Turnout and Mail Voting in Colorado or: How I Learned to Stop Worrying and Love Voter Registration Files

---

A Thesis
Presented to
the Interdivisional Committee for Mathematics and Natural Sciences,
History and Social Sciences
(Mathematics and Political Science)
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Theodore Dounias

December 2018

Approved for the Committee
(Mathematics and Political Science)

Paul Gronke

Andrew Bray

Preface

I started my second attempt at this preface by reading through my maternal grandfather's eulogy. In my bloodline I have researchers, engineers, teachers, social scientists, and literary scholars, all with astoundingly different approaches to stemming their curiosity about the world around them. All of them, have fought long and hard for me to have what I do now: my privilege to study at Reed College and live a financially secure life, but also my ability to determine my future in a democratic state. They all fought for Greece to be as free as it is today, even though our elections may often deliver painful results. I find it very fitting that the way I honor their legacy is by studying the system that they helped create[1].

Of course, my object of study is not Greek, but American elections. I made this choice first by necessity, since the data access and previous literature is exponentially greater. I soon realized, however, that the puzzles that US elections present are both generalizable and deeply intriguing. I was taken in by the complexity of translating the will of the people into political action for the world's greatest economic and military power, and by how the most initially mundane policy choices have massive impacts on representation. As I mention again in my introduction, democracies are based on procedures as much as on principles, and I am now very happy to have completed my first pass at adding to the scientific knowledge available for one such procedure.

Here I must mention my first thesis adviser, and in fact my academic adviser and guide throughout my years at Reed, Professor Paul Gronke. Paul as a person transmits emphatic dedication and love to the subject of his research, to the extent that it is hard not to be taken along for the ride. He helped me get back my sense of direction that I had lost before transferring to Reed, he was the first to suggest I do an interdisciplinary thesis, and he stuck with me and worked incredibly hard to guide me through the process. While at times not giving me the answer I wanted to hear, he also helped ground me and calm me down when my own sense of anxiety was getting the better of me, and I can't thank him enough for putting up with that as well.

I am, of course, an interdisciplinary major. This is partly by virtue of the academic credits I gained from accidentally trying to be an engineer for two years after graduating high school, but mostly because of the people who have taught me mathematics. My mathematical education was always deeply grounded in applicability and the elegance of statistically approximating real-world phenomena. At Reed College I met my second adviser, Andrew Bray, whose calm and deeply sincere admiration for statistical

[1] And by also veering wildly away from any of their own previous research interests, much to the occasional chagrin of my close relatives, and much in the same fashion that they seem to have steered clear of their own parents.

applications is infectious. Andrew often pushed me to accomplish tasks that I felt were beyond my capacity, but despite my protests I always managed to get done with his help. Because of him I feel that I not only know how to apply the methods I use, but that I really *understand* them (or at least understand how much I *do not* understand). He has worked hard to oversee my senior thesis, sacrificing time from his sabbatical, and I cannot thank him enough.

I would like to thank Andrew Menger, Judd Choate, and Robert Stein. Andrew Menger, Postdoctoral Fellow at the Weidenbaum Center on the Economy, Government, and Public Policy at Washington University, was gracious enough to share the data that he and his team collected and are currently working on; he saved my thesis by giving me access to voter files from 2012-2016, while I only had access to the 2017 file. Without his help, I would not have been able to complete my analysis. Judd Choate, Director of Elections for Colorado was instrumental in helping me procure data for 2017, from which I started my thesis. He also helped me troubleshoot data issues by connecting me with his team. Robert Stein, Lena Goldman Fox Professor of Political Science at Rice University, also gave me guidance when I ran into data problems common in his field.

I would like to also show my appreciation to the Reed College Computer User Services that loaned me the computer I am using to run my models, for saving me what I can safely assume to be several months worth of listening to my own laptop's fans screaming for mercy while I age, considerably. Paul Manson and my fellow Reedie Jay Lee were also helpful in solving issues with presentation, mapping, tables, and the R Markdown thesis template. The template itself was coded by Chester Ismay. I thank all of them.

There are so many people at Reed and back home that have been important to me that I will almost certainly fail to mention them all; consider the following a woefully incomplete list. To GameDEV, my dorm, you eternally empathetic, thoughtful, and warm friends, with whom I lived out my Reed years in full force. There is not much to say other than that I will forever miss being part of your community, and I am proud of you all. To the Reed Mock Trial Team, who helped me procrastinate and escape campus with good company when I needed it most. You are all awesome and I am sure you will get to ORCS this year!

To Kathy Saitas, who gave me a home away from home from my first week here. To all my Professors that took interest in me and helped me along the way. To Joanne Buzaglo and her family, that taught me to be a Philly sports fan and gave me a home for fall breaks. I thank all of you for your kindness.

To my innumerable friends back home, I apologize for not having the space for all of you here. Suffice to say, if you ever played DotA with me while I was here, if we played soccer back home, or shared a beer, or sang, or discussed this year's Fantasy Premier League, then consider this an expression of gratitude for being there for me. I miss you all.

Last, but most importantly, my family: Theodoros, Eftychia, Julius, Poly, my uncle and aunt Christos and Christina, my cousin (now Dr.) Philip and my parents George and Marina. You have given me a life worth living, a home worth missing, and a world worth fighting for. *Θοδωρής Ντουνιάς*

Table of Contents

List of Tables

List of Figures

Abstract

Mail voting in the United States was conceived and first implemented to serve absentee voters during the Civil War (Fortier, 2006) and has persisted until the present day, becoming one of the key reforms associated with "convenience voting" and the expansion of the democratic franchise (Gronke, Galanes-Rosenbaum, Miller, & Toffey, 2008). In 2013, Colorado implemented the latest in a series of in-state election reforms and became the third state in the nation with universal mail voting for all elections, after Oregon and Washington. Despite claims by policymakers that mail voting should have a strong, positive effect on voter turnout, a recent series of studies on Oregon, Washington, parts of California, and Colorado have failed to show consistent results, disagreeing both on the scale and the direction (positive or negative) that this effect has. This thesis aims at following this series of studies by examining Colorado voter registration files for recent elections (2010-2016). These files consist of a registration file with voter information and a history file with voter participation data in Colorado elections, and provide all information necessary for a comprehensive study of turnout. By describing, fitting, and interpreting multilevel general additive regression models of voter turnout based on these data, I show that there is a small positive effect of mail voting on turnout in national elections at the county level. This thesis also contributes to the literature by presenting a description of modeling and data wrangling difficulties associated with voter registration files, and giving a series of potential solutions, as well as an extensive coding library to aid future research on the subject.

Introduction

Elections shift the course of history in ways that are sometimes small, but often enormous. Particularly in the United States, but really in any democracy across the world, elections have the capacity to shape the existence of millions of people for decades to come. Elections always matter, regardless of the final outcome. It is truly wondrous that this power is in *our* hands, that every seismic shift in the course of our states comes from the tremors each of us cause with our ballots. This is a power that people have suffered and continue to suffer for, a power my own parents and grandparents risked their lives for. The way we give people access to this power, the way nations like the United States or my birthplace of Greece choose to conduct their elections and the enfranchisement and disenfranchisement such choices may cause has never ceased to be contentious.

What quickly becomes apparent is that beyond the soaring rhetoric of democratic power, the setup of our elections systems is not straightforward in the slightest. Starting from the beginning, there is still no consensus on *who* gets to vote. Do felons get the vote? Florida (as of 2018) says yes, as does much of Europe, but other states in the US say no. How about permanent residents? Should people have to pass a basic political knowledge test to vote? Is one informed vote more important than a vote from someone who lacks knowledge of current affairs and government? Should people be forced to show ID when voting, or is this restrictive and undemocratic? *When* people get to vote is similarly contentious. Is early voting something we should support, or does it dilute the effect of election day turnout? Should voting day be a national holiday? How about refferenda? Should the people be directly consulted by their governments on seminal issues? Finally, there is the question of *how* or *where* people vote. How many polling places are enough? Should we use paper ballots or voting machines? Should people be automatically registered to vote? Should they be allowed to register on election day? Do we freeze registrations before election day, and if so how long before? Should we allow voting by mail? Permanent absentee status? All-mail elections? This is all even before making choices on subjects like the allocation of representatives by state or prefecture, or the total number of parliamentarians, or how many chambers congress should have.

