AT YOUR CONVENIENCE: FACILITATING VOTING AND REGISTRATION

Thesis by

Allyson L. Pellissier

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 11, 2015)

To my family and friends, whose spiritual, intellectual, and emotional support have enriched my life beyond description.

And to the memory of my wonderful Aunt Carol.

Acknowledgements

I am eternally grateful for the advice and feedback of my Academic Advisor and chair of my committee, Mike Alvarez. I am enormously indebted to his commitment to scholarship and mentoring, and I owe a great deal of my progress to his feedback and wisdom. I also thank the other members of my committee, Rod Kiewiet, Jonathan Katz, and Robert Sherman, for their comments, professional guidance, and insight. Laurel Auchampaugh has been a wonderful source of support, and I am grateful for all the time and energy she has dedicated to graduate students. Additionally, I have benefited from the comments I have received in the Proseminar sequence from Federico Echinique and Jean-Laurent Rosenthal, and I am incredibly fortunate to have been surrounded by such wonderful colleagues, including Andi Bui, Jonathan Chapman, Matt Chao, Khai Chiong, Liam Clegg, Geoff Fischer, Chris Hagel, Jackie Kimble, Matt Kovach, Ines Levin, Samantha Myers, Lucas Nunez, Kirill Pogorelskiy, Welmar Rosado Buenfil, Tom Ruchti, Euncheol Shin, and Andy Sinclair.

And of course, I could not have fulfilled any of this program's requirements without the unwavering support of my family and friends. My dad, mom, and brother have offered love and encouragement beyond what I could describe, and I never forget how lucky I am to have them in my life. This gratitude also extends to my grandparents and the rest of my extended family, all of whom continually remind me that home is never farther than a phone call. I am likewise grateful beyond words for the many friends who have enriched my life, especially Nina, Kristen, Rachel, Mollie, and Kori. Additionally, the spiritual community I found at St. Philip has challenged and inspired me to seek renewal and progress every single day —thank you all. And finally, and most importantly, I would like to thank God for blessing me with faith,

hope, and love far beyond what I deserve.

Abstract

This work proposes answers to methodological and substantive questions related to convenience voting. The first analytical chapter surveys the various research designs that have been proposed within this literature, and concludes that the field benefits from using all in conjunction. The next chapter uses matching to identify the relationship between disability status and political participation, and considers whether any forms of convenience voting mediate in the relationship. The final two analytical chapters examine how online voter registration, one of the most recent policy innovations, affects participation and vote share in American elections. The concluding chapter summarizes the findings presented herein, and briefly discusses the natural extensions of this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Participation rates in American elections have remained nearly constant over the past several decades, and turnout in the United States perennially lags behind that of other contemporary democracies (Hanmer 2007, 2009). In response, the academic community has dedicated considerable attention to the "turnout puzzle": This lack of participation has persisted *despite* numerous efforts to make the voting process more convenient (first expressed in Brody 1978; see also Leighley and Nagler 1992). Yet according to the "rational voter" theory (formalized in Downs 1957; expanded in Riker and Ordeshook 1968), a reduction in cost should yield an increased propensity to cast a ballot.

The convenience voting literature, pioneered by Rosenstone and Wolfinger (1978), attempts to characterize how the voting-eligible respond to election policies, and how these liberalizations affect the representativeness of the electorate. Yet despite the vast amount of scholarship in this area, there are still important methodological and substantive questions that the peer-reviewed literature has neglected.

The second chapter of this manuscript is a methodological inquiry into the appropriate modeling strategy for this stream of research. Although the majority of scholarship emulates the Rosenstone-Wolfinger design (Ansolabehere and Konisky 2006), a few detractors have criticized Rosenstone and Wolfinger's model as untenable and offered various alternatives (Glynn and Quinn 2011; Keele and Minozzi 2013). More specifically, these detractors highlight the tenuousness of the exogenous selection assumption embedded within the binary response model that Rosenstone and Wolfinger (and their successors) employ. A few projects propose non-parametric bounds as a complement to (Hanmer 2007, 2009) or outright substitute for (Glynn and Quinn 2011) parametric point identification, while others rely on natural experiments for inference (Ansolabehere and Konisky 2006; Neiheisel and Burden 2012; Keele and Minozzi 2013). After reviewing these papers, I respond to the prominent critique of Glynn and Quinn (2011), and I offer evidence that the parametric estimates they obtain in their paper serve as little more than a straw man. I argue that the bounding method serves as a useful "sanity check," as argued in Hanmer (2007, 2009), but that the current body of literature has yet to dismantle the more traditional point estimation using observational data.

I subsequently employ this research design to examine a specific set of questions regarding how individuals are responding to election policy: How are individuals with disabilities responding to convenience voting procedures (Chapter 3)? How is online voter registration, one of the most recent innovations in electoral policy impacting individual and aggregate participation (Chapter 4)? Does online registration shift vote share for either one of the major parties (Chapter 5)? Below, I detail more thoroughly the purpose and scope of each chapter.

The third chapter considers a demographic subpopulation that has been underrepresented at the polls *and* in the literature: individuals with disabilities, who comprise one of the most sizeable political minorities and participate at systematically low rates (Schur and Adya 2013). One paper calls mobilization of this group "the last suffrage movement" (Schriner, Ochs, and Shields 1997). Another notes the irony that disability rights advocates were influential in shaping recent legislation, but the literature has overlooked how the changing electoral landscape has affected this demographic group (Stewart 2011). Among the small collection of papers, there is considerable disagreement about the relationship between disability and turnout; the reduction in turnout associated with disability status has been estimated to be as high as 21% and as low as 4% (Schur, Shields, and Shriner 2005; Schur and Kruse 2011). My study is the first in this area to utilize matching to reduce the role of potential confounders, such as age; I then estimate that disability status reduces the propensity to be registered by 3 to 5 percentage points, and the propensity to vote by 6 to 8 percentage points. Incorporating interactive effects, I discover that most of this relationship can be explained by employment status, supporting earlier findings by Schur and Kruse (2000) and Schur et al. (2002). And finally, although individuals with disabilities may take advantage of convenience voting procedures, these policies do not improve the representativeness of this demographic group in the overall electorate.

While the third chapter focuses on a particular demographic group but a broad array of policy, the fourth chapter spotlights how a particular policy affects various demographic groups. More explicitly, I consider how online voter registration, one of the newest forms of convenience voting, fits into the participation puzzle; the peer-reviewed literature is completely silent about its efficacy, although a few papers published elsewhere offer some insight. This omission is a glaring one, given the increasing popularity of online registration. In the fourth chapter, I estimate an individual-level model of participation (registration, turnout) as a function of demographics, state-level electoral characteristics (competitiveness and policy), and online registration. I find that this new registration alternative does stimulate participation at both stages of participation, though the magnitude of the impact is minimal. Moreover, the effect is statistically significant for the youngest age bracket of the voting eligible, as well as recent movers; the impact on ethnic minorities, meanwhile, is not statistically distinguishable from zero. In other words, online registration might render the electorate more representative in terms of age and residential mobility, but not race. Finally, the estimated influx of voters exceeded the margin of victory in the 2008 and 2012 Presidential contests for a handful of states. Unfortunately, because the *Current Population Survey* dataset that I (and most of the papers in this literature) use does not include information about political beliefs and preferences, it is difficult to determine whether the *outcome* would have changed in any of these states from this data alone.

Chapter 4 motivated the final analytical chapter in this manuscript, which considers the implications for convenience voting on vote share among presidential candidates. Nearly all of the papers in this literature have simply speculated how these policies might shape the partisan composition of the electorate, as a results of the data limitations of the *CPS* mentioned in the previous paragraph. I introduce a new method to recover these missing data, drawing on insight from the machine learning literature and incorporating another large nationally-representative dataset (the *Cooperative Congressional Election Study*). For every state, I use a Random Forest approach (introduced in Breiman 2001) to discover the relationship between the most-preferred presidential candidate and demographic profile. Random Forest is an ensemble method that is computationally efficient, robust to overfitting, and as accurate as other prominent machine learning techniques (Breiman and Cutler n.d.). I use this fitted model to estimate the vote share in each state, with and without online registration. I conclude from the current data that online registration will not have significant bearing on electoral outcomes. My method is readily applicable to other forms of convenience voting. More generally, because I offer a blueprint for merging information across surveys, the procedure broadens the set of hypotheses that can be tested using observational data.

Chapter 2

Modeling Convenience

2.1 Introduction

The convenience voting literature dates back several decades, but the current body of work remains ambivalent about the extent that convenience voting impacts the electorate, as well as the appropriate research design for this line of inquiry. The majority of the papers in this area adopt the binary response model used in Rosenstone and Wolfinger (1978), but recent papers have criticized the underlying assumptions as untenable. Given that the next several chapters of my thesis discuss how individuals respond to election policy, I preface them with this brief methodological overview of the literature. I introduce the binary response model in Rosenstone and Wolfinger, as well as a few prominent alternatives.

Although the critics of the binary response model articulate their discomfort about the identifying assumptions elegantly, most of the empirical results suggest that the model is less sensitive if the researcher includes variables that capture state-level heterogeneity. In particular, by pooling cross sections, incorporating fixed effects, and ensuring that pre-treatment data is included for the covariate of interest, much of the selection bias is mitigated. I first discuss the binary response model, and establish the criticisms levied against it. I then discuss the most noteworthy alternatives, natural experiments (which use a within-state design) and nonparametric bounds (which relax most assumptions). In the penultimate section, I replicate the analysis of Glynn and Quinn (2011), and contend that their parametric results function as a "straw man" arising from poor specification, rather than a genuine example of improper inference. I conclude that the Rosenstone-Wolfinger approach remains a viable estimation strategy, though I endorse a broader array of approaches as a valuable insight into the nuances of implementation.

2.2 Binary Response Model

The Voting Rights Act of 1965 dismantled several election practices that were widely considered to be unduly prohibitive (such as poll taxes and literacy tests). A reduction in the costs associated with voter participation should be accompanied by an increase in turnout (Downs 1957). In the wake of this effort to engage the voting-eligible population, Rosenstone and Wolfinger (1978) published a seminal investigation of how institutional features impact the turnout calculus and the composition of the electorate; they developed this framework further in Wolfinger and Rosenstone (1980). They concluded that certain institutional features significantly affect the probability of turnout, including policies related to registration —for example, closing date and office hours. And perhaps even more importantly, they established an initial statistical framework for estimating the impact of electoral policy on behavior.

To draw inference about how individuals interact with their electoral environment, we consider how different conditions alter an individual's expected *propensity* to participate. In data, however, this variable is unobserved, and we only observe a binary outcome: participation or abstention. Rosenstone and Wolfinger therefore conceptualized the propensity to participate as a latent variable, and they fit a variation of the following binary response model to the observed data:

$$\mathbb{P}(Y_{ist} = 1) = f(\alpha + X_{ist}\beta + Z_{st}\rho + \delta T_{st}\epsilon_{ist}),$$

where Y is observed participation behavior, X is a vector of demographic controls, Z is a vector of state-level controls, and T is a policy of interest.[1] Rosenstone and

[1]Due to computational limitations, Rosenstone-Wolfinger used a series of bivariate regressions, but their successors have employed a multivariate approach.

Wolfinger specify a probit link, but there is some disagreement within this general framework about the ideal functional representation. Many subsequent papers have used probit (Leighley and Nagler 1992; Alvarez and Nagler 2007, 2008, 2011), while others have used logit (Highton 1997; McDonald 2008). Notably, Nagler (1994) proposed a scobit model, which generalizes the point of inflection *a priori* and allows the program to discover the optimal location within the data.[2]

This framework has been used broadly in the convenience literature, as pointed out in Ansolabehere and Konisky (2006). An incomplete list of papers that use some variant of Rosenstone and Wolfinger's model for the primary statistical analysis includes Leighley and Nagler (1992), Nagler (1994), Mitchell and Wlezien (1995), Highton and Wolfinger (1998), Brians and Grofman (2001), McDonald (2008), Alvarez and Hall (2012), and Burden *et al.* (2014). This statistical approach is not without its detractors, however. Below I discuss the manuscripts that highlight the weaknesses of the binary response model, though their authors disagree about the severity of these concerns.

2.3 Criticism

Several important pieces of scholarship highlight the problems associated with identification of the average treatment effect and other quantities of interest when using observational data (Hanmer 2007, 2009; Glynn and Quinn 2011; Keele and Minozzi 2013). In particular, the researcher typically cannot assume that the treatment assignment mechanism is orthogonal to the outcome distribution ("exogenous selection"), which ensures balance among the observables and unobservables (at least in infinite populations). Mathematically, we can represent this as $Y \perp T$, where $Y = (Y^1, Y^0)$ represents the distribution of potential outcomes and T represents the distribution of the treatment.[3] As a prerequisite, the researcher needs a randomized treatment

[2] A drawback of the scobit model is its lack of robustness to improper specification (Hanmer 2006).

[3] In almost all cases, researchers work with finite samples, so these desirable properties do not necessarily hold, regardless. As the sample size increases, however, we approach asymptotic balance, and the estimates of interest become asymptotically consistent.

mechanism, which is often absent outside of laboratory and field experiments. To accommodate this feature of observational data, researchers employ various techniques based on pre- or post-treatment variables.

Most of the papers on convenience voting utilize the joint distributions of treatment and pre-treatment variables —*i.e.*, variables that may influence selection to the treatment group but are not themselves affected by treatment status. The Rosenstone-Wolfinger design falls into this category, as the researcher implicitly assumes that by conditioning on other variables, one can isolate the independent effect of the treatment on the propensity to vote. We can express this assumption as $\mathbb{E}[Y \mid X, T = 1] = \mathbb{E}[Y \mid X, T = 0]$. Note that this condition is actually less stringent than exogenous selection ($Y \mid X \perp T \mid X$), which requires that the distribution of $Y \mid X$ is independent from the distribution of $T \mid X$. The imposition is only on the first moment of $Y \mid X$. Even so, this assumption —"conditional mean ignorability" —is quite strong (Glynn and Quinn 2011).[4] And importantly, "it cannot be verified with observed data" (Keele and Minozzi 2013: p. 3; see also Manski 2007).

The binary response model used by Rosenstone and Wolfinger, and many of their successors, also makes certain assumptions about the functional nature of the relationship (Hanmer 2007; Keele and Minozzi 2013). As discussed in the previous section, most of the papers have specified a logit, probit, or scobit link; this entails assuming that the propensity to participate is some transformed function of a linear combination of the relevant covariates, estimated using maximum likelihood methods. The linear combination of variables includes individual-level demographic characteristics and state-level controls (sometimes including state or regional dummies). Rosenstone and Wolfinger conducted and published their research several decades ago, and so were limited by the computational efficiency of programming at the time. They could only regress on one variable at a time, and they used cross-sectional data. Subsequent work has adopted a multivariate approach, but despite the widespread availability of *CPS* data, most of the literature has continued to rely on the same cross-sectional

[4]Keele and Minozzi (2013) term this assumption "selection on observables" and focus on the importance of correctly identifying the set of covariates that intervene in the relationship between the outcome and treatment.

approach (Ansolabehere and Konisky 2006).

The cross-sectional resesarch design poses serious concerns for inference (Brians and Grofman 2001; Ansolabehere and Konisky 2006; Neiheisel and Burden 2012; Leighley and Nagler 2014). Ansolabehere and Konisky (2006) point out that this method is susceptible to omitted variable bias, given that the researcher cannot adequately control for heterogeneity across states in terms of election institutions and environment.[5] These authors also mention an unnecessary reduction in the statistical power of the model. Meanwhile, Leighley and Nagler (2014) explicitly frame their critique of cross-sectional design in the context of selection; states that offer certain liberalizations could already feature above-average turnout rates. Chapter 4 of their book additionally discusses the importance of including quality pre-treatment data on turnout for states that have altered their election policy.

To test or relax these assumption, the aforementioned papers have proposed natural experiments and nonparametric bounds.

2.4 Alternatives

2.4.1 Within-State Natural Experiments

When implementation permits, natural experiments can serve as an important insight into policy's influence. A few papers in the convenience literature utilize within-state heterogeneity of electoral policy to tease out this relationship (Ansolabehere and Konisky 2006; Neiheisel and Burden 2012; Keele and Minozzi 2013). Keele and Minozzi explain the intuition behind the natural experiment as an organic circumstance in which treatment assignment imitates a random mechanism. This design bears some similarity to matching, in which the researcher "pre-processes" data to improve balance among observed covariates artificially (see Ho *et al.* 2007), although in this case the observational data already possess this feature. Today, 49 states require that all residents register for voting eligibility, but this was not always the case.[6]

[5]They point out that select papers incorporate state fixed effects to sidestep this concern.
[6]Currently, only North Dakota does not require this step.

States began to require registration in the first part of the twentieth century, and several states initially required registration only if their jurisdiction exceed a certain population threshold. The "natural experiment" papers have used this within-state heterogeneity to identify the impact of registration law on turnout.

While most of the convenience voting literature estimates the degree that a liberalization increases turnout, Ansolabehere and Konisky (2006) consider the problem from reverse. Using county-level data from New York and Ohio, this paper considers how legislation that made registration requirements universal *depressed* turnout. They demonstrate that a cross-sectional approach implies that registration dramatically reduces turnout; the results from a panel model (with fixed effects) and difference-in-difference models, however, suggest that the true impact hovers somewhere between 3 and 5 percentage points, with a steeper drop in the first election with these new requirements. The paper makes a nice apology for using panel methods and exploiting within-state variation whenever possible.

Neihesel and Burden (2012) explore the impact of EDR on turnout in Wisconsin using a method similar to the fixed-effects panel model in Ansolabehere and Konisky (2006). Wisconsin was another state that did not universally require registration, and it introduced EDR while there was still variation in the registration requirement. As in Ansolabehere and Konisky, the dependent variable is county-level turnout, but their covariate of interest is the proportion of the county that allowed EDR. Post-estimation manipulation suggests that EDR increases turnout approximately 3.3%, which is within the neighborhood identified by Ansolabehere and Konsiky (2006).

Keele and Minozzi (2013) also consider the introduction of EDR in Wisconsin, and extend the analysis to Minnesota, as well. Keele and Minozzi, however, deviate much further from the Rosenstone-Wolfinger framework, and instead turn to Regression Discontinuity to identify the difference in participation likelihood in municipalities under and above the population threshold that dictated registration requirements. They compare the difference in participation prior to the introduction of EDR, as well, to create an "upper bound" for EDR's effect; that is, logically, we would expect the difference between registration with a closing date and no registration to exceed the difference between EDR and no registration. The authors identify a minimal and statistically insignificant affect using this research strategy, and conclude that traditional approaches overstate the impact of EDR.

The Keele-Minozzi critique is thorough and important, but I contend that their results should be interpreted with care. In fact, these authors admit, "No single study, including ours, is likely to be definitive" (p. 17). First, it is fairly dangerous to extrapolate that the loss of significance of EDR's impact in Wisconsin and Minnesota will extend to every other state, in every year. Imposing results from Wisconsin and Minnesota, two somewhat anomalous states in terms of their historic participation, on the rest of the country introduces different concerns about selection. Additionally, this project captures response for one brief period of time. It is entirely possible that a lagged effect exists as individuals become increasingly aware about the change in procedure. The authors themselves acknowledge that their scope is quite limited in footnote 18. And finally, as Keele and Minozzi point out, the RD design identifies a *local* effect; the analysis reveals the difference in turnout associated with election policy right around the threshold. It is worth noting that the unit of analysis is different; while the logistic and DID models use individual-level data, the RD model uses municipal-level data.

The authors' contention that the parametric model overstates the impact is subject to some of the concerns in the next section's analysis of Glynn and Quinn (2011). Keele and Minozzi use cross-sectional data for their parametric research design, so the regression does not include a pre-treatment wave. Thus, the coefficient for EDR does not help to identify the causal relationship between EDR and turnout; it represents the turnout associated with states that offer EDR. Because these two states typically have high turnout, the cross-sectional estimand should reflect this reality. In other words, without including pre-treatment data, it is difficult to disentangle how much of the estimand is a product of EDR and how much is attributable to a culture of participation; the coefficient on EDR is conflated with state-level characteristics of these two states. As Table 3 of their paper evidences, the DID approach (which does incorporate pre-treatment data) greatly reduces the magnitude of the estimand

(Keele and Minozzi 2013: p. 10). Finally, the authors contend that the Rosenstone-Wolfinger approach fails to account for between-state heterogeneity, but they include very few controls in their cross-sectional model to represent this source of variation; they do not include any other variables that describe the electoral institutions of the state, or any metric of electoral competitiveness.

Natural experiments are a rich source of data when available, but many implementations of convenience voting procedures do not accommodate this particular research design. Furthermore, the natural experiment requires the researcher to restrict the scope of inquiry to the states that implemented the procedure heterogeneously across municipalities. It is *a priori* unrealistic to assume that EDR will have the same effect in other states, or that its effect will be time-invariant in those states included. Thus, while this stream of literature should encourage researchers to identify other data conducive to this methodological setup, the broader perspective afforded by the Rosenstone-Wolfinger design should not be discarded.

2.4.2 Nonparametric Bounds

The proliferation of the bounding approach stems primarily from a number of influential works by Manski (1995). The bounding literature is characterized by a number of appealing features. First, research that utilizes bounds typically relies on fairly uncontroversial assumptions. Second, the approach encourages the researcher to explicate very clearly exactly which assumptions s/he is employing. This transparency affords the reader the opportunity to consider deliberately the plausibility of the assumptions, and how they affect the robustness of the results. Unfortunately, the weaker the assumptions, the less informative the inference, as Manski himself acknowledges (Manski 1995).[7]

Hanmer (2007) discusses concerns about exogenous selection, and turns to Manski bounds to ascertain whether such concerns are warranted. Stratifying by education

[7] In this subsection, I primarily discuss analyses in Hanmer (2007, 2009) and Glynn and Quinn (2011). For the interested reader, Keele and Minozzi (2013) also include a brief but nice discussion of the advantages and disadvantages of non-parametric bounds, though they spend more time on natural experiments.

group, he contrasts the treatment effect identified in a probit model with an identified region generated by bounds. Although the point estimate often falls outside of the bounded region, the 90% parametric confidence interval almost always overlaps with the non-parametrically identified region. Hanmer concludes that he cannot reject the exogenous selection assumption, and advocates the bounds as "an additional tool" (p. 21). This discussion is expanded in Hanmer (2009).

Glynn and Quinn (2011) also use non-parametric bounds, but they narrow the region of identification by using the distribution of a post-treatment variable: registration. The value of registration is itself a function of treatment status, but is realized prior to the outcome of turnout. Furthermore, registration "mediates" in the relationship between policy and turnout, because registration is almost universally required for voting eligibility. In particular, these authors propose that the influence of EDR on turnout is primarily borne through its effect on registration, a reasonable conjecture. As an illustrative exercise, Glynn and Quinn use bounds and self-reported explanations of abstention to identify a region of possible values for the average treatment effect of EDR on the control group (ATC); they restrict their sample to African Americans. More explicitly, the authors estimate the possible change in turnout of African Americans in non-EDR states had they been allowed to register on Election Day.

To accomplish this objective, they utilize a question in the *CPS* that asks non-voters why they abstained. They bifurcate abstainers into two groups: those who might have participated under different electoral conditions, and those who are simply "uninterested." First, they assume the direction of causality to be non-negative. They then posit that EDR would have minimal impact on registered non-voters (5% or lower), and that EDR would primarily affect interested non-registrants. However, even assuming that all interested non-registrants vote, and that only 5% or fewer of "uninterested" non-registrants would alter their behavior, this set of assumptions creates an upper bound on the region of possible values of 11%. They contrast this purportedly conservative upper bound with the point estimates of the ATC generated by logit models and conclude that the typical parametric approach dramatically overstates the effect of EDR, at least for this particular subgroup (*cf.* Figure 2, p. 284).

Their bounding technique is interesting, and offers a useful test for parametric results, but the authors draw an extreme conclusion: They assert that the Rosenstone-Wolfinger approach grossly exaggerates the magnitude of quantities of interest. In so doing, Glynn and Quinn unnecessarily throw out a great deal of information. The bounds estimates suggest that their parametric approach is flawed, *and it is.* The authors use a number of questionable practices in their parametric example, and accordingly, it functions as little more than a straw man. Re-examining their paper, I find that a few simple tweaks to the model produce far more reasonable estimands. Thus, the bounds —if used appropriately and effectively —do offer an extremely important advancement to the convenience voting literature.

2.5 Extended Response to Glynn and Quinn

Glynn and Quinn (2011) suggested that the typical parametric and semi-parametric approaches to causal inference yield impossibly high estimates of Election Day Registration (hereafter, EDR), one of the most scrutinized forms of convenience voting. They use a post-treatment variable and non-parametric bounds to conclude that traditional models grossly overstate the impact of EDR on turnout among African Americans. Upon closer examination, though, I contend that these results are actually an artifact of errors in their coding, as well as peculiarities within the 2004 cross section used in their analysis. Accordingly, I argue that abandoning the Rosenstone-Wolfinger (1978) model may be premature, and that many of the concerns advanced in Glynn and Quinn (2011) can be remedied by closer attention to parametric specification. At the same time, the identification strategy advanced in Glynn and Quinn (2011) offers a useful "sanity check" for the results of parametric estimation, and their overall argument serves as a worthwhile reminder to researchers to be transparent about the assumptions underlying their model. To this end, I contend that the regional identification employed in Glynn and Quinn (2011) is a highly useful complement to, rather than substitute for, more traditional parametric estimation.[8]

2.5.1 Reconsidering Glynn and Quinn (2011)

The analysis in Glynn and Quinn (2011) exaggerates the degree of bias in parametric regression and too quickly rejects this methodological approach as untenable. Importantly, although the *point estimates* of their quantities of interest lie outside of the non-parametrically identified region, *the parametric confidence intervals* overlap it. In an earlier paper, Hanmer (2007) concludes from similar results that he cannot reject the exogenous selection results. This in and of itself suggests that Glynn and Quinn may have been overly hasty in rejecting parametric design. Moreover, after consulting their replication materials, available in the *Political Analysis* public Dataverse collection (hdl: 1902:1/15920), I found that their results are adversely affected by researcher-induced measurement error and an unnecessarily restricted sample. After fixing the coding and constructing a panel dataset, the output of the logistic models suggests a far more reasonable estimand.

Table 2.1: Bias in ATC Estimates Induced by Coding Error

	Original Coding	*Correct Coding*	*Bias*
Model 1	12.9 %	11.0 %	14.7 %
Model 2	12.9 %	10.9 %	15.5 %
Model 3	12.9 %	10.7 %	17.1 %
Model 4	13.2 %	9.7 %	26.5 %
Model 5	13.3 %	11.2 %	15.8%
Model 6	12.7 %	8.9 %	29.9 %
Model 7	13.1 %	9.5 %	27.5 %
Model 8	9.7 %	9.5 %	2.1 %
Model 9	9.6 %	10.9 %	-13.5 %

Column 1 lists the results reported in Table 2 of Glynn and Quinn (2011). Column 2 lists the adjusted results after fixing their coding error. Column 3 reports the proportion of the original estimate that is directly attributable to the coding error. See also Tables 2.A1 and 2.A2 of the Supplementary Materials.

Firstly, Glynn and Quinn mistakenly code Michigan as an EDR state and Maine as

[8]Part of this analysis is reproduced from Pellissier (forthcoming).

a control state in their dataset. The state variable in the CPS dataset is *GESTFIPS* (or alternatively, *GESTCEN*); the authors create an EDR variable, and assign a value of "1" to the states with *GESTFIPS* codes of 16, 27, 26, 33, 55, and 56. Referencing the 2004 CPS codebook, these values correspond to Idaho, Minnesota, Michigan, New Hampshire, Wisconsin, and Wyoming. As evidenced by the authors' commentary within their code, they actually intended to include Maine and omit Michigan to reflect the true collection of states that offered EDR in 2004. I reproduce their findings exactly when I retain the coding error (Table 2.A1 of the Supplementary Materials). After correcting this oversight, it becomes evident that their misclassification in and of itself biased the estimates of the parametrically-identified ATC upwards. Table 2.1 presents the original and adjusted parametric estimates of the ATC for each of the nine logit models that Glynn and Quinn specify. Across the nine models, the flawed coding induces a bias of 1.9 percentage points in the ATC. And for all but the final (most parsimonious) model, the original exceeds the adjusted ATC.[9] That being said, the revised non-parametric upper bound is 10.4%, so the authors' concern of "impossibly large" results at this point remains (p. 284).

And secondly, these concerns of bias might be abated simply by adopting a superior parametric approach. It is instructive to consider other electoral cross sections when trying to draw broad conclusions about response to a particular treatment, in this case EDR. The final three columns of Table 2.2 report the ATC estimates for the 2000, 2008, and 2012 cross sections, respectively (see Tables 2.A3 - 2.A5 for full results).[10] In all of these years, the point estimate for each model's estimate of the ATC is far lower than the corresponding estimate in the 2004 panel. Moreover, in 2000, each point estimate for the ATC is actually *negative*. This peculiarity is certainly worth mentioning, given the popularity of cross sectional data in the convenience voting literature.[11] It is undesirable that an estimate exhibit such instability

[9] I have reproduced Glynn and Quinn's results in Table 2.A1, and I report the full set of results using the proper coding in Table 2.A2.

[10] Of note, Iowa and Montana added EDR prior to the 2008 election, and Connecticut and D.C. added EDR prior to the 2012 election; the coding of EDR states for these cross sections reflects the changes in legislation.

[11] The papers using a cross sectional approach are too numerous to mention, but include Rosenstone

Table 2.2: ATC Estimates from Alternative Cross Sections

	2004 Cross Section	*2000 CS*	*2008 CS*	*2012 CS*
Model 1	11.0 %	-1.3 %	1.8 %	3.2 %
Model 2	10.9 %	-1.2 %	1.9 %	3.5 %
Model 3	10.7 %	-1.2 %	1.8 %	3.5 %
Model 4	9.7 %	-3.1 %	1.0 %	3.3 %
Model 5	11.2 %	-1.9 %	1.7 %	3.5 %
Model 6	8.9 %	-3.4 %	0.3 %	3.9 %
Model 7	9.5 %	-3.3 %	0.3 %	3.7 %
Model 8	9.5 %	-1.7 %	0.5 %	3.6 %
Model 9	10.9 %	-1.4 %	0.2 %	4.5 %

Column 1 gives the results for models using the 2004 cross section (Column 2 of Table 2.1), mimicking the approach of Glynn and Quinn (but fixing the coding error for EDR states); the next three columns report the results for the same regressions using the 2000, 2008, and 2012 cross sections. See also Tables 2.A3, 2.A4, and 2.A5 of the Supplementary Materials.

across cross sections unless we have theoretical or substantive reason to believe that something about the 2004 election in and of itself encouraged African Americans to respond to the treatment markedly differently.[12] Furthermore, as pointed out in each of the papers that considers the bounding approach, it is unlikely that the availability of EDR *discourages* turnout (Hanmer 2007, 2009; Glynn and Quinn 2011; Keele and Minozzi 2013). Such results should give the discerning researcher pause.

As pointed out in Leighley and Nagler (2014), "the causal inferences from any cross-sectional analysis are suspect" (p. 98). Accordingly, researchers should utilize a time-series cross-section design when such data are readily available, as is the case with the CPS. A panel dataset improves the statistical power of the model, and more importantly, allows the researcher to capture a broader perspective of the treatment's impact. I merge the CPS data for 2000, 2004, 2008, and 2012 cross sections and again execute the nine logistic regressions, also including year fixed effects. As shown in

and Wolfinger (1978) and the majority of follow-up investigations.

[12]In the 2004 CPS data, the difference in African American turnout between treated and control groups is particularly large: 10.9%. In the other cross sections, it is -1.4% (2000), 0.3% (2008), and 4.5% (2012). A cursory scan of advertising patterns in the weeks preceding the 2004 election suggests that the Kerry campaign spent heavily in Minnesota, Wisconsin, and Iowa (`http://www.cnn.com/ELECTION/2004/special/president/campaign.ads/`). Given the high levels of Democratic support among African Americans, it is quite possible that the 2004 cross section reflects the concentrated mobilization efforts in several EDR states. Future research should explore the substantive factors driving the 2004 results.

Table 2.3, the emergent estimates of the ATC are far more consistent with the persistent stability of turnout rates, as well as the increasing skepticism that current reform efforts target the *relevant* costs (*cf.* Berinsky 2005); full results are displayed in Table 2.A6. Moreover, the 95% confidence interval around the ATC is much narrower for

Table 2.3: ATC Estimates for Cross-Sectional and Panel Data

	2004 CS	*Panel*	*Difference*
Model 1	11.0 %	3.6 %	7.4 %
Model 2	10.9 %	3.8 %	7.1 %
Model 3	10.7 %	3.7 %	7.0 %
Model 4	9.7 %	3.4 %	6.3 %
Model 5	11.2 %	3.9 %	7.3 %
Model 6	8.9 %	3.6 %	5.3 %
Model 7	9.5 %	3.4 %	6.1 %
Model 8	9.5 %	3.6 %	5.9 %
Model 9	10.9 %	4.4 %	6.5 %

Column 1 gives the results for models using the 2004 cross section (Column 2 of Table 2.1). Column 2 gives the results for a panel model incorporating the 2000, 2004, 2008, and 2012 data; it includes year fixed effects. The final column reports the difference in percentage points between the estimates using the panel dataset (Column 2) and the 2004 cross section (Column 1). See also Table 2.A6 of the Supplementary Materials.

the panel estimates than any of the cross-section estimates, reflecting the additional statistical power that the larger sample yields. The CPS generally maintains a good deal of consistency across implementations, rarely making major changes to the questions it asks and the answers it allows; therefore, a panel dataset is quickly attainable for research in this area. A panel approach ensures that the resulting estimands are more robust to idiosyncrasies in electoral cycles. And furthermore, if the researcher is still interested in isolating a particular cross-section for analysis, he or she can simply condition on membership of this subgroup when generating the estimate.

2.5.2 Discussion

It is worthwhile to consider and explicate the assumptions underlying the statistical models of one's research. Sound science requires transparency, and it should be a norm within the methodological literature to delineate clearly and precisely the assumptions underlying one's approach. And certainly, it is fair to question the conditional ignorability assumption covertly embedded within the Rosenstone-Wolfinger approach. That being said, I contend that it is premature to reject the parametric approach in light of current evidence. Instead, researchers should pay greater attention to the specification of their parametric models and consider using panel rather than cross-sectional data.

Glynn and Quinn exacerbate the degree of bias in parametric models of convenience voting with researcher-induced measurement error and less-than-ideal parametric design. My findings echo the earlier conclusions of Hanmer (2007) that the nonparametric approach fails to discredit the binary outcome regression, though the researcher should pay close attention to the functional form of the model. And as Manski himself acknowledges, the cost of weaker assumptions is less informative inference (Manski 1995; *cf.* Keele and Minozzi 2013).

It is not my intention to discourage researchers from non-parametrically identifying over a region, or from using post- instead of pre-treatment variables, but rather, to emphasize that the existing body of literature has yet to offer an adequate dismantling of the conventional approach. In fact, I believe that the bounding approach can be instructive as an initial inquiry into the validity of one's parametric model, as well as a reminder to be attentive to and transparent about the role of assumptions in causal analysis. The discipline would greatly benefit from methodological innovations that consider the selection process underlying the introduction of convenience voting procedures. It is my hope in presenting this re-analysis that researchers will not prematurely abandon point identification in favor of less controversial, but also less informative, regional identification.

2.6 Conclusion

The papers discussed in this chapter bring to light important concerns about the Rosenstone-Wolfinger research design. This scholarship offers alternative approaches to inference about convenience voting procedures, but both the RD approach and the bounds method come with a cost. Regression Discontinuity localizes the quantity of interest, and can lead the researcher to draw more general conclusions than warranted. The nonparametric bounds typically rely on transparent, uncontroversial assumptions, but they also yield results that are typically unhelpful as a standalone. A recent paper by Burden *et al.* (2014) acknowledges that challenges regarding exogenous selection are valid and important, but that researchers can accommodate such concerns by including a host of variables that capture the state-level electoral environment. These authors also point out that for most convenience voting procedures, the current body of states that have adopted a particular liberalization embody very different characteristics, and the legislation arose in heterogeneous ways.

One crucial takeaway from this methodological overview is that cross sectional data can produce biased results when examining electoral policy. Ansolabehere and Konisky (2006) and Leighley and Nagler (2014) highlight the dangers of using a cross-section design, and encourage researchers to select a panel approach instead whenever possible. In their critique of the Rosenstone-Wolfinger framework, Glynn and Quinn (2011) and Keele and Minozzi (2013) recover unreasonably high estimands when they execute logistic regression. My replication of the parametric estimation of Glynn and Quinn (2011), however, lends credence to the advice to use panel data; when I incorporate other cross sections, the logistic models produce more plausible estimates, in the neighborhood of 3 to 4 percentage points (see Table 2.3). This magnitude is similar to the effect identified in Ansolabehere and Konisky (2006).

