approach addresses an area of concern for this research question: the former, addresses how to get the most of the sparse data to answer important policy questions, particularly about subgroups; the latter, addresses the proper modeling of the ordinal treatment variable. Ideally, these two approaches would be combined.[21]

Ordinal variables—categorical variables with a defined order to the categories, but without equal spacing between them—are quite common in social science applications. Some examples include political ideology (liberal, moderate, conservative), partisan identification (Democrat, Independent, Republican), education and income (categorized differently across survey designs), the closing date of voter registration (30 days, 28 days, 21 days, election day registration), or the degree of local flooding caused by hurricane Katrina (see Alvarez et al., 2007b for an interesting example of this). Although a good deal of research exists on the proper modeling of ordinal response variables, there is not a clear directive as to how to model ordinal treatment variables. A good deal of the confusion arises from measurement difficulty. Many ordinal variables in social science have a true underlying structure that may be interval or even continuous that the measurement instrument fails to capture. This may be due to poor design, such as in the case of survey questions, or perhaps due to sparse data. Voter identification requirements, for example, are coded based on "sensible" breakpoints in the data, even though the effect of implementation differences—both across poll sites and across individual voters—results in practically an infinite array of requirement levels. Even once the researcher deals with measurement issues, the coding scheme applied varies greatly—whether contrasts are made to a mean value or a base category, or if some degree of "threshold" levels are contrasted. Admittedly, the choice is generally clear based on ease of interpretation. Classical modeling options for ordinal variables generally consist of either fully unconstrained, though additive, ordinal group indicators or a numeric predictor constrained to be continuous. Typically it is difficult to employ group indicators and still constrain the variable structure in that setting. Generalized additive models are an exception, however, and adding nonlinearities to the classical constrained model can be useful. This paper proposes the use of Bayesian shrinkage (or empirical Bayes) estimators to model the ordinal nature of the voter identification variable. Essentially, empirical Bayes estimators allow the model to contain both individual group level indicators and a continuous predictor. Thus, each individual effect can be arbitrary, but the model "shrinks" the estimates toward a linear framework according to the data.[22] The next sections present the general methodology and the data utilized (the 2000 and 2004 Current Population Survey Voter Supplements, from the U.S. Census Bureau). Then the specific model formulation for this example is presented. The final section presents the empirical results.

4.1 Data

The data for the analysis consists of two levels—state and individual. The state level institutional data were obtained from state election administration websites and *electionline.org*. This data consists of the state voter identification laws in 2000 and 2004. The laws in place take a variety of forms, but can be classified into the following groups:

- Voter must state his/her name.
- Voter must sign his/her name in a pollbook.

- Voter must sign his/her name in a pollbook and it must match a signature on file.

- Voter must present proof of identification or voter registration card.[23]

This coding is based on "sensible" breakpoints in the data even though the effect of implementation differences—both across poll sites and across individual voters—results in practically an infinite array of requirement levels. The level of measurement is at the state level, based on state election codes, so we have at least 50 possible implementations, but probably more. The sparseness of the data, however, requires some form of parametric or interval restrictions on the variable. While identification requirements are certainly ordered in terms of strictness (or ease of compliance), beyond that many representations could be arguably sensible. There is likely some linear structure to the impact of requirements on turnout: the requirement of stating one's name requires very little extra effort on behalf of the voter, while remembering to bring an identification card to the polls is certain to dissuade a least a handful of voters. On the other hand, perhaps the difficulty of matching one's signature to a version on file from one's registration (perhaps many years prior) is a sufficient burden (and certainly frustrating) that additional requirements have little to no effect on turnout. It is this nature of the variable that suggests a model which incorporates both structures and allows the data to "speak" as to which structure best fits.

The heterogeneity both between states and across time can be seen most clearly by comparing Figures 4.1 and 4.2.[24] The maps depict the voter identification laws in the United States in 2000 and 2004. The lightest shade represents the "state name" requirement, while the darkest shade represents the requirement for some form of ID card. In 2000, nine states had the weakest identification requirement, 18 states required a signature to vote, nine states required a matching signature, and 15 states required a recognized identification card. In 2004, seven states only required voters to state their name at the polls, 13 states required a signature, eight states required a matching signature, and 23 states had the strictest requirement. In addition to the heterogeneity between states, eight states strengthened their identification requirements between 2000 and 2004—all switched to requiring government identification cards. No states weakened their requirements.

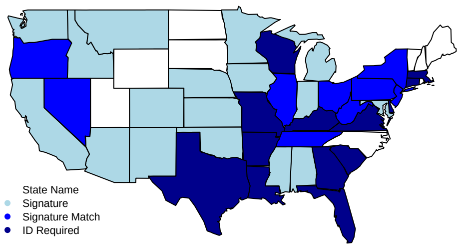

Figure 4.1: Voter Identification Laws, 2000

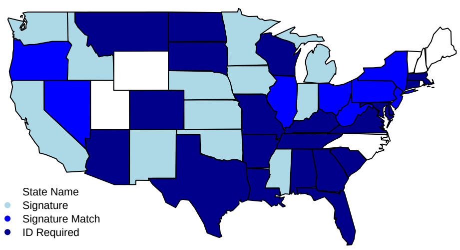

Figure 4.2: Voter Identification Laws, 2004

The individual-level data is based on a subset of the questions asked in the 2000 and 2004 Current Population Survey Voter Supplements. In addition to socio-economic and demographic questions, the respondents were asked if they were registered to vote in that year's presidential election, and, conditional on registration, if they did vote in the election. One of the major benefits of using the CPS for this analysis is the large sample size and coverage of major demographic groups. Tables C.1 and C.2 present contingency tables describing the distribution of respondents across various demographic groups for each year.

4.2 Model

Measuring the impact of voter identification laws on voter turnout is a difficult problem for several reasons: the data are sparse, there is heterogeneity both in implementation and in population, the treatment variable is not randomly assigned and is ordinal in nature. There are several options as to how to model the problem. A strictly state-level model utilizing aggregate election returns is useful as it does not suffer any of the reporting error that plagues survey data, but is subject to aggregation bias, questions of ecological inference, and a thin data matrix. In addition, the denominator in measuring turnout rates is largely inflated due to insufficient purging of voter rolls. Individual data from the Current Population Survey is beneficial because of the large sample size and excellent coverage (with around 120,000 respondents per year). Individual responses provide information on important subgroups for policy questions, as we are able to isolate effects by education level, age group, income level, or racial and ethnic background. There is concern, however, as to the level of misreporting of turnout in survey data.

The analysis begins with a logistic model of turnout from the CPS. Because we are interested in the effect of identification requirements *at the polls* and not the various unobserved barriers to voting associated with the registration process, the estimation is conditioned on the subset of respondents who are registered to vote.

$$\Pr(Y_{it} = 1) = \text{logit}^{-1}(\alpha + \beta_0 Z_{it} + \beta_1 X_{it}), \quad \text{for } i = 1, \ldots, N; \quad t = 1, 2; \tag{4.1}$$

where i indexes the respondents and t indexes years. The variable Y_{it} is binary and equal to unity if the respondent reported voting in that year's presidential election.

The variable α is an intercept term. The variable Z_{it} is the variable of interest (*VoterID*) has four levels, $j = 1, \ldots, 4$, the specification of which varies by the model and is discussed in detail below. The vector of covariates, X_{it}, includes the following:

- *South*: an indicator equal to unity if the respondent resides in a southern state;
- *Female*: an indicator equal to unity if the respondent is female;
- *Education*: a ordinal variable indicating the reported level of education—'some high school,' 'high school graduate,' 'some college,' or 'college graduate'—centered at its mean;
- *Education*^{2}: the squared value of *Education* centered at its mean;
- *Age*: the respondent's age in years centered at its mean;
- *Age*^{2}: the squared value of *Age* centered at its mean;
- *Income*: an ordinal variable indicating the reported level of education that takes on 13 values—ranging from 'Less than $5,000' to 'More than $75,000'—centered at its mean;
- *Non-White*: an indicator equal to unity if the respondent reported a race other than White.

4.2.1 Modeling the impact of voter ID on individuals

The first approach models the impact of voter identification upon subgroups of interest, namely lower educated, minority, and elderly voters. To that end, a series of interaction terms are included in the vector of covariates, X_{it}:

- *Education*VoterID*: the interaction between the respondent's reported level of education and the voter identification requirement in the respondent's state;
- *Education*^{2}**VoterID*: the interaction between the squared value of *Education* and the voter identification requirement in the respondent's state;

- *Age*VoterID*: the interaction between the respondent's age and the voter identification requirement in the respondent's state;
- *Age*^{2}**VoterID*: the interaction between the squared value of *Age* and the voter identification requirement in the respondent's state; and
- *NonWhite*VoterID*: the interaction between the indicator denoting a reported race other than White and the voter identification requirement in the respondent's state.

To allow for unobserved effects on turnout at the state-level and yearly shocks, we allow for a random intercept term that is shared by respondents in a given (state, year) pair.

$$\alpha = \alpha_{s[i]} + \alpha_{t[i]} + \alpha_{s*t[i]}, \quad \text{for } i = 1, \ldots, N;\ s = 1, \ldots, 51;\ t = 1, 2. \tag{4.2}$$

In words, each respondent has an intercept term that consists of a random intercept, $\alpha_{s[i]}$ that is shared by all respondents i in state s, $\alpha_{t[i]}$ that is shared by all respondents i in year t, and $\alpha_{s*t[i]}$ that is shared by all respondents i in state s in year t.

It may also be the case that a voter identification requirement is implemented differently across the states. To allow for this possibility, the model includes a random coefficient on voter identification that varies by state.

$$\begin{aligned}\Pr(Y_i = 1) = \text{logit}^{-1}(\alpha_i + (\beta_0 + \gamma_{s[i]})Z_i + \beta_1 X_i + \epsilon_i),& \\ \text{for } i = 1, \ldots, N;\ s = 1, \ldots, 51;\ t = 1, 2.& \quad (4.3)\end{aligned}$$

and α is as in Equation 2. That is, the effect of a voter identification requirement on the probability that two individuals i and j will vote has a common mean, β_0, but varies by a random term that is indexed by state, $\gamma_{s[i]}$ and $\gamma_{s[j]}$.