The 2000 Presidential election between Al Gore and George W. Bush was decided in Florida by a razor thin margin of ballots. Following the final result, which occurred in the halls of the United States Supreme Court whose decision in *Bush v. Gore* halted all Florida recounts that were in progress, researchers claimed that the voting machines and ballot cards used were confusing and misleading. Many voters, for example, are theorized to have mistakenly voted for a different candidate because of the confusing layout of outdated punch-card voting machines(Wand et al., 2001). 22,000 voters in Duval County had their ballots rejected due to "overvoting", because they were given ballots that implied the necessity to vote on multiple pages for the same candidate: an action that spoiled their ballot(Saltman, 2009).

A democratic system is based on procedures as much as on principles. Elections are about translating the vote of the people into political power, government action, and fair representation by chosen representatives. The way we vote can often be critical to the outcome. Thus the design and implementation of voting systems is far from being neutral; the decisions made on who votes, and how, when, and where they do so often serves to alter the course of history. Underlying those decisions is a nebulous, inconclusively answered question: are elections fair, and how can we make them more so?

The United States has long grappled with translating this goal into policy. During the Civil War, several states like Virginia pioneered the use of absentee mail ballots to serve military personnel and displaced residents; this same policy was expanded nationwide in 1942, to accommodate for soldiers and factory workers serving US efforts in World War II. In 1871 Congress passed the Enforcement Act, which clearly defined acts like voter impersonation or intimidation as crimes, and set up a federal structure of supervisors and marshals tasked with overseeing all elections. This was the first in what was to become over a century of sweeping federal election reforms, in an attempt to at least partially centralize an election process whose control had been left to localities. The Civil Rights Acts of '57, '60, and '64 along with the Voting Rights Act of '65 collectively aimed to prevent discriminatory policies at the local level. They abolished literacy tests, made tampering with voter registration rolls a federal crime, and mandated federal oversight of local authorities with a history of discriminatory voter registration practices.

In 1993 Congress passed the National Voter Registration Act (or NVRA), which increased oversight of local registration policies by mandating by-mail registration and registration at the DMV (otherwise known as the "motor voter" program), and by strictly regulating local registration forms. The NVRA also imposed strict regulations against the practice of "purging" registered voters from the rolls, and mandated that states routinely check their registration data to ensure integrity. After the events of 2000 in Florida, the NVRA was followed by the Help America Vote Act (HAVA) of 2002. HAVA included sweeping changes to how American election administration was conducted. For the first time in history, the federal government started spending money on local election administration. The act also banned certain types of voting machines, mandated that local officials accept provisional ballots, and created the Elections Administration Commission (EAC) to gather information and assist state and national elections. Quite importantly for researchers and voters alike (and for this thesis), HAVA mandated that states compile and maintain voter registration and history data *at the state level*, providing some centralized oversight for elections(Ewald, 2009).

Apart from serving local election administration, voter files consolidated by HAVA have allowed researchers to make concrete inferences of individual characteristics

(E. D. Hersh, 2015). Voting related theories derived from political science are now commonly tested using advanced statistical methods and huge amounts of data; both disciplines tackle these data to face joint problems such as quantifying the quality of voter registration files (Ansolabehere & Hersh, 2010), or linking disparate voter records (Ansolabehere & Hersh, 2017). In my thesis I take advantage of such centralized files from the state of Colorado[2].

The purpose of research into policies like the ones described above is to both model and understand the behavior of the voters, and to formulate an idea on what the optimal policy is. Research conclusions on their own obviously cannot sway policy; such decisions are made, fittingly, with the input of the people or their representatives through processes of policy-making, whose study and function is beyond the scope of this thesis. In my thesis, I wish to focus on just one of the multitude of elections policies that are either discussed or enacted in the US at present: Vote By Mail. My purpose is to add to the existing literature of quantitative studies on how Vote By Mail affects voter turnout, and through this process draw conclusions on what behavioral model of voter choice best fits the reality of mail voting.

Apart from the data, such research is also dependent on the statistical methods used to draw inferences. In my thesis, I construct both county- and individual-level models of turnout by using methods such as logistic regression, hierarchical modelling, and natural splines. The use of hierarchical modelling here is particularly significant, as the vast majority of previous studies on mail voting have employed fixed effects models. Hierarchical modelling is particularly salient when data exhibit different "levels" of grouping, here represented by the counties in which individual voters are registered. While fixed-effects regression would assume independence of effects between counties, hierarchical modeling will allow me to do away with that assumption, thus also vastly increasing the inferential potential of my models(Gelman & Hill, 2006).

This thesis should be viewed as a combination of three factors: questions, data, and methods. My fundamental question is one of the most common for elections science[3]: how do we increase electoral participation in the form of turnout, and consequently how do voters choose when to vote? My data comes from Colorado Voter Registration Files: a complete, all-inclusive collection of all current registrants in the State along with their voter histories. To these data, with the purpose of answering the fundamental question I set, I apply hierarchical models; a practice commonly applied in statistical studies but relatively new and exciting as a development in mail voting research.

[2]Voter files commonly include two distinct parts: Voter Registration Files (VRF) and Voter History Files. The former contain an entry for each registered individual with all necessary demographic and personal data needed to correctly identify which ballots they should complete, like where they live, their party registration, their local legislative district, precinct, school board etc. The later contain information on each *ballot* cast, including who cast it, what election it was cast in, how it was cast (mail vote, absentee, etc.), and what county it was cast in.

[3]Sometimes referred to as *psephology*, the study of the *psephos* (*ψήφος*) or "vote" in Greek.

Chapter 1

The State of the Literature

The most basic form of regression is $y \sim x$, where y is the response and x is the predictor. By the end of this chapter, I aim to clarify the theoretical framework behind turnout (response), mail voting (main predictor), and the predicted correlation between the two (value of regression coefficients).

The chapter starts with an examination of turnout: what it is, its use as a metric of participation, and how it is estimated. I then turn to presenting a comprehensive list of current theories on voter decision and participation; these offer conflicting descriptions as to what variables are important when trying to predict the turnout effects of elections policy. I provide a run-down of current voting methods in the US, with a particular focus on mail voting. I make a series of predictions on what the expected effect of mail voting on turnout is, based on each of the aforementioned theories. Finally, the chapter ends on a presentation of past studies that have tried to statistically estimate this effect.

1.1 Turnout and Political Participation

In their seminal work *Participation in America*, Verba and Nie divide the modes of political participation into electoral activity (voting or campaigning), and non-electoral activity (cooperative activity or citizen-initiated contact of political operatives). As Verba and Nie point out, of these forms of participation voting is the most widespread and regularized. All voting-eligible citizens are given specific instructions on how to vote, where to do so, and at what time. Verba and Nie refer to voting as a "high pressure" concern for elected officials, since their continued service in their positions is directly dependent on this form of participation. They do, however, point out that voting does not convey as much information to government actors as other forms of participation; interest groups and direct contact are significantly more targeted. The conclusion they draw is that while voting may not be the most specific of ways to participate, but it is the most measurable, uniform, and general indicator of political engagement(Verba & Nie, 1972).

In subsequent portions of their book they use two measures of voting: turnout and frequency(Verba & Nie, 1972). Frequency refers to the proportion of elections that a particular voter participated in over the total amount of elections they were eligible to vote in. Turnout is the ratio of ballots cast over a measure of the voting population, as in the following equation:

$$\%\ Turnout = \frac{Total\ Ballots\ Cast}{Measure\ of\ Voting\ Population} \times 100\%$$

In this thesis, I will use turnout as a measure of electoral political participation. I make this choice for three reasons: first, there is substantial literature on the relationship between mail voting (the subject of my study) and turnout; second, there is also substantial literature behind the choice that individuals make between turning out to vote and not participating, meaning that I have a strong theoretical background from which to build hypotheses; third, the data I have available includes complete individual electoral history, making calculating turnout possible.