In subsequent analysis, I use a longitudinal approach similar to Brians and Grofman (2001), and related to the panel approaches of Ansolabehere and Konisky (2006), Neiheisel and Burden (2012), and Leighley and Nagler (2014), though I keep the unit of inference the individual, rather than the county or state.[13] It also bears some resemblance to the variant of difference-in-difference presented in Chapter 3 of Hanmer (2009). Although the scathing criticisms of Glynn and Quinn (2011) and Keele and Minozzi (2013) point to unreasonable estimands generated by the Rosenstone-

[13]Hanmer (2009) offers some justification for keeping the analysis at the individual level (p. 30).

Wolfinger design, both papers use cross-sectional results as their example. When I replicated the models in Glynn and Quinn, but incorporated longitudinal data, the point estimates fell within the plausible neighborhood identified in Ansolabehere and Konisky. In the subsequent chapters, I find a fairly small impact of convenience voting, which may further obviate potential concern regarding research design. The next chapter focuses on voters with disabilities, and whether any of the policy relaxations effectively target the barriers faced by this group. The following two chapters pay closer attention to online registration specifically, since this new form of convenience has yet to be explored in the peer-reviewed literature.

2.7 Supplementary Materials

Table 2.A1: Glynn and Quinn Replication

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	-7.227	-7.642	-10.103	-10.047	-1.660	-6.959	-9.658	-11.441	0.974
	[-10.903, -3.552]	[-11.261, -4.022]	[-11.257, -8.949]	[-11.178, -8.915]	[-1.947, -1.373]	[-7.928, -5.989]	[-10.779, -8.537]	[-12.457, -8.537]	[0.922, 1.026]
EDR	-9.923	1.811	1.868	0.919	0.895	0.844	0.901	0.622	0.522
	[-29.323, 9.478]	[-4.266, 7.888]	[-4.132, 7.868]	[0.580, 1.257]	[0.562, 1.229]	[0.514, 1.175]	[0.563, 1.238]	[0.334, 0.909]	[0.279, 0.825]
Family Income	0.168	0.163	0.081	0.081	0.130	0.084	0.072		
	[0.116, 0.220]	[0.112, 0.215]	[0.065, 0.096]	[0.066, 0.097]	[0.116, 0.144]	[0.069, 0.099]	[0.057 ,0.088]		
Sex	-1.474	-1.201	0.430	0.419	0.500	0.434		0.319	
	[-3.774, 0.825]	[-3.459, 1.057]	[0.307, 0.553]	[0.298, 0.540]	[0.382, 0.619]	[0.316, 0.553]		[0.210, 0.428]	
Age	0.041	0.041	0.026	0.025	0.019		0.026	0.027	
	[0.028, 0.054]	[0.029, 0.054]	[0.022, 0.029]	[0.022, 0.029]	[0.015, 0.022]		[0.022, 0.619]	[0.023, 0.030]	
Education	0.112	0.124	0.221	0.220		0.168	0.228	0.275	
	[0.019, 0.205]	[0.032, 0.215]	[0.192, 0.250]	[0.192, 0.248]		[0.142, 0.194]	[0.200, 0.256]	[0.250, 0.299]	
EDR*Family Income	-0.065	0.023	0.026						
	[-0.353, 0.223]	[-0.064, 0.109]	[-0.061, 0.113]						
EDR*Sex	7.492	-0.380	-0.304						
	[-4.600, 19.584]	[-1.077, 0.318]	[-1.003, 0.395]						
Family Income*Sex	-0.017	-0.009							
	[-0.090, 0.056]	[-0.030, 0.011]							
EDR*Age	0.325	-0.002	-0.010						
	[-0.171, 0.822]	[-0.157, 0.152]	[-0.031, 0.010]						
Sex*Age	-0.056	-0.053							
	[-0.088, -0.024]	[-0.084, -0.022]							
EDR*Education	-0.010	-0.010	-0.007						
	[-0.017, -0.002]	[-0.017, -0.002]	[-0.160, 0.152]						
Sex*Education	0.071	0.063							
	[0.013, 0.129]	[0.006, 0.120]							
EDR*Family Income*Sex	0.062								
	[-0.115, 0.239]								
EDR*Sex*Age	0.004								
	[-0.039, 0.047]								
EDR*Sex*Education	-0.218								
	[-0.528, 0.091]								
ATC	0.129	0.129	0.129	0.132	0.133	0.127	0.131	0.097	0.096
	[0.089, 0.129]	[0.090, 0.171]	[0.090, 0.168]	[0.092, 0.170]	[0.091, 0.171]	[0.087, 0.166]	[0.090, 0.171]	[0.057, 0.136]	[0.056, 0.137]
N	6302	6302	6302	6302	6302	6379	6302	7390	7486

This table presents the point estimates and 95% confidence intervals for the coefficients of a logistic regression of African American voter turnout in the year 2004, replicating Table 2 of Glynn and Quinn (2011). I retain their error in the classification of EDR states: explicitly, Michigan is mistakenly labeled an EDR state, and Maine a control state. The ATC and 95% bootstrapped confidence intervals estimates in the penultimate row represent the predicted effect of Election Day Registration on turnout among African Americans in non-EDR states, and are generated using the results of each fitted model.

Table 2.A2: 2004 Cross Section (Adjusted)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	-7.130	-7.254	-9.966	-9.927	-1.608	-6.893	-9.546	-11.373	0.989
	[-10.734, -3.526]	[-10.847, -3.660]	[-11.100, -8.833]	[-11.054, -8.799]	[-1.894, -1.323]	[-7.860, -5.926]	[-10.663, -8.429]	[-12.387, -10.358]	[0.938, 1.041
EDR	-1.784	6.875	5.567	0.638	0.732	0.557	0.618	0.611	0.653
	[-59.182, 55.614]	[-4.956, 18.706]	[-6.267, 17.400]	[0.074, 1.201]	[0.173, 1.291]	[0.005, 1.109]	[0.057, 1.179]	[0.087, 1.135]	[0.146, 1.160
Family Income	0.166	0.164	0.080	0.080	0.129	0.083	0.072		
	[0.114, 0.217]	[0.113, 0.215]	[0.064, 0.095]	[0.065, 0.096]	[0.115, 0.143]	[0.068, 0.098]	[0.056, 0.087]		
Sex	-1.466	-1.384	0.422	0.414	0.496	0.430		0.319	
	[-3.725, 0.794]	[-3.635, 0.867]	[0.301, 0.544]	[0.293, 0.535]	[0.378, 0.614]	[0.311, 0.548]		[0.210, 0.428]	
Age	0.039	0.040	0.025	0.025	0.018		0.025	0.026	
	[0.027, 0.052]	[0.028, 0.052]	[0.021, 0.029]	[0.021, 0.029]	[0.015, 0.022]		[0.022, 0.029]	[0.023, 0.030]	
Education	0.113	0.116	0.219	0.218		0.167	0.227	0.274	
	[0.022, 0.204]	[0.025, 0.207]	[0.192, 0.248]	[0.190, 0.247]		[0.142, 0.193]	[0.198, 0.255]	[0.249, 0.298]	
EDR*Family Income	-0.253	0.042	0.039						
	[-0.782, 0.276]	[-0.101, 0.185]	[-0.104, 0.182]						
EDR*Sex	4.432	-0.971	-0.814						
	[-26.597, 35.462]	[-2.169, 0.227]	[-1.996, 0.368]						
Family Income*Sex	-0.055	-0.054							
	[-0.087, -0.024]	[-0.085, -0.023]							
EDR*Age	0.169	0.008	0.009						
	[-0.078, 0.415]	[-0.032, 0.049]	[-0.032, 0.049]						
Sex*Age	-0.009	-0.009							
	[-0.017, -0.001]	[-0.017, -0.002]							
EDR*Education	0.010	-0.135	-0.108						
	[-1.486, 1.505]	[-0.431, -0.162]	[-0.405, 0.189]						
Sex*Education	0.069	0.068							
	[0.012, 0.126]	[0.011, 0.124]							
EDR*Family Income*Sex	0.166								
	[-0.142, 0.475]								
EDR*Sex*Age	-0.089								
	[-0.218, 0.041]								
EDR*Sex*Education	-0.094								
	[-0.896, 0.709]								
ATC	0.110	0.109	0.107	0.1097	0.112	0.089	0.095	0.095	0.109
	[0.019, 0.174]	[0.025, 0.182]	[0.018, 0.176]	[0.016, 0.168]	[0.045, 0.179]	[0.013, 0.161]	[0.016, 0.163]	[0.015, 0.164]	[0.034, 0.173
N	6302	6302	6302	6302	6302	6379	6302	7390	7486

This table presents the point estimates and 95% confidence intervals for the coefficients of a logistic regression of African American voter turnout in the year 2004, replicating Table 2 of Glynn a Quinn (2011) but fixing the error in the classification of EDR states; Maine is now correctly coded as an EDR state, and Michigan as a control state. The ATC and 95% bootstrapped confiden intervals estimates in the penultimate row represent the predicted effect of Election Day Registration on turnout among African Americans in non-EDR states, and are generated using the resu of each fitted model. In 40 of the initial 1000 bootstrap relications used to generate the ATC confidence intervals, the logit model did not converge after 20 iterations; I drop these and use the fi 1000 simulations that achieved convergence.

Table 2.A3: 2000 Cross Section

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	-7.740	-7.796	-10.541	-10.548	-2.378	-5.736	-10.242	-12.086	0.756
	[-11.456, -4.025]	[-11.496, -4.095]	[-11.690, -9.393]	[-11.691, -9.405]	[-2.676, -2.080]	[-6.653, -4.819]	[-11.376, -9.108]	[-13.102, -11.070]	[0.706, 0.80
EDR	-8.375	-0.234	-0.314	-0.161	-0.097	-0.167	0.172	-0.089	-0.063
	[-48.639, 31.889]	[-11.703, 11.235]	[-11.771, 11.143]	[-0.659, 0.338]	[-0.587, 0.392]	[-0.647, 0.314]	[-0.670, 0.326]	[-0.572, 0.394]	[-0.519, 0.39
Family Income	0.101	0.107	0.092	0.092	0.148	0.091	0.084		
	[0.044, 0.158]	[0.051, 0.163]	[0.076, 0.109]	[0.076, 0.109]	[0.133, 0.163]	[0.075, 0.107]	[0.067, 0.100]		
Sex	-1.464	-1.428	0.350	0.355	0.445	0.353		0.208	
	[-3.757, 0.830]	[-3.710, 0.854]	[0.229, 0.471]	[0.235, 0.476]	[0.328, 0.563]	[0.238, 0.468]		[0.100, 0.316]	
Age	0.046	0.046	0.037	0.038	0.030		0.038	0.038	
	[0.033, 0.059]	[0.032, 0.059]	[0.033, 0.041]	[0.034, 0.042]	[0.026, 0.034]		[0.034, 0.563]	[0.035, 0.042]	
Education	0.133	0.133	0.216	0.216		0.133	0.224	0.278	
	[0.039, 0.227]	[0.040, 0.227]	[0.187, 0.245]	[0.187, 0.244]		[0.109, 0.158]	[0.195, 0.252]	[0.254, 0.303]	
EDR*Family Income	0.573	0.006	0.006						
	[-0.038, 1.183]	[-0.148, 0.159]	[-0.148, 0.159]						
EDR*Sex	5.045	0.576	0.614						
	[-18.823, 28.912]	[-0.556, 1.708]	[-0.519, 1.747]						
Family Income*Sex	-0.006	-0.009							
	[-0.040, 0.028]	[-0.043, 0.024]							
EDR*Age	0.019	0.043	0.042						
	[-0.153, 0.192]	[-0.005, 0.090]	[-0.005, 0.089]						
Sex*Age	-0.005	-0.005							
	[-0.014, 0.003]	[-0.013, 0.003]							
EDR*Education	0.047	-0.063	-0.062						
	[-0.961, 1.055]	[-0.349, 0.222]	[-0.348, 0.090]						
Sex*Education	0.053	0.053							
	[-0.004, 0.111]	[-0.004, 0.011]							
EDR*Family Income*Sex	-0.347								
	[-0.699, 0.005]								
EDR*Sex*Age	0.017								
	[-0.087, 0.121]								
EDR*Sex*Education	-0.059								
	[-0.657, 0.539]								
ATC	-0.013	-0.012	-0.012	-0.031	-0.019	-0.034	-0.033	-0.017	-0.014
	[-0.108, 0.074]	[-0.104, 0.072]	[-0.107, 0.073]	[-0.127, 0.057]	[-0.115, 0.071]	[-0.139, 0.066]	[-0.128, 0.061]	[-0.107, 0.080]	-0.122, 0.09
N	5991	5991	5991	5991	5991	6134	5991	73019	7203

This table presents the point estimates and 95% confidence intervals for the coeffcients of a logistic regression of African American voter turnout in the year 2000. The ATC and 95% bootstrap confidence intervals estimates in the penultimate row represent the predicted effect of Election Day Registration on turnout among African Americans in non-EDR states, and are generated using results of each fitted model. In 26 of the initial 1000 bootstrap relications used to generate the ATC confidence intervals, the logit model did not converge after 20 iterations; I drop these and use first 1000 simulations that achieved convergence.

Table 2.A4: 2008 Cross Section

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	- 8.267	-8.435	-10.713	-10.558	-1.465	-8.110	-11.341	1.399	
	[-12.337, -4.198]	[-12.455, -4.415]	[-12.024, -9.403]	[-11.848, -9.269]	[-1.784, -1.146]	[-9.243, -6.978]	[-11.313, -8.769]	[-12.490, -10.192]	[-1.341, 1.45
EDR	1.237	4.345	4.471	0.070	0.122	0.018	0.024	0.037	0.015
	[-21.843, 24.318]	[-3.428, 12.118]	[-3.250, 12.192]	[-0.347, 0.488]	[-0.287, 0.532]	[-0.392, 0.428]	[-0.390, 0.439]	[-0.354, 0.427]	[-0.359, 0.38
Family Income	0.111	0.115	0.080	0.079	0.126	0.079	0.067		
	[0.055, 0.166]	[0.060, 0.170]	[0.062, 0.097]	[0.062, 0.097]	[0.110, 0.141]	[0.062, 0.096]	[0.051, 0.084]		
Sex	-1.030	-0.923	0.597	0.605	0.700	0.584		0.508	
	[-3.630, 1.570]	[-3.486, 1.641]	[0.456, 0.738]	[0.466, 0.743]	[0.564, 0.835]	[0.448, 0.719]		[0.385, 0.631]	
Age	0.034	0.033	0.023	0.023	0.018		0.023	0.023	
	[0.021, 0.048]	[0.020, 0.046]	[0.019, 0.027]	[0.019, 0.027]	[0.014, 0.022]		[0.019, 0.835]	[0.019, 0.026]	
Education	0.162	0.167	0.244	0.240		0.203	0.252	0.280	
	[0.059, 0.266]	[0.065, 0.270]	[0.211, 0.277]	[0.207, 0.272]		[0.173, 0.233]	[0.220, 0.285]	[0.252, 0.308]	
EDR*Family Income	0.170	-0.032	-0.027						
	[-0.148, 0.487]	[-0.132, 0.069]	[-0.127, 0.073]						
EDR*Sex	3.088	0.182	0.244						
	[-12.751, 18.927]	[-0.670, 1.034]	[-0.608, 1.096]						
Family Income*Sex	-0.077	-0.023							
	[-0.174, 0.019]	[-0.058, 0.011]							
EDR*Age	-0.001	0.013	0.015						
	[-0.583, 0.582]	[-0.016, 0.043]	[-0.015, 0.045]						
Sex*Age	-0.020	-0.007							
	[-0.055, 0.015]	[-0.015, 0.002]							
EDR*Education	-0.008	-0.122	-0.130						
	[-0.016, 0.001]	[-0.314, 0.070]	[-0.321, 0.043]						
Sex*Education	0.054	0.051							
	[-0.012, 0.120]	[-0.014, 0.116]							
EDR*Family Income*Sex	-0.130								
	[-0.339, 0.080]								
EDR*Sex*Age	0.066								
	[-0.006, 0.137]								
EDR*Sex*Education	-0.106								
	[-0.510, 0.297]								
ATC	0.018	0.019	0.018	0.010	0.017	0.003	0.003	0.005	0.002
	[-0.038, 0.073]	[-0.037, 0.076]	[-0.036, 0.073]	[-0.049, 0.067]	[-0.043, 0.069]	[-0.056, 0.059]	[-0.057, 0.062]	[-0.049, 0.059]	[-0.058, 0.06
N	5945	5945	5945	5945	5945	6028	5945	7118	7226

This table presents the point estimates and 95% confidence intervals for the coeffcients of a logistic regression of African American voter turnout in the year 2008. The ATC and 95% bootstrap confidence intervals estimates in the penultimate row represent the predicted effect of Election Day Registration on turnout among African Americans in non-EDR states, and are generated us the results of each fitted model. In 1 of the initial 1000 bootstrap relications used to generate the ATC confidence intervals, the logit model did not converge after 20 iterations; I drop this iterat and use the first 1000 simulations that achieved convergence.

Table 2.A5: 2012 Cross Section

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	-8.520	-9.327	-10.144	-10.031	-1.785	-6.493	-9.590	-11.359	1.357
	[-12.306, -4.734]	[-12.918, -5.736]	[-11.357, -8.932]	[-11.172, -8.890]	[-2.058, -1.511]	[-7.469, -5.517]	[-10.718, -8.462]	[-12.461, -10.258]	[1.299, 1.415]
EDR	-5.653	1.511	1.499	0.245	0.250	0.278	0.266	0.259	0.301
	[-17.404, 6.098]	[-2.104, 5.126]	[-2.109, 5.107]	[0.046, 0.444]	[0.054, 0.446]	[0.086, 0.471]	[0.068, 0.464]	[0.061, 0.458]	[0.115, 0.487]
Family Income	0.093	0.087	0.070	0.067	0.109	0.072	0.057		
	[0.044, 0.141]	[0.041, 0.133]	[0.055, 0.086]	[0.052, 0.081]	[0.096, 0.123]	[0.058, 0.086]	[0.043, 0.071]		
Sex	-0.497	0.049	0.611	0.592	0.680	0.584		0.518	
	[-2.920, 1.926]	[-2.229, 2.326]	[0.485, 0.737]	[0.473, 0.711]	[0.563, 0.796]	[0.469, 0.699]		[0.401, 0.635]	
Age	0.041	0.043	0.031	0.031	0.027		0.031	0.032	
	[0.029, 0.053]	[0.031, 0.054]	[0.027, 0.035]	[0.027, 0.035]	[0.023, 0.030]		[0.028, 0.796]	[0.028, 0.036]	
Education	0.161	0.181	0.218	0.216		0.160	0.230	0.266	
	[0.066, 0.257]	[0.090, 0.272]	[0.188, 0.249]	[0.188, 0.245]		[0.135, 0.186]	[0.201, 0.258]	[0.240, 0.293]	
EDR*Family Income	-0.080	-0.026	-0.027						
	[-0.228, 0.068]	[-0.072, 0.019]	[-0.072, 0.019]						
EDR*Sex	4.453	-0.171	-0.170						
	[-2.747, 11.653]	[-0.072, 0.019]	[-0.072, 0.019]						
Family Income*Sex	0.022	0.004							
	[-0.019, 0.062]	[-0.009, 0.016]							
EDR*Age	0.151	-0.024	0.003						
	[-0.147, 0.449]	[-0.114, 0.066]	[-0.009, 0.015]						
Sex*Age	-0.015	-0.011							
	[-0.046, 0.016]	[-0.040, 0.018]							
EDR*Education	-0.007	-0.008	-0.023						
	[-0.015, 0.001]	[-0.015, -0.001]	[-0.113, 0.066]						
Sex*Education	0.039	0.025							
	[-0.023, 0.100]	[-0.032, 0.083]							
EDR*Family Income*Sex	0.035								
	[-0.339, 0.080]								
EDR*Sex*Age	-0.012								
	[-0.057, 0.127]								
EDR*Sex*Education	-0.113								
	[-0.036, 0.127]								
ATC	0.032	0.035	0.035	0.033	0.035	0.039	0.037	0.036	0.045
	[0.005, 0.057]	[0.009, 0.061]	[0.009, 0.061]	[0.007, 0.058]	[0.008, 0.062]	[0.016, 0.065]	[0.011, 0.061]	[0.011, 0.062]	[0.018, 0.071]
N	7812	7812	7812	7812	7812	7974	7812	7812	7974

This table presents the point estimates and 95% confidence intervals for the coeffcients of a logistic regression of African American voter turnout in the year 2012. The ATC and 95% bootstrappe confidence intervals estimates in the penultimate row represent the predicted effect of Election Day Registration on turnout among African Americans in non-EDR states, and are generated usi the results of each fitted model.

Table 2.A6: Panel Data

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	-7.993	-8.333	-10.582	-10.518	-2.137	-7.029	-10.093	-11.802	0.752
	[-9.875, -6.110]	[-10.180, -6.485]	[-11.179, -9.985]	[-11.102, -9.934]	[-2.291, -1.983]	[-7.525, -6.533]	[-10.671, -9.515]	[-12.334, -11.270]	[0.703, 0.802]
EDR	-6.171	2.112	2.052	0.219	0.244	0.223	0.221	0.230	0.256
	[-15.649, 3.306]	[-0.828, 5.052]	[-0.881, 4.985]	[0.057, 0.380]	[0.085, 0.403]	[0.067, 0.380]	[0.060, 0.382]	[0.072, 0.387]	[0.107, 0.406]
Family Income	0.117	0.116	0.080	0.079	0.127	0.081	0.069		
	[0.090, 0.143]	[0.090, 0.142]	[0.072, 0.088]	[0.071, 0.087]	[0.120, 0.134]	[0.073, 0.088]	[0.061, 0.077]		
Sex	-1.230	-1.005	0.484	0.482	0.568	0.480		0.375	
	[-2.414, -0.046]	[-2.163, 0.153]	[0.421, 0.547]	[0.420, 0.544]	[0.508, 0.629]	[0.420, 0.540]		[0.318, 0.431]	
Age	0.040	0.040	0.029	0.029	0.023		0.029	0.030	
	[0.033, 0.046]	[0.034, 0.046]	[0.027, 0.031]	[0.027, 0.031]	[0.022, 0.025]		[0.028, 0.629]	[0.028, 0.032]	
Education	0.137	0.146	0.223	0.221		0.163	0.231	0.273	
	[0.089, 0.185]	[0.099, 0.192]	[0.208, 0.238]	[0.206, 0.236]		[0.150, 0.176]	[0.217, 0.246]	[0.261, 0.286]	
EDR*Family Income	-0.049	-0.027	-0.026						
	[-0.169, 0.071]	[-0.064, 0.010]	[-0.063, 0.011]						
EDR*Sex	5.347	-0.051	-0.026						
	[-0.478, 11.173]	[-0.368, 0.265]	[-0.342, 0.291]						
Family Income*Sex	0.013	0.008							
	[-0.021, 0.046]	[-0.003, 0.018]							
EDR*Age	0.161	-0.048	0.008						
	[-0.079, 0.402]	[-0.122, 0.025]	[-0.002, 0.018]						
Sex*Age	-0.024	-0.023							
	[-0.040, -0.008]	[-0.039, 0.007]							
EDR*Education	-0.007	-0.007	-0.048						
	[-0.011, -0.003]	[-0.011, -0.003]	[-0.122, 0.025]						
Sex*Education	0.056	0.050							
	[0.026, 0.086]	[0.021, 0.080]							
EDR*Family Income*Sex	0.014								
	[-0.060, 0.088]								
EDR*Sex*Age	-0.003								
	[-0.024, 0.017]								
EDR*Sex*Education	-0.137								
	[-0.285, 0.011]								
ATC	0.036	0.038	0.037	0.034	0.039	0.036	0.034	0.036	0.044
	[0.012, 0.060]	[0.012, 0.061]	[0.013, 0.061]	[0.010, 0.056]	[0.013, 0.063]	[0.012, 0.060]	[0.011, 0.060]	[0.0112, 0.059]	[0.019, 0.067]
N	26,050	26,050	26,050	26,050	26,050	26,515	26,050	29,339	29,889

This table presents the point estimates and 95% confidence intervals for the coeffcients of a logistic regression of African American voter turnout in the years 2000, 2004, 2008, and 2012. I includ year fixed effects, with the 2000 election serving as the reference category. The ATC and 95% bootstrapped confidence intervals estimates in the penultimate row represent the predicted effect o Election Day Registration on turnout among African Americans in non-EDR states, and are generated using the results of each fitted model.

Chapter 3

The Impact of Disability on Political Participation

3.1 Summary

Individuals with disabilities are consistently underrepresented in American elections, despite legislative efforts to make civic participation more accessible. A small but steady stream of literature has consistently identified a pronounced, negative effect of disability status on turnout and other forms of civic participation. However, the magnitude of the estimated effect on turnout has varied considerably, ranging from 4% to 21%, creating ambiguity about the degree of the relationship. Moreover, because disability is strongly correlated with other demographic features, notably age, there is a danger that the existing estimates suffer from confounding bias. I present the first analysis that involves matching, and I show that ignoring the imbalance among the observables can bias the estimates of disability's impact upwards. Nevertheless, I still identify a reduction in the propensity to be registered of 2 to 5 percentage points, and a reduction in the propensity to cast a ballot of approximately 5 to 8 percentage points. When I incorporate interactive effects, though, it becomes clear that this impact is borne primarily by individuals who report disabilities and are not currently employed, supporting earlier findings by Schur and Kruse (2000) and Schur *et al.* (2002). And finally, I find no evidence that the most popular convenience voting procedures effectively target this demographic subgroup.

3.2 Introduction

The demographic distribution of the American electorate differs markedly from that of the voting-eligible population (Schlozman, Verba, and Brady 2012). Because individuals of different demographic backgrounds vary systematically in their political preferences, the degree of representativeness (or lack, thereof) has meaningful implications for electoral results. Moreover, the political system incentivizes officeholders to respond to the electorate, rather than the entire constituency, and policy is shaped accordingly (Verba, Schlozman, and Brady 1995). Due to the historical incidence of racial disenfranchisement (both overt and indirect), political scientists have paid particular attention to racial representativeness (*cf.* Alvarez and Nagler 2007, 2008, 2011; Glynn and Quinn 2011). The lack of participation among young voters is also widely noted, both in the press and in the academic literature (see, for example, Leighley and Nagler 1992).

One demographic feature that has received considerably less scrutiny is disability status, even though individuals with disabilities constitute one of the most sizable political minorities, and their participation rates are significantly —and *persistently* —lower (Schur and Adya 2013). In fact, 8.3% of respondents indicate that they have some form of disability in both the 2008 and 2012 cross sections of the *Current Population Survey Voting and Registration Supplement*; in other words, roughly 1 out of every 12 individuals reports some kind of disability (ICPSR 25643, 31082).[1] Shriner, Ochs, and Shields (1997) label the advocacy for those with cognitive and behavioral disabilities the "last suffrage movement" and point out that discrimination based on disability commands a lower level of judicial scrutiny than discrimination on race or gender. Nevertheless, the corpus of literature examining the impact of disability on electoral participation remains fairly limited. Perhaps this de-emphasis is due to a lack of salience, or simply a lack of data. Few large, nationally-representative surveys ask about both disability status and political participation and beliefs (Alvarez and

[1] The *CPS* allows individuals to report auditory, visual, cognitive, and ambulatory difficulties, as well as impediments to typical activities such as bathing or running errands. An individual is considered disabled if he or she reports any of the aforementioned conditions.

Hall 2012; Schur *et al.* 2002).

Over the past several decades, several pieces of federal legislation have attempted to facilitate electoral participation among individuals with disabilities. The effectiveness of these laws depends (at least in part) on compliance, and there is evidence that execution is lacking (Schur, Shields, and Shriner 2005). Harrington (1999), for example, argues that precincts in Texas had broadly disregarded the spirit, if not the letter, of the legal provisions within Americans with Disabilities Act of 1990. And at the national level, accessibility continues to represent a real threat to turnout among citizens with disabilities (Alvarez and Hall 2012). A recent report issued by the Government Accountability Office suggested that a mere 27% of polling places "had no potential impediments in the path from the parking to the voting area" (Bovbjerg 2009, p. 1). Although this estimate actually represented *improvement* from the 2000 election, it suggests that American elections are perenially characterized by a lack of accessibility. Moreover, technology is playing an increasingly large role in American elections, and it remains unclear how this development is affecting voters with disabilities (Baker, Roy and Moon 2005). Stewart (2011) points out the irony of this "silence," since advocates for citizens with disabilities were instrumental in shaping and passing the Help America Vote Act of 2002.

Given the fungible nature of compliance, and our continued uncertainty about how voters with disabilities will respond to new technologies, it is important to continue to estimate the relationship between disability status and turnout. The *CPS* altered its approach to measuring disability in 2008, and recent papers have emphasized the need for further exploration as additional cross sections of data become available (Alvarez and Hall 2012; Schur and Adya 2013). To my knowledge, my paper is the first to incorporate the 2012 cross section using multivariate analysis, and the first to explore concerns regarding selection. Disability status is correlated with several well-known predictors of turnout, including age, education, and employment status (Baker, Roy, and Moon 2005; Schur and Adya 2013). It is important to address whether these other demographic factors are driving the under-representation, or whether being disabled in and of itself lessens the propensity to vote. Many papers in this area rely on differences in means between the subgroups, perhaps stratified by one other demographic factor. The more statistically intricate adopt a multivariate approach (Schur *et al.* 2002; Alvarez and Hall 2012; Schur and Adya 2013). Yet no papers in this area explicitly consider the danger of confounders, and my analysis is the first to implement statistical matching as a safeguard against this potential source of bias.

The chapter proceeds as follows. First, I review the literature on the continued underrepresentation of voters with disabilities, and the electoral policies that have been introduced to encourage participation among those who face higher barriers, including citizens with disabilities. I then discuss briefly the theoretical motivation behind matching, an econometric technique that researchers can employ when they suspect imbalance among the observed characteristics across the stratification of interest (in this case, the disability indicator). Next, I describe the binary response model that I use to identify and estimate the effect of disability status on registration turnout. I extend the model to consider how age, employment status, and various convenience voting procedures may mediate the relationship between disability and turnout. To this end, Alvarez and Hall (2012) offer an excellent initial inquiry into these research questions, though the authors highlight the need for more data and further analysis in the conclusion of their book. I follow their statistical approach fairly closely, but I modify it to incorporate new data —namely, the 2012 cross section —and to reduce any bias introduced by confounders.

I find that without matching, the estimated effect of disability is artificially inflated, particularly with regard to turnout, highlighting the sensitivity of these models to the composition of the data. Still, however, the impact of disability is significant both statistically and substantively; it reduces the propensity to be registered by 3 to 5 percentage points, and the propensity to cast a ballot by approximately 6 to 8 percentage points. The interactive effects reveal that employment status heavily drives this reduction, substantiating earlier work by Schur and Kruse (2000) and Schur *et al.* (2002). Unfortunately, my findings temper the optimism of Alvarez and Hall (2012) that convenience voting procedures target this particular subpopulation effectively. Altogether, my analysis reiterates the concern in the literature that the current tactics to enfranchise this group of voters do not adequately address the barriers they face.

3.3 Higher Hurdles

Over the last half century, several pieces of federal legislation have sought to ease the electoral process for voters in general, with acts and provisions directly considering the interests of individuals with disabilities. The first landmark bill is the Voting Rights Act of 1965, which rendered illegal a number of practices. Among these were "literacy tests," which played some role in disenfranchising individuals with cognitive disabilities (Schriner, Ochs, and Shields 1997). Additionally, the VRA explicated that voters with disabilities may designate an individual to assist them at the polls (Schriner and Batavia 2001). Nearly two decades later, the Voting Accessibility for the Elderly and Handicapped Act of 1984 assigned the state an additional role in ensuring accessibility by requiring polling places to offer "auxiliary aids" (Schriner and Batavia 2001).

The Americans with Disabilities Act of 1990 and its 2008 Amendments reflected a belief among the legislative branch that individuals with disabilities do not enjoy the same legal protections as other minority groups (Schriner, Ochs, and Shields (1997). This legislation purportedly sought to encourage civic engagement of individuals with disabilities, and remove some of the existing barriers (Schur *et al.* 2002), though its efficacy has been questioned (Harrington 1999). The ADA included provisions targeting discrimination and accessibility in all aspects of civic life; yet while its focus was broad in nature, it also specifically highlighted the importance of accessibility in elections (Bovbjerg 2013; see also Baker, Roy, and Moon 2005). The federal government maintains a list of features that polling places must possess to be in compliance with the provisions of the document, available at `http://www.ada.gov/votingchecklist.htm`.

The National Voter Registration Act of 1993 liberalized several electoral policies; perhaps most prominently, in a provision labeled "motor voter," it insisted that states allow citizens to register as voters at DMVs and public assistance agencies, or offer Election Day Registration (Hanmer 2009). While these policies do not explicitly mention voters with disabilities, they expand the pool of registration alternatives. Because access is a chief concern for many individuals who report disability (Alvarez and Hall 2012), this expansion has important implications for this subpopulation. And most recently, the Help America Vote Act of 2002 sought to modernize the technology used in election administration (see Stewart 2011). HAVA requires that every polling site offer an accessible voting mode; it also insists that "each state allow electronic voter registration at disability agencies, all voting-related materials are available in alternative formats, and poll workers are provided disability etiquette training" (Schur, Shields, and Schriner 2005: 1618).

Nonetheless, individuals with disabilities still face substantially higher costs of participation, particularly if their states are strict about in-person voting. As a motivating example, Los Angeles County conducted a focus groups for several demographic groups, including voters with disabilities; many of the participants expressed a preference for voting absentee, citing various frustrations with in-person voting (Alvarez and Hall 2012).[2] For many, the polling location may be difficult to navigate. Others may have trouble obtaining transportation to and from the polling site. The inverse relationship between cost and participation is well-established in the voting literature. Since Downs (1957), the norm within the literature is to treat the voting eligible as utility-maximizers who will cast a ballot if and only if the doing so offers greater expected utility than abstention.[3]

Qualitative and observational data confirm that these costs affect voting experiences. In the 2008 *Survey of the Performance of American Elections*, the raw frequencies of individuals who report difficulty with registration, voting technology, and casting a ballot are significantly higher among the disabled (Alvarez and Hall 2012). A separate survey conducted by the authors of Baker, Roy, and Moon (2005) indicates that voters with disabilities report less satisfaction with their overall voting

[2] For a more extensive discussion of these interesting qualitative data, see Alvarez and Hall (2012).

[3] Riker and Ordeshook (1967) further developed the "rational agent" framework by adding a psychological component, and most subsequent papers have adopted this framework.

experience, poll workers and site, and voting machines. Other surveys reveal that individuals with disabilities not only experience a less enjoyable voting experience, but also *expect* to (Schur *et al.* 2002; Schur and Kruse 2011).[4] This finding is critical, as the Downsian calculus considers *anticipated* costs and benefits. If citizens with disabilities assume the voting experience will be more costly, even if it is not in actuality, their (expected) propensity to turn out will still be reduced.

Quantifying that reduction is a difficult but important task. In my review of the extant literature, every study has identified a negative and statistically significant effect, but the degree varies quite dramatically by dataset and research design. Different surveys and investigations have concluded that the effect is as low as 4% and as high as 21% (*cf.* Schur, Shields, and Shriner 2005; Schur and Kruse 2011). What drives this relationship is even murkier, as disability status is correlated with a host of other factors that influence behavior; individuals with disabilities are more likely to be older, and they report lower levels of income, education, and employment. The "net effect" of these correlates is ambiguous *a priori*, as age is typically positively related to turnout while the other features are inversely related (*cf.* Schur and Adya 2013). Additionally, it is difficult to disentangle how much of the reduced turnout among the disabled is attributable to these other correlates, and how much is due to features more specific to the disabled community (*e.g.*, physical or psychological inaccessibility).

The timing of the onset of disability could influence turnout patterns, as well (Schur, Shields, and Shriner 2005; Schur *et al.* 2002). Schur and Kruse (2000) focus specifically on individuals with spinal cord injuries. Because the occurrence of this kind of disability is more independent from other demographic characteristics, this research design (at least partially) controls for some other features that might mediate in the relationship between disability and turnout, such as age and education. The authors find that employed individuals with such injuries behave quite similarly to individuals without any disabilities, but the *un*employed are substantially less likely to

[4] The latter two surveys have sample sizes of 563 and 1240, respectively, so their results should be interpreted with care.

vote than their counterparts. A multivariate approach by Schur *et al.* (2002) echoes the importance of the interaction between disability and employment, but adds a caveat that participation is reduced for retirement-age individuals with disabilities, even if they are employed.