4.2.2 Modeling the ordinal nature of voter ID

The second framework explores the ordinal nature of the voter identification variable. The vector of covariates, X_{it}, is as follows:

- *South*: an indicator equal to unity if the respondent resides in a southern state;
- *Female*: an indicator equal to unity if the respondent is female;
- *Education*: a ordinal variable indicating the reported level of education—'some high school,' 'high school graduate,' 'some college,' or 'college graduate'—centered at its mean;
- *Education*^{2}: the squared value of *Education* centered at its mean;
- log(*Age*): the log of the respondent's age in years centered at its mean;
- *Income*: an ordinal variable indicating the reported level of education that takes on 13 values—ranging from 'Less than $5,000' to 'More than $75,000'—centered at its mean;
- *Non-White*: an indicator equal to unity if the respondent reported a race other than White.

In the first model, there will be no constraints put on Z_{it} and it will appear in the specification as three indicator variables—$Z^2_{it}, Z^3_{it}, Z^4_{it}$—interpreted relative to the base category Z^1_{it}. These correspnd to Z_{itj} = State Name, Sign Name, Match Signature, Present ID. In the second model, Z_{it} is constrained numerically to be a continuous variable ranging from 1 to 4. An alternative coding scheme, which is not considered here, would be to have three indicator variables where Z_{itj} = State Name, State Name or Sign Name, State Name or Sign Name or Match Signature, State Name or Sign Name or Match Signature or Present ID. This would be more useful if we had many more categories and wished to look for appropriate breakpoints in the data. As there are only four categories, however, this coding is not used.

For the multilevel model, again starting with a logistic model of turnout, letting the intercept term vary by group (*VoterID*) level.

$$\Pr(Y_{it} = 1) = \text{logit}^{-1}(\alpha_{j[i],t} + \beta_1 X_{it}), \quad \text{for } i = 1, \ldots, N; \quad t = 1, 2; \quad j = 1, \ldots 4. \quad (4.4)$$

$$\alpha_{jt} = \gamma_0 + \gamma_1 Z_{jt} + \nu_{jt}, \quad \nu_{jt} \sim N(0, \sigma_\alpha^t). \quad (4.5)$$

The parameters to estimate from the model are $\theta = \{\alpha, \beta, \gamma, \sigma_\alpha, \sigma_\mathbf{y}\}$, where $\alpha = (\alpha_1, \ldots, \alpha_j)'$, $\beta = (\beta_1, \ldots \beta_k)'$, and $\gamma = (\gamma_0, \gamma_1)$. Specifically for this data, the parameters estimated are a 4 x 1 matrix of group level intercepts, a 9 x 1 matrix of coefficients on the 9 control variables, and a 1 x 2 matrix of coefficients on the group-level predictor variable for each of the years in the data. A final consideration is about interpretation of the γ_1 parameter. This parameter is partially not identified between it and the linear trend in the ν_j parameters. The identification is partial, as the ν_j parameters are pooled toward zero, but with only $J = 4$ groups, converging the algorithm is time consuming. To correct for this problem, after estimation, the data is "post-processed" to obtain finite population slope parameters based on the regression of α_j on Z_j. This is equivalent to constraining the ν_j parameters to have mean zero and slope zero (Gelman and Hill, 2006).

The estimation is implemented with a Gibbs sampling algorithm via the statistical software *JAGS* (Plummer 2007). Independent conjugate priors are assumed for each element of β and γ and for the variances. Specifically, each β and γ is assumed to be distributed normally, with mean zero and precision parameter 0.0001. The parameters σ_y and σ_α are assumed to be uniformly distributed between 0 and 100. Given the large size of the CPS data and the computational intensity of the program, the results presented below are based on only one year, 2000.[25] Prior estimation on the years separately indicates very little difference in the estimates across years, however. I let the algorithm run for 25,000 iterations as a burn-in, and then run for 50,000 iterations, keeping every 10^{th} iteration. The estimates are based on the resulting 5000 draws from the posterior distributions.[26]

4.3 The impact of voter identification on subgroup turnout

Figure 4.3 displays the marginal effect in the odds of an individual voting in each year, for the covariates in the model. Living in the South decreases the odds that an individual votes, while being female increases the odds of voting. Being a minority does not increase the probability of voting. Each additional category of schooling greatly increases the odds of an individual turning out to vote, whereas age and income increase the odds slightly. These effects all are consistent with the previous literature on turnout, lending credence to the model specification (e.g., Nagler, 1991). It is interesting to note, however, that voter identification requirements have no impact on voter turnout—even within subgroups that might be disproportionately affected. One possible explanation for this result is that once an individual has taken the steps to register to vote, the further imposition of a voter identification requirement is no extra burden.

Figures 4.4 and 4.5 examine the impact of voter identification on subgroups further. In each figure, the probability that an average individual turnouts to vote, conditional on voter identification regime, education level, and minority status is plotted. The state chosen for the average individual is Ohio, as it is generally representative. The blue lines represent the probability of voting by education level for nonwhite respondents, the black lines for white respondents. The dashed lines denote the 95% confidence regions around the estimates. Not only is there no significant difference in turnout probabilities between white and nonwhite respondents, there is no significant difference across the voter identification regimes as well.

4.4 Comparing estimates of the ordinal models

Figure 4.6 presents the estimates from Equation 4.2 both with the voter identification variable, Z_{it}, constrained and unconstrained, and the estimates from Equations 4.4 and 4.5 of the Bayesian shrinkage model. The horizontal axis of the graph ranges

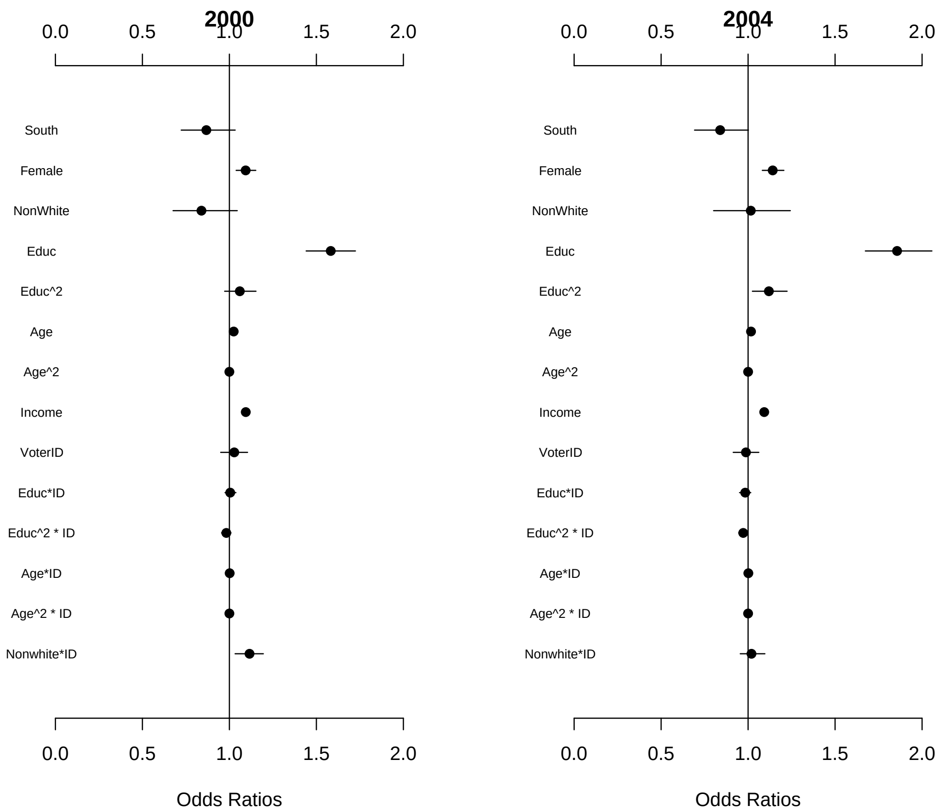

Figure 4.3: Marginal Odds of Voting Relative to the Mean Observation

from 0.5 to 2.5 and represents the coefficient estimates from the logistic regressions described above. The point estimates are the black dots on the graph, and the black lines represent the 95% intervals. The first column corresponds to the constrained model where the voter identification variable is constrained to be continuous; the second to the unconstrained model where the voter identification variable is treated as indicator variables; the third column is the Bayesian shrinkage estimate. The estimates in Figure 4.6 are based on models where the covariates are demeaned and the constrained model is forced to have no intercept term. This makes comparison across the three models clear. The constrained model forces a linear relationship on the identification coefficient that results in a large consistent increase for each step up the identification regime ladder. In the unconstrained model, there is little difference at all between the four coefficients, and the relationship across the regime is nonlinear. The Bayesian estimate looks quite similar to the unconstrained model. There are two