1.1.1 Calculating Turnout

There are two issues to consider when calculating the numerator for turnout. In any given election, there can be several candidates and issues on the ballot, ranging from US Senate races to local school-board contests. Therefore a first issue is "undervoting", or only completing some parts of a ballot and not voting in every contest(Saltman, 2009). This means that there can be different turnout counts produced by electoral contest for any given election date: one for US Senate, one per state legislative contest, etc. Alternatively, turnout can be calculated for that election day, and not for any particular race; here the numerator is every ballot cast. I use this second conception of turnout, calculating at the election level. A second issue is that a fair amount of voters may turn out (by physically going to a polling place or mailing their ballot), but have their ballot rejected for registration issues, concern for authenticity, or a range of other complications. Since these voters did take action to participate, should a metric of turnout as political participation include them? I would tend to answer yes, but the data I have available do not include information on voters having their ballots rejected; this means that my final calculations only include participants whose ballots were officially counted.

The three main statistics used for the denominator are the total voting age population (VAP), voting eligible population (VEP), and the number of registered voters in a certain geographical location. The total voting age population (all individuals over 18 years of age) can be measured using data from the US Census. However, such an interpretation of VAP counts individuals of age that are not allowed to vote, like people with severe mental illnesses or felons, and does not count oversees voters or military personnel. Michael McDonald offers an alternative to VAP he calls "voting eligible population", which corrects for such individuals (McDonald, 2007).

Counts of registered voters are also a useful tool for calculating turnout, as they usually require no estimation. These counts can simply be extracted from voter registration files. Such a metric is only representative of the electoral political participation level of a group that has already taken a step towards participating, by registering to vote. Researchers should be careful to also include registration level by income, race, gender, or other metrics if using this form of turnout calculation to infer levels of participation by social or demographic characteristics.

The punch line here is that calculation of turnout is not an obvious choice, and will have an impact on what conclusions are drawn. To give one example, consider Oregon's Motor Voter program, that automatically registers voters when they interact with the DMV. It is conceivable that this reform will *decrease* turnout when measured as a percentage of the total registered voter count, but *increase* turnout when measured against total population. This happens if more people register to vote, but do not actually do so–in other words, both number of registrants and number of ballots cast are increasing, but the former increases at a larger rate than the latter. In my thesis I will use registered voter counts as the turnout denominator.

1.1.2 Turnout and Voting Probability Models

Statistical models of turnout can be constructed at either the individual or group level. At the individual level, a model is built to predict the probability of voting for every member of a group, which then can sum over the members to create an estimate for turnout. This is a classification problem that can be solved with models such as Probit or Logit generalized linear regressions. Turnout can be calculated for several groups: counties, precincts, racial characteristics, etc. The unit of observation is a single group, and a regression model is fit on turnout as a continuous variable. In this thesis, I estimate models both at the county and the individual level.

All of these models include a standard set of societal variables at the individual and/or group level, administrative variables (whether the district uses Postal Voting, whether Voter ID requirements are particularly strict), election-specific variables (closeness of election or campaign expenditure) and sometimes time-series data like previous levels of turnout. This type of analysis is used either to predict turnout or draw inferences on the effects that certain explanatory variables have on electoral participation.

Meta-analyses conducted by Geys (2006), Geys and Cancela (2016), and Smets (2013) provide a clearer picture on which predictors are relevant to turnout. Geys includes 83 studies of national US elections in his initial meta-analysis (Geys, 2006), later increasing that number to 185 (Geys and Cancela, 2016) and adding local elections. On aggregate-level models for national elections they conclude that competitiveness, campaign financing, and registration policy have the most pronounced effects, while on the sub-national level there are more pronounced effects for societal variables and characteristics of election administration (spending, voting policy, etc.). Smets and Van Ham (2013) examine individual-level predictors for turnout in a similar meta-analysis, and conclude that "age and age squared, education, residential mobility, region, media exposure, mobilization (partisan and nonpartisan), vote in previous election, party identification, political interest, and political knowledge" (Smets & Ham, 2013) are the most significant explanatory variables for turnout, along with income and race. I will specify the model I will use for turnout in the second chapter.

1.1.3 Deciding to Vote

Here I take one step back from turnout, and examine the theories surrounding individual choices to vote or abstain. There are four main theories outlined in the literature on why individuals chose to vote:

- *Decision "at the margins"*: In his 1993 study, Aldrich posits that voting is a low cost/low benefit behavior. Voting is a decision that individuals make "at the margins"; in most people, the urge to vote is not overwhelmingly strong, so individuals will vote when it is convenient to them, when they are motivated by a competitive race, when policies are put in place to help them, or when they are subjected to GOTV (Get-Out-the-Vote) efforts. For Aldrich, this is corroborated by the fact that most turnout models present consistent, yet weak, relational variables: if decisions are made "at the margins", then no single predictor would have an overwhelming effect. This is also supported by Matsusaka (1997), and Burden & Neiheisel (2012). Matsusaka expresses support for a more "random" process of voting, where turnout models are ambiguous because of the difficulty that predicting "at the margins" entails (Matsusaka & Palda, 1999). Burden & Neiheisel (2012) use data from Wisconsin to calculate a net *negative* effect of 2% on turnout following the expansion of early voting access in the state. While this can be read as an Aldrich effect on turnout, the authors claim that their findings should be attributed to a lack of "election-day effects"(Aldrich, 1993; Neiheisel & Burden, 2012).

- *Habitual Voting*: While Aldrich supports that there is no single overwhelming predictor of turnout, Fowler (2006) posits that future voting behavior can be strongly predicted using individual voting history. This leads to the conclusion that individuals are set to either be habitual voters, or habitual non-voters (Plutzer, 2002) by their upbringing and social circumstances, locking them into distinct groups (Fowler, 2006).

- *Social/Structural Voting*: Social and structural voting presents several common characteristics with Habitual Voting, but includes the claim that the decision to vote or not is deeply rooted in socioeconomic factors. This in turn is interpreted to mean that the divide between traditionally voting and non-voting groups can only be bridged by directly dealing with the socioeconomic divide between them (Berinsky, 2005; Edlin, Gelman, & Kaplan, 2007).

- *Resources and Organization*: To some extent growing from structural theories of voting, resources and organizations theory emphasizes the interaction of personal political and societal characteristics of voters, and actions taken by politicians to mobilize participation. This theory is very broad in the inputs it assesses for voter participation, ranging from practical issues of access and resources (how easy it is for someone to vote if they so choose), to public policy feedback effects and signaling (how the government's policies affect the people and how they react), to how political parties and groups choose to mobilize and approach voters (Rosenstone, 2003). Apart from Rosenstone and Hansen's work (2013), there have been several studies examining voter participation based on resources and organizations theory, a lot of which come from the public policy side of political science. Some examples are Chen's study of how distributive benefits like federal emergency aid affect participation among recipients, after controlling for partisan characteristics (2012), or Mettler and Stonecash's examination of correlation between welfare program participation and political mobilization (2008), or Campbell's analysis of social security recipients and their voting patterns (2002). The punchline in all these studies is that public policy is correlated with trends in participation, either because recipients of benefits wish to protect such programs, or because of the interaction between partisanship and government support, or because of access related to resources and voting laws (Campbell, 2002; Chen, 2013; Mettler & Stonecash, 2008).

1.2 From Theory to Policy

1.2.1 Voting Methods

I have already flagged in my introduction the reason why theories behind voting choice matter: each constructs an image of the electorate that reacts differently to policy change around voting. They are all an answer to the fact that elections policy is not value neutral, but has implications for turnout, which in turn has implications for the franchise of democracy.

In trying to respond to the issues set up by theoretical paradigms, different states–both in the global and US contexts–have adapted to different ways of conducting elections. In the US, voting styles can be simplified into three categories:

- *In-Person Election Day*, for which all individuals are required to vote at a polling place on election day. Standard accommodations for overseas or excused absentee voters apply, but the vast majority of people will have to be present to vote in a particular time frame.