In this paper, I promote a different strategy to address concerns regarding selection: matching. In the next section, I discuss the technical literature and my specific implementation, but a brief introduction to the intuition may be instructive. In observational studies, the researcher often has reason to believe that individuals do not experience a particular condition randomly (or even conditionally randomly). Matching stratifies a dataset into subgroups according to the condition of interest, and then pairs "similar" individuals across groups. For this project, I subdivide individuals into two groups: those who report disabilities and those who do not. Each individual with a disability is matched to an individual without disabilities of a similar demographic background; I then parse the dataset to include only these matched observations. It is important to note that this procedure does not render selection exogenous, as the selection process was not organically orthogonal to the joint distribution of the other covariates. Matching merely imposes restrictions on the dataset so that it *mimics* random selection.

3.4 Matching to Achieve Balance

The previous attempts to quantify the relationship between disability and participation have used strictly parametric approaches. Ideally, researchers encounter datasets that are characterized by balance among both the observables or unobservables. When the independent variable of interest is associated with other covariates that also affect the outcome, parametric estimates can exhibit bias; these covariates are called "confounders." In laboratory and field experiments, the researcher can ignore such concerns by assigning a "treatment" randomly. With observational data, however, it is harder to disentangle the variable of interest from their confounders. Such concern certainly seems warranted for the purposes of this paper, given that age in particular is a significant driver of both disability status and turnout (*cf.* Schur *et al.* 2002).

As a means of reducing bias from confounders, there is growing support for a semi-parametric technique called matching, which extracts a subset of the original dataset so that the variable of interest is distributed "as if" randomly; essentially, the researcher "pre-processes" the dataset to induce balance among the observables before executing regression analysis (Ho *et al.* 2007).[5] I implement three different matching techniques: matching on propensity score, matching on Mahalanobis distance, and genetic matching. For propensity score matching, the probability of being disabled is modeled as a function of the other demographic covariates; each disabled individual is then paired with a non-disabled individual based on the proximity of the propensity score. This approach is appealing, because it works around the "curse of dimensionality," and in infinite populations, asymptotically induces balance among the observables (Rosenbaum and Rubin 1983).[6] The second alternative uses Mahalanobis distance, which measures the space between any two vectors of covariates, weighting the familiar Euclidean distance by the variance-covariance matrix of the covariates (Ho *et al.* 2007; Diamond and Sekhon 2013). Finally, Genetic Matching is a variation of both procedures that interposes an initially arbitary weighting matrix between the decomposition of the variance-covariance matrix. The program iteratively converges to a matrix that locally optimizes balance across specified covariates, including (if desired) the propensity score (Diamond and Sekhon 2013; Mebane and Sekhon 1998).

I execute each kind of matching with the R package 'Matching' (Sekhon 2011). Ho *et al.* (2007) recommend that the researcher use the entire set of pre-treatment covariates in the matching procedure. In this case, it is impossible to identify fully the variables that were realized prior to treatment, given that the dataset does not

[5] There is a healthy literature on the most effective techniques for matching, and when matching might be appropriate. A full review is beyond the scope of this paper, but prominent discussions of matching include Rubin (1973); Rosenbaum and Rubin (1983); Dehejia and Wahba (1999); Abadie and Imbens (2006); Ho *et al.* (2007); Gelman and Hill (2007); Angrist and Pischke (2008); Diamond and Sekhon (2013); and Morgan and Winship (2014), though this list is far from exhaustive. For a recent review, see Stuart (2010).

[6] However, if the sample is small, this method may be unable to facilitate balance among the observables (Diamond and Sekhon 2013).

chart the onset of disability. Accordingly, I have selected the demographic variables that are certainly determined prior to (or at least, concurrently with) disability: age, gender, race, and Hispanic origin. I also include education, employment, and family income to capture the socio-economic background of the individual, which sometimes plays a role in determining disability status.[7] And finally, in my implementation of genetic matching, I include the propensity score as a covariate, as well (*cf.* Diamond and Sekhon 2013). In all of the matching executions, I match exactly on state and cross-section; that is, I induce balance within each state and for each electoral year.

Prior to matching, the dataset exhibits great imbalance among the observable characteristics. Figure 3.1 displays the density plots of age distributions for disabled and non-disabled individuals in the unmatched dataset, the dataset matched on propensity score, the dataset matched on Mahalanobis distance, and the genetically-matched dataset, respectively.[8] It is immediately obvious that each of the matching methods dramatically improves the balance in age distributions between these two subsets. A cursory glance at the unmatched dataset (top, left) reveals that the disabled subpopulation is significantly older than the non-disabled. Propensity score matching improves this disparity, though not to the degree of Manahalobis matching or genetic matching. In the latter types of matched datasets, the plotted points barely deviate from the line of symmetry. Tables 3.A1 and 3.A2 of the Supplementary Materials describe the distribution for each covariate before and after matching (for the 2008 and 2012 cross sections, respectively), demonstrating that matching can assuage concerns about the orthogonality of "treatment" assignment.

Having created datasets that are less susceptible to bias from confounders, I can turn to the more traditional parametric models used in this literature to explain voter behavior. In particular, I aim to estimate the degree to which disability status affects voter registration and turnout, and whether any of the most prominent forms

[7] Admittedly, the choice to include these variables is more controversial, because the direction of causality is not fully identified. However, given the established association, and the existing claims that these factors can intervene in the relationship between disability and participation (Schur *et al.* 2002), I elect to include them.

[8] These graphs use the 2012 cross section; the same graphs are displayed for the 2008 cross section in Figure A1 of the Supplementary Materials.

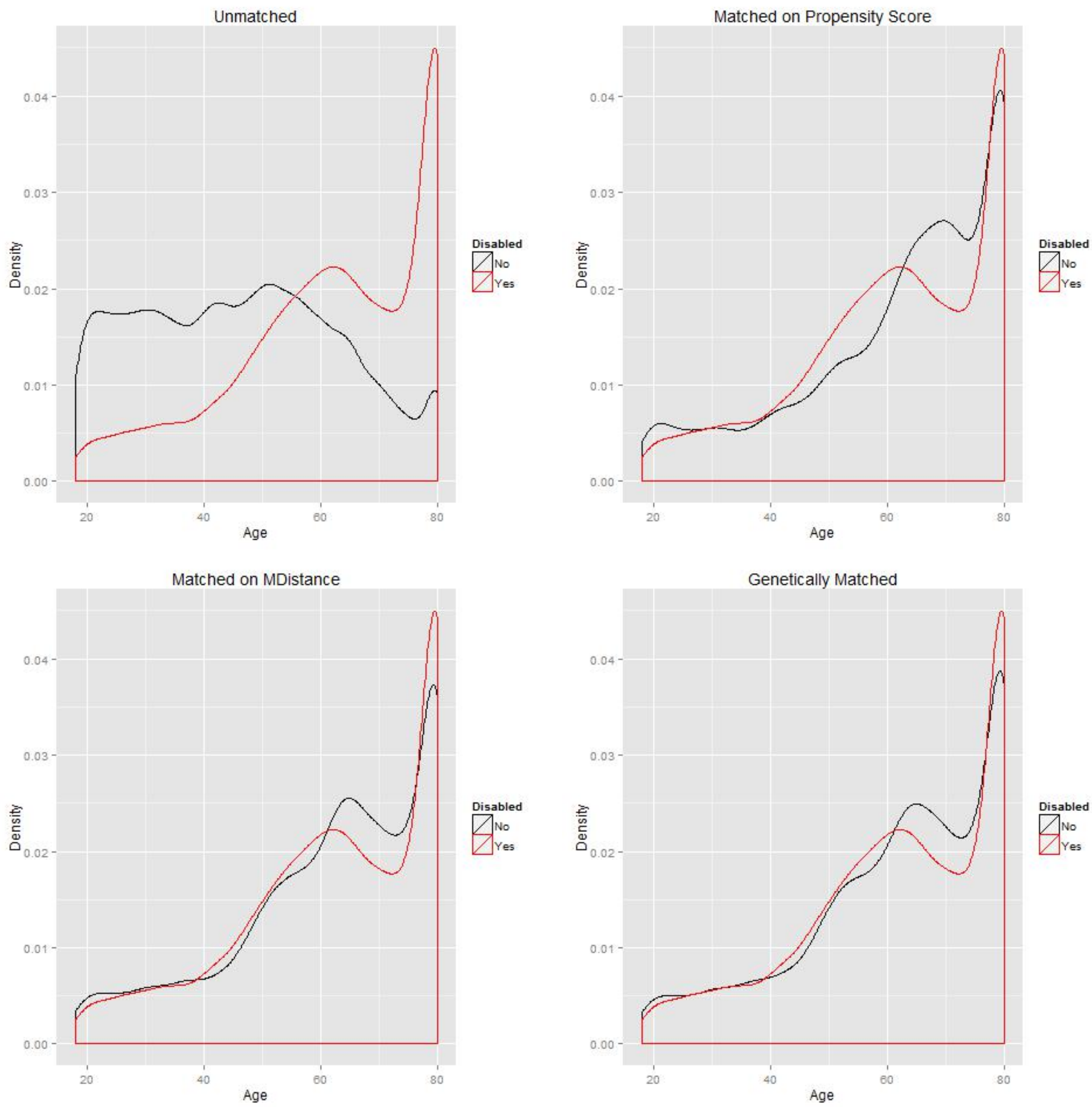


Figure 3.1: Density of Age Distribution by Disability Status and Dataset (2012)

of convenience voting intervene in those relationships.

3.5 Convenience Voting

The academic literature on the efficacy of convenience voting is quite extensive, albeit ambivalent (Alvarez, Levin, and Sinclair 2012).[9] The initial works offer optimism that policy liberalizations can yield a far more mobilized electorate (Rosenstone and Wolfinger 1978; Wolfinger and Rosenstone 1980). Many papers have highlighted the potential of Election Day Registration (hereafter, EDR) to improve turnout, in particular among racial minorities (Alvarez and Nagler 2007, 2008, 2011; Knack and White 2000; Brians and Grofman 2001; Burden *et al.* 2014; but see Glynn and Quinn 2011 for a competing perspective). Additionally, a few studies suggest that online registration, one of the most recent innovations in election law, increases turnout generally, and may improve the historical under-representation of young voters (Baretto *et al.* 2010; García Bedolla and Veléz 2013; Pellissier 2015).[10] Gronke, Galanes-Rosenbaum, and Miller (2007) and Berinsky, Burns, and Traugott (2005) find that all-postal voting increases turnout, though the latter paper argues that it does so by retaining existing voters, and eventually leads to a *less* representative electorate.

Several researchers caution that the research on election procedures may overstate their impact (Ansolabehere and Konisky 2006; Hanmer 2007, 2009; Glynn and Quinn 2011; Neiheisel and Burden 2012; Keele and Minozzi 2013). Meanwhile, a couple of papers find that early voting is actually associated with *lower* participation, perhaps due to a less patriotic electoral environment (Burden *et al.* 2014; see also Gronke, Galanes-Rosenbaum, and Miller 2007).[11] Berinsky (2005) expresses skepticism that any form of convenience voting significantly increases turnout, and argues instead that such liberalizations merely streamline the process for individuals who would vote regardless. These policies, though well-intentioned, may not address the true reasons

[9] Gronke *et al.* (2008) and Highton (2004) offer nice reviews.

[10] I will explore online registration's role within this literature more extensively in the next chapter.

[11] Another possible explanation is that regression results could be conflated by lower incidences of participation among the states that offer early voting, though such investigation is beyond the scope of this paper.

that people do not participate. Berinsky hypothesizes that the cognitive costs involved in obtaining political information may be the prevailing factor for many abstainers. If this is indeed the case, then we are targeting the wrong mechanism; and moreover, these individuals are less likely to be aware of convenience voting procedures. For individuals with cognitive disabilities, this possibility is particularly relevant; and more generally, Alvarez and Hall (2012) find that individuals with disabilities are less likely to read a political blog, read the newspaper, and listen to the news on the radio.

Only a few papers have explicitly considered how citizens with disabilities are responding to attempts to make voting less costly. Of those who voted in the 2008 and 2010 elections, individuals with disabilities reported voting by mail at significantly higher rates (Schur and Adya 2013; Schur and Kruse 2000; Alvarez and Hall 2012; see also Alvarez, Levin, and Sinclair 2012).[12] Notably, individuals with disabilities did not seem to have more difficulty procuring absentee ballots, though they did need more assistance (Alvarez and Hall 2012). Alvarez and Hall (2012) offer the most extensive insight into how individuals with disabilities respond to various convenience voting procedures. They regress both registration and turnout on a number of factors, including indicators for whether the individual's state allows particular forms of convenience voting; in some models, they also interact disability status with these state-level variables. EDR increases the probability of registration, but the magnitude of the impact may be larger for voters without disabilities. Additionally, no-excuse absentee and permanent absentee states are associated with increased likelihood of turnout, and there is additional boost for voters with disabilities. Altogether, these findings suggest that convenience voting procedures can improve participation among the disabled, even if the impact is more muted. It is important to continue to test this interaction given the persistent concerns about accessibility (*cf.* Bovbjerg 2013) and the role of advocates for the disabled in shaping electoral policy (*cf.* Stewart 2011).

[12]It is worth mentioning that this gap might be partially explained by states' varying eligibility requirements for this voting mode.

3.6 Data and Model

The data come from the 2008 and 2012 November cross sections of the *CPS*, an extensive, nationally-representative survey implemented by the Census Bureau. [13] Prior to 2008, the Census Bureau addressed disability differently (Alvarez and Hall 2012; Schur and Kruse 2011; Schur and Adya 2013), so I begin with this cross-section to minimize measurement error. The Census Bureau asks each individual a battery of questions about whether they experience various kinds of disability, and the disability indicator flags individuals who answer affirmatively to at least one of these questions. As evidenced by Table 3.A3 of the Supplementary Materials, individuals with disabilities report lower levels of participation for almost every type of demographic stratification. Obviously, disabilities vary widely in type and degree, so it may not be appropriate to treat disability status as homogeneous (Alvarez and Hall 2012). To ensure an adequate sample of disabled individuals for each state-year, though, I elect to include a single disability indicator; conceptually, the estimated coefficient and effect will be the weighted average across individuals with all types of disability. Matching already reduces the size of the dataset, and consequently, the power of the model; disaggregating the disability measure would further reduce the statistical power.[14] And finally, to measure competitiveness in a particular state for a given election, I use the margin of victory from the narrowest high-profile race available (President, Governor, U.S. Senator). My indicators of convenience voting come from a dataset originated by Cemenska *et al.* (2009), which I updated to include

[13] The size and representativeness of this dataset render it an appealing option for modeling participation. However, there is evidence that significant measurement error plagues unvalidated data on voter participation (see Katz and Katz 2010; Ansolabehere and Hersch 2012; Hur and Achen 2013). Individuals may respond incorrectly, due to imperfect recall or satisficing, or they may choose not to respond at all. Social scientists accommodate these concerns differently; some execute listwise deletion, while others attempt to estimate the responses via multiple imputation. I follow the Census Bureau's procedure, counting all those who do not answer affirmatively to the participation questions as abstainers. I listwise delete observations that are missing demographic information.

[14] My focus in this chapter is fairly broad. If future research projects prefer to consider specific kinds of disability, they can perhaps combat concerns about power by pooling additional cross-sections, or not restricting the match to state-year subsets.

Online Registration and recent legal changes.[15]

The primary estimand of interest is the analog of the Average Treatment Effect on the Treated.[16] Mathematically, this is equivalent to $\mathbb{E}[Y^1 \mid D = 1, X] - \mathbb{E}[Y^0 \mid D = 1, X]$. The first term of the expression is identified with observational data, and I impute the second using matching and regression analysis.

As described in Chapter 2, I conceptualize the propensity to participate as a latent variable, which we only observe in censored form as participation (or lack thereof).[17] Following Alvarez and Hall (2012), I consider both registration and turnout (separately). My base specification is as follows:

$$\mathbb{P}[Y_{ist}] = \Phi(\beta X_i + \zeta Z_{st} + \delta D_i + \gamma_s + \theta_t + \epsilon),$$

where Y is the binary indicator of participation, X is a standard list of demographic controls, Z is state-level covariates, D is an indicator of disability status, and γ and θ are state and year fixed effects, respectively. The vector X includes age, gender, race, Hispanic heritage, education, family income, employment, and residential mobility. The vector Z includes the metric of competitiveness described above, as well as indicators of convenience voting procedures.[18] I measure most of these variables conventionally, though it is worth noting that I categorize employment somewhat unusually. Most papers consider employment as a binary variable, with a value of unity indicating full-time employment and a value of null indicating otherwise. Instead, I treat employment as a categorical variable, classifying individuals as "employed," "unemployed," or "not in the labor force." Individuals who declare themselves out of the labor force might participate in elections systematically differently from in-

[15] This information is available at the National Council of State Legislatures website: `http:\\www.ncsl.org`.

[16] It seems inappropriate to label disability status a treatment, given that it is not a condition that researchers or policymakers ever impose. Accordingly, I am also not interested in the Average Treatment Effect of the entire population, since I am not looking to introduce a policy or impose a particular condition. I want to identify the relationship between disability and participation, conditioned on reporting disability.

[17] I specify a probit model, though other papers have used logit and scobit specifications.

[18] For the model of registration, this includes EDR and online registration. For the model of turnout, this includes EDR, online registration, no-fault early voting, no-excuse absentee voting, and all-postal voting.

dividuals who are unemployed but seeking work. This may be particularly true for individuals who report disabilities, given that their disabilities might prohibit them from the possibility of viable employment. My specification is similar to that of Alvarez and Hall (2012), though I pool cross-sections (and include year fixed effects) and modify the set of covariates slightly.[19]

I later extend the model to consider interactive effects that previous papers have identified as significant. More specifically, I interact disability with age, employment, and the convenience voting procedures. Schur and Kruse (2000) found that the deterrence of disability is primarily driven by the relationship between disability and employment, though they focused exclusively on spinal cord injuries. Schur *et al.* (2002) identified significant interactive effects between disability and age, as well as employment. More recently, using a difference-in-means test, Schur and Kruse (2011) identified similar participation behavior among the employed, whether they reported disability or not. The same study established a consistent gap between the disabled and non-disabled subgroups when stratified by age bracket. For the state-level covariates, Alvarez and Hall (2012) simulated a "typical" (modal) individual in the dataset, and found interesting patterns in how individuals with disabilities engage with their electoral environments. Those with disabilities were less likely to take advantage of EDR, though the effect of EDR was still positive. They were more likely to benefit from absentee voting, whether or not the state required an excuse. The total effect for early voting was negative, though less so for the subset of individuals with disabilities.[20]

Table 3.1: Effect of Disability on Participation by Dataset

	Registration	Turnout
Unmatched Dataset	-4.45% [-5.11%, -3.83%]	-8.32% [-9.03%, -7.60%]
Matched on PS	-3.63% [-4.55%, -2.63%]	-7.57% [-8.63%, -6.51%]
Matched on MD	-3.26% [-4.18%, -2.34%]	-6.34% [-7.35%, -5.30%]
Genetically Matched	-3.17% [-4.06%, -2.24%]	-6.25% [-7.41%, -5.20%]

3.7 Results

3.7.1 The Necessity of Matching

To understand how the imbalance in the original dataset can bias estimates, I fit the base model to each of the four datasets: unmatched, matched on propensity score, matched on Mahalanobis distance, and genetically matched. Recall that the primary quantity of interest for this study is the overall impact of disability on participation *among the subset of disabled voters.* The literature has identified a persistent depressive effect of disability on participation, but the magnitude of the impact has varied considerably (*cf.* Schur and Adya 2013). This lack of robustness may be an artifact of the alternate specifications or data samples, but it could also be driven partially by variation in how individuals with disability interact with their electoral environments over time. To my knowledge, all of the studies in this area have used unmatched datasets, and accordingly, the point estimates could be biased by confounders. Table 3.1 lists the point estimates and 95% bootstrapped confidence intervals for the effect of disability suggested by each dataset, and Figure 3.2 graphs the associated distributions.[21]

[19]Pooling improves the statistical power of the mode and reduces the influence of electoral idiosyncrasies (Leighley and Nagler 2014; Ansolabehere and Konisky 2006).

[20]Early voting states are often associated with reduced participation, perhaps due to civic atmosphere or other commonalities among the states that allow individuals to vote in person prior to Election Day (*cf.* Burden *et al.* 2014).

[21]Tables 3.A4 and 3.A5 in the Appendix display the full set of results for registration and turnout, respectively. Note that the effect on balance is not uniform across covariates; matching greatly improves the balance across disability status improves for certain characteristics (*e.g*, age) but

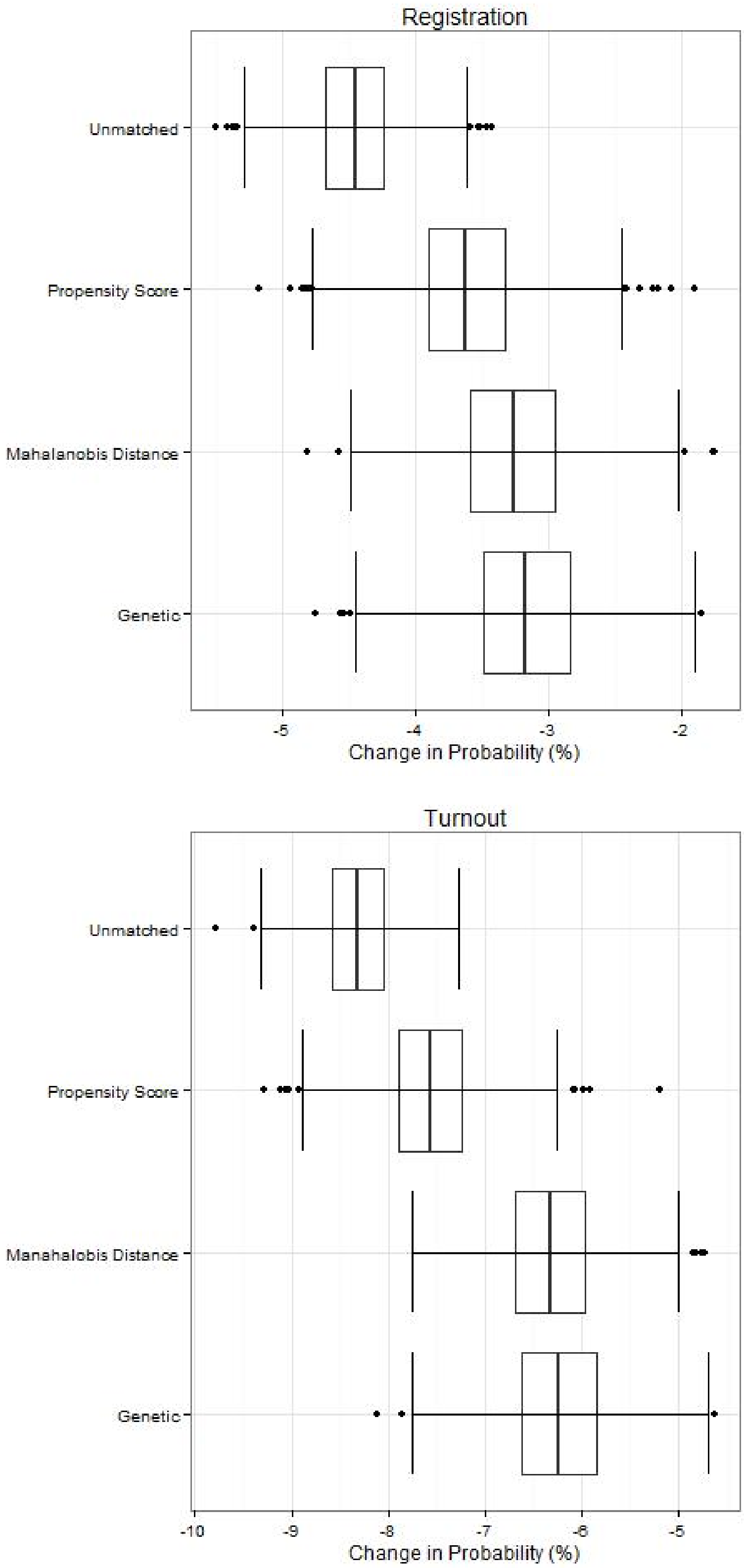


Figure 3.2: Effect of Disability on Participation by Dataset

These results lend credence to concerns that imbalance among the observables might bias the model, and in turn, postestimation. For both registration and turnout, the matched datasets produce lower point estimates of disability's effect than the unmatched dataset. It is worth mentioning that for both stages of participation, the point estimate for the unmatched data lies outside the 95% confidence interval for the genetically-matched data and the data matched on Mahalanobis distance, and near the boundary for the data matched on propensity score, as well. Notably, the dataset matched on propensity score produces higher estimates than the other matched datasets, and a couple of explanations exist. It is possible that the propensity score model is incorrectly specified, though if either the propensity score model or the participation model is correctly specified, the results will be valid (Ho *et al.* 2007). It should also be mentioned that matching on the (correctly-specified) propensity score induces balance on the observables *asymptotically*; in finite samples, this result does not hold (Diamond and Sekhon 2013).[22]

These results suggest important substantive conclusions, as well. For all datasets, disability's impact on turnout considerably exceeds its impact on registration. This result fits the intuitive narrative of how disability intervenes in participation; because disability often occurs or (at least intensifies) with age, it is quite possible that many individuals with disabilities are already registered voters by the time of onset. For one, the individuals might not have experienced cognitive and/or physical barriers to participation when they considered registration. Additionally, because disability status coincides with reduced civic participation more generally (Schur *et al.* 2002), those who were not disabled as they became eligible to vote may have experienced the typical level of mobilization efforts. Nonetheless, it noteworthy that even so, disability significantly reduces the propensity to be registered, as suggested in Alvarez and Hall (2012) and Schur, Shields, and Shriner (2005). Moreover, even the most conservative estimates suggest a persistent depressive effect of disability on partici-

actually reduces it for some of the less common characteristics (*e.g.*, Asian, other race indicators). This is probably exacerbated by my restriction that matching occur within state-year.

[22] It is worth noting that a more thorough exercise in propensity score matching would iteratively repeat the procedure until some objective balance criterion had been satisfied.

pation. Controlling for all other factors, and matching to reduce confounding bias, disability status is associated with an approximately 5 to 8% reduction in turnout. Given that certain studies suggest an independent effect of nearly 20% (Schur and Kruse 2011), my results suggest that these estimates are indeed artificially inflated by the statistical approach and bias from confounders. Even after matching, it is clear that despite legislative efforts to curtail the threats to accessibility that individuals with disabilities may face, for many the costs of voting are still unduly high.

3.7.2 Interactive Effects

To further probe the substantive implications, I incorporate interaction terms into the participation models.[23] For intuitive interpretation of binary response models, researchers can either identify a (hypothetical) individual of interest or average across all individuals in the sample. Following Hanmer and Kalkan (2013), I choose the latter approach; unless there is *a priori* reason to specify a particular realization of the demographic covariates, a broader perspective is more instructive. Table 3.2 lists the discrete differences for each of the independent variables; that is, for each entry, the quantity is obtained by the expression $\mathbb{P}(Y \mid X_{-I}, X_I = x1) - \mathbb{P}(Y \mid X_{-I}, X_I = x0)$, where Y is the outcome, X_I is the covariate of interest, evaluated at values $x1$ and $x0$, and X_{-I} is all other covariates (at their observed values). All values in the table are percentage points. It is worth mentioning that for the variables that are interacted, the effect measured below is the *total* effect.

Most of the individual-level variables exhibit the expected effect.[24] The propensity to participate increases with age, and females are also more likely to participate. All other factors held constant, Asians are less likely to register and vote than whites, and Hispanics are less likely than non-Hispanics. Interestingly, though, blacks are *more* likely to register and vote than whites, reversing the historical trend. A full investigation of this relationship is beyond the scope of this paper, though a report issued by the Census Bureau speculates that the uptick in participation among blacks

[23]For simplicity of exposition, I focus exclusively on the genetically matched dataset.
[24]For a thorough discussion of demographic predictors, see Leighley and Nagler (1992).

Table 3.2: Discrete Differences

	Registration		Turnout	
Disability: No → Yes	−1.28	(-4.15, 1.36)	−1.04	(-4.37, 2.26)
Age: 18 → 25	3.81	(3.31, 4.27)	3.69	(3.27, 4.10)
Age: 25 → 45	9.88	(8.63, 11.05)	10.16	(8.98, 10.07)
Age: 45 → 65	7.99	(7.11, 8.79)	9.11	(8.12, 10.07)
Gender: Male → Female	3.50	(2.75, 4.25)	4.38	(3.53, 5.24)
Race: White → Black	8.38	(6.96, 9.72)	13.27	(11.68, 14.95)
Race: White → Asian	−9.56	(-12.20, -7.05)	−11.19	(-14.11, -8.29)
Race: White → Other	0.61	(-1.29, 2.49)	0.09	(-1.75, 2.11)
Hispanic: No → Yes	−4.14	(-5.80, -2.67)	−4.25	(-6.05, -2.45)
Income: Level 8 → Level 11	3.32	(3.00, 3.66)	5.11	(4.73, 5.49)
Income: Level 11 → Level 14	2.97	(2.72, 3.24)	4.67	(4.36, 4.98)
Education: High School → Some College	0.87	(-0.23, 1.89)	0.00	(-1.17, 1.24)
Education: Some College → College	12.51	(11.43, 13.60)	15.23	(13.98, 16.46)
Education: College → Postgraduate	0.20	(-0.95, 1.31)	0.86	(-0.46, 2.28)
Employment: Employed → Unemployed	−6.27	(-7.48, -5.03)	−7.69	(-9.13, -6.17)
Employment: Employed → Not in Labor Force	−4.64	(-5.79, -3.59)	−4.87	(-6.14, -3.58)
Recent Mover: No → Yes	−6.50	(-7.32, -5.59)	−7.97	(-9.03, -6.93)
Margin of Victory: 1% → 5%	0.38	(-0.02, 0.84)	−0.02	(-0.50, 0.47)
EDR: No → Yes	−3.81	(-10.00, 1.37)	−1.54	(-8.52, 4.48)
Online Registration: No → Yes	0.06	(-1.81, 1.84)	1.25	(-0.99, 3.45)
Early: No → Yes			−4.74	(-11.98, 3.06)
Absentee: No → Yes			2.88	(-3.89, 9.45)
Postal: No → Yes			3.08	(-5.30, 10.32)

This table delineates the discrete differences in the outcome associated with each independent variable in the regression model, along with the 95% bootstrapped confidence intervals.

is due to a shock to the mobilization patterns among African Americans in response to the Obama campaign (2013). Less surprisingly, the propensity to participate increases with each level of income. For education level, the only significant leap occurs between "some college" and "college," suggesting that (in this subsample) the primary driver might be a binary indicator of a college degree. Those who are employed are more likely to engage than both those are unemployed and those who have are not in the labor force, and the gap is significantly wider for the unemployed; the estimated effects for membership of these two groups, however, are not statistically distinguishable, offering evidence against my hypothesis that these two subpopulations would behave differently. The effect of the last demographic variable, a metric of residential stability, is in the expected direction; those who have moved recently are less likely to be registered and to participate (McDonald 2008).

None of the coefficients on the state-level covariates (aside from the state fixed effects) exhibit statistical significance. Matching reduces the size of the dataset, and consequently, the statistical power of the model; it is possible that a larger dataset might shed more light on these associations. Additionally, some of the variation could be absorbed or conflated by the state-level fixed effects. In any case, these results should be interpreted with care. It is important to note that the *causal* effect of these convenience voting procedures is not fully identified, for any policy. These estimates represent the (weighted) change in turnout associated with states offering that particular method. To identify the causal effect, I would need to include at least one pre-treatment wave for each each state that offers the procedure, and all of these convenience voting procedures were implemented in at least one state prior to the 2008 election. Given that the Census only recently altered its survey design to capture disability more effectively (Schur and Kruse 2011; Alvarez and Hall 2012; Schur and Adya 2013), I can only include the 2008 and 2012 implementations.

For the purposes of this paper, the most interesting result is that the *total* effect of disability is not statistically distinguishable from the null in the extended model. In other words, disability status in and of itself is not a significant determinant of turnout. The negative impact of disability on participation identified in Table 3.1 is actually borne primarily by a specific subset of individuals who report disabilities. To isolate these interactive effects, I take the cross-partial derivative of the entire expression, with respect to disability status and the other interacted covariate of interest (for an extended discussion of identifying these effects in non-linear models, see Ai and Norton 2003).[25] Figure 3.3 depicts the distribution of these interactive effects.

These results strongly suggest that the primary mechanism through which disability reduces participation is in some way associated with employment status, echoing the conclusions of Schur and Kruse (2000) and Schur *et al.* (2002). In my model, the coefficient on the main effect of disability is greatly reduced in magnitude, and loses statistical significance (see Table 3.A6 in the Supplementary Materials), and the interactive effects of disability with unemployment and absence from the labor force are quite pronounced. A few different explanations exist.[26] The first explanation involves accessibility. Individuals who are disabled but currently employed have been able to adopt a lifestyle similar to those who do not report disabilities. The second relates to resources. Individuals who are employed are more likely to be able to afford monetary costs that participation may entail —for example, transportation or childcare. The final considers mobilization. Individuals who are employed may be more engaged in their communities, and therefore, more likely to encounter typical get-out-the-vote stimuli. Without further descriptive information about the individuals who are not employed, it is difficult to tease out which explanation plays a dominant role.

The other takeaway is somewhat disheartening for those working toward improving the representativeness of the electorate: There is no evidence from this investigation that convenience voting procedures particularly mitigate the heightened barriers faced by individuals with disabilities, for either registration or turnout. The interactive effect is not statistically distinguishable from the null for any of the electoral procedures I consider. This outcome accords with the hypothesis in Berinsky (2005)

[25] The following identity is relatively straightforward: $\frac{\Delta Pr}{\Delta X \Delta D} = \Phi(x = x1, d = 1) - \Phi(x = x0, d = 1) - \Phi(x = x1, d = 0) + \Phi(x = x0, d = 0)$.

[26] Much of the scholarship in this area focuses on accessibility, resources, and mobilization, per an excellent inquiry into participation by Verba, Schlozman, and Brady (1995).

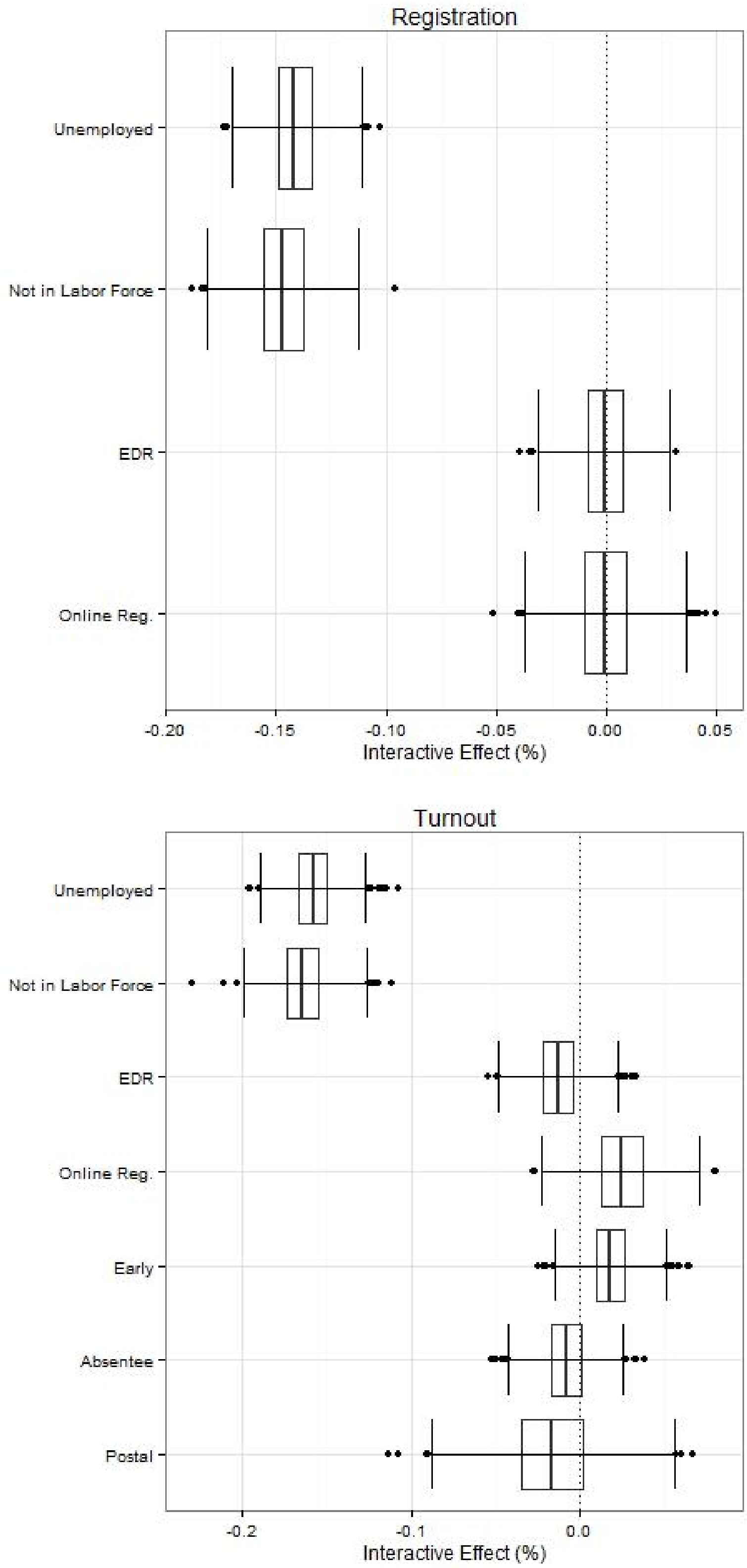


Figure 3.3: Interactive Effects on Participation

that these reforms do not improve the representativeness of the electorate, but simply expedite the process for individuals who would participate regardless. Notably, these results imply a different level of success of convenience voting than the slightly more optimistic findings of Alvarez and Hall (2012), who identified a marginally positive interactive effect with early and absentee voting on turnout. These authors also note that an individual with a disability is less likely to gain from EDR, though importantly, the *total* effect of EDR is still positive.