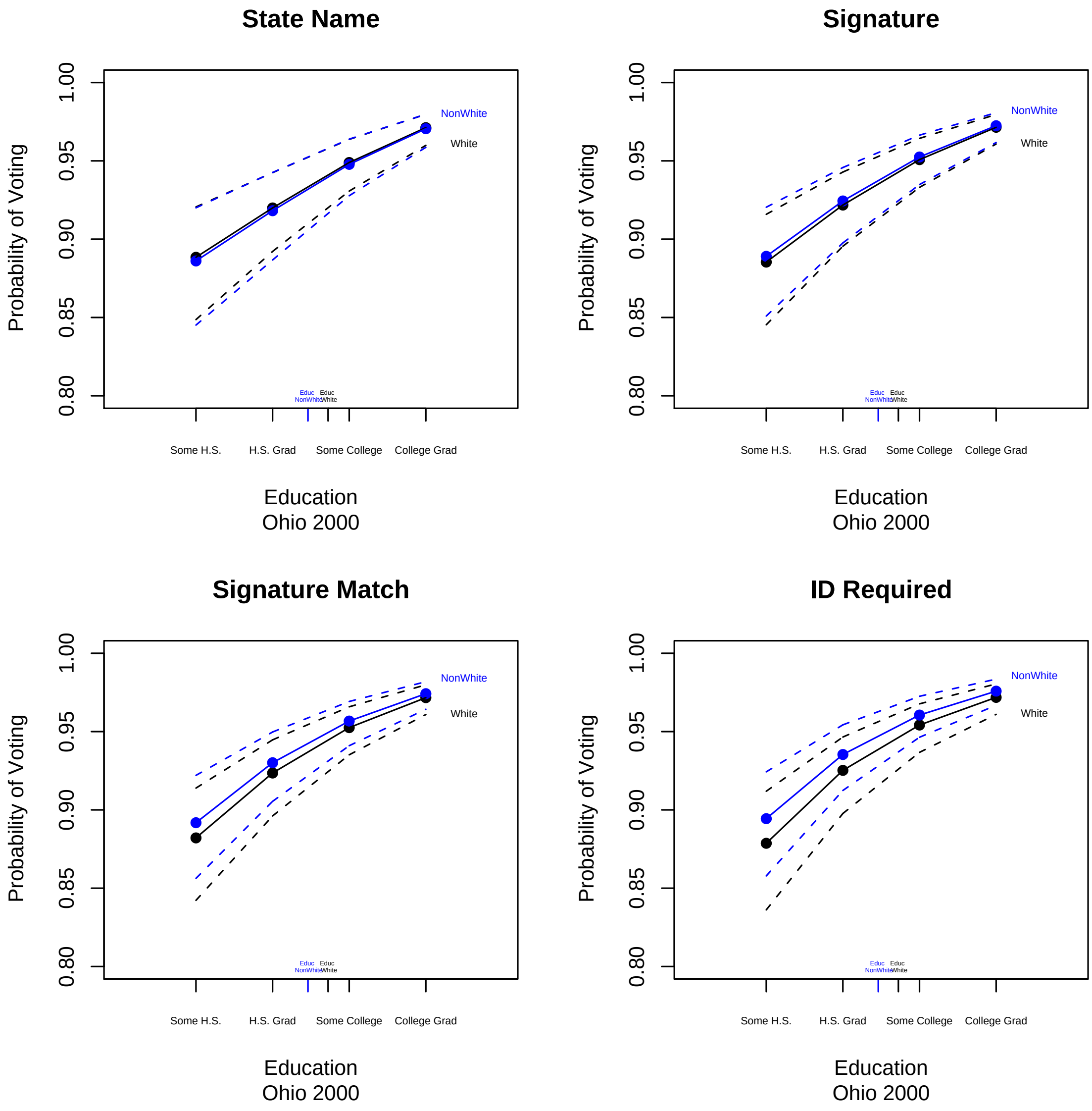

Figure 4.4: Predicted probability of turnout by ID regime, education level and minority status—Ohio, 2000

important differences, however. First, the standard errors on the shrinkage estimates are smaller, reflecting the efficiency gained by the addition of more information to the model. Second, as the shrinkage estimate we employ is relative to a linear specification, we can clearly see that the data do not support a linear relationship (as the individual random effects are quite large) and the constrained model is obviously the wrong model for this problem. To gain intuition for this result, consider first two models: one, a line (the constrained model); two, four points (the unconstrained

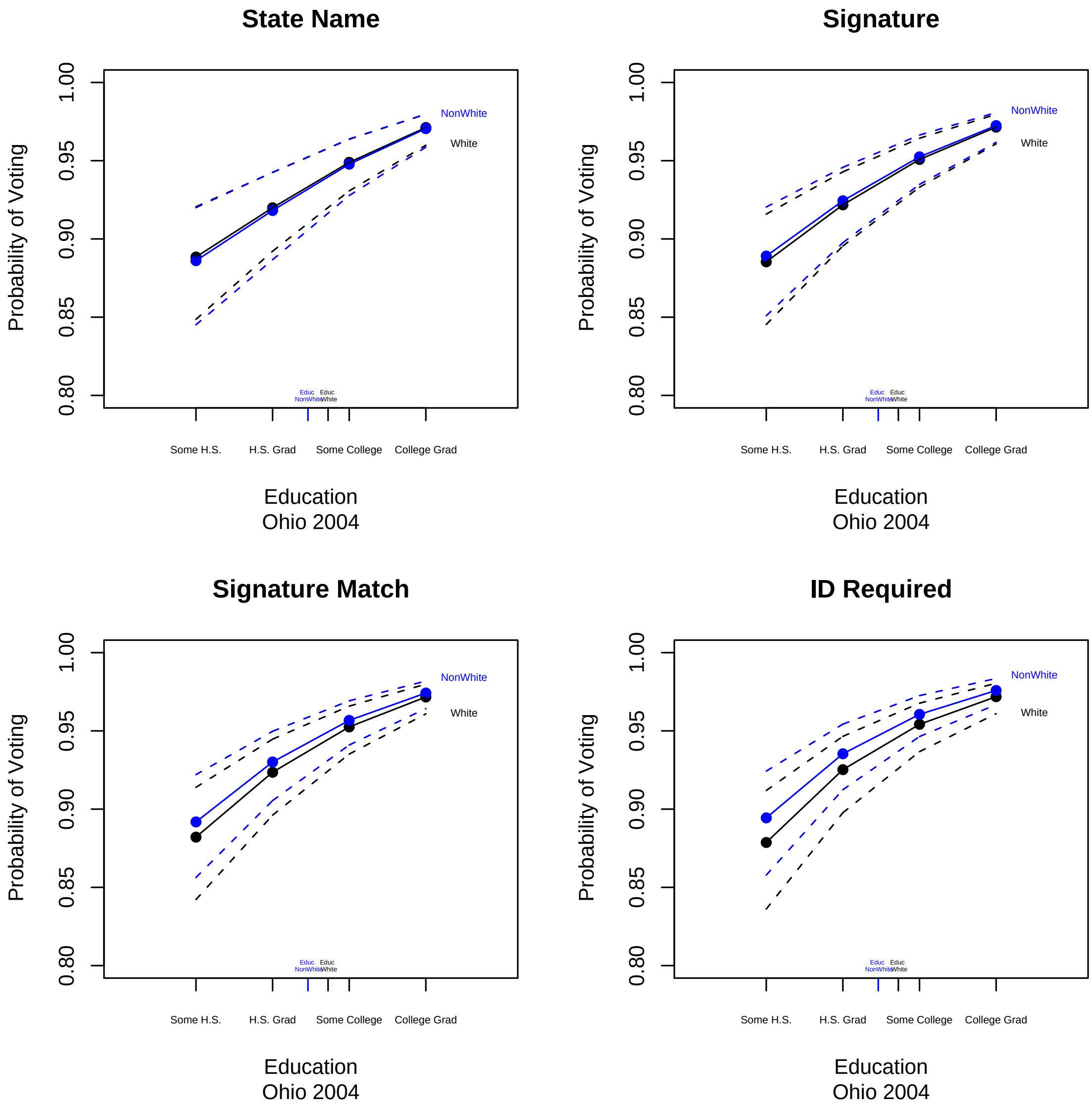

Figure 4.5: Predicted probability of turnout by ID regime, education level and minority status—Ohio, 2004

model). The Bayesian shrinkage estimator essentially computes a weighted average between the line and the point that corresponds to each ID regime. The weights are derived from the data—both in terms of the number of observations within a category (which is reflected in the standard errors of the constrained and unconstrained models) and in terms of the strength of the relationship across categories.

To further explore the predictions of the Bayesian model, Figure 4.7 plots the estimated probability of a survey respondent turning out to vote, as a function of the

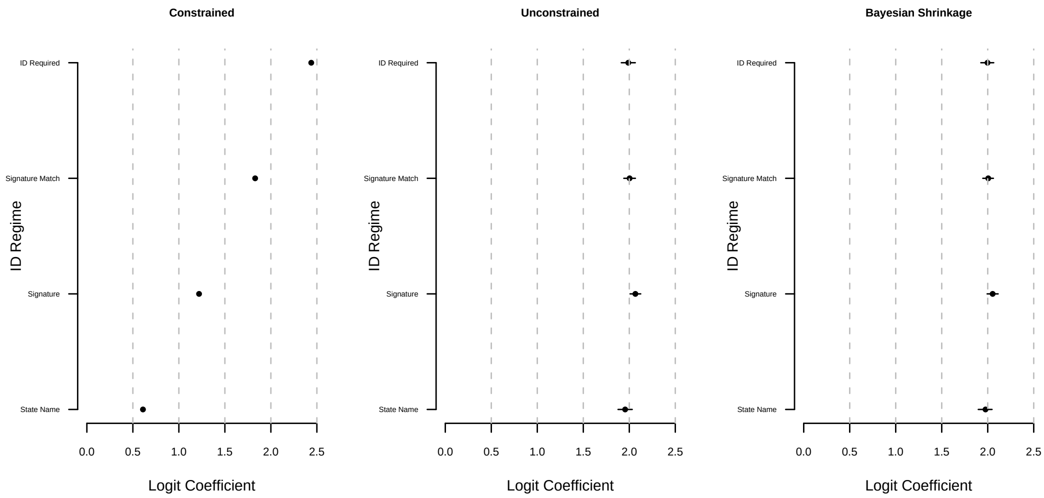

Figure 4.6: Point estimates and 95% credible intervals for the three ordinal variable models

linear predictor for demographics, under each ID regime. The dots show the data (y-jittered for visibility), and the heavy and light lines show the median estimate and 20 random simulation draws from the estimated model.[27] Under each regime, the probability of turning out to vote ranges from around 0.3 to 1.0 and there is a logistic relationship between the linear demographic predictor variable and the probability of turnout. In this model, the logistic curve is shifted by an intercept for each identification regime—the difference between this model and the unconstrained model is that these intercepts are pulled toward a group linear trend. It is interesting to note that there is no clear pattern of the effect of voter identification regime on the probability of turning out to vote. Similarly to the model above, this may be due to the fact that once an individual has registered to vote all other barriers are easily met.