- *In-Person Early Voting*, for which all individuals must vote in person at a polling place, but the timeframe for voting extends for around two weeks and not a single day.

- *Mail Voting*, for which individuals have a no-excuse-necessary[1] option for not being present when they vote, or for filling in a mailed ballot and dropping it off at designated locations.

Note that some form of voting via mail ballot exists in all three cases; the distinction between Mail Voting and the other two is that there is no excuse necessary to cast a

[1]Individuals do not need to present proof that they will be out of state or unable to physically be present at a polling place, but can just ask to be mailed their ballot.

mail ballot. In such systems mail voting is not an exception, but exists as something between a common practice and a universal means of casting a ballot[2].

In-person election day voting is, historically speaking, the way the vast majority of democracies have conducted their elections. Therefore it has been of interest for researchers to examine if other systems can outperform that baseline for some metric of participation. Specifically, it is most interesting to examine voting styles that are heralded for their expansion of turnout, to see whether popular beliefs on their benefits and drawbacks hold: are such policies different from the base model of conducting American elections? Do they present new challenges and unique selling points? In this context, mail voting is particularly interesting because it is quickly becoming a trend in state-level elections administration.

1.2.2 What is VBM?

Vote-By-Mail (or as it is commonly referred to by most in the field, VBM) is a process by which voters receive a ballot delivered by mail to their homes. Voters then have a variety of options on how to return their ballot, ranging from dropping it off at a pre-designated location, to mailing it in, to bringing it to a polling place. The two first options are most commonly implemented, with a very small number of states still operating polling places for mail ballots. This varies across states that have implemented VBM. Some common forms of the VBM policy are:

- *Postal Voting*: All-Mail elections. All voters receive a ballot by mail, which can then be returned to a pre-designated location or mailed in to be counted. All-mail elections currently occur in Oregon, Washington, and Colorado.

- *No-Excuse Absentee*: Voters can choose to register as absentee voters without giving any reason related to disability, health, distance to polling place etc. This is the case in 27 states and the District of Columbia.

- *Permanent No-Excuse Absentee*: This is similar to the previous system, but allows voters to register as absentees indefinitely, without having to renew their registration each year; they become de facto all-mail voters. This is in place in a very large number of the no-excuse absentee voting states like Utah, California, Montana, Arizona, New Jersey, and others.

- *Hybrid Systems*: In hybrid systems, voters receive a mail ballot but can choose to disregard it and vote conventionally. This is the case in Colorado. In other states postal voting is not mandatory for all counties, as some are still allowed to conduct their elections conventionally; these states keep an in-person option open as a part of their elections administration system. This is the case in California, Utah, and Montana.

[2] It is also commonly argued that full postal voting is a 4th category in and of itself. I don't make this distinction here, but in the following section I break down Mail Voting into different categories, one of which is full postal voting.

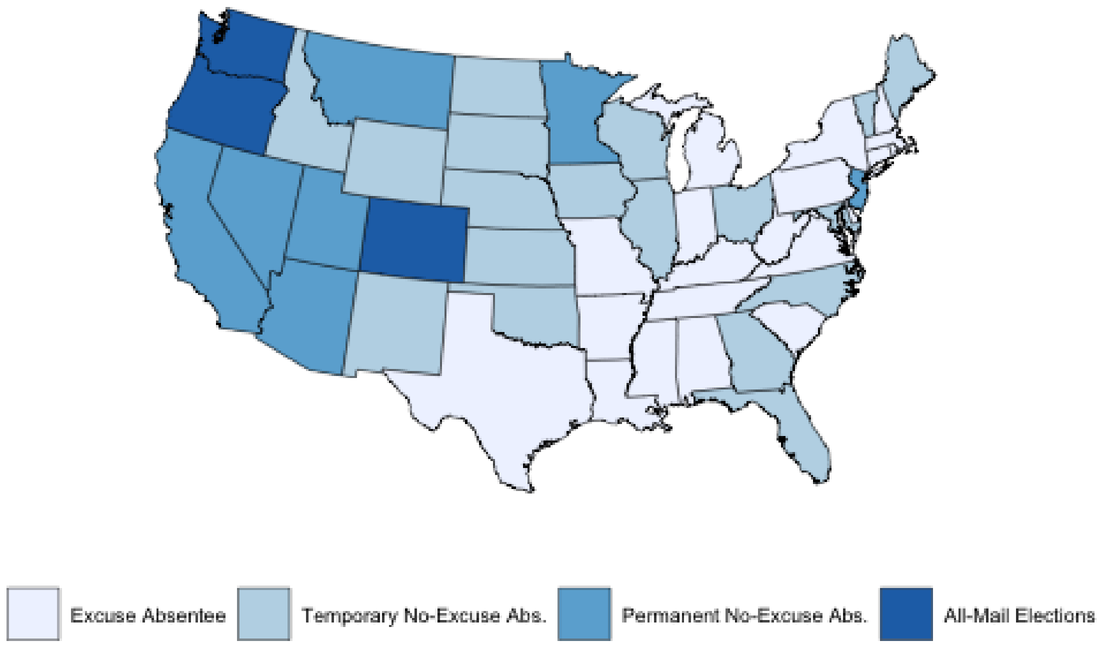

Figure 1.1: Mail Voting Status by State

Vote-By-Mail is also commonly considered a type of early voting, since voters receive their ballots around two weeks in advance of election day; they are also able to return that ballot whenever they wish within that time-frame. This means that Vote-By-Mail can be counted as a "convenience voting" reform (Gronke et al., 2008). These are usually implemented by state and local governments with the argument that they either expand the democratic franchise by bringing in new voters, or by making it more likely that current registered voters participate in the electoral process ("Absentee and Early Voting," 2018).

Figure 1.1 presents current mail voting status by state. Not included in this map is the fact that several states have provisions for all-mail elections to be conducted under certain circumstances; what these circumstances are varies by state. Such states are: Alaska, Arizona, Arkansas, California, Florida, Hawai, Idaho, Kansas, Maryland, Minnesota, Missouri, Montana, Nebraska, Nevada, New Jersey, New Mexico, North Dakota, and Utah("Absentee and Early Voting," 2018).

1.2.3 How Theories Apply to VBM

Under Aldrich's paradigm, vote by mail would not effect significant change in voting behavior. The whole concept of a decision "at the margins" is that the forces at play when an individual decides to vote are overwhelmingly strong both ways, so any effect that policy can have will minimally shift these margins. In a presidential election, for example, the forces at play include the media, national committees, social effects etc. In this environment, some added convenience does not significantly add to an individual's decision to turn out. However, this would indicate that at a local level, where national and media effects are less strong, the effect of VBM on turnout might be more significant. The effect would be present for all groups, not only those currently registered, since voting would be easier uniformly.

If we assume habitual voting, the conclusion on VBM would differ significantly. In this case, the effect to be considered is how VBM impacts already formed habits around voting. It could be argued that VBM has no effect, which follows if we assume that voting habits formed do not shift if the mode of voting changes. It could also be argued that VBM might have a negative effect on turnout in the short term, because it disrupts the habit of election day, before people adapt to the new policy context and settle into new groups of habitual voters and non-voters.

Under social and structural voting contexts, VBM retains rather than stimulates new voters (Berinsky, 2005). This means that already registered and semi-active voters are more likely to participate, but there is no significant change in the amount of new voters entering the franchise. This would mean that traditional forms of voting policy that emphasize access to the polls will do nothing to bring in disenfranchised people, and potentially hide the problem under an inflated turnout statistic calculated on registered voters. Berinsky in particular emphasizes the need for a shift towards voter education, rather than early voting or VBM policies (Berinsky, 2016).

Vote-by-Mail is obviously not a welfare or spending program, but it does increase individual resources in terms of voting capacity. Filling out a ballot at your kitchen table does not include spending time to go to a polling center, or standing in line to vote, or factoring in significant changes to your daily schedule for election day. A mail ballot can usually be dropped off at several locations, or can be mailed in along with any other mail a voter may have to send. Less time and effort is spent on casting your vote. This in turn has both a practical effect (building capacity) and a more behavioral effect such as a feeling of inclusion, or an interaction with the process of voting that comes through a ballot at your doorstep. This effect would not exist if you had chosen not to go to a polling place (Schneider & Ingram, 1990). Under a resources and organizations framework, such practical and behavioural effects would lead to increased political participation and as such would predict a strong, positive effect of VBM on turnout.