My results do not directly conflict with those of Alvarez and Hall (2012), however. Although matching has a number of appealing features, explained above, it does come with one important cost: By extracting a subset of the original dataset, it inevitably reduces the statistical power of the model. Alvarez and Hall consider each electoral cross section separately, but even though I pool across two cross sections, the number of individuals in a single cross section of CPS data nearly doubles the number of individuals in my matched (pooled) dataset. Furthermore, those authors measured absentee voting differently; more explicitly, they distinguish states that allow permanent absentee status. My data consider a single indicator of whether a state allows a resident to vote absentee without an excuse. And finally, Alvarez and Hall set a (singular) specified list of values for the other covariates. I take a broader scope, focusing not on a prototypical individual but using the observed values of the remaining covariates and averaging across the sample, as prescribed by Hanmer and Kalkan (2013).[27]

3.8 Discussion

The first contribution of this chapter is methodological. The matching estimates offer evidence that concerns about selection are both theoretically and empirically justified. Although the estimated effects of disability status on participation achieve statistical significance for all datasets, matching typically reduced the magnitude. Substantively,

[27]More technically, those authors focus on the marginal interactive effect, evaluated at a specific point on the plane. I average the discrete double difference over all points occupied in the plane.

my results suggest that the average effect of disability status on participation is negative and statically significant, but a closer investigation reveals that this finding is being driven almost entirely by individuals who are not employed. Thus, disability status mainly affects participation in its interaction with employment, a result that substantiates the conclusions in Schur and Kruse (2000) and Schur *et al.* (2002).

It is extremely important for future research to evaluate the mechanisms driving this result, since there are a variety of possible explanations. Chiefly, there is fairly broad concern among advocates for the disabled that despite legislative efforts, the barriers to voting (particularly in person) remain prohibitively high (Baker, Roy, and Moon 2005). Accessibility is improving, but certainly still characterizes modern elections (Bovbjerg 2009). However, election officials are typically subject to tight budget constraints, and even the most sympathetic to accessibility concerns may be logistically prohibited from implementing policy changes. Advocates interested in policy reform should consider local pilot studies and qualitative feedback to prioritize which features to target. Secondly, the barriers may be cognitive rather than physical; that is, even if the polling place is physically accessible, information about the campaigns and political process may not be (Alvarez and Hall 2012). Finding a means of disseminating political information more effectively might be an important next step in stimulating turnout among all under-represented subgroups, including the disabled (*cf.* Berinsky 2005). And finally, it remains possible that those with disabilities are simply less interested in politics; they may feel alienated by their current representatives, jaded by the political system, or simply disinclined to participate. Apathy is fairly widespread among non-voters, and those interested in mobilization must find some way to capture the interest of abstainers. Any combination of these factors —and others —could contribute to the reduced likelihood of participation among the disabled, and subsequent research should find some way to quantify the composition of these competing mechanisms' roles. To improve the participation rate, we must identify why the existing system and culture induce low mobilization.

3.9 Supplementary Materials

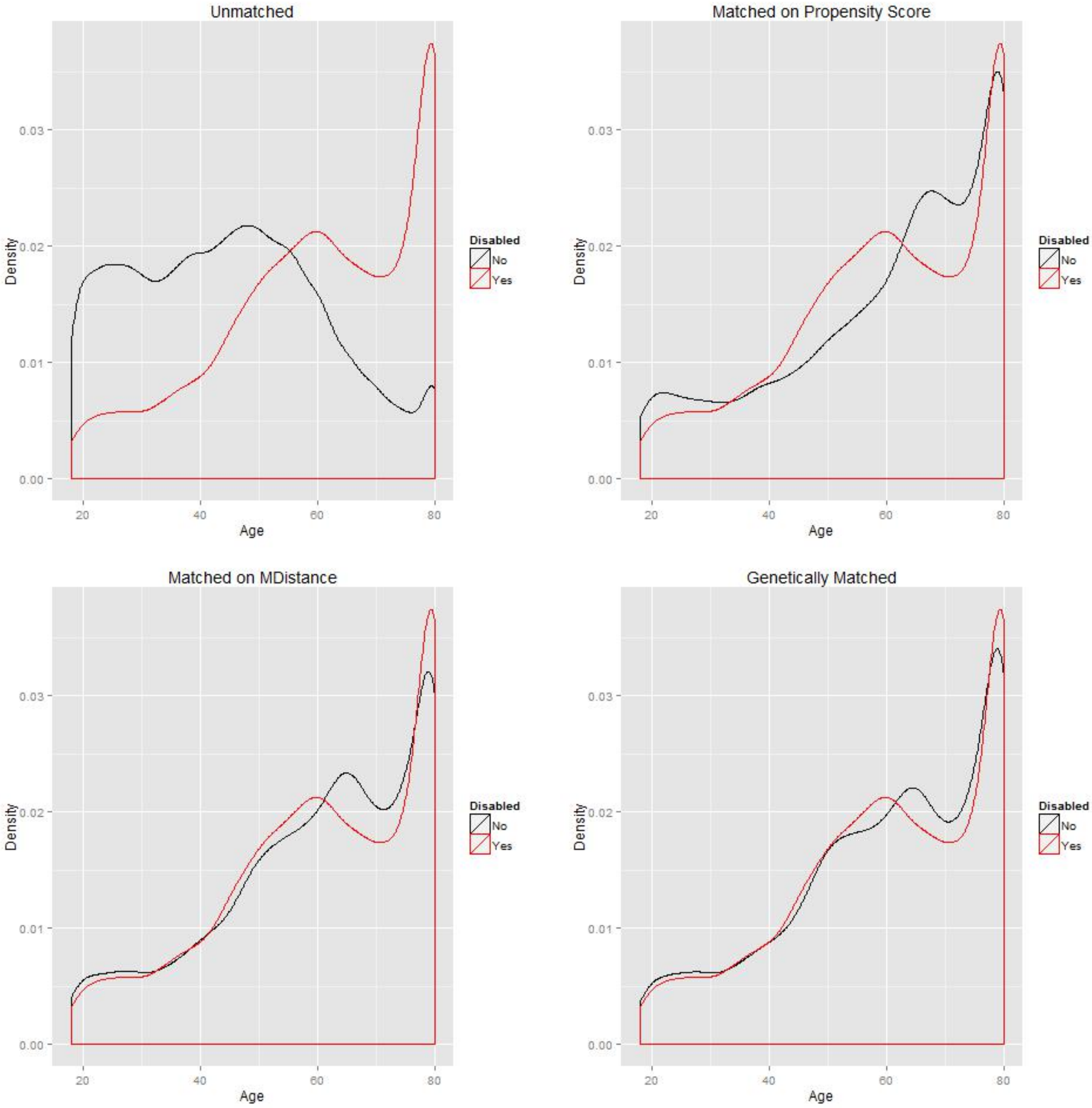


Figure 3.A1: Density of Age Distribution by Disability Status and Dataset (2008)

Table 3.A1: Balance Statistics, Before and After Matching (2008)

	Standard Mean Difference			
	Before Matching	*Propensity Score*	*Mahalanobis Distance*	*Genetic*
Age	88.95	−1.29	2.83	1.79
Female	2.69	−2.07	0.02	−0.35
Black	3.71	2.82	−0.77	0.44
Asian	5.64	2.72	6.85	8.27
Other Race	−0.52	−4.05	−2.87	−4.42
Hispanic	−6.69	1.80	0.00	0.04
High School	−34.89	−0.97	1.29	2.84
Some College	43.12	−1.40	−3.97	−10.15
Postgraduate	−7.93	1.93	2.77	8.01
Unemployed	114.53	−1.87	0.64	0.07
Not in Labor Force	−9.36	−1.92	0.40	0.27
Income	−73.14	−3.88	−4.99	−3.64

	Variance Ratio (Treated/Control)			
	Before Matching	*Propensity Score*	*Mahalanobis Distance*	*Genetic*
Age	1.06	0.89	0.99	0.99
Female	1.00	1.00	1.00	1.00
Black	1.12	1.09	0.98	1.01
Asian	1.41	1.16	1.55	1.75
Other Race	1.01	1.09	1.06	1.10
Hispanic	0.81	1.07	1.00	1.00
High School	0.67	0.98	1.02	1.05
Some College	1.01	1.01	1.02	1.05
Postgraduate	0.89	1.03	1.05	1.16
Unemployed	1.04	1.02	0.99	1.00
Not in Labor Force	0.65	0.90	1.02	1.02
Income	1.39	1.05	1.06	1.04

Table 3.A2: Balance Statistics, Before and After Matching (2012)

	Standard Mean Difference			
	Before Matching	*Propensity Score*	*Mahalanobis Distance*	*Genetic*
Age	94.09	−2.78	3.26	1.85
Female	3.38	−3.94	−0.09	0.07
Black	4.44	3.40	0.57	1.20
Asian	−13.21	0.76	−0.82	−0.44
Other Race	4.46	−0.05	1.18	1.97
Hispanic	−9.71	3.69	0.07	2.12
High School	44.18	2.73	−2.85	−3.81
Some College	−7.14	−3.04	1.03	1.78
Postgraduate	−24.55	0.65	1.49	1.45
Unemployed	−11.87	−1.44	2.05	4.54
Not in Labor Force	120.40	−1.62	0.36	0.83
Income	−67.46	−5.14	−4.24	−4.77

	Variance Ratio (Treated/Control)			
	Before Matching	*Propensity Score*	*Mahalanobis Distance*	*Genetic*
Age	0.94	0.95	0.99	1.00
Female	1.00	1.01	1.00	1.00
Black	1.13	1.09	1.01	1.03
Asian	0.53	1.05	0.95	0.97
Other Race	1.30	1.00	1.06	1.11
Hispanic	0.74	1.16	1.00	1.09
High School	1.05	0.99	1.01	1.01
Some College	0.90	0.95	1.02	1.03
Postgraduate	0.53	1.03	1.06	1.06
Unemployed	0.59	0.92	1.14	1.37
Not in Labor Force	0.88	1.02	1.00	0.99
Income	1.28	1.03	1.05	1.06

Table 3.A3: Participation Rates by Subgroup (Unweighted)

	Registered			Voted		
	Disability	No Disability	Gap	Disability	No Disability	Gap
Age 17 - 25	41.32	61.41	−20.09	28.10	48.72	−20.62
Age 26 - 45	65.65	79.51	−13.86	46.82	67.90	−21.08
Age 46 - 65	74.69	86.97	−12.28	64.76	78.88	−14.12
Age 66+	85.53	91.36	−5.83	74.23	86.25	−12.02
Male	76.18	79.16	−2.98	65.67	68.49	−2.82
Female	78.21	83.78	−5.57	65.19	75.36	−10.17
White	77.53	81.91	−4.38	65.65	72.24	−6.59
Black	80.13	85.54	−5.41	71.79	78.42	−6.63
Asian	52.54	67.39	−14.85	33.90	58.19	−24.29
Other Race	75.82	73.93	1.89	57.14	60.43	−3.29
Not Hispanic	77.60	82.75	−5.15	65.78	73.45	−7.67
Hispanic	71.51	67.69	3.82	59.30	55.92	3.38
High School	70.90	70.08	0.82	57.83	58.18	−0.35
Some College	84.30	82.42	1.88	71.96	71.85	0.11
College	88.82	89.96	−1.14	79.19	82.11	−2.92
Postgraduate	90.67	94.82	−4.15	88.00	90.69	−2.69
Employed	81.49	82.68	−1.19	71.90	72.88	−0.98
Unemployed	71.43	70.53	0.90	61.90	58.22	3.68
Not in Labor Force	76.38	80.70	−4.32	63.88	72.43	−8.55
Did Not Move	81.28	86.36	−5.08	70.56	78.17	−7.61
Recent Mover	68.06	73.65	−5.59	53.66	62.01	−8.35
No EDR	77.52	81.15	−3.63	65.54	70.96	−5.42
EDR	76.07	83.57	−7.50	64.82	77.22	−12.40
No Online Reg	77.01	82.03	−5.02	64.32	72.08	−7.76
Online Reg	78.01	80.32	−2.31	68.74	72.17	−3.43
No Early				65.35	74.42	−9.07
Early				65.45	70.54	−5.09
No Abs				64.16	71.81	−7.65
Abs				66.76	72.41	−5.65
No Mail				65.14	71.98	−6.84
Mail				72.57	76.19	−3.62

Table 3.A4: Effect of Disability on Registration by Dataset

	Full Dataset		**Matched on PS**		**Matched on MD**		**Genetic Matching**	
	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)
Constant	0.378	(0.049)	0.483	(0.108)	0.567	(0.109)	0.574	(0.108)
Disabled	−0.179	(0.013)	−0.211	(0.024)	−0.176	(0.024)	−0.194	(0.024)
Age	0.016	(0.000)	0.017	(0.001)	0.017	(0.001)	0.017	(0.001)
Female	0.175	(0.008)	0.184	(0.021)	0.165	(0.016)	0.174	(0.015)
Black	0.366	(0.017)	0.377	(0.034)	0.400	(0.034)	0.366	(0.033)
Asian	−0.329	(0.023)	−0.307	(0.047)	−0.317	(0.046)	−0.338	(0.046)
Other Race	0.076	(0.017)	0.051	(0.036)	0.053	(0.035)	−0.161	(0.035)
Hispanic	−0.213	(0.014)	−0.194	(0.029)	−0.172	(0.029)	−0.180	(0.029)
High School	−0.511	(0.013)	−0.543	(0.025)	−0.529	(0.025)	−0.549	(0.025)
Some College	−0.581	(0.013)	−0.574	(0.026)	−0.594	(0.026)	−0.590	(0.026)
Postgraduate	−0.133	(0.016)	−0.121	(0.032)	−0.149	(0.031)	−0.161	(0.032)
Unemployed	−0.201	(0.012)	−0.202	(0.024)	−0.198	(0.024)	−0.208	(0.024)
Not in Labor Force	−0.165	(0.012)	−0.156	(0.023)	−0.164	(0.022)	−0.154	(0.022)
Income	0.052	(0.001)	0.051	(0.002)	0.050	(0.003)	0.048	(0.002)
Recent Mover	−0.247	(0.009)	−0.241	(0.017)	−0.240	(0.017)	−0.230	(0.017)
Margin of Victory	−0.001	(0.001)	−0.001	(0.002)	−0.002	(0.002)	−0.001	(0.002)
EDR	−0.070	(0.057)	−0.047	(0.117)	−0.097	(0.115)	−0.034	(0.118)
Online Registration	0.016	(0.018)	0.010	(0.037)	0.031	(0.037)	0.055	(0.090)
N	147,922		39,341		39,341		39,341	
Pseudo R^2								
Effect of Disability	-5.03		-5.21		-4.25		-4.85	
95% CI	[-5.88, -4.27]		[-6.78, -3.66]		[-5.90, -2.61]		[-6.44, -3.26]	

Table 3.A5: Effect of Disability on Turnout by Dataset

	Full Dataset		**Matched on PS**		**Matched on MD**		**Genetic Matching**	
	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)
Constant	−0.996	(0.074)	−0.774	(0.153)	−0.564	(0.156)	−0.510	(0.156)
Disabled	−0.298	(0.016)	−0.283	(0.030)	−0.225	(0.030)	−0.245	(0.030)
Age	0.022	(0.000)	0.023	(0.001)	0.023	(0.001)	0.024	(0.001)
Female	0.126	(0.010)	0.123	(0.019)	0.150	(0.019)	0.160	(0.020)
Black	0.443	(0.019)	0.445	(0.037)	0.503	(0.037)	0.474	(0.036)
Asian	−0.490	(0.027)	−0.462	(0.056)	−0.453	(0.054)	−0.489	(0.056)
Other Race	−0.070	(0.030)	−0.072	(0.063)	−0.053	(0.063)	−0.060	(0.062)
Hispanic	−0.117	(0.019)	−0.099	(0.037)	−0.112	(0.037)	−0.122	(0.038)
High School	−0.675	(0.013)	−0.656	(0.025)	−0.664	(0.025)	−0.686	(0.025)
Some College	−0.208	(0.015)	−0.206	(0.029)	−0.211	(0.029)	−0.195	(0.029)
Postgraduate	0.287	(0.022)	0.293	(0.043)	0.270	(0.043)	0.265	(0.043)
Unemployed	−0.104	(0.024)	−0.046	(0.047)	−0.086	(0.047)	−0.079	(0.047)
Not in Labor Force	−0.113	(0.013)	−0.100	(0.024)	−0.127	(0.024)	−0.121	(0.024)
Income	0.048	(0.002)	0.051	(0.003)	0.050	(0.003)	0.048	(0.003)
Margin of Victory	−0.010	(0.005)	−0.033	(0.013)	−0.047	(0.013)	−0.048	(0.013)
EDR	0.254	(0.126)	0.749	(0.289)	0.921	(0.288)	0.937	(0.291)
Online Registration	0.029	(0.035)	−0.013	(0.083)	0.045	(0.082)	0.041	(0.083)
Early Voting	0.220	(0.076)	0.320	(0.165)	0.356	(0.169)	0.336	(0.168)
Absentee Voting	0.211	(0.044)	0.240	(0.102)	0.315	(0.105)	0.331	(0.105)
Postal Voting	0.192	(0.067)	−0.066	(0.149)	−0.227	(0.151)	−0.277	(0.150)
N	147,922		39,341		39,341		39,341	
Pseudo R^2								
ATT	-9.04		-8.54		-6.83		-7.39	
95% CI	[-10.01, -8.14]		[-10.43, -6.71]		[-8.65, -5.06]		[-9.08, -5.62]	

Table 3.A6: Participation Models, with Interactions

	Registration		Turnout	
	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)
Constant	0.283	(0.149)	−0.210	(0.149)
Disabled	0.061	(0.062)	0.152	(0.062)
Age	0.018	(0.001)	0.019	(0.001)
Female	0.142	(0.015)	0.145	(0.014)
Black	0.387	(0.035)	0.495	(0.032)
Asian	−0.341	(0.047)	−0.344	(0.045)
Other Race	0.024	(0.037)	0.003	(0.034)
Income	0.045	(0.002)	0.055	(0.002)
High School	−0.588	(0.026)	−0.521	(0.023)
Some College	−0.557	(0.026)	−0.521	(0.024)
Postgraduate	0.012	(0.034)	0.036	(0.029)
Unemployed	−0.223	(0.030)	−0.200	(0.028)
Not in Labor Force	−0.152	(0.026)	−0.100	(0.024)
Recent Mover	−0.256	(0.017)	−0.257	(0.016)
Margin of Victory	0.004	(0.002)	0.000	(0.002)
EDR	−0.152	(0.116)	−0.038	(0.105)
Online Registration	0.003	(0.038)	0.017	(0.038)
Early Voting			−0.182	(0.131)
Absentee Voting			0.106	(0.114)
Postal Voting			0.125	(0.139)
Interactions				
Age	−0.002	(0.001)	−0.004	(0.001)
Unemployed	−0.113	(0.046)	−0.179	(0.042)
Not in Labor Force	−0.140	(0.053)	−0.207	(0.048)
EDR	0.006	(0.045)	−0.039	(0.043)
Online Registration	0.000	(0.057)	0.076	(0.059)
Early Voting			0.073	(0.043)
Absentee Voting			−0.033	(0.042)
Postal Voting			−0.061	(0.094)
N	39,341		39,341	
Pseudo R^2	0.123		0.127	

Chapter 4

In Line or Online? American Voter Registration in the Digital Era

4.1 Summary

Within the past decade, an increasing number of states have begun to allow their residents to register as voters electronically. Like other efforts to increase political participation, though, the actual impact on registration and turnout remains unclear. Although other voting liberalizations have received a fair amount of scrutiny, the peer-reviewed literature does not include a systematic exploration of how individuals are responding to online registration. This chapter estimates an individual-level model of the impact of online registration on the propensities to register and vote. The results suggest that online registration may be one of the more successful implementations of convenience voting, though its impact is still fairly minimal. Perhaps even more importantly, its effects seem to be concentrated most highly among young adults and those who have moved recently, two subgroups that are consistently under-represented at the polls. At the aggregate level, several states in both the 2008 and 2012 Presidential elections could have experienced different outcomes had they offered online registration.[1]

[1] A version of this chapter appears in Pellissier (2015).

4.2 Introduction

Online voter registration is the byproduct of sustained interest in stimulating participation and the modernization of the election process. It is the latest innovation in a series of policies that aim to reduce the costs associated with electoral participation. Beginning with Arizona in 2002, a number of states now allow their residents to complete and submit their registration paperwork over the Internet. Media outlets and election administrators have enthusiastically announced that thousands of citizens have taken advantage of this option.[2]

Without further inquiry, however, we cannot know if these individuals would have registered regardless, and if online registration will merely have a substitutive impact. A few recent papers address this question (notably, Baretto *et al.* 2010 and García Bedolla and Veléz 2013), and several news outlets have published articles, but the body of peer-reviewed literature is entirely silent. Meanwhile, legislators, election administrators, and the press have touted online registration as a natural evolution in an increasingly technological election environment, and an innovation that will encourage participation from historically underrepresented demographic subgroups, particularly young adults.

Yet other voting reforms have been accompanied by similar promises, and failed to produce the eagerly-anticipated results. Instead, there is evidence that these reforms simplify the process for the politically engaged, but fail to stimulate turnout from those who typically abstain (Berinsky 2005). For decades, turnout in American presidential elections has hovered around 60% of the voting-eligible population (Leighley and Nagler 2014). Turnout in midterm and local elections typically has dipped even lower; recent turnout in midterm elections has hovered between 40% and 42%.[3] Despite concentrated efforts to mobilize eligible voters, low turnout rates have proved pervasive, as the United States perennially lags behind other democracies in participation rates (Leighley and Nagler 1992; Hanmer 2007, 2009).

[2]See, for example, Merl (2012).

[3]More specifically, 40.5% in 2002, 41.3% in 2006, and 41.7% in 2010, according to McDonald's estimates, per `http://elections.gmu.edu/voter_turnout.htm`.

To assess whether online registration reshapes the turnout puzzle, I use an individual-level model that estimates the relationship between online registration and political behavior, at both the registration and turnout stages. Although its enduring legacy remains to be seen, I find evidence that online registration *does* somewhat effectively stimulate participation in the voting process. Furthermore, the effect seems to become more pronounced as it becomes institutionalized within the state. I also execute regressions on subsamples of theoretical interest —young adults, ethnic minorities, and those who moved recently. The results suggest that young adults and recent movers are utilizing this registration opportunity to some degree, but it is unlikely to generate a more ethnically representative electorate. And in the final component of analysis, I use the fitted model to predict counterfactual aggregate turnout in states that did not allow online registration in 2008 and 2012. The predicted influx of voters exceeds the margin of victory in the Presidential contest for a handful of states, though it is difficult to know whether electoral outcomes would change without additional data. I explore this question in the next chapter.

In the following sections, I first present a brief review of the broader literature on convenience voting. I then offer an overview of the limited history of online registration in American elections, engaging the few inquiries into its efficacy. The next section details the econometric methods I use to tackle this question of causal inference. After presenting the results at the individual level, I use the fitted model to predict aggregate turnout for the 2008 and 2012 elections if control states had introduced online registration. I conclude with a discussion of the results, and the implications for this and other mobilization efforts.

4.3 Convenience Voting

Since the initial inquiry of Rosenstone and Wolfinger (1978), a number of voices have contributed to the debate of the efficacy of various voter reforms.[4] A tension has emerged within the literature, as some scholars have identified a meaningful impact

[4] For nice reviews of the literature, see Gronke *et al.* (2008) and Highton (2004).

while others have painted a murkier picture. The most optimistic papers suggest that certain liberalizations (particularly EDR) can substantially increase turnout, and perhaps more importantly, yield an electorate that more accurately reflects the demographic distribution of voting-eligible citizens (*e.g.*, Alvarez and Nagler 2007, 2008, 2011; Knack and White 2000; Brians and Grofman 2001; Burden *et al.* 2014). These papers acknowledge that no reform is a panacea for the turnout problem, but they express hope that certain reforms can make participation more accessible for some subgroups that are systematically under-represented.

Other papers on convenience voting arrive at more conservative estimates and more ambiguous conclusions. These papers tend to emphasize that while specific implementations have yielded isolated success stories, by and large, we simply have not seen the promised and anticipated results (Hanmer 2007, 2009; Neiheisel and Burden 2012; Keele and Minozzi 2013). Gronke, Galanes-Rosenbaum, and Miller (2007) contend that all-postal voting stimulates turnout, but point out that early voting is correlated with *reduced* participation. Ansolabehere and Konisky (2006) approach this question somewhat differently, asking how the introduction of registration laws *depressed* turnout in New York and Ohio; the authors conclude that the effect was actually fairly muted, leading them to suggest that registration barriers are not an overwhelming factor in the turnout calculus. Importantly, even local successes of convenience voting might warrant a caveat; convenience voting procedures might simply aid retention, rather than increase engagement (Berinsky, Burns, and Traugott 2005).

Finally, the most dismal studies suggest that all these reforms simply streamline the process for those who are already interested in politics, and that to engage non-voters, we must entirely re-conceptualize our efforts to mobilize voters (Berinsky 2005; see also Glynn and Quinn 2011). Berinsky (2005) aptly points out that reform efforts have targeted the more obvious costs, but have thus far failed to address the barriers introduced by cognitive costs. That is, to entice non-voters, we must find some means of communicating information about politics in a more accessible manner. And furthermore, all these attempts to stimulate participation might have pernicious unintended consequences (Berinsky 2005; Berinsky, Burns, and Traugott 2005). One potential drawback of EDR, for instance, is that delayed registration might hinder parties in their mobilization efforts, since those who register on Election Day are not included in any registration databases (Neiheisel and Burden 2012). Another paper suggests that early voting actually alters the civic atmosphere surrounding the electoral process and diminishes the "Get Out the Vote" effort (Burden *et al.* 2014).

Thus, although there is a healthy literature on barriers to participation, the debate on the efficacy of convenience voting procedures remains unresolved. Moreover, the peer-reviewed literature has yet to establish a niche for online registration, and so it is unclear whether its implementation in any way shifts the debate. Given the increasing popularity of this registration alternative, this chapter offers an important initial investigation. Fortunately, the literature on convenience methods, particularly EDR, offers a nice methodological blueprint for this line of inquiry.

4.4 Registration in the Digital Era

Before proceeding further, it is useful to consider online registration in the context of the modern electoral landscape, a landscape whose frontier is a function of the available technology. In the wake of the high-profile technological failures of the 2000 presidential election, political scientists have devoted considerable attention to how we can modernize the technology in election administration (Alvarez and Hall 2014). Academics across several disciplines collaborated to form the Caltech/MIT Voting Technology Project, which advances research that addresses obsolete technology. Proposals to incorporate new forms of technology are often met with public resistance, with parties worrying that their opposition will exploit the new technology (Alvarez and Hall 2008).

Registration problems are a consistent source of concern in American elections (Ansolabehere and Hersh 2014). The advent of online registration in America coincides nicely with legislative efforts to utilize technology to improve the registration process. In 2002, Congress passed the Help America Vote Act, which requires states to maintain electronic registered voter databases; this feature allows potential voters to verify their registration status prior to Election Day, and perhaps more importantly, the registration closing date for their states of residence. Online registration is a somewhat natural extension, then, of the digitalization of registration records.

Recent work has suggested that online registration will allow more accurate registration databases: Individuals verify their own personal information on registration forms, and digitalized depositories will streamline the process of cross-referencing pollbooks across jurisdictions (PCEA 2014). Moreover, there is evidence that the Internet is a viable tool for engaging the citizenry. With information about candidates, issues, and campaigns accessible online, citizens may find obtaining political knowledge less costly, and may consequently become more politically active (Tolbert and McNeal 2003). Future elections could feature the widespread use of online registration drives at schools and libraries. As online registration gains traction, the literature should explore how this policy influences mobilization, as other forms of convenience voting have had important implications for campaigns (Gronke *et al.* 2008).

It is important to clarify just what online registration entails, and how it differs from other registration mechanisms. Citizens in every state may fill out registration paperwork online, but in most, they must print and return their paperwork via fax or postal mail. The states that offer online registration allow residents to *submit* the forms online. Typically, this option is limited to citizens with a driver's license or other form of state-issued identification. The website retrieves the citizen's signature from that form of identification (which is stored electronically) for the voter registration paperwork.

Arizona spearheaded this movement in 2002, and eighteen states have since followed suit:[5] Washington (2008); Kansas (2009); Colorado, Indiana, Louisiana, Oregon, and Utah (2010); New York (2011); California, Maryland, Nevada, and South Carolina (2012); Minnesota, Virginia (2013); Connecticut, Delaware, Georgia, Illinois, Missouri (2014); and Hawaii, Massachusetts, Nebraska, and West Virginia (legislation passed, but not yet implemented). This voting initiative has received support

[5] For a current list, see `http://www.ncsl.org/legislatures-elections/elections/electronic-or-online-voter-registration.aspx`.

and criticism from both major political parties. Typically, the majority party in the state legislature has supported the legislation, and some members of the minority party have vocalized concerns of fraud.[6] By and large, however, policymakers of both major political parties have embraced online registration (PCEA 2014).

Proponents point out that this form of registration is much cheaper for the state and leads to fewer clerical errors, since individuals —rather than bureaucrats or third-party registrants —verify the accuracy of their own information. In Arizona, for example, the estimated difference in cost for the state is approximately \$ 0.80 per registrant (Barreto *et al.* 2010). Such features make online registration an appealing alternative from an administrative perspective, given the inaccuracies that perennially plague registration databases (Alvarez and Hall 2014). By adding an online alternative, potential registrants have an option that is often more convenient and less time-intensive. On the other hand, online registration may "crowd out" proven methods of encouraging citizens to register. Furthermore, voters may fall into the "procrastination trap": Registration appears so costless that they put it off until another time, and realize too late that they have neglected the process (*cf.* Bennion and Nickerson 2011). The primary concern among the popular press and dissenting legislators, however, appears to be potential fraud; detractors worry that there will not be enough oversight to ensure that only voting-eligible citizens are using this option.[7]

As online registration has gained traction, it has naturally attracted a brighter spotlight in the public debate on election administration. The 2012 Presidential election featured many news articles that questioned how online registration would affect turnout. To date, though, very little research directly examines whether online registration encourages *additional* participation. Studies that have been published outside of academic journals give grounds for optimism about the potential impact of online registration, particularly for young adults, who traditionally participate at lower rates. Baretto *et al.* published a report of how citizens are utilizing online

[6] See Achohido (2012).

[7] As just one example, see Perlroth (2012).

registration in the two states that implemented it first, Arizona and Washington (2010). Those who are registering via this mechanism are disproportionately young. The authors laud the fact that those who register online actually cast ballots at higher rates than their counterparts, suggesting that this method of registration might do more to stimulate actual turnout than other alternatives. And in a study of online registrants in California prior to the most recent election, García Bedolla and Veléz (2013) find that most belonged to the youngest cohort of eligible voters, though a large number also belonged to other age brackets. Additionally, many of the online registrants in San Diego and Alameda Counties belonged to low- or middle-income brackets.[8]

While the above papers consider case studies of online registration in particular states, I am interested in gauging the impact on the decision-making process at an individual level. Furthermore, I want to predict how online registration could shift participation rates in states that have not yet implemented it. Before I describe my identification strategy for these tasks, it is worth quickly reviewing the theoretical decision-making process for electoral participation.

4.5 The Voting Calculus

Although the act of voting in today's American elections is technically "free," the actual process is nonetheless accompanied by certain inconveniences. As described in Chapter 2, these costs depress participation (Downs 1957; Riker and Ordeshook 1968). Because registration is a prerequisite for voting, we can also expect these costs to affect behavior. Currently, all states except North Dakota require that citizens register prior to (or in some states, concurrently with) casting a ballot. The United States is somewhat unusual among Western democracies in that registration is both voluntary and passive (Alvarez and Hall 2014); that is, individuals do not face any kind of penalty for neglecting to register, and the burden of registration is on the

[8] The authors examined only these counties, selected for their volume of online registrants and socioeconomic heterogeneity.

citizen rather than the government. Embedded within this utilitarian framework, then, are any costs that the individual must incur to complete the registration process.

We can expect online registration to increase aggregate turnout if it provides a substantial reduction in cost for a subset of non-registrants. More explicitly, for some individuals, the introduction of online registration must reduce the cost of registration enough that the expected utility crosses the threshold from negative to positive. Put differently, the cost of the entire voting process must be sufficiently low with the advent of online registration and prohibitively high otherwise.

It is reasonable to infer that the advent of online registration will reduce the overall costs associated with electoral participation. Citizens can complete this stage of the voting process from their homes, offices, or schools without the inconvenience of driving to a government building or waiting in line; they can register at their most convenient time, without having to consider the business hours of registration sites. And although many states address these concerns by allowing their residents to mail or fax registration forms, even this additional step might be exceedingly cumbersome. Additionally, there is some evidence that online registration reduces the incidence of clerical error among registration pollbooks (PCEA 2014). While states often have provisional voting procedures that accommodate such inaccuracies, these kinds of clerical errors at best increase the hassle of the voting process and at worst effectively disenfranchise would-be voters (*cf.* Alvarez and Hall 2014; Ansolabehere and Hersh 2014).

As is the case with EDR, it is unlikely that online registration will in and of itself stimulate participation among registered non-voters (*cf.* Glynn and Quinn 2011). Certainly, online registration will reduce the overall cost of voting for some subset of non-registrants, but the task at hand is to distinguish those who would register regardless from those who would register only if the online alternative is available. From a public policy perspective, it might be normatively desirable to reduce the burden of registration independent of the impact on turnout, but the lingering question within the literature remains whether we can elicit participation from perennial non-voters (Berinsky 2005).

4.6 Model

With this intellectual history in mind, I turn to the primary questions of this chapter: Does online voter registration actually fulfill its promise to capture new segments of the electorate, or does it simply make the registration process easier for those would participate regardless? Does it do anything to stimulate participation among the historically underrepresented? To answer these questions, I need a model that identifies and estimates the probability that an individual votes with and without the opportunity to register online. As with any question of causal inference, I encounter the challenge that I do not observe the counterfactual. In this particular case, I cannot simultaneously know an individual's voting behavior with and without treatment exposure, since for each election the individual experiences only one of these two states of the world. To derive estimates of how individuals would behave in the counterfactual, I can use an experimental design that is within-unit (if their treatment status changes over time) or between-unit. To my knowledge, there does not exist a longitudinal survey that asks individuals about their registration and voting behavior before and after exposure to the online registration treatment. Fortunately, though, I can exploit the heterogeneity in the availability of online registration across states and time to accommodate a between-unit approach.