4.5 Discussion and future research

Clearly the Bayesian model presented here is the better choice for modeling the ordinality of the voter identification variable, as it exploits the information from both an

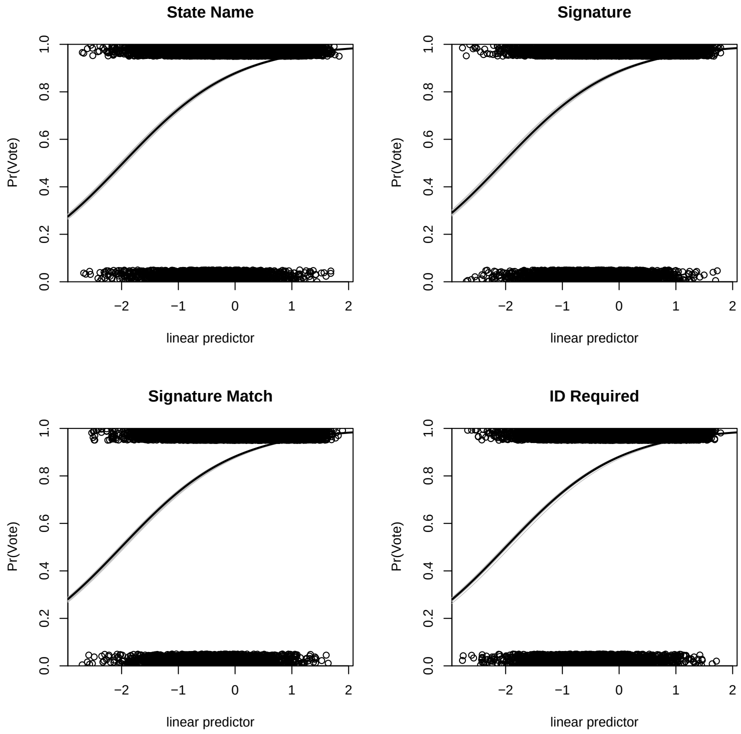

Figure 4.7: Estimated probability of voting by by ID regime

additive and a linear structure. It is relatively certain, however, that the modeling of the nationwide election trends and state-implementation is specified correctly in the first model only. The next step in this research agenda is to combine the Bayesian model here with a proper model of the implementation of voter ID and its effects on subpopulations. First, interaction terms between the voter identification variable and the interesting covariates would be included into the model. The next step in the research would be to allow for unobserved effects on turnout at the state level and yearly shocks, by allowing for a random intercept term that is shared by respondents in a given (state, year) pair. As it may also be the case that a voter identification requirement is implemented differently across the states the model will include a random coefficient on voter identification that varies by state. Then the model eventually would include a voter identification group mean and individual indentification level intercepts, in addition to a state group level mean for the voter identification levels, and state, year, and state*year random intercepts. Another benefit to using the Bayesian framework, which could be exploited, would be to impute the missing responses in the survey data simultaneously with the model estimation. In addition, the level of parsing in the identification variable could be greatly relaxed. Though seemingly overcomplex, this model is, I believe, the correct way to model the problem. Given the large availability of data and the computational convenience of Bayesian methods and imputation methods, these extentions are the next step in the research agenda. gamm deparse id

4.6 Notes

[20]Indiana implemented photo ID requirements in 2006. The data will be available soon and will be analyzed in future research.

[21]Due to computational difficulties, the combined approach is left for future research.

[22]For an interesting non-political science example of multilevel regression with an ordinal explanatory variable, see Gelman and Hill (Section 21.3).

[23]The range of acceptable proof of identification ranges across the states, but in addition to a form of government-issued photo identification, other acceptable pieces of identification include utility bills, social security cards, student ID cards, paychecks, and bank statements, as well as hunting and fishing licenses and gun permits. Combinations of the above requirements are often in place, such as requiring a voter to both state *and* sign his/her name. These cases are coded at the level of requirement that is more stringent. In this example, the case would be coded as a signature requirement. Most states in 2004 require first-time voters who registered by mail to present identification (per HAVA requirements), but for this analysis we coded states based on the requirements for *all* voters. In addition, four states require a photo identification card of some sort, but are coded as merely a general identification card because the classification of photo IDs is quite broad. For instance, Florida allowed any type of photo identification, including student IDs and Sam's Club cards; South Carolina required photo identification only if the freely provided voter registration card was not presented.

[24]Alaska and Hawaii both required government-issued identification cards in 2000 and 2004.

[25]The major next step in the research is utilizing C++ coding to greatly speed computation time and increasing the size of the sample to include the entire relevant

CPS data for both years.

[26]See Gelman et al. (2004) pp. 277-278 for a discussion of the appropriate number of draws.

[27]See Gelman and Hill (2007) for R code to make these graphs.

Appendix A

Tables for Chapter 2

Package Name	Developer URL	Description
lmer{Matrix}	Pinheiro and Bates	Fit (Generalized) Linear Mixed-Effects Models in *R*
	`http://rweb.stat.umn.edu/R/library/Matrix/html/lmer.html`	
Match	Abadie et al.	Matching Estimators for Average Treatment Effects in *STATA* and *Matlab*
	`http://elsa.berkeley.edu/~imbens/estimators.shtml`	
Matching	Sekhon, Jasjeet S.	Algorithms and Software for Multivariate and Propensity Score Matching with Balance Optimization via Genetic Search in *R*
	`http://sekhon.berkeley.edu/matching`	
MatchIt	Ho et al.	Nonparametric Preprocessing for Parametric Causal Inference in *R*
	`http://gking.harvard.edu/matchit`	
optmatch	Hansen, Ben B.	Flexible, Optimal Matching Observational Studies in *R*
	`http://www.stat.lsa.umich.edu/~bbh/optmatch.html`	

Table A.1: Useful Software for Implementing Various Causal Inference Techniques

State Name	Counties	1988	1992	1996	2000	2004	Total
Alabama	67	0	0	0	0	0	0
Alaska	40	0	0	0	0	0	0
Arizona	15	15	15	15	15	15	75
Arkansas	75	0	0	0	27	0	27
California	58	57	58	58	58	58	289
Colorado	63	62	63	63	0	62	250
Connecticut	8	8	8	8	8	6	38
D.C.	1	1	1	1	1	1	5
Delaware	3	3	3	0	0	0	6
Florida	67	0	66	66	67	67	266
Georgia	159	0	0	154	159	159	472
Hawaii	5	4	4	4	4	4	20
Idaho	44	44	44	43	44	44	219
Illinois	102	102	101	102	102	91	498
Indiana	92	90	86	89	83	71	419
Iowa	99	0	82	98	99	99	378
Kansas	105	0	82	79	94	0	255
Kentucky	120	116	115	112	107	96	546
Louisiana	64	0	55	64	62	0	181
Maine	16	0	0	0	0	0	0
Maryland	24	23	23	24	24	20	114
Massachusetts	14	0	0	0	0	0	0
Michigan	83	19	20	20	29	53	141
Minnesota	87	56	76	78	79	0	289
Mississippi	82	0	60	2	3	0	65
Missouri	115	0	0	0	113	0	113
Montana	57	54	55	56	51	56	272
Nebraska	93	93	93	91	91	93	461
Nevada	17	17	17	16	17	17	84
New Hampshire	10	7	7	6	0	0	20
New Jersey	21	15	17	19	21	20	92
New Mexico	33	27	28	31	33	33	152
New York	62	61	61	61	62	62	307
North Carolina	100	0	25	32	29	97	183
North Dakota	53	53	53	53	53	53	265
Ohio	88	88	88	88	88	88	440
Oklahoma	77	76	77	0	7	0	160
Oregon	36	29	36	36	36	36	173
Pennsylvania	69	0	0	0	1	0	1
Rhode Island	5	0	0	0	0	0	0
South Carolina	46	45	39	43	45	0	172
South Dakota	66	0	0	62	65	66	193
Tennessee	95	0	11	11	10	91	123
Texas	254	0	0	0	145	0	145
Utah	29	29	29	29	29	27	143
Vermont	14	8	0	8	0	0	16
Virginia	135	0	0	0	133	132	265
Washington	39	39	38	39	37	38	191
West Virginia	55	55	0	55	0	54	164
Wisconsin	72	0	0	0	0	0	0
Wyoming	23	19	20	21	22	23	105
Total		1,315	1,656	1,837	2,153	1,832	8793

Table A.2: States Included in the Residual Vote Rate Analysis

	Percent of Counties Using Technology				
Machine Type	1988	1992	1996	2000	2004
Punch Card	25.0%	24.3%	22.5%	20.9%	9.7%
Lever Machine	29.3%	27%	18.5%	14.8%	7.9%
Paper	32.7%	22.5%	15.1%	8.8%	10.7%
Optical Scan	5.3%	18%	30.7%	41.2%	49.9%
Electronic	1.9%	3.5%	8.7%	11.5%	19.0%
Mixed	5.8%	4.7%	4.5%	2.8%	2.8%

Table A.3: Usage of Voting Equipment in the 1988-2004 Presidential Elections, by Percent of Counties

	Percent of Population Using Technology				
Machine Type	1988	1992	1996	2000	2004
Punch Card	43.1%	41.5%	37.4%	36.2%	10.6%
Lever Machine	32.8%	29.7%	21.6%	16.2%	13.0%
Paper	6.1%	3.2%	1.8%	0.5%	0.7%
Optical Scan	7.0%	13.9%	21.6%	30.7%	40.0%
Electronic	3.1%	4.3%	11.3%	12.3%	30.4%
Mixed	7.9%	7.4%	6.3%	4.1%	5.3%

Table A.4: Usage of Voting Equipment in the 1988-2004 Presidential Elections, by Percent of Population