1.2.4 General Results on VBM

I will start with studies that show a negative effect on turnout. Bergman (2011) uses a series of logistic regression models of individual voting probability in California, during a period where only some parts of the state conducted VBM elections. This is called a natural experiment, and it is common throughout the literature. Bergman's results show a statistically significant drop in voting probability in VBM counties (Bergman & Yates, 2011). Using a similar method, Keele (2018) takes a single city in Colorado (Basalt City) which is divided into two different voting districts which impliment different voting systems. The conclusion is, again, a 2-4% drop in turnout along the VBM part of the city (Keele & Titiunik, 2017). Burden et al. (2014) take a different approach, using country-wide election data from 2004 and 2008 presidential elections, and compares districts based on early voting practices. Their results show a significant drop in turnout, which can be associated to VBM as well due to its closeness to EV (Burden, Canon, Mayer, & Moynihan, 2014).

In contrast, Gerber et al. (2013), applying both individual and county-level models for the state of Washington, reach the conclusion that VBM increases turnout by around 2-4%; they use the same natural experiment model that offers itself to researchers in states that are under transitional systems (Gerber, Huber, & Hill, 2013). R.M. Stein also reaches a similar conclusion when examining Colorado's practice of "vote centers", which are non-precinct attached polling places servicing multiple counties (Stein & Vonnahme, 2008). I include this paper here due to the link that voting centers have with VBM, as they serve as drop-off points for mail-in ballots. Richey (2008) examines the effects that Oregon's VBM program has on turnout by using past elections data, concluding a 10% positive trend associated with the policy (Richey Sean, 2008). This effect is studied again by Gronke et al.(2012) who find a similar positive effect with much lower magnitude, which might point to a novelty effect: the existence of diminishing returns in turnout after the implementation of this policy (Gronke & Miller, 2012). Gronke et al. (2017), again studying Oregon but focusing on Oregon's Motor Voter program, find evidence of positive association to turnout. I include these effects due to Oregon being an exclusively VBM state, and because this paper uses a "synthetic control group" model, a particularly interesting statistical technique(Griffin, Gronke, Wang, & Kennedy, 2017). Lastly, a study conducted by Pantheon Analytics on Colorado, which compares actual turnout to predicted levels for VBM counties in Colorado. The results show a positive effect of approximately 3.3% due to VBM (Edelman & Glastris, 2018).

The conclusion to be drawn is that results on VBM vary significantly. There are multiple studies, using multiple methods, on multiple states, with multiple results. This only adds to the importance of being careful when constructing models and hypotheses to test VBM's effects on turnout, as assumptions made in the process can critically impact the results.

Chapter 2

Hypotheses and Methods

The previous chapter served as an introduction to the literature surrounding turnout, mail voting, and voter participation. Based on the conclusions of past researchers and the frameworks they have presented, I move to outlining the hypotheses that motivate my own study. I conclude this chapter with a brief presentation of the statistical methods employed.

2.1 Hypotheses

2.1.1 Questions

Before outlining my hypotheses, the first step necessary is to frame a series of questions which the hypotheses will flow from. Based on Chapter 1, the most obvious first question to ask is:

> Q1: *What is the effect of mail voting on turnout?*

I went through this question substantially in the previous chapter; it should be clear that depending on which paradigm of participation choice is present, the answer here can be radically different. In order to best answer the previous question, it is necessary to establish some conditions on importance of effect. Therefore it is also necessary to ask the following question:

> Q2: *Does this effect persist when accounting for other relevant predictors of turnout?*

The last question asked in this thesis is more specific to a particular formulation of Aldrich's hypothesis on voting "at the margins". I mentioned in the previous section that VBM could be theorized to have a more significant effect when discussing elections at the local level, or the regional level, rather than national general elections. Therefore a third question is:

> Q3: *Is the effect of VBM on turnout more pronounced as significant, national determinants become less strong?*

2.1.2 Hypotheses

Using the above questions I can now move on to formulate more clear hypotheses. The hypotheses in this section are meant to test theories of voter choice from the perspective of the theory of voting "at the margins" as introduced by Aldrich.

In response to Q1, Q2, a first hypothesis is:

> H1: *Mail voting is another incremental effect on voting decisions, and therefore does not significantly affect turnout*

The alternative hypothesis would be:

> H1′: *Mail voting significantly affects turnout, even compared to other metrics*

Similarly, for the third question, a corresponding hypothesis derived from Aldrich's paradigm is:

> H2: *The effect of VBM on turnout is larger as national effects become less pronounced*

The alternative hypothesis is:

> H2′: *The effect of VBM on turnout is either independent of or proportional to the presence of national effects*

2.1.3 Criteria

A first, glaring issue that needs to be clarified is the apparent contradictions between my two hypothesized results. This becomes clear, however, if I define "significant effect" in the context of my first hypothesis. Aldrich's paradigm does state that "conveniences" like mail voting should not have significant effects, but those effects are defined in the context of huge, clashing forces that vastly outweigh them. This does not necessarily mean that they are literally non-existent, but that they are poor indicators of consistently increased turnout. Therefore, I will confirm my first hypothesis not only if the effect of mail voting on turnout is vanishingly small, but also if it is relatively small in comparison to the effects of other variables I include. I will confirm the alternative hypothesis if, across multiple models, VBM retains a consistent, significant effect on turnout. If the effect is negative, this may point to a habitual or structural voting effect. If the effect is positive, this may be a signifier that issues of convenience in voting–having a mail delivered ballot, voting from your kitchen table etc.–have a particularly strong effect in the examined elections.

Moving on to the second hypothesis. It is extremely hard to correctly operationalize and account for all variables going into turnout. Therefore, instead of trying to include all possibly relevant effects into a model and try to see how they interact with VBM, I will test my hypothesis on different levels of elections: local, midterm, presidential, and primary. National effects on turnout should be especially present in presidential races, since a specific set of candidates is running across the whole nation. These effects should also be present in midterm and primary elections to some extent, as the results of local races are aggregated in control of congress or high-profile governorships. Apart from a ballot measure that garners national interest, or a singularly high-profile race, local off-year election turnout should have a negligible relation to national effects. Therefore I will use election level as a stand-in for the prominence of national turnout effects. The following is an alternative formulation of the second hypothesis, made more specific to the criteria I have set:

H3: *The effect of VBM on turnout is more pronounced in local or off year elections*

I will confirm this hypothesis if mail voting has substantially larger positive effects on turnout in smaller, local elections. Based on Section 1.2.3, Table 2.1 presents what each theory predicts for the hypotheses.

Table 2.1: Predicted Outcomes from Hypotheses

	H1	H2	Effect on Turnout
Decision at the margins	Y	Y	Marginal
Habitual Voting	?	N	No Effect/Decrease
Social/Structural Voting	Y	N	No Effect
Resources and Organizations	N	N	Increase

2.1.4 Importance of Hypotheses

The importance of these hypotheses is intrinsically tied to the importance of different theories of electoral participation. Confirming or rejecting each hypothesis–even when only applied to a single state–serves as an argument for or against one of the aforementioned theories. The theories in and of themselves are important, since they form a part of a broader literature on elections, democracy, and electoral processes, that can be said to be foundational to political science as a whole. As mentioned in my introduction, elections themselves are significant, since they are the process by which political power and representation is allocated according to the will of the people.

Additionally, from a public policy perspective, these hypotheses are significant since they are connected to the effectiveness of mail voting as an electoral reform. Whether, in general, mail voting increases turnout is directly connected to whether it is successful in expanding the democratic franchise. If it is not, questions can be raised as to the effectiveness of expanding voter access through elections administration, rather than education, or even measures like voting-day-holidays or local transportation to polling places. For local elections in particular, significant effects of mail voting could be precursors to more general involvement of individuals in their local politics. This may open the way to numerous comparative studies on local politics between states that apply VBM and states that do not.