To this end, I rely on the Rosenstone-Wolfinger design discussed in Chapter 2. The outcome of interest —turnout, and in this case, registration as well —is modeled as a probabilistic function of various demographic features, electoral characteristics, and the treatment (online registration):[9]

$$\mathbb{P}(Y_{ist} = 1) = \Phi(\alpha + X_{ist}\beta + Z_{st}\rho + \delta T_{st} + \gamma_s + \theta_t + \epsilon_{ist}).$$

[9] I follow the conventional approach in the literature for modeling turnout, and the precedent established in Mitchell and Wleizen (1995) and Alvarez and Hall (2012) for registration. For a discussion of modeling the joint distributions of registration and turnout, and an alternative approach to modeling registration, see Achen (2008). Furthermore, it is quite possible that changes in registration policy can shift GOTV efforts (Gronke *et al.* 2008); as online registration becomes more prominent within American elections, it will be worth considering how mobilization groups alter their strategies in response.

Previous papers have used probit (Rosenstone and Wolfinger 1978; Leighley and Nagler 1992; Alvarez and Nagler 2007, 2008, 2011), logit (Highton 1997; McDonald 2008), and scobit (Nagler 1994) regressions; I select the probit link but include other structural forms in the Supplementary Materials as a robustness check.[10]

In this specification, Y is a binary variable indicating whether or not the individual participated (registered or voted, depending on the outcome of interest); X is a vector of demographic covariates; Z is a vector of controls for the local political climate; and T is a binary variable indicating whether the state offered online registration for that particular election. State and year fixed effects (γ and θ, respectively) account for heterogeneity in electoral trends dictated by geography and time. The demographic covariates are taken from a long stream of literature that has explored heterogeneity in voting behavior: age, gender, ethnicity, income, education, employment, and residential mobility. The electoral controls are the competitiveness of the election and the availability of the most prominent forms of convenience voting; I include EDR in the model of registration behavior and EDR, early voting, no-excuse absentee voting, and vote by mail in the model of turnout.[11]

Like many other papers in this area, I rely primarily on data from the *Current Population Survey* (CPS), a survey jointly administered by the Census Bureau and the Bureau of Labor Statistics.[12] Every presidential and midterm election, the CPS also includes a Voter Supplement, which asks respondents a series of questions about their registration and voting behavior. The CPS is quite extensive, with each implementation surveying tens of thousands of respondents. Thus, the overall sample is quite large, and more importantly, I observe a large number of individuals for each state, for each election. Such a rich source of data offers us the opportunity to explore political participation using a model with adequate statistical power.

[10] See Tables 4.A11 and 4.A12. The direction and significance of the main effects are not affected by the variation in functional form.

[11] I see no strong theoretical reason to include early voting, no-excuse absentee voting, and vote by mail when registration is the outcome of interest.

[12] For example, see Rosenstone and Wolfinger (1978), Mitchell and Wlezien (1995), Alvarez and Nagler (2007, 2008, 2011), Knack and White (2000), Highton (1997), Hanmer (2007), and Glynn and Quinn (2011).

4.7 Data

Although the *CPS* does not track the same individuals across elections, I can leverage the behavior of those who do not have access to online registration to draw inference about the policy's efficacy. Each cross-section of the survey samples tens of thousands of respondents and lends itself quite naturally to a time-series cross-section design (see Leighley and Nagler 2014; Ansolabehere and Konisky 2006; and particularly, Brians and Grofman 2001). My sample includes each electoral cross-section from 2000 until the most recent election in 2012.[13] For the distribution of demographics and access to convenience voting, refer to Tables 4.A1 and 4.A2 of the Supplementary Materials.

Somewhat problematically, a sizable minority of individuals refuse to answer questions regarding political participation, due to privacy concerns or simply memory lapse. Furthermore, those who do respond may give incorrect information due to imperfect recall or social desirability concerns. There is reason to be skeptical that measurement error and missingness are distributed as-if randomly among respondents, and accordingly, some attention has been paid as to how to measure participation appropriately (notably, see Katz and Katz 2010; Ansolabehere and Hersch 2012; Hur and Achen 2013). The Census Bureau and many of the papers in this literature adopt a somewhat conservative approach, coding all those who reply in the affirmative to registration status and/or turnout as registrants and/or voters, and everyone else as non-participators. I follow this precedent, but perform robustness checks both coding non-responders as participants and dropping non-respondents.[14]

The literature on political participation considers both demographic and institutional predictors (Rosenstone and Wolfinger 1978). The model employs a host of demographic covariates that can intervene in the relationship between online registration and participation: age, gender, ethnicity, income, education, and employment status. I also include an indicator for those who moved within the last four years

[13]As described earlier, Arizona pioneered the treatment prior to the 2002 election. I therefore begin with the 2000 cross-section to ensure that the sample contains residents who were and were not allowed to register online for each state that adopted online registration over this time period.

[14]The estimates of the coefficients are actually the most conservative when using the Census Bureau's approach. For results using alternate coding, see Tables 4.A13 and 4.A14.

as a measure of residential mobility, a well-known factor of predicting political behavior; moreover, there is evidence that recent movers may be particularly receptive to convenience voting procedures (*cf.* McDonald 2008). Given that citizens have to re-register every time they move, it is hardly surprising that those who moved within the last four years are less likely to be registered and to vote. The *CPS* requests such information from all respondents, and implements the Voter Supplement for all voting-age citizens.[15]

Although demographic variables explain a substantial proportion of the variation in registration and turnout, I would be remiss if I did not also control for institutional factors that influence the electoral environment. Because turnout is systematically higher in presidential elections, I include an indicator variable for these occurrences. Additionally, it is important to consider the electoral competitiveness of the ballots; it is well-documented that more competitive elections elicit greater likelihood of participation. Even so, the discipline has yet to agree upon a benchmark measure of competitiveness. To incorporate this feature of the electoral environment, I use the margins of victory between the top two candidates in each state for its presidential, U.S. senatorial, and gubernatorial elections (availability permitting); for each state, I then take the narrowest of these three margins of victory and use it as the gauge of competitiveness for that election. And lastly, I include other measures of convenience voting: Election Day Registration, no-excuse absentee voting, no-fault early voting, and vote by mail.[16] Table 4.A3 in the Supplementary Materials delineates the convenience voting procedures allowed by each state for each electoral cross-section.

[15] A non-trivial proportion of individuals refuse to report income and length at current address. The results presented within the chapter use Complete Case Analysis.

[16] To construct this table, I relied primarily on an extensive data collection project by Cemenska *et al.* (2009) as well as information published by the National Council of State Legislatures. The Cemenska *et al.* dataset, although thoroughly researched and well-codified, has a couple of drawbacks: It does not explicitly consider midterm elections, and it tracks legislation only until 2008. The NCSL datasets are more current, but fail to report dates of implementation for early and absentee voting procedures. I cross-referenced these two databases and searched state legislation when necessary to delineate the convenience voting procedures in each state for each federal electoral cycle in this chapter.

4.8 Analysis

4.8.1 Primary Results

Table 4.1 displays estimates of each variable's impact on participation. For the full probit results, refer to Figures 4.A1 and 4.A2 and Table 4.A4 of the Supplementary Materials. A recent paper by Hanmer and Kalkan (2013) eloquently advocates that researchers working with nonlinear models estimate quantities of interest using observed values, rather than consider the effect on the "average case." Occasionally, the average case is of theoretical interest, but in this case (as in many others), the scope lies far beyond the impact on individuals belonging to any particular demographic background. Accordingly, I examine discrete differences for each variable while using the observed values for the other covariates. In general, the signs on the coefficient point estimates accord with the general consensus in the literature. Unless otherwise noted in the discussion, the sign of the coefficient and significance level are the same for both the registration and turnout models.

As expected, the discrete difference for each pair of ages is positive and statistically significant, confirming that older voters have higher propensities to participate. Females are more likely than males to register and vote. For employment status, those who are employed serve as the reference group; the negative effects on the variables Unemployed and Not in the Labor Force indicate that individuals in these groups are less likely to participate than those who are employed. Likewise, the relationships among the education variables are predictable; the expected change in propensity to vote is positive and statistically significant for each additional level of education. The estimated effect of moving is significantly negative, supporting the general wisdom that residential stability is an important correlate of political behavior.

One of the only surprising demographic results is that all else held constant, a Black person has a greater likelihood of registering and voting than a White person. This result could reflect a recent trend of increasing political participation among minorities; indeed, the 2012 CPS reports that for the first time in American history, blacks are participating at higher levels than whites (Census Bureau 2013). Thus,

Table 4.1: Average Effect on Political Participation

Variable	**Registration**	**Turnout**
Age: 18 $\rightarrow$ 25	0.049	0.063
Age: 25 $\rightarrow$ 35	0.062	0.084
Age: 35 $\rightarrow$ 45	0.053	0.073
Age: 45 $\rightarrow$ 55	0.044	0.062
Age: 55 $\rightarrow$ 65	0.035	0.051
Male $\rightarrow$ Female	0.033	0.028
White $\rightarrow$ Black	0.050	0.083
White $\rightarrow$ Asian	-0.160	-0.171
White $\rightarrow$ Other Race	-0.028	-0.047
Non-Hispanic $\rightarrow$ Hispanic	-0.046	-0.049
Income: Level 3 $\rightarrow$ Level 7	0.051	0.069
Income: Level 7 $\rightarrow$ Level 10	0.035	0.050
High School $\rightarrow$ Some College	0.137	0.150
Some College $\rightarrow$ College	0.046	0.070
College $\rightarrow$ Postgraduate	0.042	0.065
Employed $\rightarrow$Unemployed	-0.028	-0.033
Employed $\rightarrow$ Not in Labor Force	-0.038	-0.033
Recent Mover: 0 $\rightarrow$ 1	-0.083	-0.100
Midterm $\rightarrow$ Presidential	0.029	0.145
Margin of Victory: 1% $\rightarrow$ 5%	-0.001	-0.003
EDR: 0 $\rightarrow$ 1	-0.002	0.008
No-Fault Early: 0 $\rightarrow$ 1		-0.010
No-Excuse Absentee: 0 $\rightarrow$ 1		0.007
Vote by Mail: 0 $\rightarrow$ 1		0.010
Online Registration 0 $\rightarrow$ 1	0.014	0.018

This table delineates the estimated effects for each variable on the outcome of interest. All estimates are statistically significant at the 5% level.

the positive estimated effect for African American heritage is perhaps less surprising than it appears at first blush. The negative estimated effects of the other ethnicity covariates indicate that Asians and those who belong to another race are less likely to participate than Whites, and Hispanics are less likely to participate than non-Hispanics.

The electoral variables are also somewhat in line with what we would expect. An individual has a higher predicted probability of voting in a Presidential election than a midterm election. Note, too, that the relative magnitude is particularly high for the model of turnout; this result intuitively makes sense given that registration typically carries over from other elections, so those who register during a Presidential election are often still registered for the next midterm election. The estimated effect of the margin of victory moving from 1% to 5% is negative, as we should expect; in more competitive races, turnout is usually higher. The relative magnitude, however, is quite small; it is possible that the intensity of the political climate is already effectively captured by the state and year fixed effects.

Figure 4.1 displays the empirical distribution of discrete differences associated with the most popular forms of convenience voting, including online registration. Before discussing the differences associated with other forms of election policy, it is worth mentioning that this model does not fully identify the causal relationship between these variables and participation, as the dataset does not contain a pre-treatment cross-section for each state offering these liberalizations. Instead, the measure captures the difference in participation associated with states that offer a particular policy. The discrete difference associated with EDR is fairly minimal, in contrast to cross-sectional designs that have suggested a pronounced, positive impact (Alvarez and Nagler 2007, 2008, 2011; Brians and Grofman 2001; Burden *et al.* 2014). Other forms of convenience voting, such as the opportunities to vote absentee (without an excuse) and to vote by mail, exhibit similarly weak results. Meanwhile, the negative effect associated with early voting lends credence to the concern in the literature that this opportunity might reduce participation (Burden *et al.* 2014); at the very least, there seems to be a culture of lower participation associated with states that allow

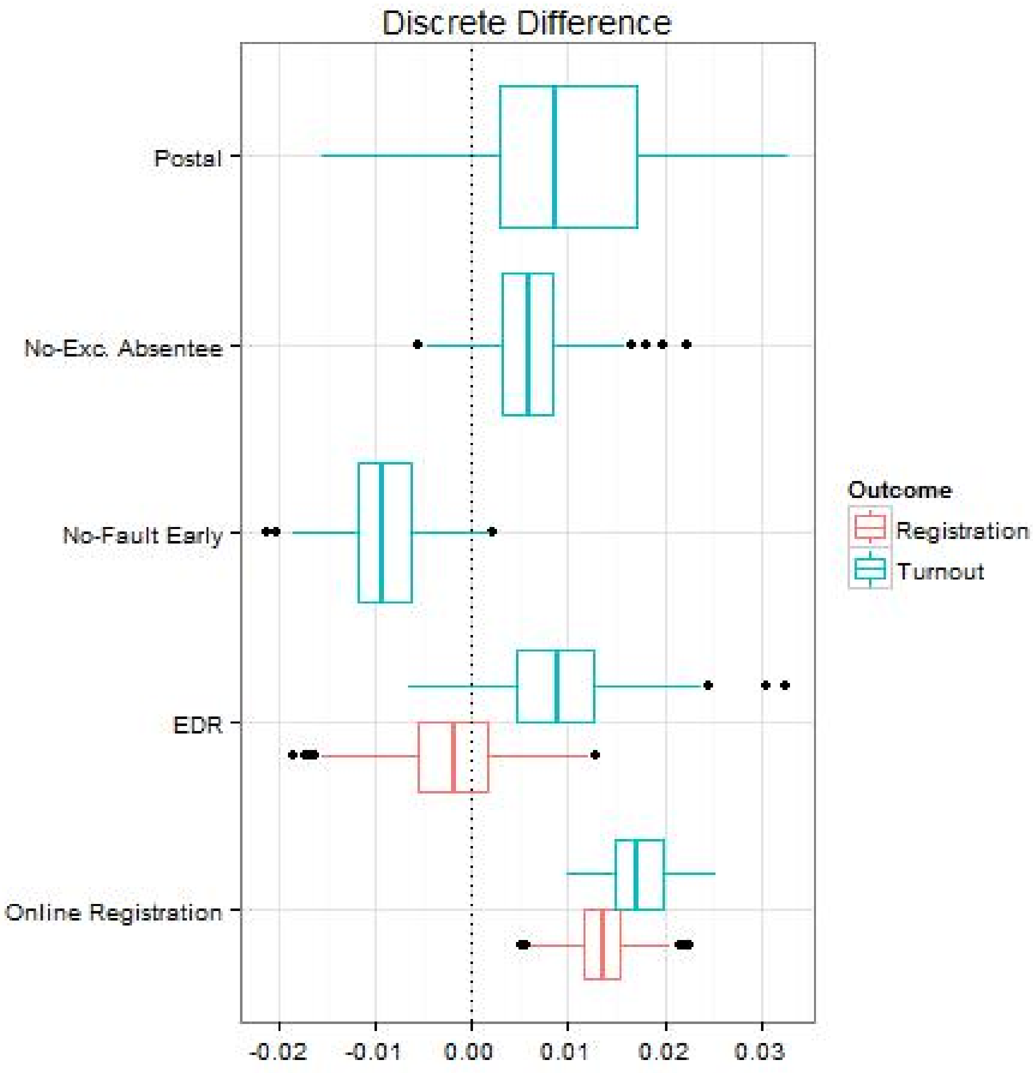


Figure 4.1: The Relationship Between Participation and Election Policies

their residents to vote early.

And of course, the variable of primary interest is the availability of online registration. The estimated effect of online registration is positive and statistically significant in both the registration and turnout models, indicating that individuals who live in OR states are more likely both to register and to vote. Those who promote online registration as a promising mechanism for stimulating turnout will no doubt be pleased to see evidence that individuals are taking advantage of this liberalization. Yet although the nature of the relationship gives reason for optimism, I point out that the relative magnitude is somewhat small, approximately 1.4% for registration and 1.8% for turnout.[17] Even so, it still seems to exert influence that matches or even exceeds that of other forms of convenience voting.

Moreover, the discrete differences in Table 4.2 suggest that the impact of online registration might become more pronounced over time.[18] In the regression model, treatment is no longer binary. Instead, states that are introducing online registration for the first time are coded as "Wave 1," and states that have allowed online registration for at least one previous election are coded as "Wave 2."[19] As an example, Arizona introduced online registration prior to the 2002 midterm election. Thus, for residents of Arizona in 2000, Wave 1 = Wave 2 = 0; in 2002, Wave 1 = 1 and Wave 2 = 0; and in 2004 —2012, Wave 1 = 0 and Wave 2 = 1. Because it might take time for residents to become aware of the opportunity to register online, and/or states to work out the kinks in the websites, there is theoretical reason to believe that the impact might vary over time. For example, South Carolina introduced online registration only four days prior to the closing date for the 2012 cross-section (Pew Center on the States 2012); it is *a priori* unrealistic to assume that exposure to the treatment is homogeneous across South Carolina, and Arizona, which had allowed online registra-

[17]The confidence intervals overlap, for the reader who may feel concerned that the impact on turnout is higher than that on registration.

[18]For the full results, refer to Table 4.A5.

[19]Only two states (AZ and WA) allowed online registration for more than two electoral cross-sections, so a model with additional waves runs the risk of conflating the effects of online registration with electoral idiosyncrasies of these states. As we obtain data for additional cross sections, it will be worthwhile to consider finer partitions.

Table 4.2: Average Effect on Political Participation (Probit, Treatment by Wave)

Variable	**Registration**	**Turnout**
Age: 18 → 25	0.049	0.063
Age: 25 → 35	0.062	0.084
Age: 35 → 45	0.053	0.073
Age: 45 → 55	0.044	0.062
Age: 55 → 65	0.035	0.051
Male → Female	0.033	0.028
White → Black	0.050	0.083
White → Asian	−0.160	−0.171
White → Other Race	−0.028	−0.047
Non-Hispanic → Hispanic	−0.046	−0.049
Income: Level 3 → Level 7	0.051	0.069
Income: Level 7 → Level 10	0.035	0.050
High School → Some College	0.137	0.150
Some College → College	0.046	0.070
College → Postgraduate	0.042	0.065
Employed → Unemployed	−0.028	−0.033
Employed → Not in Labor Force	−0.038	−0.033
Recent Mover: 0 → 1	−0.083	−0.100
Midterm → Presidential	0.029	0.145
Margin of Victory: 1% → 5%	−0.001	−0.003
EDR: 0 → 1	−0.002	0.009
No-Fault Early: 0 → 1		−0.010
No-Excuse Absentee: 0 → 1		0.008
Vote by Mail: 0 → 1		0.011
Control → Wave 1	0.008	0.010
Wave 1 → Wave 2	0.014	0.026
Control → Wave 2	0.022	0.036

This table delineates the estimated effects for each variable on the outcome of interest. All estimates are statistically significant at the 5% level.

tion for ten years. The results of this model indeed suggest the presence of a lagged effect, as evidenced by the positive (and statistically significant) estimated effect of switching from a Wave 1 to a Wave 2 state. It is possible, then, that the estimated effect of online registration is actually biased *downward* when the treatment is coded as binary.

4.8.2 Spotlighting Under-represented Subgroups

Those who promote various forms of convenience voting often express a desire for an electorate that is more representative of the voting-eligible population. Historically, certain demographics have registered and voted at lower rates, and many policies seek to reduce the institutional barriers that might discourage participation. In this vein, I consider how individuals of particular demographic backgrounds interact with the online registration treatment. In this case, I test for an interactive effect of online registration and (1) age, (2) ethnicity, and (3) residential mobility. Fortunately, the size of the *CPS* allows me to partition individuals into subgroups and still achieve adequate statistical power. I execute the probit model using three separate subsamples and estimate the average treatment effect of online registration for each: (1) individuals under the age of 30, (2) racial minorities, and (3) recent movers. Table 4.3 displays the average treatment effect of online registration for each subsample; Tables 4.A6, 4.A7, and 4.A8 report the full set of results.

Table 4.3: Average Effect of Online Registration on Political Participation (Subsamples)

Subsample	**REGISTRATION**	**TURNOUT**
Young	0.022	0.027
Non-White	0.015	0.009
Recent Movers	0.022	0.021

This table delineates the estimated treatment effect of online registration on the outcome of interest for different subsamples of interest. For the full set of results, see Tables 4.A6, 4.A7, and 4.A8.

I will first discuss the impact on young voters. We have theoretical reason to suspect that young adults might be particularly affected by this registration opportunity, given their exposure to and utilization of technology. Indeed, the estimated effect of online registration is positive and statistically significant in both outcome models, and the estimated average treatment effect is 2.2% for registration and 2.7% for turnout. Although these numbers may fall far short of advocates' hopes, it is nonetheless promising that at least some portion of this demographic is responding to the opportunity. Interestingly, the coefficient on EDR achieves statistical significance in the regression of turnout (unlike the analogous regression using the entire sample), suggesting that perhaps this subgroup might reap greater benefits from convenience voting in general.

The results for ethnic minorities are less promising. Although the estimated effect of online registration achieves statistical significance at the 5% level in the registration model, it fails to reach any conventional level of significance in the turnout model. The estimated average treatment effect is approximately 1.5% for registration and less than 1% for turnout, suggesting that the gains —if existent —are minimal. We have little reason to believe *a priori* that ethnic minorities will take particular advantage of online registration, so these results are not especially surprising, but reformers seeking a more ethnically representative electorate may have to pursue other approaches to stimulate participation from this demographic subgroup.

And finally, as with young voters, there is evidence that recent movers respond to this treatment at the registration and turnout stages. Because registration is tied to residence, this subgroup would have to go through the registration process even if they were registered prior to moving. Members of new communities may have more reason to seek online resources rather than utilize the civic services, given their lack of familiarity with and connection to the area. Additionally, online registration might better inform residents of their assigned precincts, resulting in fewer mishaps on Election Day. The average treatment effect is similar to that of young voters —just over 2% for both registration and turnout, and significantly bounded away from null in both models. And notably, the results are weaker for other forms of convenience voting, suggesting that online registration might be the most effective attempt yet to capture this portion of the voting-eligible population. Again, these results hardly suggest that online registration is bringing in new participants *en masse*, but they do give cause for hesitant optimism.

4.8.3 Aggregate Results

As its advocates attempt to implement online registration, many will want to know whether this reform will have any practical impact on the election. I have presented the estimated impact at the individual level, but it remains unclear whether or not online registration will actually have any bearing on the actual election. To explore the real-world significance, for each state that did not allow online registration, I estimate the number of additional votes if it had implemented such a system prior to the 2008 and 2012 Presidential elections. To do so, for each control state, I estimate the Average Treatment Effect on the Control (or the ATC), and then multiply this quantity by the voting-eligible population.

I arrive at the measure of the ATC by using the point estimates from the fitted model to predict the probability that each individual in a control state votes, with and without OR available, and then averaging the difference over all respondents. More explicitly, for each individual in a control state, I impute the following quantities:

$$\hat{\mathbb{P}}^0(Y_{ist} = 1) = \Phi(\hat{\alpha} + X_{ist}\hat{\beta} + Z_{st}\hat{\rho} + \hat{\gamma}_s + \hat{\theta}_t),$$

which is simply the fitted probability from the main effects model (recall that $T_{st} = 0$ for individuals in control states), and

$$\hat{\mathbb{P}}^1(Y_{ist} = 1) = \Phi(\hat{\alpha} + X_{ist}\hat{\beta} + Z_{st}\hat{\rho} + \hat{\delta} + \hat{\gamma}_s + \hat{\theta}_t),$$

where $\hat{\delta}$ is the point estimate for the coefficient on online registration. I take their difference to estimate the change in predicted probability:

$$\hat{\Delta}_{ist} = \hat{\mathbb{P}}^1(Y_{ist} = 1) - \hat{\mathbb{P}}^0(Y_{ist} = 1).$$

For each state-year, I average over Δ and multiply by the total voting-eligible population to arrive at the estimated influx of votes that online registration would have generated.[20] To broaden the scope of inference, I repeat this process twice, using the upper and lower bounds of the 95% confidence interval for the coefficient on online registration.

Table 4.4: Estimated Influx of Voters

State	**Lower Bound**	**Estimate**	**Upper Bound**	**Margin of Victory**
NC 2008	70.832	108.119	144.965	14.177
MO 2008	45.861	69.996	93.841	3.903
IN 2008	53.557	81.830	109.824	28.391
FL 2008	129.294	197.235	264.290	236.148
MT 2008	6.908	10.530	14.100	11.723
FL 2012	146.480	223.729	300.158	74.309
NC 2012	77.987	119.124	159.832	92.004
OH 2012	97.774	149.368	200.436	166.272

I use the probit results in Table 4.A4 to estimate the additional votes that online registration would have produced in the 2008 and the 2012 Presidential contests; the final column of this table lists the actual margin of victory in that election. For these states, the upper bound on the estimated influx exceeded the realized margin of victory. Tables 4.A9 and 4.A10 report the results for every control state in 2008 and 2012, respectively.

Table 4.4 lists the states for which the upper bound on the estimated influx of votes exceeds the actual margin of victory in the 2008 and 2012 Presidential contests; Tables 4.A9 and 4.A10 report the information for each control state in the 2008 and 2012 election cycles. In each year, for a small collection of states, the margin of victory between the top two presidential slates is contained within (or lies to the left of) the projected interval. In 2008, for three states (IN, MO, NC), the margin of victory is less than the lower bound for the estimated influx; the same is true of FL in 2012.

Hypothetically, then, it is within the realm of possibility for online registration to influence the outcome. The Voter Supplement does not detail partisan affiliation, so it is difficult to estimate the partisan makeup of these additional votes. Consider,

[20] I utilize Michael McDonald's datasets for estimates of the voting-eligible population, available at http://elections.gmu.edu/Turnout_2012G.html and http://elections.gmu.edu/Turnout_2008G.html.

though, Missouri in 2008. The margin of victory was 3903 votes, while the estimated increase in turnout with online registration is 69,996; it is hardly inconceivable that the partisan makeup of these additional voters could alter the outcome of the election. This comparison gives merely a glimpse of the potential aggregate impact of this form of convenience voting, but Missouri's results in 2008 suggest that politicians and campaign activists may want to consider the potential electoral implications. In the next chapter, I will advance a method to predict the impact of online registration on vote share.

4.9 Conclusion

Despite my skepticism that online registration would have any meaningful impact on individual or aggregate turnout, the results of this investigation suggest that this particular form of convenience voting may actually live up to the promises of its enthusiasts. Indeed, the results of this investigation suggest that this reform may be especially important for young adults and recent movers. Given these implications, perhaps we can don our rose-colored glasses with fewer misgivings. On the other hand, it is fair to point out that the results are driven primarily by the two most recent cross-sections of data. As we amass additional data, future projects should continue to evaluate whether online registration is part of the solution to the turnout puzzle, or simply a distraction from the true barriers to universal participation.

Of course, conservative results are still entirely compatible with a normative case for the online registration movement. Firstly, states save money and reduce error when residents register themselves online (Baretto *et al.* 2010). Secondly, we can expect that the effect, however minimal, will be positive; it would be extremely surprising if voters became less likely to participate if given the opportunity to register online (*cf.* Hanmer 2007, 2009; Glynn and Quinn 2011). And lastly, even if the effect of online registration is purely one of substitution, we can still assume that overall welfare has increased; the citizens who utilize online registration presumably find it more convenient and less costly than other registration alternatives.

The true effect of online voter registration remains to be seen, as the "waves" treatment suggests (Table 4.2). In the future, more citizens may become aware of this opportunity, and knowledge of the existence of online registration is a prerequisite to its utilization. As more states' election administrations consider online registration, news coverage of the movement will increase. Others may learn through word of mouth, and it is quite possible that voter mobilization groups will shift resources to promoting this method. Decades passed in California before a substantial proportion of citizens took advantage of no-excuse absentee voting (Alvarez, Levin, and Sinclair 2012). Furthermore, people are already completing more of their daily activities online. And finally, as the voting base naturally shifts with each generation, we can hope that the influx of new voters, who have greater technological exposure, will be more likely to seek and pursue registration online.

On the other hand, while proponents of online voter registration have argued that it will foster a new sense of political activism, the current project gives reason to temper those expectations. It is fair to celebrate the potential impact on young adults and recent movers, but we should bear in mind that there is currently little evidence that online registration will capture other underrepresented segments of the electorate, such as racial minorities. Consequently, the results of this study seem to echo the concern that despite persistent reform efforts, we still have difficulty encouraging fully representative participation (Berinsky 2005).

There is still a great deal of work to be done if we want to develop an accurate metric for assessing the performance of these reforms. An assumption of conditionally exogenous selection is implicitly embedded within my research design, and this assumption is tenuous at best (see Hanmer 2009). In other words, this method assumes that whether or not an individual is treated is independent from how she would respond to treatment. We might reasonably infer that if this assumption does not hold in reality, our results might incorporate some bias, but just how this bias might manifest itself remains unclear. Some papers have attempted to sidestep this assumption by instead using nonparametric bounds (Glynn and Quinn 2011, Hanmer 2007), but Hanmer (2007) admits that the results of this method are substantively uninformative. Other papers have utilized an RDD approach (Keele and Minozzi 2013, Ansolabehere and Konisky 2006), but opportunities to exploit this design are limited by the structure of implementation. Furthermore, the literature does not fully account for the possibility that individuals may interact with the treatment differently, and perhaps as a function of their demographic backgrounds. It seems inherently plausible, for example, that only politically engaged citizens take advantage of these reforms because they are the only ones aware of the liberalizations. Future papers should consider and address the importance of salience in these reform efforts.

And finally, even with the fairly promising results presented in this paper, the turnout puzzle remains unresolved. Despite numerous pieces of legislation designed to reduce the costs of voting and millions of dollars poured into "Get Out the Vote" campaigns, American elections are still characterized by low levels of participation (Leighley and Nagler 1992). Even if it helps, online registration is not a panacea; the most generous estimates of its promise do not suggest anything near universal participation. This chapter acknowledges the concerns about the futility of mobilization efforts and naturally begs the question of why these reforms are not having the promised effect. Are non-voters so apathetic about politics that they are indifferent? Are they convinced of the improbability of pivotality and unwilling to vote if there is any cost? Or are they perhaps unaware of the reforms that could enhance their voting experiences? Future research should disentangle these potential explanations, as they imply dramatically different forecasts of the upper bound on participation.

4.10 Supplementary Materials

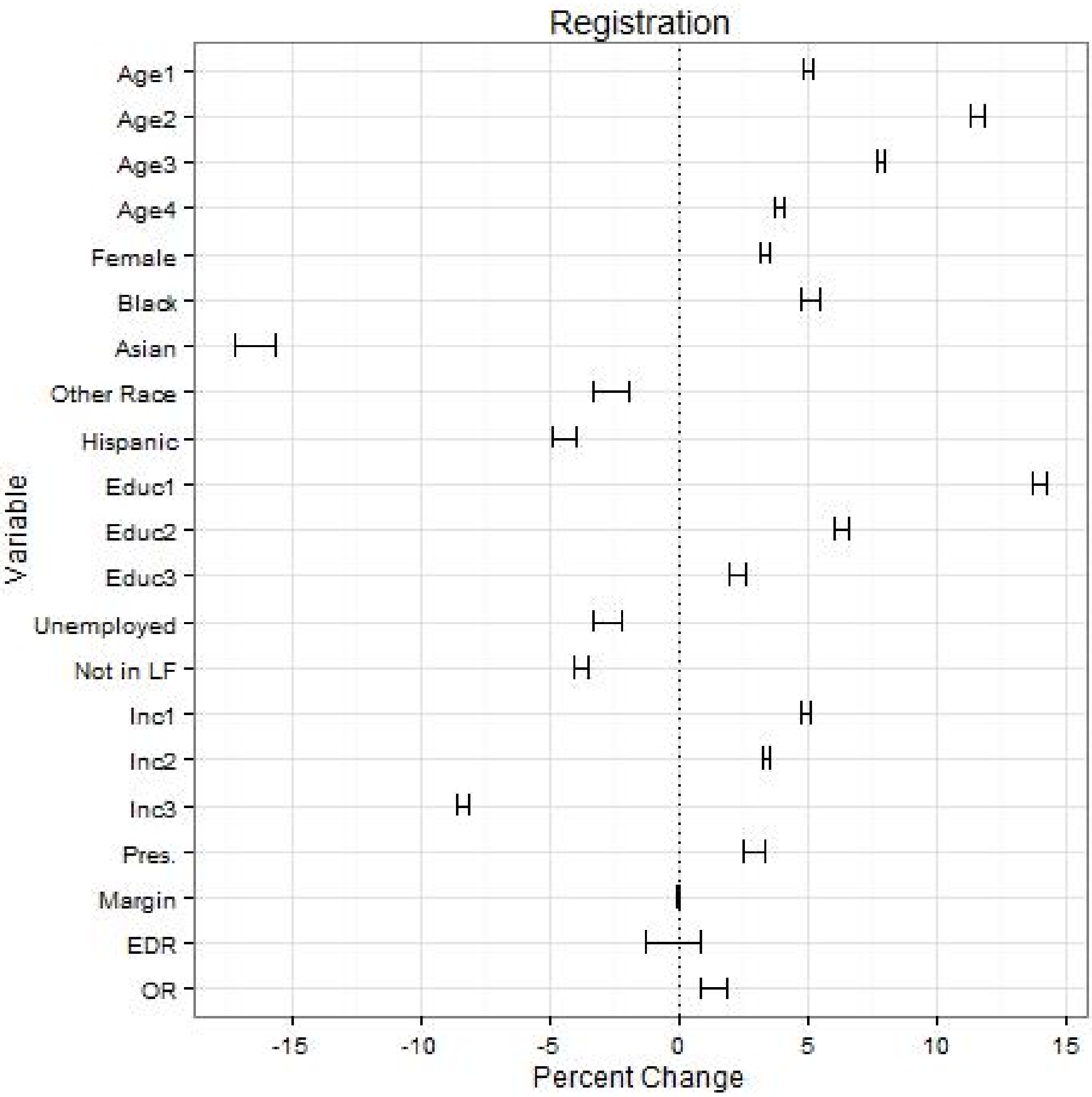


Figure 4.A1: Confidence Intervals for Discrete Differences, Registration

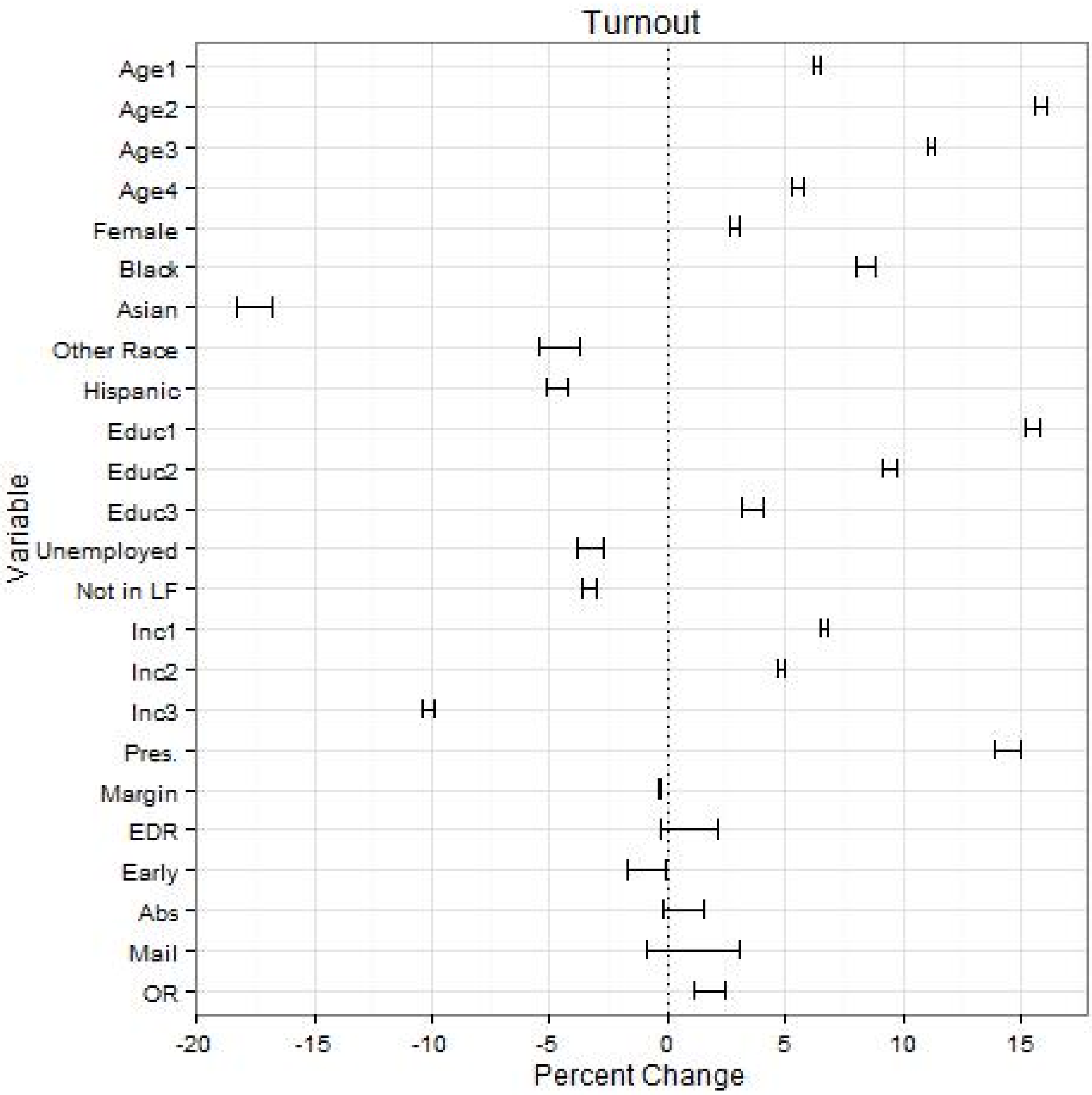


Figure 4.A2: Confidence Intervals for Discrete Differences, Turnout

Table 4.A1: Demographic Distribution

Variable	Frequency	Percent
Female	335.063	52.72%
Male	300.448	47.28%
Black	59.111	9.30%
Asian	21.559	3.39%
Other Race	14.152	2.23%
White	540.689	85.08%
Hispanic	45.486	7.16%
High School or Less	280.821	44.19%
Some College	127.076	20.00%
College	170.678	26.86%
Postgraduate	56.936	8.96%
Unemployed	22.399	3.52%
Not in Labor Force	209.829	33.02%
Employed	403.283	63.46%
Recent Mover	228.824	40.31%

This table gives the demographic distribution for the sample. The variables in boldface serve as reference categories in the econometric models. Age is treated as a continuous variable, with mean 47.23, minimum 18, and maximum 80.