	Counties				Voters				
Machine Type	Mean (Std. Dev.)	1^{st} Quartile	Median	3^{rd} Quartile	Mean (Std. Dev.)	1^{st} Quartile	Median	3^{rd} Quartile	
Punch Card	2.9% (2.1%)	1.8%	2.4%	3.4%	2.5% (1.5%)	1.7%	2.3%	2.8%	[illegible]
Lever Machine	1.9% (1.8%)	0.8%	1.3%	2.3%	1.7% (1.7%)	0.8%	1.2%	1.8%	[illegible]
Paper	2.4% (2.2%)	1.3%	1.9%	2.9%	2.2% (2.2%)	1.1%	1.8%	2.6%	[illegible]
Optical Scan	2.1% (2.7%)	0.8%	1.4%	2.4%	1.5% (2.0%)	0.6%	1.0%	1.8%	[illegible]
Electronic	2.4% (2.6%)	1.1%	2.0%	3.2%	1.6% (2.6%)	0.5%	1.1%	2.1%	
Total	2.4% (2.3%)	1.1%	1.8%	2.9%	2.0% (1.9%)	0.9%	1.6%	2.5%	[illegible]

Table A.5: Residual Vote by Machine Type in U.S. Counties, 1988-2004 Presidential Elections

	Counties					Voters				
Machine Type	1988	1992	1996	2000	2004	1988	1992	1996	2000	2004
Punch Card	3.5%	2.7%	3.1%	2.6%	2.0%	3.0%	2.2%	2.5%	2.6%	1.6%
Lever Machine	1.8%	1.7%	2.2%	2.2%	1.1%	2.0%	1.7%	1.8%	1.9%	0.9%
Paper	2.7%	1.9%	2.6%	2.2%	2.2%	2.5%	1.7%	2.4%	1.9%	2.1%
Optical Scan	3.1%	3.1%	2.4%	2.1%	1.4%	2.6%	2.4%	1.7%	1.3%	1.1%
Electronic	3.6%	3.8%	3.3%	2.4%	1.6%	3.5%	3.3%	3.6%	1.8%	1.0%
Total	2.9%	2.4%	2.7%	2.3%	1.6%	2.6%	2.1%	2.3%	1.9%	1.1%

Table A.6: Average Residual Vote by Machine Type and Year in U.S. Counties, 1988-2004 Presidential Elections

Period	log($\mathcal{F}$(Residual Vote Rate)) $\widehat{ATE}$
1988-1992	13.9%
	(16)
1992-1996	-28.3%
	(5.7)
1996-2000	-43.0%
	(3.4)
2000-2004	-23.8%
	(4.6)

Table A.7: Difference-in-Differences Estimates: Estimated percentage change in residual votes for an average county switching from punch cards to optical scan machines in the specified time period

Coefficients and standard errors are transformed in order to interpret the difference as a percentage change in residual votes. Approximate standard errors are in parentheses.

Control: Punch Cards

	$\widehat{ATE}$				
	Fixed Effects		Matching		
	Eq. (1)	Eq. (2)	Difference in Means	Eqn. (1)	Eqn. (2)
Paper Ballots	-33.7%	-48.9%	-31.0%	-31.0%	-36.9%
	(2.7)	(3.1)	(2.1)	(2.1)	(8.8)
Lever Machines	-30.2%	-31.6%	-32.3%	-23.7%	-8.5%
	(1.4)	(2.1)	(1.4)	(1.5)	(21.6)
Optical Scan	-21.3%	-24.4%	-37.5%	-26.7%	-42.9%
	(0.8)	(1.5)	(1.3)	(1.5)	(2.3)
Electronic	-25.9%	-30.3%	-41.1%	-23.7%	-20.0%
	(1.5)	(2.1)	(1.8)	(3.1)	(5.6)

Table A.8: Fixed Effects and Matching Estimates: Estimated percentage change in residual votes for an average county switching from punch cards to treatment technology

Coefficients and standard errors are transformed in order to interpret the difference as a percentage change in residual votes. Approximate standard errors are in parentheses.

Control: Punch Cards

	$\widehat{ATE}$		
	FE	Matching: Means	Matching: FE
Central Count OS	-8.7%	5.0%	-12.4%
	(4.6)	(5.2)	(5.3)
Precinct Count OS	-19.8%	-36.3%	-5.6%
	(3.2)	(3.2)	(11.3)
Electronic DRE	5.0%	-19.1%	-23.9%
	(5.2)	(4.9)	(5.3)
Electronic TS	-28.9%	-58.0%	-
	(3.6)	(9.5)	(-)

Table A.9: Fixed Effects and Matching Estimates: Estimated percentage change in residual votes for an average county switching from punch cards to treatment technology, 2004

Coefficients and standard errors are transformed in order to interpret the difference as a percentage change in residual votes. Approximate standard errors are in parentheses.

Independent Variable	log($\mathcal{F}$(Residual Vote Rate)) (1)	(2)
Punch Cards	-	-
	-	-
Paper Ballots	-0.41	-0.67
	(0.04)	(0.06)
Lever Machines	-0.36	-0.38
	(0.02)	(0.03)
Optical Scan	-0.24	-0.28
	(0.01)	(0.02)
Electronic	-0.30	-0.36
	(0.02)	(0.03)
Gov or Sen on ballot	-0.01	-0.002
	(0.01)	(0.01)
Shift in Voting Tech.	-0.08	-0.04
	(0.02)	(0.02)
log(Turnout)	-0.02	-0.33
	(0.01)	(0.05)
% Black	0.01	-
	(0.001)	-
% Hispanic	0.01	-
	(0.001)	-
% Other Origin	0.01	-
	(0.001)	-
% aged 18-24	-0.01	-
	(0.002)	-
% 65 and older	0.01	-
	(0.002)	-
Median Income	-0.14	-
in 10,000's	(0.03)	-
(Median Income)^2	0.004	-
in 10,000's	(0.003)	-
Constant	-3.0	0.38
-	(0.09)	(0.56)
Observations	8793	8793
R-squared	0.48	0.66
Fixed Effect: (not shown)	State & Year	County & Year
Number of categories	45	2646
F test	F(44, 8730) = 57.8 ($p < .0001$)	F(2645, 6136) = 2.94 ($p < .0001$)

Standard errors in parentheses.

Table A.10: Residual Vote Multivariate Estimation, 1988-2004 Presidential Elections

Appendix B

Tables for Chapter 3

	Before Matching	Matching - Specification 1	Matching - Specification 2
Year	1904		
$v_{i,t-1}$			
Mean	0.101	0.040	0.018
Median	0.098	0.024	0.013
Max	0.287	0.134	0.099
$P_{i,t-1}$			
Mean	0.054	0.067	0.022
Median	0.054	0.067	0.022
Max	0.108	0.134	0.043
P_{it}			
Mean	0.112	0.000	0.002
Median	0.112	0.000	0.002
Max	0.224	0.000	0.003
$I_{i,t-1}$			
Mean	0.042	0.041	0.082
Median	0.013	0.014	0.010
Max	0.114	0.110	0.235
Year	1906		
$v_{i,t-1}$			
Mean	0.073	0.045	0.041
Median	0.077	0.039	0.039
Max	0.153	0.142	0.104
$P_{i,t-1}$			
Mean	0.046	0.037	0.021
Median	0.046	0.037	0.021
Max	0.091	0.074	0.042
P_{it}			
Mean	0.066	0.031	0.023
Median	0.066	0.031	0.023
Max	0.132	0.061	0.045
$I_{i,t-1}$			
Mean	0.042	0.067	0.044
Median	0.059	0.087	0.061
Max	0.067	0.113	0.071
Year	1908		
$v_{i,t-1}$			
Mean	0.029	0.032	0.041
Median	0.027	0.026	0.036
Max	0.090	0.125	0.143
$P_{i,t-1}$			
Mean	0.034	0.035	0.018
Median	0.034	0.035	0.018
Max	0.068	0.071	0.036
P_{it}			
Mean	0.044	0.024	0.049
Median	0.044	0.024	0.049
Max	0.087	0.048	0.097
$I_{i,t-1}$			
Mean	0.040	0.050	0.035
Median	0.019	0.074	0.006
Max	0.102	0.077	0.097
Year	1910		

$v_{i,t-1}$			
Mean	0.028	0.020	0.017
Median	0.022	0.017	0.013
Max	0.104	0.067	0.059
$P_{i,t-1}$			
Mean	0.012	0.002	0.002
Median	0.012	0.002	0.002
Max	0.024	0.003	0.003
P_{it}			
Mean	0.043	0.002	0.000
Median	0.043	0.002	0.000
Max	0.086	0.003	0.000
$I_{i,t-1}$			
Mean	0.018	0.009	0.013
Median	0.015	0.007	0.003
Max	0.039	0.020	0.036
Year	1914		
$v_{i,t-1}$			
Mean	0.038	0.016	0.013
Median	0.029	0.015	0.011
Max	0.157	0.069	0.059
$P_{i,t-1}$			
Mean	0.024	0.010	0.026
Median	0.024	0.010	0.026
Max	0.047	0.019	0.051
P_{it}			
Mean	0.024	0.010	0.002
Median	0.024	0.010	0.002
Max	0.049	0.019	0.004
$I_{i,t-1}$			
Mean	0.025	0.024	0.034
Median	0.015	0.012	0.022
Max	0.059	0.062	0.081
Year	1916		
$v_{i,t-1}$			
Mean	0.057	0.040	0.029
Median	0.057	0.032	0.019
Max	0.143	0.131	0.095
$P_{i,t-1}$			
Mean	0.073	0.011	0.000
Median	0.073	0.011	0.000
Max	0.145	0.022	0.000
P_{it}			
Mean	0.024	0.008	0.000
Median	0.024	0.008	0.000
Max	0.049	0.016	0.000
$I_{i,t-1}$			
Mean	0.091	0.019	0.019
Median	0.126	0.006	0.022
Max	0.148	0.051	0.035
Year	1918		
$v_{i,t-1}$			
Mean	0.025	0.035	0.019