Lastly, from a narrower perspective specific to the study of early and mail voting, my first hypothesis can still be said to be quite important, yet mundane. It does its job according to the particular state I chose to look at–in this case Colorado–to add to existing literature on mail voting effects in different parts of the country. However, my second and third hypotheses are much more unique in their scope. There have not been many studies that look at VBM at a local level, and any addition to the literature on this front could be significant.

2.2 Methodology

In the next chapters, I will introduce the data and fully outline my models. Before that, the following section should serve as a general introduction to the methods. I will not extensively go through the statistics behind linear or multiple regression, but will assume that it is common knowledge. For an extensive introduction to such methods, James et al.(2017) or Chihara and Hesterberg (2011) are particularly useful.

2.2.1 Logistic Regression

Let function $f : [0, 1] \to \mathbb{R}$ be defined as:

$$f(p) = \text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

This is called the logit function or, when p refers to a probability, the log-odds function. When modelling a binary response Y, which follows a Bernoulli distribution:

$$Y \sim \text{Bernoulli}(p),$$

the logit function can be used as a link function to model Y in a generalized linear model. The generic form of a generalized linear model looks like:

$$f(Y) = X\beta,$$

where Y is a vector of response variable values, X is a matrix of predictors, and B is a vector of coefficients to be estimated. The function f is called a link function, because it "links" the response variable with the set of predictors included in the model. This is typically done to ensure that the range of values outputted by the model are consistent with the range of the response variable.[1] When wanting to compute a model on a binary response through its corresponding Bernoulli distribution probability parameter, the inverse logit function should be a perfect fit for a link function, since it maps values from all real numbers to a range between 0 and 1. Using

[1] Or, in this case, the range of the parameter defining the distribution of the response, which is p for the Bernoulli distribution

the inverse logit function, we arrive at the final form of logistic regression, which is:

$$\mathbb{P}(Y = 1) = \text{logit}^{-1}(XB)$$

Conveniently, despite the use of a link function, there is an easy way to interpret the coefficients of such a regression. While obviously individual values from the B vector will not be particularly helpful, e^B can be used as a vector of multiplicative, one-unit shifts in the value of the probability that $Y = 1$. This means that a one unit increase in any predictor will cause an effect equal to multiplying p by the exponent of the corresponding coefficient[2]. (James, Witten, Hastie, & Tibshirani, 2017)

2.2.2 Generalized Additive Models

In simple logistic or linear regression, there is an assumption made on the functional form of the relationship between predictors and response variable. These are called parametric models, where the data is exclusively used to estimate values for coefficients. Non-parametric models, on the other hand, use the data to estimate both coefficients and the function that serves to connect response to predictors. While on the surface this seems like a great idea (more reliance on your data and fewer assumptions!), such an exclusively non-parametric model would suffer greatly from the curse of dimensionality–where the addition of multiple predictors or over-reliance on data leads to substantial over-fitting.

One solution is the Generalized Additive Model, or GAM. This model lets us fit a different functional form to each predictor, allowing for assumptions to be made on the data where it is safe to do so, and for non-parametric fitting when it is necessary. This model looks like:

$$y_i = \alpha + \sum_{j=1}^{p} \beta_j f_j(x_{ij}), \quad j \in \{1, 2, ...p\}, i \in \{1, 2, ...n\}$$

where y_i the i-th response variable, α is the intercept term, f_j, β_j a series of p functions and coefficients, and x_{ij} the i-th observation for the j-th predictor. Note that for $f(x_{ij}) = x_{ij}$, this is a multilinear regression! (James et al., 2017)

A type of most commonly fit functions–and the type I will make use of–are smoothing splines. These are functions connected at specific points called "knots", with the limitation that the full function must be continuous and smooth, and have a continuous first and second derivative. Between knots, different functional forms are fit to the data, within some constraints; they may, for example, all have to be cubic polynomials. These are particularly useful when modeling time variables, as they can be fitted to variables like years or months in order to distinguish a secular trend from a general trend over time (Barr, Diez, Wang, Dominici, & Samet, 2012). In terms of this thesis, this will help when responding to Q2 as it was framed earlier in this chapter.

[2]This can be simplified even more, since exponentiation can be approximated by dividing the coefficient by 4. Crude, yet effective for a quick scan of the results

2.2.3 Multilevel Models

Multilevel models (otherwise known as hierarchical or "mixed effects" models) can be intuitively pictured in two ways: either as a set of models working on different "levels", where one is calculated first, with its effects having implications for the second, or as a model where some of the parameters are estimated under a particular series of constraints. Multilevel models are, in essence, a compromise between levels of "pooling" data. If the dataset on which parameters are being estimate operates in different units of observation–say on the individual and county level–you could run a model that treats all individuals as coming from the same larger group; this would be a complete pooling model. You could also add indicator variables for each and every group, de facto estimating n different models for n groups; this would be a no pooling model. Multilevel modelling offers partial pooling (Gelman & Hill, 2006).

To consider what this model looks like, let's assume a dataset comprising of a vector of values for the response variable Y, a matrix of i individual level predictors X, a matrix of j group level predictors U, intercept terms α, individual level coefficients B, and group level coefficients Γ. Based on this, a multilevel model with intercept terms varying by group looks like:

$$Y_i = \alpha_{[i],j} + X_i B \ , \qquad \alpha_{[i],j} \sim N(U_{j[i]}\Gamma, \sigma_\alpha^2)$$

Multilevel models can be fit using the `lme4` *R* package that uses restricted maximum likelihood calculations for estimating coefficients (Bates, Mächler, Bolker, & Walker, 2015). Multilevel modelling also works perfectly well with general additive models! In *R* this can be accomplished with the `gamm4` package (S. Wood & Scheipl, 2017).

2.2.4 Model Accurracy and Quality of Fit

Mean Squared Error (MSE)

For all generalized linear regression models (including GAMs, mixed and fixed effects models) I use Mean Squared Error to assess the accuracy of the fit. Assuming a dataset $\{(y_0, x_0^1, x_0^2, ..., x_0^m), ..., (y_n, x_n^1, x_n^2, ..., x_n^m)\}$ of n observations and m predictors, with X_i a vector of the predictors for the i-th observation, and $f : R^m \to R$ the true multivariate function connecting the predictors and response, mean squared error is calculated as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(X_i))^2$$

MSE can be calculated either using the same dataset used in estimating the model coefficients, or on a new dataset. In the later case it is called predictive or test MSE. Despite prediction not being the purpose of the models presented in this thesis, I use test MSE because of the independence such a calculation method brings from the data used for the fit, compensating in a way for over-fitting(James et al., 2017). To calculate test MSE I use five-fold cross-validation, which will be analyzed shortly.

Area Under the Curve (AUC)

Logistic regression models estimate the probability of a binary variable being equal to 1, or alternatively an indicator variable taking a "TRUE" value. The predictive output of such a model will be a series of probabilities. These probabilities can then be used to approximate a dataset of positive and negative values for the response variable (in my case, voting). Based on the true values of the response, one can calculate the counts of true positive, true negative, false positive, and false negative predictions. To make this calculation, a probability threshold is set over which the prediction for the response is positive. Positive predictive values of the response are assigned based on the following statement:

$$\mathrm{P}(y_i = 1|X_i) > p$$

where y_i, X_i can be assumed to be the same as in the previous section, and p is the threshold. A common and intuitive threshold is 0.5, but any number in $(0, 1)$ can be used. After getting counts for true/false negative/positive values, one can then calculate *specificity* and *sensitivity* for the model. These are:

$$\text{specificity} = \frac{\text{True Positive}}{\text{False Negative} + \text{True Positive}}$$

$$\text{sensitivity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}}$$

Using sensitivity, specificity, and probability threshold it's possible to create an ROC curve, which is one of the most widely used diagnostic plots for classification models[3]. The ROC curve has $1 -$ specificity on the x-axis, sensitivity on the y-axis, and each point describes a pair of x-y values for each value of the probability threshold. Using this plot, it's possible to measure the *area under the (ROC) curve*, or AUC, which serves as a goodness-of-fit measure for classification models. The AUC is a number in the $[0, 1]$ range and should be maximized; a .5 AUC is representative of an ROC curve on the $y = x$ line, which is a coin-toss no-information classifier(James et al., 2017). Plot 2.1 is an example of an ROC curve.