Table 4.A2: Access to Convenience Voting by State and Year

Convenience Voting	**2000**	**2002**	**2004**	**2006**	**2008**	**2010**	**2012**
EDR	8.44 %	10.73 %	11.40 %	12.71 %	14.87 %	14.68 %	16.87 %
No-Fault Early	44.86 %	46.50 %	51.41 %	58.50 %	60.47 %	63.03 %	63.03 %
No-Excuse Abs	38.67 %	39.13 %	42.26 %	48.49 %	48.09 %	53.69 %	53.49 %
Vote by Mail	2.60 %	3.45 %	3.35 %	3.29 %	3.27 %	3.25 %	3.40 %
Online Registration	0.00 %	1.27 %	1.26 %	1.30 %	2.92 %	12.00 %	25.33 %

This table illustrates the distribution of access to each type of convenience voting for each yearly cross-section of the survey sample. Of particular interest is access to online registration, which has been steadily increasing.

Table 4.A3: Convenience Voting by State and Year

State	EDR	No-Excuse Absentee	No-Fault Early	Mail	Online Registration
AL					
AK		X	X		
AZ		X	X		2002
AR			X		
CA		X	X		2012
CO		X	X		2010
CT	2012				
DE					
FL		X	X		
GA		2004	2004		
HI		X	X		
ID	X	X	X		
IL			2006		
IN			X		2010
IO	2008				
KS		X	X		2010
KY					
LA			2008		2010
ME	X	X	X		
MD		2010	2010		2012
MA					
MI					
MN	X				
MS					
MO					
MT	2006	X	X		
NE		X	X		
NV		X	X		2012
NH	X				
NJ		2006			
NM		X	X		
NY					2012
NC		X	X		
OH		2006	2006		
OK		X	X		
OR		X	X	X	2010
PA					
RI					
SC					2012
SD		2004	2004		
TN			X		
TX			X		
UT		X	2004		2010
VT		X	X		
VA					
WA		X		X	2008
WV			2002		
WI	X	X	X		
WY	X	X	2008		

This table lists the convenience voting procedures for each electoral cross-section by state. An "X" denotes that the state offered that particular form of convenience voting in all electoral cross-sections of this dataset. If the entry is a year, that is the first electoral cross-section that a particular state offered that form of convenience voting. And finally, if the entry is blank, the state did not allow that form of convenience voting in any electoral cross-sections in this dataset.

Table 4.A4: Effects on Political Participation (Probit, Binary Treatment)

	Registration		**Turnout**	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.024	(0.001)	0.031	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.129	(0.004)	0.089	(0.004)
Black	0.216	(0.008)	0.283	(0.007)
Asian	−0.550	(0.013)	−0.531	(0.012)
Other Race	−0.107	(0.014)	−0.150	(0.013)
Hispanic	−0.175	(0.008)	−0.155	(0.008)
Income	0.043	(0.001)	0.051	(0.001)
High School	−0.073	(0.006)	−0.682	(0.005)
Some College	−0.215	(0.007)	−0.233	(0.006)
Postgraduate	0.248	(0.011)	0.248	(0.008)
Unemployed	−0.111	(0.011)	−0.107	(0.010)
Not in Labor Force	−0.147	(0.006)	−0.106	(0.005)
Moved	−0.319	(0.005)	−0.316	(0.004)
Margin of Victory	−0.001	(0.000)	−0.003	(0.000)
Presidential	0.115	(0.008)	0.461	(0.008)
EDR	−0.009	(0.021)	0.027	(0.019)
No-Fault Early			−0.033	(0.014)
No-Excuse Absentee			0.022	(0.015)
Vote by Mail			0.034	(0.032)
Online Registration	0.055	(0.011)	0.059	(0.011)
Intercept	0.147	(0.033)	−0.939	(0.033)
N	509,439		509,439	
Pseudo R^2	0.151		0.172	

This table represents the point estimates and associated (robust) standard errors for the probit model fitted to the entire sample, with online registration modeled as a binary treatment.

Almost all coefficients are statistically significant at the 1% level. The exception in the registration model is EDR (no significance); exceptions in the turnout model are No-Fault Early (5% significance) and EDR, No-Excuse Absentee, and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A5: Effects on Political Participation (Probit, Treatment by Wave)

	Registration		**Turnout**	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.024	(0.001)	0.031	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.129	(0.004)	0.089	(0.004)
Black	0.216	(0.008)	0.283	(0.007)
Asian	−0.550	(0.013)	−0.531	(0.012)
Other Race	−0.107	(0.014)	−0.150	(0.013)
Hispanic	−0.175	(0.008)	−0.155	(0.008)
Income	0.043	(0.001)	0.051	(0.001)
High School	−0.703	(0.006)	−0.682	(0.005)
Some College	−0.215	(0.007)	−0.233	(0.006)
Postgraduate	0.248	(0.011)	0.249	(0.008)
Unemployed	−0.111	(0.011)	−0.107	(0.010)
Not in Labor Force	−0.147	(0.006)	−0.106	(0.005)
Moved	−0.319	(0.005)	−0.316	(0.004)
Margin of Victory	−0.001	(0.000)	−0.003	(0.000)
Presidential	0.114	(0.008)	0.461	(0.008)
EDR	−0.008	(0.021)	0.030	(0.019)
No-Fault Early			−0.033	(0.014)
No-Excuse Absentee			0.026	(0.015)
Vote by Mail			0.036	(0.032)
Wave 1	0.031	(0.012)	0.031	(0.011)
Wave 2	0.088	(0.017)	0.117	(0.016)
Intercept	0.147	(0.033)	−0.944	(0.033)
N	509,439		509,439	
Pseudo R^2	0.151		0.172	

This table represents the point estimates and associated (robust) standard errors for the probit model fitted to the entire sample, with Online Registration modeled in two binary waves.

Almost all coefficients are statistically significant at the 1% level. The exception in the registration model is EDR (no significance); exceptions in the turnout model are No-Fault Early (5% significance), No-Excuse Absentee (10% significance), and EDR and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A6: Effects on Political Participation (Probit, Young Voters)

	REGISTRATION		**TURNOUT**	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Female	0.113	(0.008)	0.101	(0.008)
Black	0.215	(0.015)	0.324	(0.014)
Asian	−0.383	(0.024)	−0.386	(0.025)
Other Race	−0.079	(0.024)	−0.107	(0.025)
Hispanic	−0.180	(0.014)	−0.148	(0.015)
Income	0.019	(0.001)	0.028	(0.001)
High School	−0.817	(0.011)	−0.802	(0.011)
Some College	−0.284	(0.012)	−0.335	(0.011)
Postgraduate	0.250	(0.027)	0.323	(0.024)
Unemployed	−0.122	(0.016)	−0.149	(0.017)
Not in Labor Force	−0.158	(0.010)	−0.103	(0.010)
Moved	−0.167	(0.009)	−0.196	(0.009)
Margin of Victory	−0.001	(0.000)	−0.003	(0.000)
Presidential	0.145	(0.016)	0.543	(0.017)
EDR	0.044	(0.041)	0.092	(0.040)
No-Fault Early			−0.071	(0.029)
No-Excuse Absentee			0.067	(0.032)
Vote by Mail			0.094	(0.066)
Online Registration	0.065	(0.021)	0.080	(0.021)
Intercept	0.651	(0.056)	−0.219	(0.063)
N	106,843		106,843	
Pseudo R^2	0.094		0.134	

This table represents the point estimates and associated (robust) standard errors for the probit model fitted to the subsample of individuals less than 30 years of age, with Online Registration modeled as a binary treatment.

Almost all coefficients are statistically significant at the 1% level. The exceptions in the registration model are Margin of Victory (10% significance) and EDR (no significance); exceptions in the turnout model are EDR, No-Fault Early, and No-Excuse Absentee (5% significance), and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A7: Effects on Political Participation (Probit, Minority Voters)

	REGISTRATION		**TURNOUT**	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.029	(0.001)	0.034	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.176	(0.009)	0.153	(0.008)
Income	0.017	(0.001)	0.026	(0.001)
High School	−0.565	(0.012)	−0.559	(0.011)
Some College	−0.099	(0.014)	−0.121	(0.013)
Postgraduate	0.186	(0.023)	0.184	(0.020)
Unemployed	−0.029	(0.018)	−0.042	(0.018)
Not in Labor Force	−0.216	(0.011)	−0.176	(0.011)
Moved	−0.274	(0.009)	−0.296	(0.009)
Margin of Victory	−0.001	(0.000)	−0.003	(0.000)
Presidential	0.089	(0.016)	0.377	(0.016)
EDR	0.010	(0.060)	0.087	(0.059)
No-F ault Early			−0.031	(0.033)
No-Excuse Absentee			−0.012	(0.034)
Vote by Mail			−0.103	(0.109)
Online Registration	0.051	(0.022)	0.027	(0.021)
Intercept	−0.086	(0.107)	−1.024	(0.106)
N	106,001		106,001	
Pseudo R^2	0.114		0.139	

This table represents the point estimates and associated (robust) standard errors for the probit model fitted to the subsample of non-White individuals, with Online Registration modeled as a binary treatment.

Almost all coefficients are statistically significant at the 1% level. The exceptions in the registration model are Margin of Victory and Online Registration (5% significance) and Unemployed, EDR, and the intercept (no significance); exceptions in the turnout model are Unemployed (5% significance) and EDR, No-Excuse Absentee, No-Fault Early, Vote by Mail, and Online Registration (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A8: Effects on Political Participation (Probit, Recent Movers)

	REGISTRATION		**TURNOUT**	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.019	(0.001)	0.024	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.126	(0.006)	0.091	(0.006)
Black	0.229	(0.011)	0.298	(0.010)
Asian	−0.478	(0.018)	−0.452	(0.018)
Other Race	−0.084	(0.019)	−0.131	(0.019)
Hispanic	−0.172	(0.011)	−0.157	(0.011)
Income	0.038	(0.001)	0.047	(0.001)
High School	−0.696	(0.008)	−0.711	(0.008)
Some College	−0.217	(0.009)	−0.249	(0.008)
Postgraduate	0.244	(0.014)	0.260	(0.012)
Unemployed	−0.127	(0.014)	−0.141	(0.014)
Not in Labor Force	−0.140	(0.008)	−0.089	(0.008)
Margin of Victory	0.000	(0.000)	−0.003	(0.000)
Presidential	0.147	(0.012)	0.522	(0.012)
EDR	−0.030	(0.031)	−0.013	(0.030)
No-Fault Early			−0.052	(0.021)
No-Excuse Absentee			0.045	(0.024)
Vote by Mail			−0.026	(0.048)
Online Registration	0.070	(0.016)	0.064	(0.016)
Intercept	−0.028	(0.047)	−1.016	(0.051)
N	208,995		208,995	
Pseudo R^2	0.114		0.152	

This table represents the point estimates and associated (robust) standard errors for the probit model fitted to the subsample of recent movers, with Online Registration modeled as a binary treatment.

Almost all coefficients are statistically significant at the 1% level. The exceptions in the registration model are Margin of Victory, EDR, and the intercept (no significance); exceptions in the turnout model are No-Fault Early (5% significance), No-Excuse Absentee (10% significance) and EDR and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A9: Estimated Influx of Voters, 2008

State	Lower Bound	Estimate	Upper Bound	Margin of Victory
AK	5,789	8,854	11,896	70,247
AL	32,054	48,875	65,461	453,067
AR	22,327	34,079	45,691	215,707
CA	238,888	364,830	489,415	3,262,692
CO	32,394	49,400	66,172	215,004
CT	25,203	38,451	51,529	368,345
DE	6,283	9,581	12,832	103,085
FL	129,294	197,235	264,290	236,148
GA	68,743	104,969	140,792	204,636
HI	11,406	17,449	23,446	205,305
IA	21,608	32,949	44,131	146,561
ID	11,241	17,171	23,040	166,572
IL	96,157	146,818	196,907	1,388,169
IN	53,557	81,830	109,824	28,391
KS	21,875	33,411	44,825	184,890
KY	34,017	51,943	69,669	296,477
LA	33,063	50,435	67,579	365,286
MA	43,450	66,222	88,655	795,244
MD	39,227	59,840	80,184	669,605
ME	9,900	15,093	20,211	126,650
MI	69,313	105,674	141,519	823,940
MN	29,080	44,242	59,125	297,945
MO	45,861	69,996	93,841	3,903
MS	22,008	33,570	44,979	169,935
MT	6,908	10,530	14,100	11,723
NC	70,832	108,119	144,965	14,177
NE	13,107	19,998	26,803	119,660
NH	10,017	15,276	20,463	68,292
NJ	63,124	96,352	129,186	602,215
NM	14,307	21,838	29,281	125,590
NV	19,482	29,771	39,961	120,909
NY	142,828	218,097	292,528	2,052,174
OH	90,312	137,839	184,791	262,224
OK	29,411	44,936	60,308	457,669
OR	24,099	36,713	49,128	298,816
PA	106,866	163,214	218,957	620,478
RI	7,976	12,173	16,319	131,180
SC	36,691	56,034	75,168	172,447
SD	5,502	8,387	11,230	32,130
TN	53,988	82,531	110,820	391,741
TX	176,170	269,532	362,221	950,695
UT	20,362	31,135	41,818	268,360
VA	58,223	88,850	119,099	234,527
VT	4,955	7,561	10,134	120,288
WI	37,396	56,978	76,258	414,818
WV	17,937	27,448	36,893	93,609
WY	4,256	6,495	8,706	82,090

This table estimates the additional votes in the 2008 Presidential contest had these states offered online registration; the final column lists the actual margin of victory in that election.

Table 4.A10: Estimated Influx of Voters, 2012

State	Lower Bound	Estimate	Upper Bound	Margin of Victory
AK	5,220	7,967	10,683	42,036
AL	39,839	60,861	81,668	460,229
AR	26,565	40,682	54,723	253,335
CT	27,143	41,425	55,534	270,210
DE	6,702	10,225	13,702	77,100
FL	146,480	223,729	300,158	74,309
GA	77,314	118,192	158,709	304,861
HI	11,901	18,217	24,495	185,643
IA	23,818	36,347	48,720	91,927
ID	12,524	19,145	25,708	208,124
IL	100,002	152,758	204,968	884,296
KY	38,350	58,639	78,758	407,820
MA	47,515	72,480	97,117	733,301
ME	10,204	15,562	20,845	109,030
MI	74,378	113,475	152,072	449,313
MN	33,276	50,678	67,797	225,942
MO	48,775	74,491	99,931	258,644
MS	23,512	35,902	48,153	147,797
MT	7,512	11,459	15,354	66,089
NC	77,987	119,124	159,832	92,004
NE	14,681	22,421	30,076	172,983
NH	10,687	16,309	21,863	39,643
NJ	66,954	102,277	137,234	647,861
NM	15,330	23,422	31,434	79,547
NY	152,974	233,794	313,862	1,995,381
OH	97,774	149,368	200,436	166,272
OK	32,328	49,452	66,447	447,778
PA	113,658	173,697	233,165	309,840
RI	8,440	12,893	17,300	122,473
SD	6,367	9,722	13,038	65,571
TN	57,251	87,584	117,693	501,621
TX	197,945	303,093	407,652	1,261,719
VA	64,929	99,165	133,035	149,298
VT	5,223	7,972	10,688	106,541
WI	40,603	61,886	82,854	213,019
WV	18,276	27,981	37,629	179,386
WY	4,773	7,293	9,788	101,676

This table estimates the additional votes in the 2012 Presidential contest had these states offered online registration; the final column lists the actual margin of victory in that election.

Table 4.A11: Effects on Political Participation (Logit)

	Registration		**Turnout**	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.039	(0.001)	0.051	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.224	(0.007)	0.149	(0.007)
Black	0.389	(0.014)	0.484	(0.013)
Asian	−0.942	(0.022)	−0.893	(0.021)
Other Race	−0.179	(0.023)	−0.251	(0.022)
Hispanic	−0.288	(0.014)	−0.258	(0.013)
Income	0.073	(0.001)	0.086	(0.001)
High School	−1.234	(0.010)	−1.143	(0.009)
Some College	−0.395	(0.012)	−0.397	(0.010)
Postgraduate	0.502	(0.021)	0.443	(0.015)
Unemployed	−0.185	(0.018)	−0.178	(0.017)
Not in Labor Force	−0.261	(0.010)	−0.180	(0.009)
Moved	−0.549	(0.008)	−0.530	(0.007)
Margin of Victory	−0.001	(0.000)	−0.004	(0.000)
Presidential	0.194	(0.014)	0.769	(0.013)
EDR	−0.023	(0.037)	0.046	(0.033)
No Fault Early			−0.055	(0.023)
No Excuse Absentee			0.038	(0.026)
Vote by Mail			0.059	(0.054)
Online Reg	0.096	(0.020	0.101	(0.018))
Intercept	0.313	(0.057)	−1.548	(0.056)
N	509,439		509,439	
Pseudo R^2	0.150		0.172	

This table represents the point estimates and associated (robust) standard errors for the logit model fitted to the entire sample, with Online Registration modeled as a binary treatment.

Almost all coefficients are statistically significant at the 1% level. The exception in the registration model is EDR (no significance); exceptions in the turnout model are No Fault Early (5% significance) and EDR, No Excuse Absentee, and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A12: Effects on Political Participation (Scobit)

	Registration		Turnout	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.029	(0.001)	0.046	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.149	(0.005)	0.130	(0.006)
Black	0.246	(0.011)	0.412	(0.012)
Asian	−0.657	(0.017)	−0.787	(0.020)
Other Race	−0.131	(0.016)	−0.223	(0.019)
Hispanic	−0.213	(0.010)	−0.233	(0.012)
Income	0.052	(0.001)	0.076	(0.001)
High School	−0.800	(0.015)	−0.983	(0.014)
Some College	−0.238	(0.009)	−0.333	(0.010)
Postgraduate	0.263	(0.013)	0.354	(0.014)
Unemployed	−0.138	(0.013)	−0.159	(0.015)
Not in Labor Force	−0.167	(0.007)	−0.154	(0.008)
Moved	−0.371	(0.008)	−0.458	(0.008)
Margin of Victory	−0.001	(0.000)	−0.004	(0.000)
Presidential	0.135	(0.010)	0.669	(0.014)
EDR	−0.009	(0.024)	0.038	(0.028)
No Fault Early			−0.048	(0.020)
No Excuse Absentee			0.031	(0.022)
Vote by Mail			0.048	(0.046)
Online Registration	0.063	(0.013)	0.087	(0.015)
Intercept	−1.261	(0.088)	−1.958	(0.062)
ln α	1.131	(0.067)	0.414	(0.039)
α	3.099	(0.208)	1.513	(0.060)
N	509,439		509,439	

This table represents the point estimates and associated (robust) standard errors for the scobit model fitted to the entire sample, with Online Registration modeled as a binary treatment.

Almost all coefficients are statistically significant at the 1% level. The exception in the registration model is EDR (no significance); exceptions in the turnout model are No Fault Early (5% significance) and EDR, No Excuse Absentee, and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A13: Effects on Political Participation (Probit, Alternate Coding of Non-Response to Participation)

	Registration		Turnout	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.017	(0.001)	0.025	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.109	(0.004)	0.070	(0.004)
Black	0.252	(0.008)	0.317	(0.007)
Asian	−0.518	(0.013)	−0.483	(0.012)
Other Race	−0.079	(0.014)	−0.123	(0.013)
Hispanic	−0.160	(0.008)	−0.132	(0.008)
Income	0.043	(0.001)	0.051	(0.001)
High School	−0.690	(0.006)	−0.668	(0.005)
Some College	−0.201	(0.007)	−0.218	(0.006)
Postgraduate	0.250	(0.011)	0.248	(0.009)
Unemployed	−0.125	(0.011)	−0.115	(0.010)
Not in Labor Force	−0.150	(0.006)	−0.102	(0.005)
Moved	−0.326	(0.005)	−0.329	(0.004)
Margin of Victory	−0.001	(0.000)	−0.003	(0.000)
Presidential	0.073	(0.008)	0.432	(0.008)
EDR	0.007	(0.022)	0.035	(0.019)
No-Fault Early			−0.039	(0.014)
No-Excuse Absentee			0.035	(0.015)
Vote by Mail			0.043	(0.032)
Online Registration	0.062	(0.012)	0.062	(0.011)
Intercept	0.436	(0.034)	−0.701	(0.033)
N	509,439		509,439	
Pseudo R^2	0.143		0.162	

This table represents the point estimates and associated (robust) standard errors for the same model as Table 4.A4, but registration and turnout are measured differently. For each outcome variable, all those who did not answer the question are coded as participants. (In Table 4.A4, these respondents are coded as non-participants, following the rubric of the Census Bureau.)

Almost all coefficients are statistically significant at the 1% level. The exceptions in the registration model are Age^2 and EDR (no significance); exceptions in the turnout model are No-Excuse Absentee (5% significance), EDR (10% significance), and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Table 4.A14: Effects on Political Participation (Probit, Omitting Non-Respondents to Participation Questions)

	Registration		Turnout	
Variable	**Coefficient**	(Std. Err.)	**Coefficient**	(Std. Err.)
Age	0.020	(0.001)	0.029	(0.001)
Age^2	0.000	(0.000)	0.000	(0.000)
Female	0.120	(0.004)	0.082	(0.004)
Black	0.252	(0.009)	0.312	(0.008)
Asian	−0.548	(0.013)	−0.524	(0.013)
Other Race	−0.091	(0.014)	−0.142	(0.013)
Hispanic	−0.172	(0.008)	−0.149	(0.008)
Income	0.044	(0.001)	0.053	(0.001)
High School	−0.715	(0.006)	−0.691	(0.005)
Some College	−0.209	(0.007)	−0.230	(0.006)
Postgraduate	0.252	(0.011)	0.251	(0.009)
Unemployed	−0.126	(0.011)	−0.115	(0.010)
Not in Labor Force	−0.153	(0.006)	−0.107	(0.005)
Moved	−0.334	(0.005)	−0.330	(0.004)
Margin of Victory	−0.001	(0.000)	−0.003	(0.000)
Presidential	0.087	(0.008)	0.457	(0.008)
EDR	0.001	(0.022)	0.030	(0.020)
No-Fault Early			−0.038	(0.014)
No-Excuse Absentee			0.031	(0.015)
Vote by Mail			0.040	(0.032)
Online Registration	0.061	(0.012)	0.063	(0.011)
Intercept	0.325	(0.034)	−0.857	(0.034)
N	496,837		499,463	
Pseudo R^2	0.153		0.173	

This table represents the point estimates and associated (robust) standard errors for the same model as Table 4.A4, but registration and turnout are measured differently. For each outcome variable, all those who did not answer the question are dropped from the sample. (In Table 4.A4, these respondents are coded as non-participants, following the rubric of the Census Bureau.)

Almost all coefficients are statistically significant at the 1% level. The exception in the registration model is EDR (no significance); exceptions in the turnout model are No-Excuse Absentee (5% significance) and EDR and Vote by Mail (no significance).

I include both state and year fixed effects. Year 2000, Year 2010, and Maine are dropped due to perfect collinearity. In the Turnout model, Oregon is also dropped due to perfect collinearity.

Chapter 5

Who Votes, and How?

5.1 Summary

The literature on convenience voting has devoted considerable attention to identifying its impact on turnout, and to a lesser extent, the demographic representativeness of the electorate. Most of the papers in this area, however, merely speculate how voting liberalizations will affect electoral outcomes. This de-emphasis is likely a result of the lack of ideological data in the *Current Population Survey*, rather than disinterest. This paper offers a new approach for estimating the partisan composition of the electorate under various electoral conditions. After identifying an electorate using the Rosenstone-Wolfinger approach, a Random Forest model (generated by out-of-sample data) simulates how these individuals would have voted in the 2012 Presidential contest. The distributions of vote share barely shift, offering empirical support for the conjecture that convenience voting procedures will minimally affect electoral outcomes. Although the substantive focus of discussion is on online registration, the procedure advanced herein is readily applicable to other forms of convenience voting. More broadly, this paper offers a blueprint for any project that can incorporate multiple surveys, as long as the two surveys contain a common set of questions that predict the missing outcome.

5.2 Introduction

Modern American elections are characterized by low turnout, with participation rates far lower than those of other contemporary democracies (Hanmer 2007, 2009). Moreover, the demographic distribution of the electorate differs meaningfully from that of the voting-eligible population (Schlozman, Verba, and Brady 2012). Over the past several decades, states have introduced various liberalizations to try to stimulate turnout by improving the voter experience, popularly called "convenience voting" policies. The political science literature has devoted considerable attention to the impact of these policies, beginning with a seminal paper by Rosenstone and Wolfinger (1978).

Since this important paper, political scientists have carefully scrutinized convenience voting procedures, especially Election Day Registration (hereafter, EDR) and the Motor Voter provisions (Hanmer 2009). Still, though, the literature is ambivalent about the effectiveness about specific policies, as well as the appropriate methodological approach (*cf.* Berinsky 2005). Yet while many scholars have contributed to the debate on the consequences for turnout, the literature is fairly quiet about the impact on actual election results. In a review of the convenience voting literature, Gronke *et al.* point out that "the vast bulk of political science research in the field has [instead] concentrated on voter participation" and to a lesser degree, the demographic representativeness of the electorate (2008: p. 438). A discussion of the impact on the partisan composition of the electorate is usually relegated to a quick hypothesis near the end of the manuscript, if included at all.

This is a noteworthy and unfortunate gap in the literature. Within and outside of academia, people are interested not only *whether* people are voting, but also *how* they are voting. Popular wisdom contends that voting liberalizations will aid Democrats, since those in lower socio-economic brackets vote at systematically lower rates, and these individuals tend to prefer Democratic candidates (Highton and Wolfinger 1998). But many political scientists, including Rosenstone and Wolfinger, speculate that Democrats will benefit only marginally, or not at all (Rosenstone and Wolfinger 1978;

Mitchell and Wlezien 1995; Highton and Wolfinger 1998; Knack and White 2000; Brians and Grofman 2001; Highton 2004; McDonald 2008). One explanation is that convenience voting primarily impacts the behavior of those who have moved recently and young adults, rather than those in lower income brackets (Highton and Wolfinger 1998; Knack and White 2000). Gronke *et al.* (2008) infer from the literature a positive correlation between strength of partisanship and utilization of convenience methods, suggesting that these procedures in and of themselves will impact the aggregate partisan distribution only minimally.

Other papers argue to the contrary, and suggest that convenience voting may actually aid Republicans. The narrative is as follows: Convenience voting procedures do little to encourage non-voters to alter their behavior, and instead render the process easier for those who are already politically engaged. Berinsky (2005) and Berinsky, Burns, and Traugott (2001) offer evidence that these reforms actually render the electorate *less* demographically representative, which might in turn lead to Republican gains. And in a recent paper, Neiheisel and Burden (2012) use a natural experiment in Wisconsin to conclude that EDR increased Republican vote share; they suggest that the non-voters likely to capitalize on EDR are politically resourceful and from higher socio-economic brackets, and they also posit that EDR might hinder mobilization efforts because the pool of registrants is not fully realized prior to the election.

In my review of the convenience voting literature, only two papers infer conclusions regarding partisanship from an individual-level regression model: Mitchell and Wlezien (1995) and Berinsky, Burns, and Traugott (2001).[1] This absence could be due to a lack of data: Many papers in this area rely on the *Current Population Survey*, administered by the U.S. Census Bureau, which asks respondents about participation but not about partisanship or ideology.[2] Brians and Grofman (2001) explicitly lament

[1]To derive their model, Mitchell and Wlezien use data from the *American National Election Survey*, which they admit has some shortcomings compared to the more-commonly used *CPS*. Berinsky, Burns, and Traugott use a dataset restricted to Oregonians, so the scope of inference is inherently limited.

[2]An incomprehensive list includes Rosenstone and Wolfinger (1978), Leighley and Nagler (1992), Nagler (1994), Mitchell and Wlezien (1995), Highton (1997), Highton and Wolfinger (1998), Knack and White (2000), Brians and Grofman (2001), Alvarez and Nagler (2007, 2008, 2011), Hanmer (2007), Keele and Minozzi (2013), and Glynn and Quinn (2011).

this drawback in their paper and consign their analysis to comparisons of Democrat-Republican vote share prior to and after implementation of EDR. Knack and White (2000) simply include a measure of state-level Republican vote share in their turnout model.

A similar paucity of data inspired methodological innovation in the public opinion and electoral forecasting literature (see Erikson, Wright, and McIver 1993). My paper advances a new method for estimating how a change in election policy will affect vote share, using out-of-sample data to simulate the *individual* presidential preferences of those most likely to alter their turnout behavior. Other work in this area uses an individual-level model to identify the demographic distribution of the counterfactual electorate, but these papers use aggregate opinion of demographic groups from other surveys to inform their inference (Rosenstone and Wolfinger 1978; Highton and Wolfinger 1998; Knack and White 2000). This paper is the first (to my knowledge) to incorporate an individual-level model of partisanship after parametrically identifying the electorate. To ground the analysis, I focus exclusively on how online registration, one of the most recent innovations, might impact electoral outcomes, but my method is readily applicable to other forms of convenience voting.

The paper proceeds as follows: I first discuss the qualitative evidence that suggests that online registration does not benefit a particular political party (at least uniformly). Because online registration has been supported and signed into law by politicians of both major parties, I predict that online registration will not lead to substantial electoral gains for the state's minority party. I briefly review the turnout model from Chapter 4 of my dissertation, which allows me to identify the demographic composition of the electorate with and without online registration. I then describe how and why I use a nonparametric technique called Random Forests to describe the relationship between demography and vote choice, using out-of-sample data. The Random Forest model enables me to simulate partisan distributions for conditions with and without online registration. The results confirm that this particular form of convenience voting does not dramatically alter the partisan landscape for any given state. I conclude with a discussion of the implications for further adoption of online registration, and an exploration of how this method can advance other research in this area and others.

5.3 Online Registration: A Bipartisan Movement

Online registration is one of the newest forms of convenience voting, a byproduct of perennial interest in mobilization and an increasingly tech-savvy electorate (Pellissier 2015).[3] The adoption of online registration offers an especially interesting case study for partisanship, because unlike other forms of convenience voting, it is not associated with a particular political party (see Underhill, Hernandez, and Hubler 2014). In contrast, for example, the National Voter Registration Act of 1993 was strongly backed by Democrats, and signed into law by a Democratic President (Clinton) after being rejected by a Republican (Bush Sr.); Republican leadership in the House later supported its repeal, and Republican Governors resisted its implementation (Highton and Wolfinger 1998). Ponoroff highlights the peculiarity of online registration's broad ideological appeal in a "field often subject to partisan bickering" (2010: p. 2). The qualitative evidence supports these claims: Both red and blue states have introduced online registration and the bipartisan Presidential Commission on Election Administration recently endorsed it as a valuable innovation in election administration (2014). Table 5.1 details the states that currently allow online registration, as well as the partisan affiliation of the Governor who signed the requisite legislation.

Online registration offers a number of benefits to the government. The Presidential Council on Election Administration argued that online registration streamlines the administrative duties of the state and enhances the voter experience (2014). The Council lauded this policy as a way of easing the financial and administrative burdens of running elections, and removing barriers that might discourage potential voters. Ponoroff (2010) mentions that states do need to invest a sizable sum to develop the infrastructure to allow for online registration, but the marginal cost of registering

---

[3]For an excellent resource on which states currently allow online registration, refer to the NCSL website (last updated 2014): `http://www.ncsl.org/research/elections-and-campaigns/electronic-or-online-voter-registration.aspx`.

Table 5.1: Legislative History of Online Registration

State	**Signed**	**Legislation**	**Governor (Party)**
Arizona	*no legislation required, implemented 2002*		
California	2011	SB 397	Jerry Brown (D)
Colorado	2009	HB 1160	Bill Ritter (D)
Connecticut	2012	HB 5024	Dannel Malloy (D)
Delaware			
Georgia	2014	HB 942	Nathan Deal (R)
Hawaii	2012	HB 1755	Neil Abercrombie (D)
Illinois	2013	HB 2418	Patt Quinn (D)
Indiana	2009	HB 1346	Mitch Daniels (R)
Kansas	*no legislation required, implemented 2009*		
Louisiana	2009	HB 520	Bobby Jindal (R)
Maryland	2012	HB 173	Martin O'Malley (D)
Massachusetts	2014	HB 3788	Deval Patrick (D)
Minnesota	2014	HF 2096	Mark Dayton (DFL)
Missouri	2014	HB 1739	Jay Nixon (D)
Nebraska		LB 661	
New York	2011	A08165	Andrew Cuomo (D)
Oregon	2009	HB 2386	Ted Kulongoski (D)
South Carolina	2012	H 4945	Nikki Haley (R)
Utah	2009	SB 25	Jon Huntsman, Jr. (R)
Virginia	2012	HB 2341	Bob McDonnell (R)
Washington	2007	HB 1528	Christine Gregoire (D)
West Virginia	2013	SB 477	Earl Ray Tomblin (D)

This table delineates the year that each state passed legislation allowing online registration, the specific bill, and the Governor who signed it into law. I obtained much of this information from NCSL (2013). It is important to note that the date of implementation does not always occur in the same year as the date the legislation passed.

individuals is substantially lower with an online (rather than paper-based) method. Additionally, he recounts anecdotal evidence from election administrators that suggests fewer registration errors, which have been shown to disenfranchise voters (see Alvarez and Hall 2014).

The extant literature does not offer a great deal of information about the impact of online registration on the partisan composition —at least not directly. Most of the limited discussion focuses on the distinct but related question of how this form of convenience voting will alter the demographic composition of the electorate. Baretto *et al.* (2010) find that whites, young adults, and urbanites were most likely to register online in Arizona and Washington. This study, commissioned by Pew, also claims that many residents of these states indicated that they would update their register information online should they move. Using a probit model, Pellissier (2015) draws similar conclusions: Young adults and recent movers are more likely to participate if the state allows online registration, but not racial minorities. Meanwhile, García Bedolla and Vélez (2013) focus exclusively on the response in California. The ethnic composition of the individuals who registered online reflected the composition of the state's population fairly closely. Again, young voters were most likely to utilize this method, especially among Latinas. Moreover, restricting their attention to two counties, these authors ascertain that most of the individuals who registered online lived in less affluent areas.

It is difficult to extrapolate from these results how online registration will affect election results, as young voters and the less affluent tend to be more Democratic, while white voters tend to be less so. In my reading of the literature, the discussion of the implications for partisanship is somewhat limited. In Washington and Arizona, online registrants were significantly more likely to avoid registering with a particular party (Baretto *et al.* 2010). Among online registrants in California, women and minorities tended to register with the Democratic Party at higher rates than men and whites, respectively (García Bedolla and Vélez 2013: Figure 2, p. 3). These authors determine that race and gender, and their interaction, are important determinants of the political affiliations of individuals who register online.