Median	0.024	0.035	0.014
Max	0.075	0.095	0.069
$P_{i,t-1}$			
Mean	0.014	0.003	0.024
Median	0.014	0.003	0.024
Max	0.028	0.006	0.048
P_{it}			
Mean	0.000	0.017	0.031
Median	0.000	0.017	0.031
Max	0.000	0.035	0.062
$I_{i,t-1}$			
Mean	0.008	0.004	0.050
Median	0.005	0.006	0.069
Max	0.019	0.006	0.080
Year	1920		
$v_{i,t-1}$			
Mean	0.107	0.022	0.015
Median	0.077	0.011	0.011
Max	0.294	0.115	0.096
$P_{i,t-1}$			
Mean	0.084	0.026	0.028
Median	0.084	0.026	0.028
Max	0.169	0.052	0.056
P_{it}			
Mean	0.126	0.000	0.000
Median	0.126	0.000	0.000
Max	0.252	0.000	0.000
$I_{i,t-1}$			
Mean	0.095	0.049	0.051
Median	0.107	0.029	0.027
Max	0.178	0.117	0.128
Year	1924		
$v_{i,t-1}$			
Mean	0.069	0.024	0.025
Median	0.045	0.020	0.023
Max	0.209	0.074	0.066
$P_{i,t-1}$			
Mean	0.052	0.018	0.008
Median	0.052	0.018	0.008
Max	0.103	0.037	0.017
P_{it}			
Mean	0.100	0.013	0.013
Median	0.100	0.013	0.013
Max	0.200	0.027	0.026
$I_{i,t-1}$			
Mean	0.076	0.023	0.042
Median	0.097	0.017	0.043
Max	0.132	0.054	0.083
Year	1926		
$v_{i,t-1}$			
Mean	0.039	0.037	0.036
Median	0.035	0.030	0.023
Max	0.113	0.112	0.119

$P_{i,t-1}$			
Mean	0.035	0.026	0.044
Median	0.035	0.026	0.044
Max	0.070	0.053	0.087
P_{it}			
Mean	0.005	0.025	0.039
Median	0.005	0.025	0.039
Max	0.010	0.050	0.077
$I_{i,t-1}$			
Mean	0.082	0.070	0.083
Median	0.056	0.073	0.074
Max	0.190	0.139	0.174
Year	1928		
$v_{i,t-1}$			
Mean	0.037	0.028	0.033
Median	0.030	0.025	0.032
Max	0.116	0.076	0.095
$P_{i,t-1}$			
Mean	0.009	0.000	0.009
Median	0.009	0.000	0.009
Max	0.018	0.000	0.019
P_{it}			
Mean	0.016	0.003	0.013
Median	0.016	0.003	0.013
Max	0.033	0.006	0.025
$I_{i,t-1}$			
Mean	0.069	0.045	0.052
Median	0.092	0.050	0.032
Max	0.115	0.085	0.123
Year	1930		
$v_{i,t-1}$			
Mean	0.048	0.020	0.018
Median	0.034	0.017	0.013
Max	0.169	0.080	0.070
$P_{i,t-1}$			
Mean	0.043	0.003	0.010
Median	0.043	0.003	0.010
Max	0.085	0.007	0.020
P_{it}			
Mean	0.047	0.000	0.000
Median	0.047	0.000	0.000
Max	0.093	0.000	0.000
Mean	0.056	0.019	0.021
Median	0.076	0.027	0.030
Max	0.092	0.030	0.033
Year	1934		
$v_{i,t-1}$			
Mean	0.128	0.046	0.036
Median	0.128	0.038	0.030
Max	0.232	0.138	0.114
$P_{i,t-1}$			
Mean	0.098	0.004	0.008
Median	0.098	0.004	0.008

Max	0.196	0.008	0.015
P_{it}			
Mean	0.037	0.021	0.021
Median	0.037	0.021	0.021
Max	0.074	0.042	0.042
$I_{i,t-1}$			
Mean	0.123	0.018	0.024
Median	0.141	0.023	0.019
Max	0.229	0.031	0.053
Year	1936		
$v_{i,t-1}$			
Mean	0.111	0.026	0.015
Median	0.114	0.020	0.013
Max	0.229	0.091	0.054
$P_{i,t-1}$			
Mean	0.079	0.037	0.000
Median	0.079	0.037	0.000
Max	0.158	0.074	0.000
P_{it}			
Mean	0.084	0.007	0.000
Median	0.084	0.007	0.000
Max	0.168	0.013	0.000
$I_{i,t-1}$			
Mean	0.110	0.027	0.040
Median	0.164	0.034	0.032
Max	0.165	0.047	0.088
Year	1938		
$v_{i,t-1}$			
Mean	0.031	0.032	0.022
Median	0.024	0.025	0.016
Max	0.106	0.115	0.073
$P_{i,t-1}$			
Mean	0.023	0.001	0.000
Median	0.023	0.001	0.000
Max	0.046	0.002	0.000
P_{it}			
Mean	0.006	0.005	0.000
Median	0.006	0.005	0.000
Max	0.013	0.011	0.000
$I_{i,t-1}$			
Mean	0.068	0.001	0.048
Median	0.083	0.000	0.039
Max	0.120	0.002	0.104
Year	1940		
$v_{i,t-1}$			
Mean	0.038	0.021	0.012
Median	0.034	0.018	0.008
Max	0.119	0.090	0.051
$P_{i,t-1}$			
Mean	0.025	0.005	0.000
Median	0.025	0.005	0.000
Max	0.050	0.009	0.000
P_{it}			

Mean	0.023	0.003	0.000
Median	0.023	0.003	0.000
Max	0.046	0.007	0.000
$I_{i,t-1}$			
Mean	0.007	0.006	0.010
Median	0.005	0.009	0.011
Max	0.017	0.009	0.020
Year	1944		
$v_{i,t-1}$			
Mean	0.115	0.025	0.018
Median	0.101	0.018	0.011
Max	0.281	0.112	0.103
$P_{i,t-1}$			
Mean	0.031	0.000	0.000
Median	0.031	0.000	0.000
Max	0.062	0.000	0.000
P_{it}			
Mean	0.122	0.000	0.000
Median	0.122	0.000	0.000
Max	0.244	0.000	0.000
$I_{i,t-1}$			
Mean	0.047	0.003	0.045
Median	0.022	0.003	0.000
Max	0.120	0.008	0.135
Year	1946		
$v_{i,t-1}$			
Mean	0.100	0.025	0.015
Median	0.101	0.021	0.013
Max	0.195	0.097	0.051
$P_{i,t-1}$			
Mean	0.066	0.009	0.000
Median	0.066	0.009	0.000
Max	0.132	0.018	0.000
P_{it}			
Mean	0.051	0.015	0.009
Median	0.051	0.015	0.009
Max	0.102	0.029	0.017
$I_{i,t-1}$			
Mean	0.084	0.010	0.026
Median	0.064	0.012	0.006
Max	0.189	0.018	0.071
Year	1948		
$v_{i,t-1}$			
Mean	0.094	0.053	0.026
Median	0.099	0.056	0.023
Max	0.219	0.119	0.088
$P_{i,t-1}$			
Mean	0.020	0.038	0.018
Median	0.020	0.038	0.018
Max	0.039	0.076	0.036
P_{it}			
Mean	0.042	0.050	0.018
Median	0.042	0.050	0.018

Max	0.084	0.099	0.036
$I_{i,t-1}$			
Mean	0.039	0.026	0.053
Median	0.016	0.003	0.071
Max	0.101	0.076	0.088
Year	1950		
$v_{i,t-1}$			
Mean	0.085	0.033	0.057
Median	0.078	0.026	0.053
Max	0.190	0.129	0.167
$P_{i,t-1}$			
Mean	0.018	0.028	0.048
Median	0.018	0.028	0.048
Max	0.036	0.056	0.097
P_{it}			
Mean	0.066	0.015	0.050
Median	0.066	0.015	0.050
Max	0.133	0.029	0.100
$I_{i,t-1}$			
Mean	0.046	0.020	0.040
Median	0.063	0.023	0.059
Max	0.076	0.035	0.062
Year	1954		
$v_{i,t-1}$			
Mean	0.036	0.036	0.074
Median	0.033	0.028	0.083
Max	0.104	0.147	0.158
$P_{i,t-1}$			
Mean	0.009	0.003	0.040
Median	0.009	0.003	0.040
Max	0.018	0.007	0.079
P_{it}			
Mean	0.008	0.013	0.021
Median	0.008	0.013	0.021
Max	0.016	0.026	0.043
$I_{i,t-1}$			
Mean	0.035	0.005	0.031
Median	0.045	0.007	0.017
Max	0.059	0.010	0.076
Year	1956		
$v_{i,t-1}$			
Mean	0.077	0.055	0.053
Median	0.067	0.056	0.055
Max	0.200	0.157	0.143
$P_{i,t-1}$			
Mean	0.032	0.049	0.053
Median	0.032	0.049	0.053
Max	0.063	0.099	0.106
P_{it}			
Mean	0.002	0.022	0.026
Median	0.002	0.022	0.026
Max	0.004	0.043	0.052
$I_{i,t-1}$			

Mean	0.035	0.063	0.062
Median	0.022	0.071	0.088
Max	0.083	0.117	0.097
Year	1958		
$v_{i,t-1}$			
Mean	0.076	0.030	0.023
Median	0.070	0.021	0.021
Max	0.233	0.115	0.066
$P_{i,t-1}$			
Mean	0.105	0.000	0.001
Median	0.105	0.000	0.001
Max	0.210	0.000	0.003
P_{it}			
Mean	0.096	0.018	0.010
Median	0.096	0.018	0.010
Max	0.192	0.036	0.021
$I_{i,t-1}$			
Mean	0.114	0.013	0.019
Median	0.166	0.003	0.018
Max	0.178	0.036	0.039
Year	1960		
$v_{i,t-1}$			
Mean	0.077	0.033	0.029
Median	0.078	0.024	0.021
Max	0.162	0.149	0.138
$P_{i,t-1}$			
Mean	0.014	0.015	0.009
Median	0.014	0.015	0.009
Max	0.027	0.030	0.018
P_{it}			
Mean	0.079	0.024	0.018
Median	0.079	0.024	0.018
Max	0.159	0.049	0.036
$I_{i,t-1}$			
Mean	0.037	0.034	0.035
Median	0.026	0.030	0.021
Max	0.086	0.070	0.084
Year	1964		
$v_{i,t-1}$			
Mean	0.067	0.073	0.092
Median	0.060	0.078	0.092
Max	0.185	0.169	0.211
$P_{i,t-1}$			
Mean	0.028	0.019	0.048
Median	0.028	0.019	0.048
Max	0.055	0.039	0.095
P_{it}			
Mean	0.039	0.044	0.072
Median	0.039	0.044	0.072
Max	0.079	0.088	0.145
$I_{i,t-1}$			
Mean	0.012	0.032	0.093
Median	0.003	0.006	0.125