Similarly to MSE, there is value in calculating AUC from a test dataset, rather than the dataset used to train the model. Therefore I also use 5-fold cross-validation for AUC as well[4].

k-Folds Cross Validation

The goal of statistical modeling is to approximate the true function that links predictors to response. While the final model's coefficients should be estimated using as much data as possible, when assessing how good a fit that model is there can be better uses of the power that large amounts of data give us. k-Folds cross validation allows

[3]The ROC curve takes its name from a term in communications science, the *receiver operating characteristics curve*. The name is historic, and not relevant to its statistical application.

[4]This also compensates for models not converging, as some of mine do.

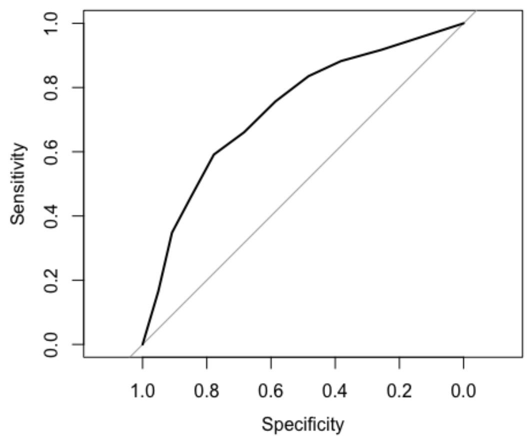

Figure 2.1: Example of an ROC curve

for better approximations of goodness-of-of-fit metrics, by partitioning the data into training datasets and test datasets. The fundamental idea is that the data is split into k different subsets, which are then sequentially used to fit the model and calculate the value of some metric(James et al., 2017). The algorithm goes as follows:

1. Partition data into k folds
2. Fit model on all but the i-th fold
3. Calculate goodness-of-fit metric on the i-th fold
4. Repeat 2 and 3 for i$\in [0, k]$
5. Calculate the average of all obtained goodness-of-fit measurements

I perform 5-fold cross validation to calculate MSE and AUC for all models which I estimate in R.

Chapter 3

Case Selection and Data

After operationalizing my Hypotheses and providing an introduction to the methods of my thesis, the natural next step is to introduce the data. In turn, the first step to this process is to introduce the *source* of the data: the State of Colorado. This chapter begins with a presentation on the demographics, politics, and electoral policy of the Centennial State, which transitions into a justification for th selection of Colorado as a case for my research. I then introduce the data itself, going through basic descriptions like unit of observation and variable specifications. I conclude by presenting some problems faced with the wrangling that was necessary to get the data into a usable form. These problems are not negligible, but are a key part of the work that is necessary to conduct elections science research with voter records.

3.1 The Centennial State and Its Voters

3.1.1 Demographics

Colorado, named the Centennial State due to assuming statehood on the centennial of the Union, lies in the Southwestern United States, with its Western half squarely atop the Rocky Mountains. Based on its estimated population of just over 5.5 million, Colorado is the 21st most populous state, and ranks 37th in population density. The vast majority of that population is gathered in a series of urban areas that comprise a North-to-South strip in the middle of the state, containing the Denver-Aurora-Lakewood Metro Area, Colorado Springs, Pueblo, and Fort Collins. Apart from the Western town of Grand Junction, the rest of the population resides in vast rural areas.

Colorado is landlocked, and far from any coastal town; in place of seaside resorts, Colorado attracts a substantial amount of tourists to its mountains every year. They also heavily depend on federal money and protection for national parks. Colorado has a median age of 34.3 and median household income of $65,685. Colorado's population is mostly white, with a higher minority group population density in its Southern regions, as shown in figure 3.1. ("U.S. Census Bureau QuickFacts," 2010). The conclusion here is that Colorado is a relatively young, mostly white, and fairly well-off state that is increasingly getting more diverse, particularly in the South. These factors are important as they serve to associate Colorado with other states; such associations

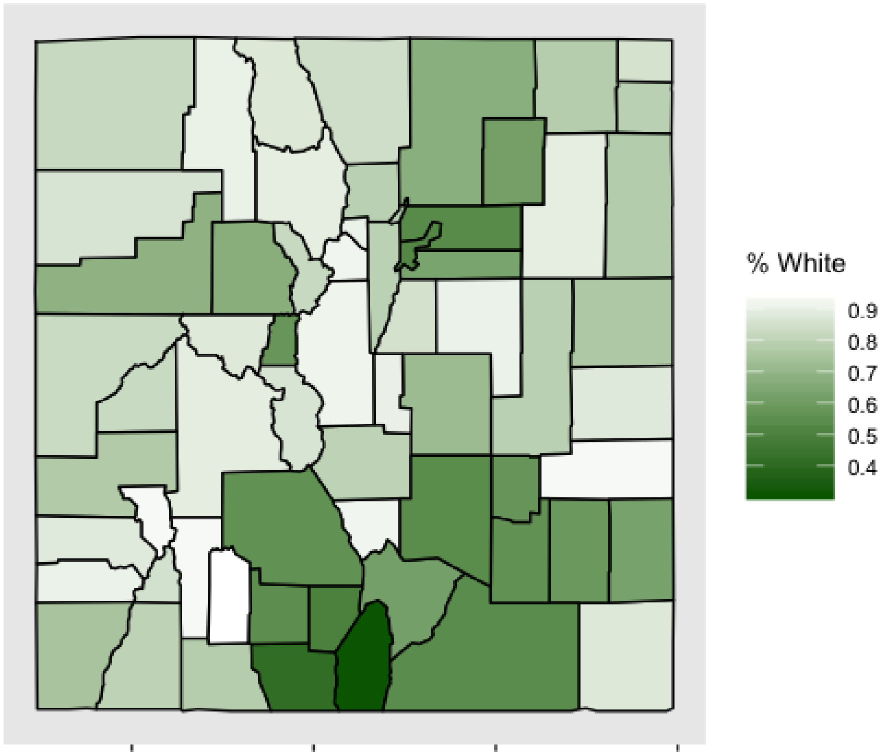

Figure 3.1: White voters per Colorado county

are useful for the replication of this study or the generalization of my results.

The State Capital is Denver; Colorado is split into 64 Counties, of which the most populous are, in no particular order, the following: El Paso, Denver, Arapahoe, Jefferson, Adams, Larimer, Boulder, and Douglas. These counties comprise 73% of the total population of Colorado.

Table 3.1: Colorado population for largest counties

County	Total Population	CO Population %	Largest Metro Area
Adams	441603	0.08781	Denver-Aurora-Lakewood
Arapahoe	572003	0.1137	Denver-Aurora-Lakewood
Boulder	294567	0.05857	Boulder
Denver	600158	0.1193	Denver
Douglas	285465	0.05676	Denver-Aurora-Lakewood
El Paso	622263	0.1237	Colorado Springs
Jefferson	534543	0.1063	Denver-Aurora-Lakewood
Larimer	299630	0.05958	Fort Collins
Other	1378964	0.2742	
Colorado	5029196	100	

3.1.2 The Politics of Colorado

Curtis Martin (1962) notes that Colorado, due to its status as a frontier state, has always been fiercely democratic and independent. He connects this fact with Colorado's past, by pointing out that its political institutions were deeply rooted in mining culture, ordinary citizens' participation,a strong feeling of being "far away" from sources of centralized power on the coasts, and a wish for the protection and preservation of Colorado's natural environment. As such, Colorado can be described as a populist state with a strong libertarian streak, that highly values democratic processes when they serve the people or protect and fund national parks, but staunchly opposes state intervention when it is unwarranted (Martin, 1962).