A priori, then, it is still somewhat ambiguous whether online registration will alter the aggregate partisan distribution of the electorate, and if so, how. It is also entirely possible that the impact of online registration on the preference distribution will be heterogeneous across states. In the Downsian framework (1957), online registration should primarily affect those who are nearly indifferent between casting a ballot and abstaining; the characteristics of this population could vary by state. If politicians are office-motivated, it is at least somewhat unlikely that they would support a piece of legislation that reduced their future electoral prospects, or their party's.[4] It is also worth noting that few of these states typically hold statewide elections that are competitive. Given this incentive structure, I speculate that online registration will not result in disproportionate gains for the state's minority party. At the very least, I expect that any gains will be minimal enough to preserve the outcome for the majority party.

Like any question of causal inference, the primary difficulty is that of missingness: We only observe one of two (or more) potential outcomes (Rubin 1974). In this context, I only know with certainty how individuals behave with or without the opportunity to register online. If a large, nationally-representative survey included information about participation and candidate preference, I would only need to identify who belonged to the electorate under each condition. The *American National Election Survey* asks about both, but the sample size for each state is relatively small for each electoral cross section. As mentioned earlier, the *CPS* is a rich source of data for studying registration and turnout, and much of the participation literature relies on this survey instrument (see footnote 2); the *CPS*, though, does not ask any questions about candidate preference, partisan identification, or ideology. Meanwhile, the *Cooperative Congressional Election Study* is a large, nationally-representative dataset that offers excellent information about political preferences (Ansolabehere 2013; hereafter, *CCES*). Yet although the researcher can recover the estimates of turnout fairly

[4]The focus of my paper is the voter perspective, so a full discussion of politicians' incentives and decision-making is beyond its intended scope. By and large, however, majority party support of a piece of legislation suggests that the party does not perceive a policy as a threat to its future electoral success.

closely using the given weights, the raw frequency of individuals who report abstaining is fairly low for some of the less-populated states. In Wyoming, for example, only 3 of the 110 individuals sampled reported abstaining; for an additional 28 individuals, the dataset does not list a response (probably due to survey attrition, given that "Skipped" and "Not Asked" were other possible values). Similarly, in Vermont, only 5 (of 159) reported not casting a ballot, with 19 missing values. Accordingly, it might be overly ambitious —if not impossible —to procure enough within-state variation to describe the counterfactual electorate.

To identify how online registration affected the electoral outcomes in the 2012 election, I need information about both participation *and* partisanship. Fortunately, the *CPS* and the *CCES* profile demography similarly, and demography explains a great deal of the variation in political preference within a given state (see Park, Gelman, and Bafumi 2004). I exploit this relationship to bridge information across these two surveys, even though they sample different individuals. More specifically, like so much of the work on voter participation, I model turnout behavior with the *CPS*. I then fit a Random Forest model to the *CCES* data to predict partisanship as a function of demographic variables. I use these results to simulate the preferred candidate for the *CPS* individuals predicted to vote by the turnout model. My method allows the researcher to take advantage of the relative merits of both surveys without having to limit the scope of investigation. The literature on convenience voting offers a nice blueprint for identifying turnout behavior with the *CPS* dataset, and I discuss that before turning my attention to my less conventional model of partisanship.

5.4 Predicting Turnout

Most of the convenience voting literature builds upon the framework introduced in Rosenstone and Wolfinger (1978), in which a binary response model recovers the propensity to participate (a latent variable). The determinants of participation include individual-level demographic variables (*e.g.* age, gender, education) and state-level electoral conditions (particularly policy).

The Rosenstone-Wolfinger approach is not without detractors. Recently, several papers have expressed concern about the assumptions embedded within this method. The binary response model assumes that the treatment distribution is independent of the first moment for the distribution of potential outcomes, at least after conditioning on a set of pre-treatment variables. But several papers have questioned whether this condition is satisfied, notably Hanmer (2007, 2009), Glynn and Quinn (2011), and Keele and Minozzi (2013). Hanmer (2007) finds that the results are not overly sensitive to the model specification, and so cannot reject the Rosenstone-Wolfinger approach, but Glynn and Quinn (2011) and Keele and Minozzi (2013) argue that the conventional approach grossly overestimates the impact of convenience voting (specifically, EDR) on turnout.

These papers helpfully advance methods that rely on less tenable assumptions, but as I argued in Chapter 2, the current body of evidence fails to discredit the traditional approach to estimation in this area of the literature. The bounding techniques employed in Hanmer (2007, 2009) and Glynn and Quinn (2011) serve as a useful sanity check, but (used alone) offer an unnecessarily restrictive scope of inference. Meanwhile, the natural experiments advanced in Keele and Minozzi (2013) and Neiheisel and Burden (2012) require a specific form of implementation, and it is somewhat precarious to extrapolate that individuals in other states will respond like residents of Wisconsin and Minnesota, two states with anomalous participation histories. Most problematically, the parametric results that Glynn and Quinn (2011) critique are unduly inflated by the specification that they employ.

Although it is important to acknowledge these concerns about the validity of the Rosenstone-Wolfinger framework, I contend that the current body of literature has yet to dismantle it. As in Chapter 1, I use a probit model to estimate the propensity to vote as a function of demographic variables (age, gender, race, education, employment status, income, and residential mobility) and state-level variables (competitiveness, EDR, no-fault early voting, no-excuse absentee voting, postal voting, and of course, online registration). While many papers in this area have used cross-sectional analysis, I pool cross-sections to create a panel dataset, as advocated in Leighley and Nagler (2014).[5] For identification purposes, I include at least one pre-treatment wave for each state that currently allows online registration; since Arizona pioneered this procedure prior to the 2002 election, I include all cross-sections from 2000 through 2012.

Table 5.2: Accuracy of the Turnout Model

State	**% Correctly Predicted**	**State**	**% Correctly Predicted**
Minnesota	80.71	Missouri	74.39
Wisconsin	80.41	Georgia	73.75
Colorado	80.40	Ohio	73.67
Massachusetts	78.11	South Carolina	73.48
Maine	78.08	Wyoming	73.48
Delaware	78.00	Alaska	73.04
Oregon	77.82	Utah	72.58
New Hampshire	77.36	Illinois	72.50
Maryland	77.01	New Mexico	72.42
Connecticut	76.52	Rhode Island	72.40
North Carolina	76.39	South Dakota	72.25
Virginia	76.38	Pennsylvania	72.13
Michigan	76.28	Kentucky	71.93
Iowa	76.27	New Jersey	71.57
Washington	76.14	California	71.50
Vermont	76.11	Tennessee	71.10
Indiana	75.77	Oklahoma	71.05
Montana	75.75	Texas	70.75
Kansas	75.63	New York	70.69
Louisiana	75.48	Arizona	70.22
Mississippi	75.37	Arkansas	69.77
Florida	75.32	Hawaii	67.93
Alabama	75.11	Nevada	67.05
Idaho	74.90	West Virginia	66.82
Nebraska	74.77		

This table presents the number of individuals correctly classified as voters or abstainers by the parametric turnout model, disaggregated by state. I use a probit specification and classify an individual as a voter if the estimated propensity to vote exceeds 50%.

For more details about the model and data, refer to Chapter 1. The fitted model yields $\hat{Y} = \Phi(\hat{\alpha} + X\hat{\beta} + Z\hat{\rho} + \hat{\delta}T + \hat{\gamma} + \hat{\theta})$, where $\hat{Y}$ is the estimated propensity to cast a ballot. To generate two different electorates, one for each online registration condition, I consider an individual as a voter if $\hat{Y} \geq 0.5$ Overall, the model performs quite well, correctly classifying nearly three-quarters of respondents (74.14%). Table 5.2 lists the percent correctly predicted by state.

[5]See Chapter 2 of my dissertation for a brief discussion of why panel datasets offer a superior statistical approach. Cross-sectional data may be biased by electoral anomalies, as I found in my replication of Glynn and Quinn (2011); Keele and Minozzi (2013) made this point, as well.

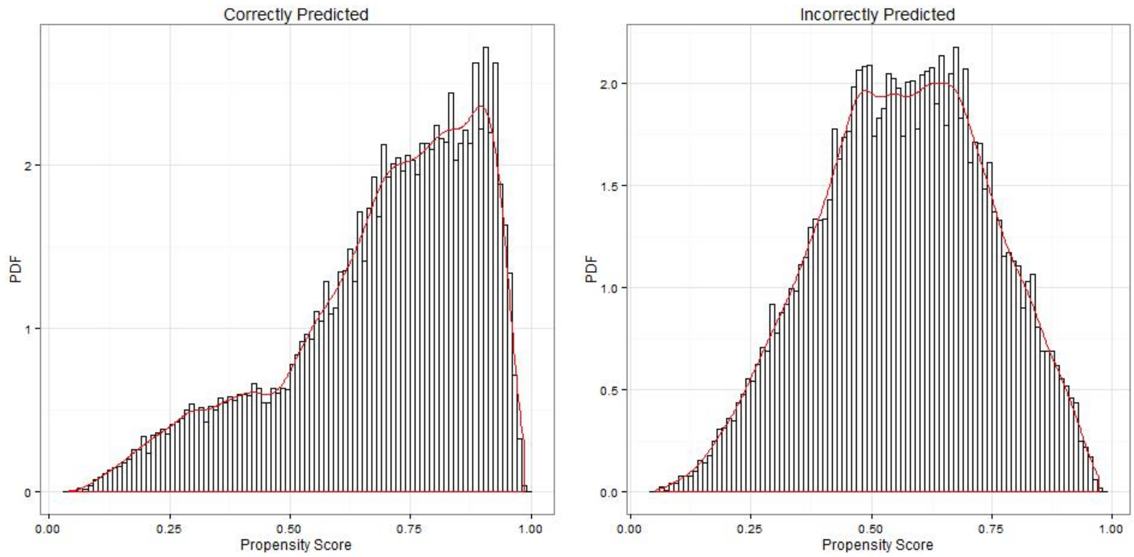


Figure 5.1: Classification Accuracy by (Estimated) Propensity Score

Stratifying by turnout behavior, the model accurately predicts turnout for 86.86% of voters but only 43.73% of non-voters. Figure 5.1 illustrates the distribution of propensity scores for those which the model predicts accurately (1a) and inaccurately (1b). Interestingly, the distribution of propensity scores is far more symmetric for individuals whom the model incorrectly classifies; this suggests that the raw frequencies of individuals the model misclassifies are fairly similar across the two groups. But because 70% of the individuals in the sample reported voting in the 2012 election, this means that the *proportion* of voters the model misclassifies is far lower than the proportion of abstainers.

Having identified electorates in each online registration condition, I need some way to estimate aggregate vote share at the state level. As mentioned earlier, the *CPS* does not describe political opinion or candidate preference. To gain insight into the political consequences of this election policy, I need an individual-level model of vote choice. The next section describes how I create this model and use it to simulate candidate preference for an arbitrary individual.

5.5 Predicting Partisanship

5.5.1 Motivation

The fundamental challenge for this particular research question is how to accommodate missingness. Most research in the participation literature uses the *CPS*, which does not ask about partisanship, vote choice, or ideology. In this case, then, the concern is not whether data are missing at random or conditionally at random, but simply that they are missing altogether. To simulate vote choice in counterfactual scenarios using the *CPS* data, I need a method to identify the relationship between characteristics described in the *CPS* and candidate preference. In other scholarship, researchers use the joint distributions of non-missing observations *within* a dataset to impute values for the missing observations ("multiple imputation"; see King *et al.* 2001). Unfortunately, though, since the desired information is missing for all observations in the *CPS*, this method will not suffice. In this chapter, I instead estimate these joint distributions using another survey entirely, leveraging the fact that both the *CPS* and the *CCES* include common demographic covariates that predict vote choice fairly accurately, and the *CCES* also includes various measures of candidate preference.

The motivation behind this paper bears some resemblance to the literature on state-level public opinion, particularly the studies that use disaggregation (Erikson, Wright, and McIver 1993) and Multilevel-Regression and Poststratification (MRP; introduced in Park, Gelman, and Bafumi 2004). Disaggregation merges national surveys and then (as the name suggests) partitions individuals into states; the partitioning is immediate, but finding enough national-level surveys and making the questions of interest compatible across surveys can invite some frustration. The related MRP technique is more involved in terms of modeling, but also more robust (Kastellec, Lax, and Phillips 2014). With this procedure, the researcher recovers state-level preferences from a national survey (or collection of surveys) by fitting a predictive model to the entire dataset, deriving within-state demographic weights via Census data, and then weighting predictions accordingly before aggregating. The MRP approach requires a smaller collection of data because it exploits the relationship between demography and public opinion and the demographic information readily available from the U.S. Census.

In both this stream of scholarship and my own project, the researcher aims to construct a measure of public opinion for each state, but lacks the necessary data. The reason the data are missing differs, though: In the public opinion literature, the researcher does not have enough data points in every state to construct an aggregate measure with confidence. The *CCES* dataset is a large enough sample to allow for such measurement, but cannot identify a counterfactual electorate. More specifically, my turnout model is necessary to predict the demographic distribution of each state's electorate both with and without online registration. The Census-derived demographic weights in MRP only inform the researcher about the demographic composition of the voting-age population, *not* the population of voters conditioned on a particular electoral policy.

One approach to estimating partisanship in a different scenario would be to use the Rosenstone-Wolfinger turnout model to identify a counterfactual electorate, and impute counterfactual weights for this group. Applying MRP would be immediate. Although MRP has enjoyed some prominence within the public opinion literature, though, I use an alternative technique from the machine learning literature called Random Forests. I first discuss CART, the intellectual predecessor to Random Forest, and I then explain why I believe this research design to be superior.

5.5.2 Introducing CART

Breiman *et al.* (1984) introduce Classification and Regression Trees, a nonparametric method for identifying relationships in data. The two types of tree are similar, but classification trees are specific to outcomes with categorical values, such as vote choice. In common parlance, a classification tree looks like a flowchart; theorists could alternatively imagine an extended game. There are an outcome of interest Y and a set of predictor variables X. The classification tree sequentially bifurcates individuals into different classes based on their covariates profiles.[6] More concretely, for the turnout outcome, a tree could first partition the sample into different age brackets, and then divide each age bracket into different racial groups. The process would continue iteratively until some stopping criterion is satisfied. (This criterion is unnecessary for the Random Forest, as explained below.) The aim is to repeatedly split the data until the classes are as homogeneous in their outcome values as possible, to maximize predictive accuracy. As Hastie, Tibshirani, and Friedman (2011) point out, a classification tree is easily interpretable, and the researcher needs no prior knowledge about the functional relationship between X and Y.

Classification trees are attractive for a number of reasons: they are intuitive and computationally efficient, and require no assumptions about the functionality of the relationship. But they are not without drawbacks. Notably, they can be incredibly unstable, as "the effect of an error in the top split is propagated down to all of the splits below it" (Hastie, Tibshirani, and Friedman 2011: p. 312). This concern, known as "overfitting," applies to almost all models, as it can lead the researcher to extrapolate conclusions that may be unwarranted, or at the very least, not quite accurate. It is especially pertinent, though, if the aim is to make out-of-sample predictions. Intuitively, overfitting means underestimating the variance of the stochastic component of the functional relationship. To accommodate this concern, a variety of techniques have modified the classification tree to integrate this uncertainty.

Random Forest, introduced in Breiman (2001) and further formalized in Giau (2014), is one such modification.[7] This method performs favorably to other approaches, such as bagging or boosting (Hastie, Tibshirani, and Friedman 2011), and it is computationally more efficient (Breiman and Cutler n.d.; Kuhn and Johnson 2014). The Random Forest generates a pre-specified number K of classification trees. The Random Forest safeguards against overfitting by estimating a model for a boot-

[6]It is possible to partition into more than two groups for each step, but Hastie, Tibshirani, and Friedman (2011) advise against this procedures, since it can affect the splits at the next nodes. They point out that collapsing in this way is necessary, since the algorithm can simply divide the data again on the same variable if it provides the optimal split.

[7]Zachary Jones maintains a collection of articles that discuss Random Forests on the following website: `https://github.com/zmjones/statisticallearning/blob/master/abf.md`.

strapped sample of the dataset for each classification tree k in the ensemble (Breiman 2001). The cross-validation, then, is built into the model, similar to bagging procedures. But bagging can still lead to highly-correlated trees, so Random Forest incorporates randomness into the covariates, as well (Hastie, Tibshirani, and Friedman 2011). To reduce statistical dependence among the classification trees in the ensemble, the Random Forest only allows the algorithm to choose among a random (strict) subset of the variables at each node. For each observation in the dataset, then, the Random Forest has K independently-drawn predictions of the individual's class (Breiman 2001).

Ensemble methods, like the Random Forest, allow "increased out-of-sample stability and the ability to capture complex functional forms with relatively simple classifiers" (Grimmer and Stewart 2013: p. 278). Although this method has become quite popular (Hastie, Tibshirani, and Friedman 2011), it is still largely absent from the political science literature (Jones and Linder 2015). For a more detailed treatment of Random Forests, see the original paper Breiman (2001), Chapter 8 of Kuhn and Johnson (2014), or Chapter 15 of Hastie, Tibshirani, and Friedman (2011). In the next subsection, I discuss my particular execution of the Random Forest model, as well its advantages over MRP.

5.5.3 The Value of Random Forests

As mentioned earlier, another design for this question would simulate a counterfactual electorate, re-weight the subset of voters, and then apply MRP. There are a number of reasons to prefer the Random Forest design. First, the Random Forest model already incorporates a safeguard against overfitting by using a bootstrapped sample for each classification tree. The version of MRP implemented in the current body of literature does not, to my knowledge, incorporate any kind of cross-validation. Of course, it is possible to adjust the bias induced by overfitting by performing some kind of cross-validation (for example, k-fold cross-validation, as presented in Section 4.4 of Kuhn and Johnson 2014), but the researcher should add this extra step.

More precariously, the MRP approach is highly parametric, and so requires all the assumptions embedded in this class of models.[8] In Park, Gelman, and Bafumi (2004), the function f linking Y and X is an inverse logistic CDF, though the authors emphasize that their approach is generalizable to other specifications. Alternatively, in the Random Forest model, f is a generic function; like other machine learning techniques, this procedure allows an algorithm to "learn" the ideal representation from the data. This functional flexibility is appealing, as it allows the researcher to have a completely diffuse prior about the relationships in the data; it imposes no assumptions about the form or distribution, but instead discovers the functional form that best explains the relationship (Jones and Linder 2015). Moreover, while Park, Gelman, and Bafumi (2004) and Kastellec, Lax, and Phillips (2014) need to include interacted variables in their specification, the Random Forest identifies interactions through natural experimentation with the data (Breiman and Cutler n.d.).

Furthermore, the parametric specification in the MRP literature fits a single model to all fifty states, although the model is multilevel in nature. But this restriction imposes a single functional form on the preference relationship, while the partisan data suggest a good deal of variance across states in how demographic inputs affect vote choice. In other words, the MRP method not only assumes a particular functional form for the joint distribution of the outcome and covariates (*e.g.* logit, probit), but it also further imposes that the joint distribution the parametric model recovers is homogeneous across all states, which is quite unrealistic. Consider the proportion of *CCES* individuals who reported casting a ballot for Obama in 2012. In Vermont, 66% of all residents indicated doing so, and 67% of white residents; meanwhile, in Louisiana, these numbers were 36% for all residents and only 21% for whites. A parametric regression on pooled data would control for other factors that could intervene, but it would estimate a single parameter (or vector of parameters) that explain(s) the relationship between race and preference. It is obvious from this example, though, that race relates to vote choice heterogeneously across states. In my Random For-

[8] Recall that I am already using a parametric design in the turnout model, which itself has raised concern in the literature, discussed in Chapter 2.

est model, alternatively, I disaggregate the individuals by state and begin with a generic function f mapping demographic profile to preference space. This allows the researcher to recover the proper functional form to describe the relationship between demography and vote choice *in each state*.

5.5.4 Data and Execution

Nonetheless, the MRP literature offers a useful insight into measuring public opinion: Demographic information can explain a good deal of the variation in political preferences. As in Park, Gelman, and Bafumi (2004) and Kastellec, Lax, and Philips (2014), I model preference as a function of Age Bracket, Gender, Race, and Education.[9] I also add a variable for Employment Status to capture the individual's socio-economic background more fully. The MRP approach specifies a multilevel model to incorporate state-level characteristics, since the typical paucity of data does not allow the researcher to specify an individual-level model for each state. For my purposes, the richness of the *CCES* does permit me to fit a separate model to the sampled individuals in each state, so I rely on the following functional form:

$$Pref = f(\mathit{AgeBracket}, \mathit{Gender}, \mathit{Race}, \mathit{Education}, \mathit{Employment} \,|\, \mathit{State}).$$

Again, allowing the algorithm to discover a different functional representation for every state allows demography to explain vote choice heterogeneously across states.

In the post-survey wave, the *CCES* asks all voters for whom they cast a ballot in the presidential contest (variable "CC410a"), and all non-voters whom they preferred (variable "CC410a_nv"). I use both of these variables to create a categorical variable P of presidential vote choice.[10] Admittedly, these two variables admit measurement

[9] The authors of Park, Gelman, and Bafumi (2004) only consider African American heritage; I partition race slightly differently, described in the next subsection.

[10] In the pre-survey wave, there is a variable "V305c," which asks individuals whom they prefer most for President. This same question is asked of all respondents, whether or not they eventually vote, so researchers could consider using this variable instead. On the other hand, in battleground states, campaign activity is heightened in the weeks leading up to the election, so individuals in the states we are most interested in analyzing may be subject to changing their minds between the pre-survey wave and Election Day.

error differently; while individuals who voted might recall their ballots imperfectly, or be reluctant to share that information, non-voters are subject to other kinds of uncertainty. By using information from both variables, however, I increase the statistical power of the model. Candidate preference P_i is treated as an individual-level variable drawn from a generalized Bernoulli distribution $\{p_i^{Obama}, p_i^{Romney}, p_i^{Other}\}$. The basic goal of my predictive model is to generate an empirical probability distribution that is specific to an individual's demographic profile and state of residence.

Individuals are partitioned into four different age groups: 18-25, 26-45, 46-65, and 66^+. The *CPS* and *CCES* measure ethnicity slightly differently. The *CCES* allows the individual to choose "Hispanic" when asked about race, but the *CPS* does not, instead asking later whether the individual is of Hispanic heritage. If an individual in the *CPS* indicates Hispanic heritage, I classify that individual as "Hispanic" (in other words, this designation overrides the other racial categories). Otherwise, the *CPS* creates finer partitions for race. Accordingly, I designate individuals as "White," "Black", "Hispanic," or "Other," using the (broader) *CCES* partitions. Gender is a binary variable, "Male" or "Female." Education is an ordinal variable, with four levels: "High School," "Some College," "College," and "Postgraduate." Employment (the only variable the cited MRP literature did not include) is a categorical variable, with three possible values: "Employed," "Unemployed," and "Not in Labor Force."

I implement the procedure with the "randomForest" package in R (Liaw and Wiener 2002; adapted from Breiman and Cutler's Fortran code). Per the recommendation of Kuhn and Johnson (2013), I create an ensemble of 1000 classification trees. At each node of every classification tree, the Random Forest model chooses the optimal split using two randomly-selected variables.[11] I subdivide the data by state, and the Random Forest model generates 1000 predictions of candidate preference for an individual of arbitrary demographic background; the model assigns the individual to whichever class receives the most votes among the individual trees.

I could use this approach to make an out-of-sample prediction, but doing so would

[11] The researcher can manually tweak the number of variables randomly selected, but Breiman (2001) recommends the square root of the total number.

disregard the probability that the model misclassifies the individual. To incorporate this information, I take 500 bootstrap samples of the *CCES* dataset for each state; I use these samples to estimate

$$\mathbb{P}(\text{Pref} = \text{Obama} \,|\, \text{prediction} = \text{Obama}),$$

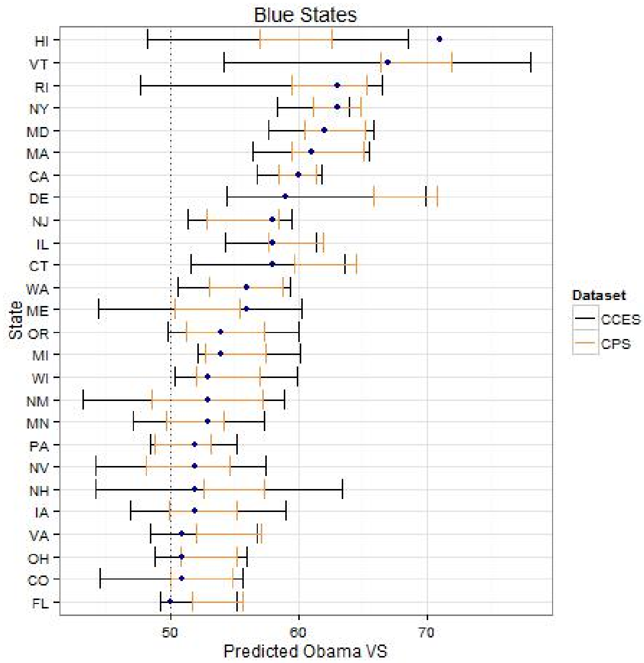
$$\mathbb{P}(\text{Pref} = \text{Romney} \,|\, \text{prediction} = \text{Obama}),$$

and

$$\mathbb{P}(\text{Pref} = \text{Other} \,|\, \text{prediction} = \text{Obama}),$$

and the analagous probabilities when the tree predicts "Romney" or "Other."[12] I average over these probabilities to generate an estimate of the multinomial distributions $\{p_j^O, p_j^R, p_j^{ot}\}$ for $j \in \{\text{Obama, Romney, Other}\}$, where the superscript indicates actual preference and the subscript indicates the Random Forest's predicted classification.[13] Table 5.3 lists the estimated conditional distributions for each state.

I can use this information to make individual predictions about preference and aggregate information about vote share. For each individual for whom the model predicts candidate j, I generate 1000 realizations of her actual preference from the empirically-generated multinomial distribution $\{\hat{p}_j^O, \hat{p}_j^R, \hat{p}_j^{ot}\}$. Then, for each of these 1000 iterations, I take the weighted average across all individuals in the state's sample to simulate that state's vote share for candidate j, using the given survey weights.[14] I do this for each state, using the electorates identified under each online registration condition (Section 2).



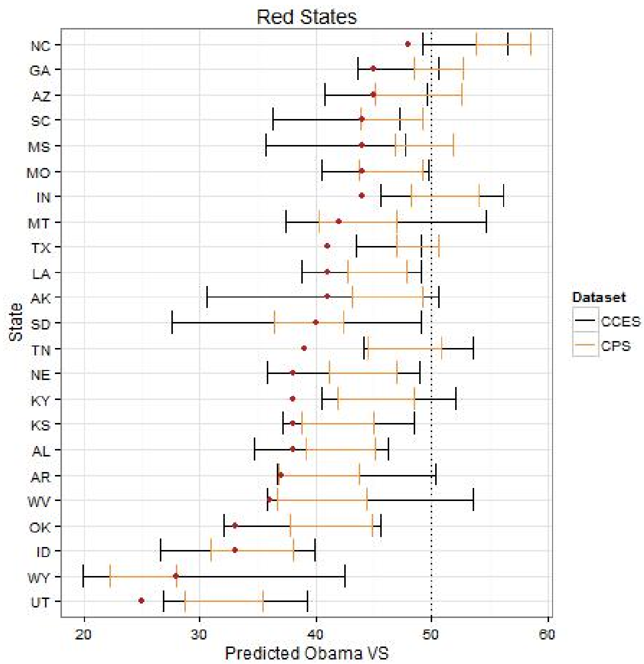


Figure 5.2: Simulated and Actual Vote Shares in the 2012 Presidential Elections

5.6 Validity

Because I need to consider the accuracy of my out-of-sample predictions, I discuss the validity of my Random Forest model. To gauge validity, I contrast the aggregate results for both the *CCES* and *CPS* with actual election results, taken from the a report issued by the Federal Elections Commission, available at `http://www.fec.gov/pubrec/fe2012/federalelections2012.pdf`. Figure 5.2 graphs the distributions of predicted Obama vote share against the true results, and Table 5.4 delineates similar information. In most cases, the model over-predicts the vote share for Obama, the winner of the 2012 election.[15] Graphically, this is represented by the point (indicating the actual election result) lying outside the 95% confidence interval for simulated vote share. If the point lies to the right of the interval(s), the simulations underestimate Obama vote share and are biased in favor of Romney; if the point lies to the left, the converse is true.

The vertical line in the graph illustrates the 50% threshold necessary for Obama to secure victory (though a vote share less than 50% can still result in victory if the total vote share for third-party candidates is sizable enough). If the distribution intersects this line, the simulations suggest some uncertainty about the electoral outcome. For the *CCES*, these states include Vermont, Maine, Oregon, New Mexico, Minnesota, Pennsylvania, Nevada, New Hampshire, Iowa, Virginia, Ohio, Colorado, Florida, North Carolina, Georgia, Indiana, Missouri, Montana, Alaska, Tennessee, Kentucky, Arkansas, West Virginia, and (surprisingly) Hawaii, the state with the greatest electoral support for Obama.[16] For the *CPS*, these states are New Mexico, Minnesota, Pennsylvania, Nevada, Iowa, Colorado, Georgia, Arizona, Mississippi, In-

[12] These probabilities are not always fully identified, given that in some bootstrap samples in some states, the ensemble never predicts "Other."

[13] Future work should incorporate additional uncertainty into the multinomial distribution using the empirical CDF.

[14] It is important to note that the weights correspond to the residential population of each state, rather than the population of voters, so there is likely some measurement error.

[15] This could be the result of psychological incentives to support the winner (see Atkeson 1999), unrepresentative sampling or weighting by the survey instrument, biases in the Random Forest model, or any combination of these factors.

[16] This could be due to the smaller sample of Hawaiians.

diana, Texas, Tennessee. The narrower *CPS* regions reflect the larger sample size of this survey.

Let us briefly consider the state-level biases in greater detail. One could also raise a concern that the model overfits *CCES* data, despite the literature that claims that Random Forest circumvents this issue (Breiman 2001; Hastie, Tibshirani, and Friedman 2011; Kuhn and Johnson 2014). However, discrepancies in the *CCES* and the *CPS* results could instead indicate different sampling biases in these surveys. Although both are large and nationally-representative, selection into the survey sample follows a different mechanism, so bias could arise. And again, the confidence intervals for the *CCES* are typically far wider (probably due to the smaller sample size), however, so the likelihood that the true vote share falls within the identified region is greater. Only two distributions favored Romney unduly: Hawaii and Maine (*CPS* only); Obama won each of these states. Meanwhile, many distributions over-predict Obama with *CPS* data: Alabama, Alaska, Connecticut, Delaware, Florida, Georgia, Indiana, Kansas, Kentucky, Louisiana, Mississippi, North Carolina, Nebraska, New Hampshire, Oklahoma, Tennessee, Texas, Utah, Vermont, and West Virginia; the distributions generated by *CCES* data favor Obama for a small subset of these states: Indiana, North Carolina, Tennessee, Texas, and Utah. Although these distributions are biased, the mean vote share still typically predicts the victor correctly. And across all states, the average deviation from the actual result is approximately 2.95% for the *CCES* estimates and 3.42% for the *CPS* estimates.

These results do pose concerns for a small subset of states. The model greatly overpredicts Obama's vote share in Kansas, Kentucky, Mississippi, Texas, and Tennnessee, though Romney won fairly handily in each of these states. Florida and Ohio warrant particular mention, given the level of competitiveness in their elections. The mean simulated vote share correctly predicts Obama for both datasets and both states, though the *CCES* distribution indicates some uncertainty about the victory, and the *CPS* slightly overestimates Obama's vote share for Florida. The most egregious concerns, however, are Indiana and North Carolina, two states that also featured very competitive elections. For North Carolina, the *CPS* distribution does not even admit the possibility of a Romney victory. The results for Indiana are somewhat less concerning, though still worth pointing out: Both surveys suggest considerable uncertainty about the victor, but lean toward Obama, while Romney succeeded in the actual election. A couple different explanations exist as to why the surveys' simulated distributions could deviate from the actual results. First, it could be that the relationship between preference and demography, or at least these particular demographic variables, exhibits instability. Second, it is possible that the surveys' sampling procedures accurately capture the overall population of these two states, but not the population of individuals who voted in the 2012 election. Regardless, it is encouraging to see that in most states, the model correctly recovers the winner, and without a strong degree of bias.

5.7 Results

Table 5.5 presents the average simulated vote share for Obama, with and without online registration. A positive number in the "Difference" column suggests that Obama would (in expectation) gain vote share if the state introduced online registration, and a negative suggests the opposite.

For all of these states, the actual margin of victory exceeds the expected increase in vote share. Averaging across blue states, Obama (in expectation) gains 0.0% of the vote; across red states, 0.12%. Returning to the original hypothesis, I speculated that politicians would be hesitant to introduce policy that diminished their probability of success in future elections, suggesting that the state would trend in the direction of the party of the Governor who signed the requisite legislation. The results are actually mixed. Of the states with Democratic Governors (at the time the legislation passed), Obama is expected to gain vote share in California, Colorado, Connecticut, Illinois, Missouri, New York, and West Virginia; and lose it in Hawaii, Maryland, Massachusetts, Minnesota, Oregon, Washington. Of the states with Republican Governors, Obama is expected to gain vote share in Georgia, Indiana, and South Carolina; and lose it in Louisiana, Utah, and Virginia. In all of these cases, it is important to remember that both the expected gains and the expected losses are quite modest.

The Democratic candidate's largest expected losses occur in Oklahoma (2012 results: 67% to 33%, Romney), Washington (56% to 41%, Obama), Alaska (55% to 41%, Romney), Arkansas (61% to 37%, Romney), and Hawaii (71% to 28%, Obama). On the other hand, his largest expected gains occur in Nevada (52% to 46%, Obama), Missouri (54% to 44%, Romney), Nebraska (60% to 38%, Romney), Indiana (54% to 44%, Romney), and North Carolina (50% to 48%, Romney). Of all of these, the only competitive contest occurred in North Carolina, and even there, the margin of victory exceeds the expected change. In the other competitive elections —Pennsylvania, Virginia, Ohio, Colorado, and Florida —the outcome would still be preserved, according to these results.

It is worthwhile to consider the most competitive elections in greater detail. Figure 5.3 plots the simulated Obama vote shares for electorates without and with the availability of online registration, focusing on the six most competitive elections of 2012 (all with a margin of victory under 5%). These distributions support the hypothesis that the partisan composition of the electorate would not change dramatically. For each state, the distributions overlap, indicating that we cannot indicate a statistically significant difference. For all but Virginia, though, the distribution shifted slightly in favor of Obama, so future research should continue to monitor these effects. Notably, the model does not do an excellent job of predicting partisanship in North Carolina, as discussed in Section 4 and evidenced by the bounds on the simulated electorates; these results, therefore, should be viewed with some degree of skepticism. In all cases, though, the estimated shift in distribution from this policy innovation does not dramatically alter the partisan preferences of the electorate.

By and large, these results support the widespread conception that online registration is a bipartisan trend (PCEA 2014; Ponoroff 2010; Underhill, Hernandez, and Hubler 2014). The distribution of Obama's simulated vote share looks nearly similar with and without the availability of online registration, and for every state. This null finding regarding partisanship may help to alleviate concerns among detractors,

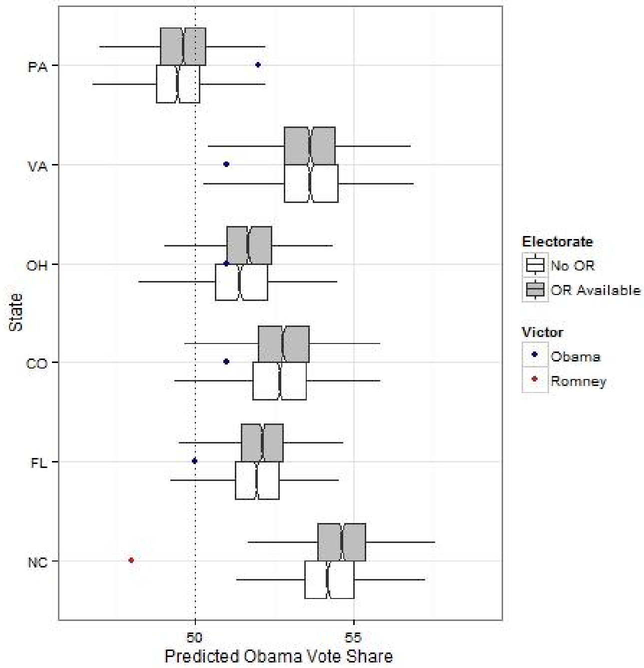


Figure 5.3: Partisanship With and Without OR

although individuals concerned about fraud may still have some hesitancy (see Underhill, Hernandez, and Hubler 2014). In Arizona and Washington, those who have registered online have evaluated this system favorably, and a supermajority of those who registered by more traditional means support this election procedure (Baretto *et al.* 2010). Given this high level of popularity among registered voters, and the financial and clerical advantages for election administration (PCEA 2014), states that can invest the money to establish a safe and secure online infrastructure may want to consider strongly this new form of convenience voting. As McDonald (2008) points out, it is far easier to pass legislation that affects election administration if it is not perceived to advantage one of the major parties over the other.