Max	0.033	0.091	0.155
Year	1966		
$v_{i,t-1}$			
Mean	0.040	0.054	0.056
Median	0.035	0.046	0.051
Max	0.110	0.189	0.179
$P_{i,t-1}$			
Mean	0.068	0.041	0.038
Median	0.068	0.041	0.038
Max	0.136	0.082	0.077
P_{it}			
Mean	0.009	0.051	0.038
Median	0.009	0.051	0.038
Max	0.018	0.102	0.077
$I_{i,t-1}$			
Mean	0.056	0.056	0.049
Median	0.082	0.036	0.031
Max	0.087	0.133	0.117
Year	1968		
$v_{i,t-1}$			
Mean	0.110	0.143	0.088
Median	0.082	0.151	0.077
Max	0.350	0.297	0.276
$P_{i,t-1}$			
Mean	0.034	0.109	0.092
Median	0.034	0.109	0.092
Max	0.068	0.219	0.184
P_{it}			
Mean	0.001	0.125	0.059
Median	0.001	0.125	0.059
Max	0.002	0.250	0.117
$I_{i,t-1}$			
Mean	0.077	0.104	0.054
Median	0.106	0.130	0.010
Max	0.126	0.182	0.153
Year	1970		
$v_{i,t-1}$			
Mean	0.071	0.032	0.030
Median	0.065	0.026	0.023
Max	0.173	0.151	0.116
$P_{i,t-1}$			
Mean	0.012	0.000	0.011
Median	0.012	0.000	0.011
Max	0.024	0.000	0.023
P_{it}			
Mean	0.040	0.005	0.003
Median	0.040	0.005	0.003
Max	0.080	0.010	0.006
$I_{i,t-1}$			
Mean	0.007	0.013	0.025
Median	0.009	0.003	0.016
Max	0.013	0.035	0.058
Year	1974		

$v_{i,t-1}$			
Mean	0.031	0.045	0.048
Median	0.028	0.044	0.041
Max	0.109	0.126	0.137
$P_{i,t-1}$			
Mean	0.028	0.005	0.036
Median	0.028	0.005	0.036
Max	0.056	0.010	0.072
P_{it}			
Mean	0.004	0.036	0.014
Median	0.004	0.036	0.014
Max	0.007	0.072	0.027
$I_{i,t-1}$			
Mean	0.039	0.050	0.052
Median	0.004	0.061	0.010
Max	0.113	0.089	0.147
Year	1976		
$v_{i,t-1}$			
Mean	0.041	0.041	0.036
Median	0.025	0.028	0.031
Max	0.158	0.131	0.128
$P_{i,t-1}$			
Mean	0.040	0.006	0.004
Median	0.040	0.006	0.004
Max	0.079	0.011	0.008
P_{it}			
Mean	0.006	0.030	0.028
Median	0.006	0.030	0.028
Max	0.011	0.060	0.056
$I_{i,t-1}$			
Mean	0.055	0.051	0.058
Median	0.003	0.063	0.067
Max	0.162	0.091	0.108
Year	1978		
$v_{i,t-1}$			
Mean	0.089	0.033	0.026
Median	0.092	0.032	0.017
Max	0.164	0.095	0.089
$P_{i,t-1}$			
Mean	0.032	0.004	0.014
Median	0.032	0.004	0.014
Max	0.065	0.009	0.029
P_{it}			
Mean	0.032	0.006	0.023
Median	0.032	0.006	0.023
Max	0.065	0.011	0.046
$I_{i,t-1}$			
Mean	0.038	0.016	0.011
Median	0.022	0.023	0.000
Max	0.091	0.026	0.034
Year	1980		
$v_{i,t-1}$			
Mean	0.059	0.026	0.015

Median	0.057	0.022	0.011
Max	0.164	0.084	0.055
$P_{i,t-1}$			
Mean	0.069	0.004	0.000
Median	0.069	0.004	0.000
Max	0.137	0.008	0.000
P_{it}			
Mean	0.030	0.027	0.021
Median	0.030	0.027	0.021
Max	0.061	0.054	0.042
$I_{i,t-1}$			
Mean	0.096	0.043	0.039
Median	0.086	0.059	0.054
Max	0.201	0.070	0.063
Year	1984		
$v_{i,t-1}$			
Mean	0.068	0.025	0.024
Median	0.047	0.020	0.020
Max	0.210	0.081	0.086
$P_{i,t-1}$			
Mean	0.074	0.006	0.000
Median	0.074	0.006	0.000
Max	0.148	0.013	0.000
P_{it}			
Mean	0.040	0.013	0.013
Median	0.040	0.013	0.013
Max	0.080	0.026	0.026
$I_{i,t-1}$			
Mean	0.111	0.010	0.017
Median	0.129	0.009	0.006
Max	0.204	0.022	0.044
Year	1986		
$v_{i,t-1}$			
Mean	0.101	0.022	0.020
Median	0.115	0.016	0.013
Max	0.248	0.089	0.084
$P_{i,t-1}$			
Mean	0.109	0.000	0.000
Median	0.109	0.000	0.000
Max	0.218	0.000	0.000
P_{it}			
Mean	0.077	0.030	0.028
Median	0.077	0.030	0.028
Max	0.155	0.060	0.055
$I_{i,t-1}$			
Mean	0.119	0.019	0.015
Median	0.107	0.006	0.019
Max	0.250	0.051	0.026
Year	1988		
$v_{i,t-1}$			
Mean	0.097	0.054	0.053
Median	0.069	0.028	0.025
Max	0.283	0.271	0.275

$P_{i,t-1}$			
Mean	0.099	0.011	0.009
Median	0.099	0.011	0.009
Max	0.198	0.022	0.019
P_{it}			
Mean	0.113	0.000	0.000
Median	0.113	0.000	0.000
Max	0.226	0.000	0.000
$I_{i,t-1}$			
Mean	0.104	0.022	0.032
Median	0.153	0.006	0.035
Max	0.160	0.059	0.060
Year	1990		
$v_{i,t-1}$			
Mean	0.139	0.033	0.021
Median	0.141	0.025	0.014
Max	0.298	0.103	0.133
$P_{i,t-1}$			
Mean	0.113	0.019	0.006
Median	0.113	0.019	0.006
Max	0.227	0.038	0.011
P_{it}			
Mean	0.121	0.011	0.000
Median	0.121	0.011	0.000
Max	0.242	0.022	0.000
$I_{i,t-1}$			
Mean	0.157	0.020	0.020
Median	0.201	0.000	0.023
Max	0.272	0.060	0.037
Year	1994		
$v_{i,t-1}$			
Mean	0.083	0.065	0.043
Median	0.065	0.068	0.041
Max	0.210	0.147	0.105
$P_{i,t-1}$			
Mean	0.022	0.068	0.055
Median	0.022	0.068	0.055
Max	0.043	0.136	0.111
P_{it}			
Mean	0.016	0.055	0.029
Median	0.016	0.055	0.029
Max	0.033	0.109	0.058
$I_{i,t-1}$			
Mean	0.054	0.056	0.043
Median	0.021	0.000	0.044
Max	0.141	0.168	0.085
Year	1996		
$v_{i,t-1}$			
Mean	0.032	0.036	0.026
Median	0.024	0.030	0.019
Max	0.128	0.141	0.132
$P_{i,t-1}$			
Mean	0.009	0.006	0.000

Median	0.009	0.006	0.000
Max	0.018	0.011	0.000
P_{it}			
Mean	0.049	0.011	0.005
Median	0.049	0.011	0.005
Max	0.097	0.022	0.011
$I_{i,t-1}$			
Mean	0.036	0.016	0.037
Median	0.053	0.019	0.027
Max	0.055	0.028	0.085
Year	1998		
$v_{i,t-1}$			
Mean	0.034	0.027	0.057
Median	0.030	0.024	0.060
Max	0.103	0.096	0.129
$P_{i,t-1}$			
Mean	0.025	0.028	0.032
Median	0.025	0.028	0.032
Max	0.051	0.057	0.063
P_{it}			
Mean	0.004	0.045	0.062
Median	0.004	0.045	0.062
Max	0.008	0.090	0.124
$I_{i,t-1}$			
Mean	0.051	0.026	0.033
Median	0.026	0.012	0.011
Max	0.128	0.066	0.086
Year	2000		
$v_{i,t-1}$			
Mean	0.091	0.043	0.032
Median	0.079	0.043	0.026
Max	0.272	0.112	0.114
$P_{i,t-1}$			
Mean	0.103	0.003	0.011
Median	0.103	0.003	0.011
Max	0.207	0.007	0.023
P_{it}			
Mean	0.111	0.002	0.000
Median	0.111	0.002	0.000
Max	0.221	0.003	0.000
$I_{i,t-1}$			
Mean	0.139	0.035	0.036
Median	0.152	0.046	0.036
Max	0.265	0.059	0.072

Table B.1: Differences in QQPlots Before and After Matching, Incumbency Advantage

Appendix C

Tables for Chapter 4

Education:	Some HS	HS Grad	Some College	College Grad	Total
Male	4.7	13.4	13.2	14.6	45.9
Female	5.4	16.9	17.0	14.8	54.1
Total	10.05	30.25	30.25	29.45	100

Age:	18–24	25–40	41–64	65 and older	Total
Male	3.8	13.2	20.8	8.1	45.9
Female	4.8	15.7	23.3	10.3	54.1
Total	8.6	28.9	44.1	18.4	100

Race:	White	Non-White	Total
Male	38.5	7.5	46.0
Female	43.7	10.3	54.0
Total	82.2	17.8	100

Income:	Less than $25,000	$25,000–$49,999	$50,000–$74,999	More than $75,000	Total
Male	8.9	13.9	10.4	12.8	46.0
Female	14.2	16.2	10.9	12.7	54.0
Total	23.1	30.1	21.3	25.5	100

These percentages are based on the 49,897 individuals with complete responses to the subset of questions of interest that were asked on the CPS.