5.8 Discussion and Conclusion

My findings echo the speculations of earlier scholarship that convenience voting will not dramatically shift the electorate toward one of the major parties (Rosenstone and Wolfinger 1978; Mitchell and Wlezien 1995; Highton and Wolfinger 1998; Knack and

White 2000; Brians and Grofman 2001; Highton 2004; McDonald 2008; Gronke *et al.* 2008). Moreover, my state-level analysis reveals that the minimal gains that do exist do not uniformly benefit one party in particular. Instead, the effect on vote share depends on the particular characteristics of the members of the voting-eligible population who are most sensitive to election law, and this group's preferences vary across states.

More broadly, this paper has important implications for research on other election procedures (particular convenience voting procedures), other projects on political behavior that are constrained by the absence of ideological information in the *CPS*, and (most generally) observational studies that require information from multiple surveys. For the convenience voting literature, my proposed method is immediately applicable to any project that can identify counterfactual electorates. Because full identification of a causal effect requires at least one pre-treatment wave for every state that allows a particular policy, this might require an extensive dataset for some of the older policies (such as EDR). But if the researcher is willing to admit some uncertainty and use a partially-identified model, it is still readily possible to extrapolate the partisan implications of EDR. This advancement can greatly benefit the convenience voting literature, which has to date been mostly limited to questions about participation and conjectures about partisan implications (Neiheisel and Burden 2012).

A natural next line of inquiry for research in this area exists in the contemporary debate over voter identification laws. Although these identification policies are widely discussed in the popular press, the scientific literature is relatively quiet about how these laws affect turnout (Alvarez, Bailey, and Katz 2007, 2011), and to my knowledge, even less about how they affect election results. This is an important question for American elections, and certainly warrants further inquiry. In a recent blog article about this topic, the author laments that none of the scientific studies "sought to measure how a decline in turnout [due to voter ID laws] could effect the Democratic and Republican candidates in particular, rather than the overall figure" (Silver 2012). To remedy this lacuna, he used some of the techniques employed in the convenience voting literature: He looked at the legislators who introduced these policies, changes in aggregate vote share, and demographic characteristics of the groups most likely to be affected, which are self-admittedly “crude.” I believe that an individual-level model would provide a superior approach, and the method that I advance in this chapter for binary convenience voting treatments would only need slight modification. The estimation of vote choice with the Random Forest approach and *CCES* data can be combined with the Bayesian ordinal treatment model advanced in Alvarez, Bailey, and Katz (2011) —rather than the Wolfinger-Rosenstone binary response model —to simulate election results given different forms of voter identification policy. This question is ripe for scientific inquiry, and the method that I advance in this paper is readily applicable.

The Random Forest model of candidate preference that I generate is immediately applicable to any *CPS* dataset, or other survey dataset that asks respondents about their age, gender, race, education, and employment. Because the Random Forest does not overfit, it is particularly useful for out-of-sample predictions (Breiman 2001). The same procedure can impute an array of missing data that might be of interest to researchers who bemoan the lack of political information in the *CPS*; the *CCES* also asks respondents about their ideology, religious backgrounds, attentiveness to the news, and policy preferences, all of which relate to demography and might be valuable inputs in a model of turnout. The Random Forest generates K classification trees, and in so doing, naturally creates a measure that admits some uncertainty. I use the empirical output of the Random Forest model to create a probability distribution for a categorical variable, but researchers have some flexibility about how to utilize the trees' information. Another alternative could be a measure that is based on some normalized scale of the potential outcomes. A discussion of all the ways a researcher can extract information from the trees lies far beyond the scope of this investigation, but it is worthwhile to note that the measurement approach that I employ in this paper is only one of an entire menu of options.

And finally, in the broadest sense, the method that I advance is viable for any project that needs to utilize information from multiple surveys, as long as the surveys contain a common set of questions that predict the missing data with some accuracy.

The Random Forest model can inform the researcher about data missing from a particular survey by drawing this information from another source, and it avoids the somewhat tenuous and controversial assumptions embedded within any parametric approach. This procedure broadens the scope of analysis by allowing a researcher to leverage information across datasets. Applied social science relies on data, and researchers often have to adjust their research questions to accommodate the data available. Many projects use observational data from pre-existing surveys, due to funding limitations. By offering a procedure that liberates the researcher from being limited to one particular dataset, the method that I propose broadens the set of questions that researchers can address.

Table 5.3: Conditional Multinomial Distributions

	Random Forest Prediction								
	Obama			**Romney**			**Other**		
State	$\mathbb{P}(Obama)$	$\mathbb{P}(Romney)$	$\mathbb{P}(Other)$	$\mathbb{P}(Obama)$	$\mathbb{P}(Romney)$	$\mathbb{P}(Other)$	$\mathbb{P}(Obama)$	$\mathbb{P}(Romney)$	$\mathbb{P}(Other)$
AL	92	6	2	24	72	4	0	0	100
AK	80	20	0	20	77	3	0	0	100
AZ	73	23	4	34	63	3	0	0	100
AR	82	14	3	24	69	7	0	0	100
CA	65	30	5	35	61	4	NA	NA	NA
CO	70	28	2	31	65	4	0	0	100
CT	72	23	5	29	70	1	8	23	69
DE	80	19	1	25	71	4	0	0	100
FL	69	27	4	38	59	3	NA	NA	NA
GA	88	8	4	24	73	3	NA	NA	NA
HI	79	17	4	18	82	0	0	0	100
ID	80	20	0	28	66	6	0	0	100
IL	71	26	3	37	58	5	NA	NA	NA
IN	65	32	4	27	67	5	0	0	100
IA	64	32	4	28	67	5	0	0	100
KS	69	26	5	30	65	5	0	0	100
KY	69	28	3	34	61	5	0	0	100
LA	94	4	3	19	77	4	NA	NA	NA
ME	66	28	6	26	73	2	NA	NA	NA
MD	77	18	5	34	64	2	0	0	100
MA	67	29	4	32	63	5	NA	NA	NA
MI	65	30	5	34	62	4	0	0	100
MN	68	29	4	36	59	5	NA	NA	NA
MS	92	5	3	16	80	4	0	0	100
MO	73	22	5	35	61	4	0	0	100
MT	70	28	3	29	68	4	0	0	100
NE	72	22	5	32	66	2	0	0	100
NV	76	24	1	33	64	3	0	18	82
NH	65	31	5	31	67	2	0	0	100
NJ	60	38	2	51	47	2	NA	NA	NA
NM	65	28	7	29	67	4	0	0	100
NY	71	25	4	37	59	4	0	32	68
NC	80	18	3	31	65	4	0	0	100
ND	72	21	6	15	74	10	0	0	100
OH	67	30	4	34	62	4	0	0	100
OK	81	16	3	31	67	2	NA	NA	NA
OR	64	32	4	28	68	3	0	0	100
PA	67	29	4	35	60	6	0	0	100
RI	68	28	4	26	74	0	0	0	100
SC	83	12	5	24	71	5	0	0	100
SD	74	26	0	27	70	3	NA	NA	NA
TN	81	17	3	32	64	5	NA	NA	NA
TX	76	20	4	29	65	5	0	0	100
UT	81	14	5	28	66	6	0	0	100
VT	74	23	3	20	80	0	NA	NA	NA
VA	69	29	3	33	62	5	0	20	80
WA	64	32	4	30	65	5	0	0	100
WV	65	29	6	25	72	3	NA	NA	NA
WI	63	35	2	32	65	3	NA	NA	NA
WY	78	11	11	20	71	9	0	0	100

For each individual i in state j, the Random Forest model predicts how that individual voted; I identify the true probability that the individual voted for Obama, Romney, or another candidate conditioned on that prediction. NA indicates that the Random Forest never predicts a particular candidate.

Table 5.4: Contrasting Simulated and Official Obama VS

State	*CCES* Mean	*CCES* Bounds	*CPS* Mean	*CPS* Bounds	*2012* Official Results
AK	41	[31, 51]	46	[43, 49]	41
AL	40	[35, 46]	42	[39, 45]	38
AR	43	[37, 50]	40	[37, 44]	37
AZ	45	[41, 50]	49	[45, 53]	45
CA	59	[57, 62]	60	[58, 61]	60
CO	50	[45, 56]	53	[50, 55]	51
CT	58	[52, 64]	62	[60, 65]	58
DE	63	[54, 70]	68	[66, 71]	59
FL	52	[49, 55]	54	[52, 56]	50
GA	47	[44, 51]	51	[48, 53]	45
HI	58	[48, 69]	60	[57, 63]	71
IA	53	[47, 59]	53	[50, 55]	52
ID	33	[27, 40]	34	[31, 38]	33
IL	58	[54, 61]	60	[58, 62]	58
IN	51	[46, 56]	51	[48, 54]	44
KS	43	[37, 48]	42	[39, 45]	38
KY	46	[41, 52]	45	[42, 48]	38
LA	44	[39, 49]	45	[43, 48]	41
MA	61	[56, 66]	62	[59, 65]	61
MD	62	[58, 66]	63	[60, 65]	62
ME	52	[44, 60]	53	[50, 55]	56
MI	56	[52, 60]	55	[53, 57]	54
MN	53	[47, 57]	52	[50, 54]	53
MO	45	[41, 50]	47	[44, 49]	44
MS	42	[36, 48]	49	[47, 52]	44
MT	46	[37, 55]	44	[40, 47]	42
NC	53	[49, 56]	56	[54, 58]	48
NE	42	[36, 49]	44	[41, 47]	38
NH	54	[44, 63]	55	[53, 57]	52
NJ	56	[51, 60]	56	[53, 58]	58
NM	51	[43, 59]	53	[49, 57]	53
NV	51	[44, 57]	51	[48, 55]	52
NY	61	[58, 64]	63	[61, 65]	63
OH	53	[49, 56]	53	[51, 55]	51
OK	39	[32, 46]	41	[38, 45]	33
OR	55	[50, 60]	54	[51, 57]	54
PA	52	[48, 55]	51	[49, 53]	52
RI	58	[48, 67]	62	[59, 65]	63
SC	42	[36, 47]	46	[44, 49]	44
SD	37	[28, 49]	39	[36, 42]	40
TN	49	[44, 54]	48	[44, 51]	39
TX	46	[43, 49]	49	[47, 51]	41
UT	33	[27, 39]	32	[29, 35]	25
VA	53	[48, 57]	55	[52, 57]	51
VT	67	[54, 78]	69	[66, 72]	67
WA	55	[51, 59]	56	[53, 59]	56
WI	55	[50, 60]	54	[52, 57]	53
WV	45	[36, 54]	40	[37, 44]	36
WY	30	[20, 42]	25	[22, 28]	28

This table describes the distributions of simulated Obama vote share generated by fitting the Random Forest model to the weighted *CCES* and *CPS* datasets. The partisanship model is discussed in detail in Section 3 of this chapter.

Table 5.5: Average Obama Vote Share Without and With Online Registration

Blue States				*Red States*			
State	**No OR**	**OR**	**Difference**	**State**	**No OR**	**OR**	**Difference**
NV	49.53	50.32	0.79	MO	45.02	45.56	0.54
IA	52.90	53.19	0.29	NE	42.77	43.28	0.51
OH	51.42	51.69	0.27	IN	49.40	49.81	0.41
MI	54.60	54.84	0.24	NC	54.25	54.63	0.38
IL	58.65	58.88	0.23	MS	45.22	45.58	0.36
DE	68.62	68.85	0.23	TN	43.08	43.44	0.36
CA	58.91	59.10	0.19	TX	45.49	45.85	0.36
FL	51.94	52.12	0.18	KY	43.82	44.14	0.32
PA	49.45	49.63	0.18	AZ	45.90	46.20	0.30
ME	52.53	52.69	0.16	GA	48.57	48.84	0.27
RI	61.96	62.10	0.14	WV	38.78	39.04	0.26
OR	53.73	53.86	0.13	KS	42.15	42.28	0.13
CO	52.66	52.78	0.12	UT	30.02	30.10	0.08
NJ	55.40	55.48	0.08	SC	44.95	45.01	0.06
WI	54.04	54.10	0.06	SD	38.92	38.94	0.02
MA	61.76	61.78	0.02	ID	33.46	33.43	-0.03
NY	61.90	61.92	0.02	LA	43.50	43.47	-0.03
MN	51.73	51.74	0.01	WY	25.37	25.31	-0.06
MD	62.42	62.41	-0.01	AL	41.30	41.18	-0.12
VA	53.62	53.60	-0.02	MT	44.80	44.68	-0.12
VT	69.67	69.64	-0.03	AR	39.59	39.40	-0.19
NH	54.30	54.25	-0.05	AK	44.93	44.56	-0.37
NM	52.17	52.12	-0.05	OK	41.24	40.67	-0.57
CT	61.32	61.22	-0.10				
HI	57.49	57.32	-0.17				
WA	56.31	55.95	-0.36				

This table presents the average Obama vote share across simulations of turnout with and without online registration. Identification of the electorate in each scenario is described in Section 2 of this chapter, and the partisanship model is discussed in Section 3. The final column presents the difference between the average vote share in the two scenarios.

Chapter 6

Conclusion

Together, these chapters address questions that remain unresolved in the convenience voting literature. Although this work offers further insight, additional scholarship should continue to explore these issues.

Chapter 2 surveys a lingering methodological question of how to design research to identify the relationships between policy and behavior. Although several prominent papers have challenged the dominant statistical approach (particularly Keele and Minozzi 2013 and Glynn and Quinn 2011), the current body of research has yet to offer a convincing demonstration of its inappropriateness. Both of these papers execute parametric regressions modeled after those of Rosenstone and Wolfinger (1978), and reject the estimands as unreasonably large in magnitude. Yet both of these papers specify regressions that neglect important convariates and utilize only cross-sectional data, which has already been shown to bias results (Ansolabehere and Konisky 2006; Leighley and Nagler 2014). The alternative methods that scholars have proposed, such as natural experiments and non-parametric bounds, offer important supplements to the binary response model. I conclude that the literature should take a holistic approach and examine these questions through all of these lenses. Future work should continue to examine the sensitivity of results to these designs and any alternative methods that scholars discover.

Chapter 3 considers the impact of election policy, but narrows the focus to a particular subset of the electorate that has been underrepresented both at the polls and in the literature. Individuals with disabilities vote at consistently low rates, but it is difficult to identify the precise relationship between disability status and participation because of the mediating influence of confounders. To accommodate the imbalance across observable across disabled and non-disabled populations, I execute three different kinds of matching: propensity score, Mahalanobis distance, and genetic. Matching typically reduces the magnitude of the affect of disability on participation, though the direction and significance do not change. After accounting for interactive effects, the impact of disability appears to be realized in relation to employment status, and future research should continue to explore this characterization. Notably, the most popular forms of convenience voting do not appear to remedy the disproportionate representation, though they might have some small mobilizing affect. Additional scholarship should continue to examine how individuals with disabilities participate in elections, and how they respond to new technologies (Stewart 2011). The research design should incorporate matching, or address possible confounders in some other way. As the *CPS* offers additional cross sections of data, it will be interesting to explore the stability of the point estimates in this work.

Chapter 4 returns to a more general focus in terms of population, but a narrower focus on policy. The peer-reviewed literature has yet to describe how online voter registration impacts the electorate, and this silence is particularly glaring given its rising popularity. My work offers insight into how this new form of convenience voting is influencing registration and turnout behavior. Online voter registration is associated with a small but statistically significant increase in participation at both stages, and this finding is robust to many alternate specifications. Moreover, there is evidence that its impact may become more pronounced over time, though its brief history means that this relationship may be driven by one or two states that introduced it first. One can use my results to benchmark how this particular form of convenience voting compares to others, but it is important to interpret these estimates with care. The dataset does not contain a full set of pre-treatment waves for the other policies, so the causal impact is not fully identified. Nonetheless, my results suggest that online registration may be one of the more successful electoral innovations, though a substantial portion of the voting-eligible population will still abstain. As more states adopt this policy and additional electoral cross sections become available, it will be worthwhile to investigate the impact more broadly. Furthermore, future work should consider whether a binary treatment is the appropriate statistical strategy; a heterogeneous treatment might be superior, or some other categorical representation.

And finally, Chapter 5 offers the most novel approach to scholarship on convenience voting. Most studies have shied away from formally investigating the impact of policy on the electorate's partisan composition, due to data limitations of the *CPS*. To circumvent this lack of data, I turn to the *CCES*, a large, nationally-representative survey that asks about both demography and vote choice. I model vote choice as a function of age, gender, race, and education using a Random Forest (Breiman 2001), and I correct the model's prediction by estimating the distribution of true vote choice conditioned on model prediction.

Bibliography

[1] Abadie, Alberton, and Guido Imbens. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica*. Vol. 74, No. 1 (Jan. 2006): pp. 235-67.

[2] Achen, Christopher. "Registration and Voting Under Rational Expectations: The Econometric Implications." Working Paper, 2008.

[3] Acohido, Byron. "Online Voter Registration Helps Bulk Up Voter Rolls." *USA Today.* 14 Oct. 2012 `<http://www.usatoday.com/story/tech/personal/2012/10/14/online-voter-registration-catches-on/1625321/>`.

[4] Ai, Chunrong, and Edward C. Norton. "Interaction Terms in Logit and Probit Models." *Economics Letters*. Vol. 80, No. 1 (Jul. 2003): pp. 123-9.

[5] Alvarez, R. Michael, Delia Bailey, and Jonathan N. Katz. "The Effect of Voter Identification Laws on Turnout." Worling Paper 1267R, California Institute of Technology and Washington University in St. Louis. Latest draft 2007.

[6] Alvarez, R. Michael, Delia Bailey, and Jonathan N. Katz. "An Empirical Bays Approach to Estimating Ordinal Treatment Effect." *Political Analysis*. Vol. 19, No. 1 (Winter 2011): pp. 20-31.

[7] Alvarez, R. Michael, and Thad E. Hall. "Defining the Barriers to Political Participation for Individuals with Disabilities." The Information Technology and Innovation Foundation. Accessible Voting Technology Initiative, Working Paper Series. Working Paper 001. Latest draft 2012.

[8] —–. *Electronic Elections: The Perils and Promises of Digital Democracy.* Princeton, NJ: Princeton University Press, 2008.

[9] —–. "Resolving Voter Registration Problems: Making Registration Easier, Less Costly and More Accurate." *Electronic Administration in the United States: The State of Reform After* Bush v. Gore. Ed. by R. Michael Alvarez and Bernard Grofman. New York: Cambridge University Press, 2014.

[10] Alvarez, R. Michael, Ines Levin, and J. Andrew Sinclair. "Making Voting Easier: Convenience Voting in the 2008 Presidential Election." *Political Research Quarterly.* Vol. 65, No. 2 (Jun. 2012): pp. 248 - 262.

[11] Alvarez, R. Michael, and Jonathan Nagler. "Election Day Voter Registration in California." *Demos.* Policy Brief (Spring 2011).

[12] —–. "Election Day Voter Registration in Massachusetts." *Demos.* Policy Brief (Jan. 2008).

[13] —–. "Same Day Voter Registration in North Carolina." *Demos.* Policy Brief (Spring 2007).

[14] Ansolabehere, Stephen. *Cooperative Congressional Election Study 2012: Common Content.* [Computer File] Release 1: April 15, 2013. Cambridge, MA: Harvard University [producer] `http://cces.gov.harvard.edu`.

[15] Ansolabehere, Stephen, and Eitan Hersch. "Voter Registration: The Process and Quality of Lists." *The Measure of American Elections.* Ed. by Barry C. Burden and Charles Stewart, III. New York: Cambridge University Press, 2014.

[16] —–. "What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis.* Vol. 20, No. 4 (Autumn 2012): pp. 437-59.

[17] Ansolabehere, Stephen, and David M. Konisky. "The Introduction of Voter Registration and Its Effect on Turnout." *Political Analysis.* Vol. 14, No. 1 (Winter 2006): 83 - 100.

[18] Atkeson, Lonna Rae. "'Sure, I voted for the Winner!' Over Report of the Primary Vote for the Party Nominee in the American National Election Studies." *Political Behavior* Vol. 21, No.3 (1999): pp.197-215.

[19] Baker, Paul M.A., Robert G.B. Roy, and Nathan W. Moon. (Nov. 2005). "Getting Out the Vote: Assessing the Technological, Social, and Process Barriers to (e)Voting for People With Disabilities." Paper presented at the APPAM Research Conference.

[20] Baretto, Matt A., Bonnie Glaser, Karin Mac Donald, Loren Collingwood, Francisco Pedraza, and Barry Pump. "Online Voter Registration (OLVR) Systems in Arizona and Washington: Evaluating Usage, Public Confidence and Implementation Processes." Washington Institute of the Study of Ethnicity and Race and Election Administration Research Center. University of Washington and University of California, Berkeley. 1 Apr. 2010.

[21] Bennion, Elizabeth A., and David W. Nickerson. "The Cost of Convenience: An Experiment Showing Email Outreach Decreases Voter Registration." *Political Research Quarterly.* Vol. 64, No. 4 (Dec. 2011): pp. 858-69.

[22] Berinsky, Adam J. "The Perverse Consequences of Electoral Reform in the United States." *American Politics Research.* Vol. 33, No. 4 (Jul. 2005): pp. 471-91.

[23] Berinsky, Adam J., Nancy Burns, and Michael W. Traugott. "Who Votes by Mail? A Dynamic Model of the Individual-Level Consequences of Voting-by-Mail Systems." *Public Opinion Quarterly.* Vol. 65, No. 2 (Jun. 2001): pp. 178-97.

[24] Bovbjerg, Barbara. "Voters with Disabilities: Challenges to Voting Accessibility." Statement Before the National Council on Disability. United States Government Accountability Office. Delivered 23 Apr., 2013.

[25] Breiman, Leo. "Random Forests." *Machine Learning.* Vol. 45, No. 1 (Oct. 2001): pp. 5-32.

[26] Breiman, Leo, and Adele Cutler. "Random Forests." `https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm`.

[27] Brians, Craig Leonard, and Bernard Grofman. "Election Day Registration's Effect on U.S. Voter Turnout." *Social Science Quarterly.* Vol. 82, No. 1 (Mar. 2001): pp. 170 - 183.

[28] Burden, Barry C., David T. Canon, Kenneth R. Mayer, and Donald P. Moynihan. "Election Law, Registration, and Turnout: The Unanticipated Consequences of Electoral Reform." *American Journal of Political Science.* Vol. 58, No. 1 (Jan. 2014): pp. 95-109.

[29] "Campaign Ads." *America Votes 2004.* Cable News Network LP, LLLP. 2005 `http://www.cnn.com/ELECTION/2004/special/president/campaign.ads/`.

[30] Cemenska, Nathan, Jan E. Leighley, Jonathan Nagler, and Daniel P. Tokaji. "Report on the 1972-2008 Early and Absentee Voting Dataset.' The Pew Charitable Trusts. *Non-Precinct Voting in the States: An Extensive Dataset of State Laws and Related Resources.* 14 Dec. 2009.

[31] Census Bureau. "The Diversifying Electorate - Voting Rates by Race and Hispanic Origin in 2012 (and Other Recent Elections." *Current Population Survey.* May 2013.

[32] Dehejia, Rajeev H., and Sadek Wahba. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *American Statistical Association.* Vol. 94, No. 448 (1999): pp. 1053-62.

[33] Diamond, Alexis, and Jasjeet S. Sekhon. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics.* Vol. 95, No. 3 (Jul. 2013): pp. 932-45.

[34] Downs, Anthony. *An Economic Theory of Democracy.* New York: Harper and Row, 1957.

[35] Erikson, Robert S., Gerald C. Wright, and John P. McIver. *Statehouse Democracy: Public Opinion and Policy in the American States.* Cambridge, NY: Cambridge University Press, 1993.

[36] Federal Elections Commission. "Federal Elections 2012: Election Results for the U.S. President, the U.S. Senate and the U.S. House of Representatives." Jul. 2013 `http://www.fec.gov/pubrec/fe2012/federalelections2012.pdf`.

[37] García Bedolla, Lisa, and Veronica N. Veléz. "Differences Among Latina/o, Asian American, and White Online Registrants in California." Center for Latino Policy Research. University of California, Berkeley. Mar. 2013.

[38] Gelman, Andrew, and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge, UK: Cambridge University Press, 2007.

[39] Glynn, Adam N., and Kevin M. Quinn. "Why Process Matters for Causal Inference." *Political Analysis.* Vol. 19, No. 3 (Jul. 2011): pp. 273 - 286.

[40] —–. 2011. "Replication data for: Why Process Matters for Causal Inference," `http://hdl.handle.net/1902.1/15920`. *IQSS Dataverse Network* [Distributor] V1 [Version].

[41] Grimmer, Justin, and Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis.* Vol. 21, No. 3 (Summer 2013): pp. 267-97.

[42] Gronke, Paul, Eva Galanes-Rosenbaum, Peter A. Miller, and Daniel Toffey. "Convenience Voting." *Annual Review of Political Science.* Vol. 11 (Nov. 2008): pp. 437-55.

[43] Hanmer, Michael J. "An Alternative Approach to Estimating Who Is Most Likely to Respond to Changes in Registration Laws." *Political Behavior.* Vol. 29, No. 1 (Mar. 2007): pp. 1 - 30.

[44] —–. *Discount Voting: Voter Registration Reforms and Their Effects.* Cambridge, NY: Cambridge University Press, 2009.

[45] Hanmer, Michael J., and Kerem Ozan Kalkan. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variables." *American Journal of Political Science.* Vol. 57, No. 1 (Jan. 2013): pp. 263-77.

[46] Harrington, James C. "Pencils Within Reach and a Walkman or Two: Making the Secret Ballot Available to Voters Who Are Blind or Have Other Physical Disabilities." *Texas Journal on Civil Liberties and Civil Rights.*

[47] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. Springer, NY: Springer Science and Business Media, LLC, 2009.

[48] Highton, Benjamin. "Easy Registration and Voter Turnout." *Journal of Politics.* Vol. 59, No. 2 (May 1997): pp. 565 - 575.

[49] —–. "Voter Registration and Turnout in the United States." *Perspectives on Politics.* Vol. 2, No. 3 (Sept. 2004): pp. 507-14.

[50] Highton, Benjamin, and Raymond E. Wolfinger. "Estimating the Effects of the National Voter Registration Act of 1993." *Political Behavior.* Vol. 20, No. 2 (Jun. 1998): pp. 79 - 104.

[51] Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis.* Vol. 15, No. 3 (Summer 2007): pp. 199-236.

[52] Hur, Aram, and Christopher H. Achen. "Coding Voter Turnout Responses in the Current Population Survey." *Public Opinion Quarterly.* Vol. 77, No. 4 (Winter 2013): pp. 985 - 993.

[53] ICPSR 25643. *Current Population Survey, November 2008: Voting and Registration Supplement* Codebook. Bureau of the Census. Bureau of Labor Statistics. Inter-university Consortium for Political and Social Research. Anne Arbor, MI.

[54] ICPSR 31082. *Current Population Survey, November 2010: Voting and Registration Supplement* Codebook. Bureau of the Census. Bureau of Labor Statistics. Inter-university Consortium for Political and Social Research. Anne Arbor, MI.

[55] Kastellec, Jonathan P., Jeffrey R. Lax, and Justin Phillips. "Estimating State Public Opinion With Multi-Level Regression and Poststratification using R." Working Paper. Latest Draft: 8 Jan. 2014 `http://www.princeton.edu/~jkastell/MRP_primer/mrp_primer.pdf`.

[56] Katz, Jonathan, and Gabriel Katz. "Correcting for Survey Reports Using Auxiliary Information with an Application to Estimating Turnout." *American Journal of Political Science.* Vol. 54, No. 3 (Jul. 2010): pp. 815-35.

[57] Keele, Luke, and William Minozzi. "How Much Is Minnesota Like Wisconsin? Assumptions and Counterfactuals in Causal Inference with Observational Data." *Political Analysis.* Vol. 21, No. 2 (Spring 2013): pp. 193 - 216.

[58] King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review.* Vol. 25, No.1 (Mar. 2001): pp. 49-69.

[59] Knack, Stephen, and James White. "Election-Day Registration and Turnout Inequality." *Political Behavior.* Vol. 22, No. 1 (Mar. 2000): pp. 29 - 44.

[60] Kuhn, M., and K. Johnson. *Applied Predictive Modeling.* New York, NY: Spring Science, 2013.

[61] Leighley, Jan E., and Jonathan Nagler. "Individual and Systematic Influences on Turnout: Who Votes?" *Journal of Politics.* Vol. 54, No. 3 (Aug. 1992): pp. 718 - 740.

[62] ——. *Who Votes Now? Demographics, Issues, Inequality, and Turnout in the United States.* Princeton, NJ: Princeton University Press, 2014.

[63] Liaw, Andy, and Matthew Wiener. "Classification and Regression by randomForest." *R News.* Vol. 2, No. 3 (Dec. 2002): pp. 18-22.

[64] Manski, Charles F. *Identification Problems in the Social Sciences.* Cambridge: Harvard University Press, 2005.

[65] McDonald, Michael P. "Portable Voter Registration." *Political Behavior.* Vol. 30, No. 4 (Dec. 2008): pp. 491 - 501.

[66] —–. United States Election Project. George Mason University. `http://elections.gmu.edu/Turnout_2008G.html` and `http://elections.gmu.edu/Turnout_2012G.html`.

[67] Merl, Jean. "Online Registration Boosts Voter Rolls Sharply, Officials Say." *L.A. Times.* 3 Oct. 2012 <http://latimesblogs.latimes.com/california-politics/2012/10/online-registration-boosts-voter-rolls-sharply-area-officials-say.html>.

[68] Mitchell, Glenn E., and Christopher Wlezien. "The Impact of Legal Constraints on Voter Registration, Turnout, and the Composition of the American Electorate." *Political Behavior.* Vol. 17, No. 2 (Jun. 1995): pp. 179 - 202.

[69] Morgan, Stephen L., and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* 2nd ed. Cambridge: Cambridge University Press, 2014.

[70] Nagler, Jonathan. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science.* Vol. 38, No. 1 (Feb. 1994): pp. 230 - 255.

[71] "Online Voter Registration." National Conference of State Legislatures. Nov. 2013 `https://www.supportthevoter.gov/files/2013/12/Online-voter-registration.pdf`.

[72] "Online Voter Registration." National Conference of State Legislatures. 10 Dec. 2014 `http://www.ncsl.org/research/elections-and-campaigns/electronic-or-online-voter-registration.aspx`.

[73] *National Council of State Legislatures.* `http:\\www.ncsl.org.`

[74] Neiheisel, Jacob R., and Barry C. Burden. "The Impact of Election Day Registration on Voter Turnout and Election Outcomes." *American Politics Research.* Vol. 40, No. 4 (Jul. 2012): pp. 636 - 664.

[75] Park, David K., Andrew Gelman, and Joseph Bafumi. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis.* Vol. 12, No. 4 (Nov. 2004): pp. 375-85.

[76] Pellissier, Allyson L. "In Line or Online? American Voter Registration in the Digital Era." *Working Paper.* Prepared for the 2015 *Midwest Political Science Association* Conference. Latest Draft: 18 Mar. 2015.

[77] ——. “Erratum: Glynn and Quinn.” Forthcoming, *Political Analysis.*

[78] Perlroth, Nicole. “Voter Registration Rolls in 2 States Are Called Vulnerable to Hackers.” *New York Times.* 21 Oct. 2012 <http://www.nytimes.com/2012/10/13/us/politics/cracks-in-maryland-and-washington-voter-databases.html?_r=0>.

[79] Pew Center on the States. “Online Voter Registration Now Offered in 13 States —Driving Thousands of New Registrants and Updates.” Pew Charitable Trusts. 11 Oct. 2012 http://www.pewtrusts.org/en/about/news-room/press-releases/2012/10/11/online-voter-registration-now-offered-in-13-statesdriving-thousands-of-new-registrants-an

[80] Ponoroff, Christopher. “Voter Registration in a Digital Age.” Ed. by Wendy Weisman. *Brennan Center for Justice.* NYU School of Law. 2010 https://www.brennancenter.org/sites/default/files/legacy/Democracy/Paperless_Registration_FINAL.pdf.

[81] Presidential Commission on Election Administration. *The American Voting Experience: Report and Recommendations of the Presidential Commission on Election Administration.* Jan. 2014 https://www.supportthevoter.gov/files/2014/01/Amer-Voting-Exper-final-draft-01-09-14-508.pdf.

[82] Riker, William H., and Peter C. Ordeshook. “A Theory of the Calculus of Voting.” *American Political Science Review.* Vol. 62, No. 1 (Mar. 1968): pp. 25-42.

[83] Rosenbaum, Paul R., and Donald B. Rubin. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika.* Vol. 70, No. 1 (Apr. 1983): pp. 41-55.

[84] Rosenstone, Steven J., and Raymond E. Wolfinger. “The Effect of Registration Laws on Voter Turnout.” *American Political Science Review.* Vol. 72, No. 1 (Mar. 1978): pp. 22 - 45.

[85] Rubin, Donald B. “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology.* Vol. 56, No. 5 (1974): pp. 688-701.

[86] ——. “Matching to Remove Bias in Observational Studies.” *Biometrics.* Vol. 29, No. 1 (Mar. 1973): pp. 159-83.

[87] Schlozman, Kay Lehman, Sidney Verba, and Henry E. Brady. *The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy.* Princeton, NJ: Princeton University Press, 2012.

[88] Schriner, Kay and Andrew I. Batavia. “The Americans with Disabilities Act: Does it Secure the Fundamental Right to Vote?” *Policy Studies Journal.* Vol. 29, No. 4 (Nov. 2001): pp. 663-73.

[89] Schriner, Kay, Lisa A. Ochs and Todd G. Shields. “The Last Suffrage Movement: Voting Rights for Persons with Cognitive and Emotional Disabilities.” *Publius.* Vol. 27, No. 3 (Summer 1997): pp. 75-96.

[90] Schur, Lisa, and Meera Adya. "Sidelined or Mainstreamed? Political Participation and Attitudes of People with Disabilities in the United States." *Social Science Quarterly.* Vol. 94, No. 3 (Sept. 2013): pp. 811-39.

[91] Schur, Lisa, and Douglas Kruse. "What Affects Voter Turnout?: Lessons from Citizens with Disabilities." *Social Science Quarterly.* Vol. 81, No. 2 (Jun. 2000): pp. 571-87.

[92] Schur, Lisa, and Douglas Kruse. "Disability, Voter Turnout, and Polling Place Accessibility." Presented to the Board of Advisors, Election Assistance Commission. 7 Jun. 2011.

[93] Schur, Lisa, Todd Shields, and Kay Schriner. "Voting." *Encyclopedia of Disability.* Ed. by Gary Albrecht. Thousand Oaks, CA: Sage Publications, 2005.

[94] Schur, Lisa, Todd Shields, Douglas Kruse, and Kay Schriner. "Enabling Democracy: Disability and Voter Turnout." *Political Research Quarterly.* Vol. 55, No. 1 (March. 2002): pp. 167-90.

[95] Sekhon, Jasjeet. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R." *Journal of Statistical Software.* Vol. 42, No. 7 (Jun. 2011): pp. 1-52.

[96] Sekhon, Jasjeet Singh, and Walter R. Mebane, Jr. 1998. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models." *Political Analysis* Vol. 7, No. 1 (1998): 187-210.

[97] Silver, Nate. "Measuring the Effects of Voter Identification Laws." *New York Times.* FiveThirtyEight. 15 Jul. 2012 `http://fivethirtyeight.blogs.nytimes.com/2012/07/15/measuring-the-effects-of-voter-identification-laws/?_r=0`

[98] Stewart, Charles III. "Voting Technologies." *Annual Review of Political Science.* Vol. 14 (2011): pp. 353–78.

[99] Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science.* Vol. 25, No. 1 (Feb. 2010): pp. 1-21.

[100] Tolbert, Caroline J., and Ramona S. Mcneal. "Unraveling the Effects of the Internet on Political Participation?" *Political Research Quarterly.* Vol. 56, No. 2 (Jun. 2003): pp. 175-185.

[101] Underhill, Wendy, Michael D. Hernandez, and Katy Owens Hubler. "Online Voter Registration Grows in 2014." *The Canvass.* National Conference of State Legislatures. Pew Charitable Trusts. Issue 48 (Apr. 2014).

[102] Wolfinger, Raymond E., and Steven J. Rosenstone. *Who Votes?* New Haven, CT: Yale University Press, 1980.