Table C.1: Contingency Tables for Selected Characteristics, 2000 CPS

Education:	Some HS	HS Grad	Some College	College Grad	Total
Male	4.1	13.4	13.5	15.3	46.3
Female	4.7	15.9	17.4	15.7	53.7
Total	8.8	29.3	30.9	31.0	100

Age:	18–24	25–40	41–64	65 and older	Total
Male	4.3	12.1	22.0	7.9	46.3
Female	4.9	14.3	24.5	10.0	53.7
Total	9.2	26.4	46.5	17.9	100

Race:	White	Non-White	Total
Male	38.6	7.8	46.4
Female	43.3	10.3	53.6
Total	81.9	18.1	100

Income:	Less than \$25,000	\$25,000–\$49,999	\$50,000–\$74,999	More than \$75,000	Total
Male	7.9	12.7	10.8	14.9	46.3
Female	12.6	14.8	11.4	14.9	53.7
Total	20.5	27.5	22.2	29.8	100

These percentages are based on the 60,431 individuals with complete responses to the subset of questions of interest that were asked on the CPS.

Table C.2: Contingency Tables for Selected Characteristics, 2004 CPS

	Pooled	2000	2004
South	-0.19	-0.14	-0.18
	(0.008)	(0.008)	(0.01)
Gender	0.11	0.09	0.13
	(0.0004)	(0.0008)	(0.0008)
Non-White	-0.08	-0.18	0.008
	(0.006)	(0.01)	(0.01)
Education	0.07	0.46	0.62
	(0.02)	(0.002)	(0.003)
Education^2	0.08	0.06	0.11
	(0.0009)	(0.0002)	(0.002)
Age	0.03	0.02	0.02
	(0.00007)	(0.000005)	(0.000005)
Age^2	-0.0001	-0.0002	-0.00001
	(0.000000007)	(0.00000001)	(0.00000002)
Income	0.09	0.09	0.09
	(0.000008)	(0.00002)	(0.00002)
Voter ID	-0.36	0.03	-0.014
	(0.008)	(0.002)	(0.002)
Education * Voter ID	0.12	0.005	-0.02
	(0.003)	(0.0003)	(0.0003)
Education^2 * Voter ID	-0.02	-0.02	-0.03
	(0.0001)	(0.0002)	(0.0002)
Age * Voter ID	0.008	0.002	0.001
	(0.000007)	(0.0000006)	(0.0000005)
Age^2 * Voter ID	0.00008	-0.00009	-0.00008
	(0.0000000008)	(0.000000002)	(0.000000002)
Non-White * Voter ID	0.06	0.11	0.02
	(0.0006)	(0.001)	(0.001)
Constant	-1.16	2.09	2.45
	(0.08)	(0.01)	(0.01)
Number of Observations	110,328	49,897	60431

Table C.3: Logit Coefficients for Model of Voter Turnout as a Function of Voter Identification Regime
Standard deviations are in parentheses.
Note that the *education*, *age* and *income* variables
are centered at their means in the 2000 and 2004 regressions.

Appendix D

References

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido W. Imbens. 2001. "Matching Estimators for Average Treatment Effects in Stata." *The Stata Journal* 1:1–18.

Abadie, Alberto, and Guido Imbens. 2004. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Harvard University, University of California, Berkeley and NBER.* (Unpublished manuscript)

Alford, J.R., and J.R. Hibbing. 1981. "Increased Incumbency Advantage in the House." *Journal of Politics* 43:1042–1061.

Alvarez, R. Michael, Delia Bailey, and Jonathan N. Katz. 2007a. "The Impact of Voter Identification Laws on Voter Turnout." *California Institute of Technology.* (Unpublished manuscript)

Alvarez, R. Michael, Thad E. Hall, and Betsy Sinclair. 2007b. "Katrina's Voters: Floods and Representation." *California Institute of Technology* and *University of Utah.* (Unpublished manuscript)

Ansolabehere, Stephen, James M. Snyder, and Charles Stewart III. 2000. "Old Voters, New Voters and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage." *American Journal of Political Science* 44:17–34.

Ansolabehere, Stephen, and James M. Snyder. 2002. "Using Term Limits to Estimate Incumbency Advantage When Officeholders Retire Strategically." *Massachusetts Institute of Technology.* (Unpublished manuscript)

Ansolabehere, Stephen, and Charles Stewart III. 2005. "Residual Votes Attributable to Technology." *Journal of Politics* 67(2):365–389.

Brace, Kimball W., and Michael P. McDonald. 2005. "Final Report of the 2004 Election Day Survey." *EAC Survey Analysis Support (EAC 0524).* `http://www.eac.`

gov/election_survey_2004/pdf/EDS-Full_Report_wTables.pdf.

Caltech/MIT Voting Technology Project. 2001. *Voting—What Is, What Could Be.* Pasadena and Cambridge: Caltech and MIT.

Center for American Politics and Citizenship. 2003. "Characteristics of Contemporary Voting Machines." (Accessed 15 May 2005.) http://www.capc.umd.edu/rpts/md_evote_ContempVotingMach.pdf.

Cochran, W.G. 1965. "The Planning of Observational Studies of Human Populations (with discussion)." *Journal of the Royal Statistical Society* Series A, 128:134–155.

Cox, Gary W., and Jonathan N. Katz. 1996. "Why Did the Incumbency Advantage Grow?" *American Journal of Political Science* 40:478–497.

Deheija, Rajeev H., and Sadek Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84(1):151–161.

Duflo, Esther. "Empirical Methods." 2002. *MIT* and *Harvard University.* (Mimeo)

Erikson, R.S. 1971. "The Advantage of Incumbency in Congressional Elections." *Polity* 3:395–405.

Fox, John. 1997. *Applied Regression Analysis, Linear Models and Related Methods.* Thousand Oaks, CA: Sage Publications.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis.* 2nd ed. New York: Chapman and Hall.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Gelman, Andrew, and Zaiying Huang. 2007. "Estimating Incumbency Advantage and Its Variation, as an Example of a Before-After Study." *Journal of the American Statistical Association* (forthcoming)

Gelman, Andrew, and Gary King. 1990. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science* 34:1142–1164.

Green, Donald P., and Alan S. Gerber. 2002. "Reclaiming the Experimental Tradition in Political Science." In Helen V. Milner and Ira Katznelson, Eds., *Political Science: The State of the Discipline,* 3rd ed. New York: W.W. Norton & Co., 805–832.

Greene, William H. 2003. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Halvorsen, Robert, and Raymond Palmquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review* 70(3):474–475.

Hansen, Ben B. 2006. "Optmatch: Flexible, Optimal Matching for Observational Studies." Package for the statistical software *R*.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64(4):605–654.

Highton, Benjamin. 2004. "Voter Registration and Turnout in the United States." *Perspectives in Politics* 2(3):507–515.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stewart. 2006. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Harvard University* and *Princeton University*. (Unpublished manuscript)

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stewart. 2006. "Matchit: Nonparametric Preprocessing for Parametric Causal Inference." Package for the statistical software *R*.

Imai, Kosuke, and David A. Van Dyk. 2004. "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association* 99(467):854–866.

Imbens, Guido. 2000. "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika* 87(3):706–710.

Kennedy, Peter E. 1981. "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations." *American Economic Review* 71(4):801.

Kimball, David C., Chris T. Owens, and Katherine M. Keeney. 2004. "Unrecorded Votes and Political Representation." In Robert P. Watson, Ed., *Counting Votes: Lessons from the 2000 Presidential Election in Florida.* Gainesville: University Press of Florida, 135–150.

Knack, Stephen, and Martha Kropf. 2003. "Voided Ballots in the 1996 Presidential Election: A County-Level Analysis." *Journal of Politics* 65(3):881–898.

Leip, David. 2004. *Atlas of U.S. Presidential Elections.* http://www.uselectionatlas.org

Nagler, Jonathan. 1991. "The Effect of Voter Registration Laws and Education on U.S. Voter Turnout." *American Political Science Review* 85:1393–1405.

Pinheiro, Josè C., and Douglas M. Bates. *Mixed-Effects Models in S and S-Plus.* New York: Springer.

Plummer, Martyn. 2007. "JAGS: Just another Gibbs sampler." (Version 0.98) *International Agency for Research on Cancer.* http://www-fis.iarc.fr/~martyn/software/jags/.

R Development Core Team. 2006. "R: A language and environment for statistical computing." *R Foundation for Statistical Computing*, Vienna, Austria. http://www.R-project.org.

Rosenbaum, Paul R. 2002. *Observational Studies*, 2nd ed. New York: Springer.

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.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *American Statistician* 39(1):33–38.

Sekhon, Jasjeet S. 2004. "The Varying Role of Voter Information Across Societies." *Harvard University.* (Unpublished manuscript)

Sekhon, Jasjeet S. 2006. "Matching: Algorithms and Software for Multivariate and Propensity Score Matching with Balance Optimization via Genetic Search." Package for the statistical software *R.*

Slottje, Daniel J., Daniel L. Millimet, and Michael J. Buchanan. 2003. "Copyright Impacts on Relative Firm Sales: Evidence from the National Football League." *Southern Methodist University*, and *KPMG, LLP.* (Unpublished manuscript)

Smith, Jeffrey A., and Petra E. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125:305–353.

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.

van Garderen, Jan, and Chandra Shat. 2002. "Exact Interpretation of Dummy Variables in Semilogarithmic Equations." *Econometrics Journal* 5:149–159.

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron, and Henry E. Brady. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95(4):793–810.