This 1964 study of Colorado politics rings true to this day. One needs not search for long to see instances when Colorado honored this description. One example is TABOR, or the Taxpayer's Bill of Rights; a strongly libertarian, small-government, populist series of regulations that mandated a referendum for any measure that would increase state taxes, and caped government spending. TABOR was passed by referendum in 1992, and later amended in 2005 after the dot com economic crisis exposed the fact that inability to spend is very bad for a state government trying to jump start its economy (Staff, 2009).

Similarly, Amendment 64 passed in 2012 made Colorado one of the first states to legalize the selling, possession, and consumption of recreational marijuana; a policy advocated by progressives and libertarians alike. Colorado was also the staging ground for what has been coined the "Sagebrush Rebellion": a movement primarily consisting of ranchers in dispute with the federal government over land use laws and wildlife protection. While this "rebellion" primarily consisted of battles in local legislatures or elections in the 1970s, its echoes can be heard till today in events like the Bundy Standoff, with ranchers taking up arms against federal employees and occupying federal land (Thompson, 2016).

Setting policy aside, this description of Colorado is also confirmed by polling data and election results. While being traditionally more conservative, inflows of immigration from the South coupled with increasing urban liberalization and tourism has led the state from leaning republican to being aggressively purple: the quintessential swing state. Colorado voted both for and later against Bill Clinton, voted for G.W. Bush twice, and has supported democratic presidential candidates since (Hamm, 2017). Additionally, when polled on trust of federal or local governments, Colorado residents are systematically skeptical; in a random sample poll conducted by Cronin and Loevy (2012) in 2010, 56% stated that their state officials were lazy, wasteful, and inefficient. However–again indicating a libertarian, independent streak–most Coloradoans from 1988 to today consistently believe that their state is "on the right track."[1]

3.1.3 Voting in Colorado

Each County individually administers local, coordinated, primary, and general elections, under the supervision of the Colorado Secretary of State. This means that each county

[1]Colorado College Citizens Polls, taken from Cronin et al. (Cronin & Loevy, 2012)

individually handles the voters registered in that county. Unsurprisingly, the same eight most populous counties are also the counties with the majority of registered voters, as their registrants comprise 73% of total Colorado registered voters (as of November 2017). As table 3.2 shows, these eight counties have a registration rate between 60-80%, compared to a Colorado-wide rate of about 67%. Registration rates for all counties are also graphically depicted in figure 3.2. In terms of Party registration, Colorado as a whole leans democratic by a very narrow margin (figure 3.3).

Table 3.2: Colorado voter registration for largest counties

County	Total Registered Voters	County Registration Rate	% of Statewide Registrants
Adams	270,303	0.61	0.07
Arapahoe	410,546	0.72	0.11
Boulder	237,091	0.80	0.06
Denver	450,616	0.75	0.12
Douglas	237,659	0.83	0.06
El Paso	445,708	0.71	0.12
Jefferson	422,362	0.79	0.11
Larimer	250,626	0.84	0.06
Other	1,009,392	—	0.27
Colorado	3,734,303	0.67	1.00

In the past 25 years, there have been a series of key changes in the way Colorado administers elections, in relation to Vote By Mail and other reforms targeted and expanding the democratic franchise. In 1992, Colorado introduced no-excuse absentee voting, allowing voters to either physically pick up a mail ballot at a Vote Center or County Office, or have a ballot mailed to them prior to election day. In 2008, this reform was expanded to a permanent Vote-By-Mail system, which gave counties the option to allow voters to be permanently put on a list of addresses that received mail ballots prior to the election. The State also entered a transitional status to full mail elections, giving counties the option to make all coordinated local elections, general elections, and primary elections exclusively VBM. In 2013, the Colorado State Legislature passed HB13-1303: The Voter Access and Modernized Elections Act, which mandated that every voter currently registered receive a mail ballot for all future elections. The Act also expanded the use of Vote Centers instead of traditional polling places, instituted same-day voter registration, and revamped the way active and inactive voter status was designated on voter rolls–more on this in future sections (Hullinghorst & Pabon, 2013). These changes are summarized in Table 3.3.

Table 3.3: Key changes to Colorado elections policy

Year	Key Changes
1992	No Excuse Absentee Statewide Implementation

Year	Key Changes
2008	Permanent No-Excuse VBM Lists, Option of Full-VBM Elections
2013	Automatic Mail Ballot System Implemented Statewide, Established Vote Centers

3.1.4 Colorado as a Case for this Thesis

Colorado presents an interesting case for research on Vote By Mail exactly because it has gone through a long transitional process to reach its current elections system. It has steadily developed voting policy through a mixture of state mandates, county action, and outside policy motivations. Colorado's streak of independence and direct democracy is also very apparent in this shift in electoral practices, since they have been passing policies trying to expand participation for a very long time. It gives researchers access to approximately 22 years during which at least part of the state conducted elections by mail, making comparative, county- or individual-level case studies particularly alluring. Colorado's streak of independence and direct democracy is also very apparent in this shift in electoral practices, since they have been passing policies trying to expand participation for a very long time.

On a more general level, Colorado is interesting exactly because it is "typical" but with a wild streak. It is typical rocky mountain country, great plains country, and liberal urban city but all *in one state*. In is libertarian yet increasingly Democratic. It heavily relies on state funding for national parks, yet rebels against federal land use laws. Colorado overwhelmingly supports marijuana legalization, despite being a frontier state with traditional values. It is also a consistent purple state, with a Democratic Governor and House, but Republican Attorney General, Secretary of State and Senate. This means that Colorado is a combination of distinct national effects, but also local effects that make it significantly different from national trends as a whole. In this environment, predicting results of policy can be difficult, but extremely salient as multiple effects can be tested against each other.

3.2 The Data

This thesis relies on county and individual level models to draw conclusions on voting behaviors, and how they are affected by voting method. As such, the data I need will optimally contain the following:

- **County and individual level demographic characteristics**: race, gender, urban population
- **County and individual level voting data**: turnout, party registration, total registrants
- **Information on individual elections**: date, ballots cast, voting methods, county, election descriptions

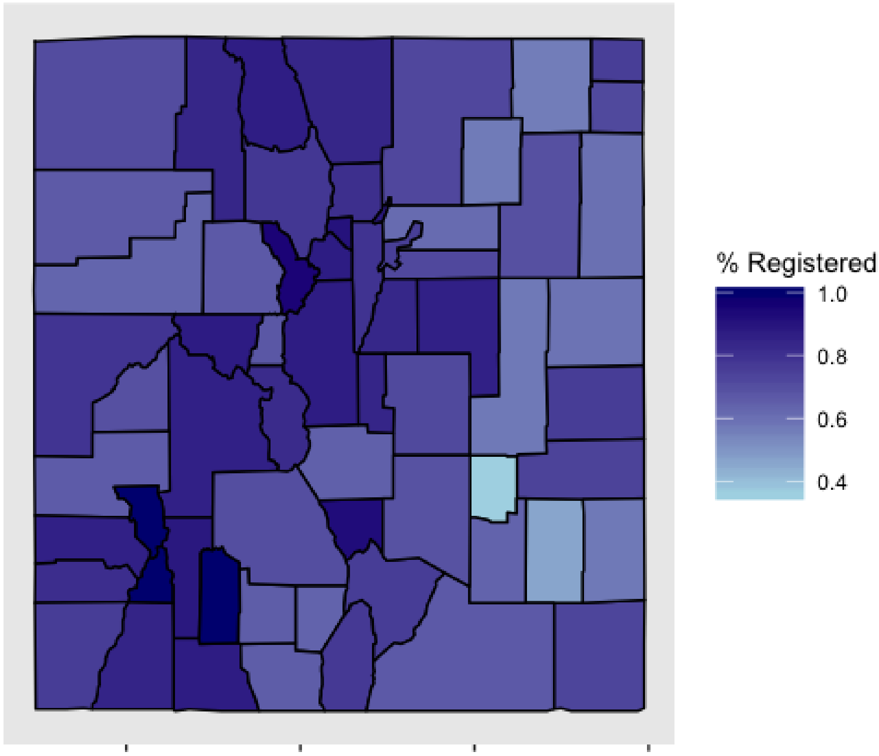

Figure 3.2: Registration rates per Colorado county

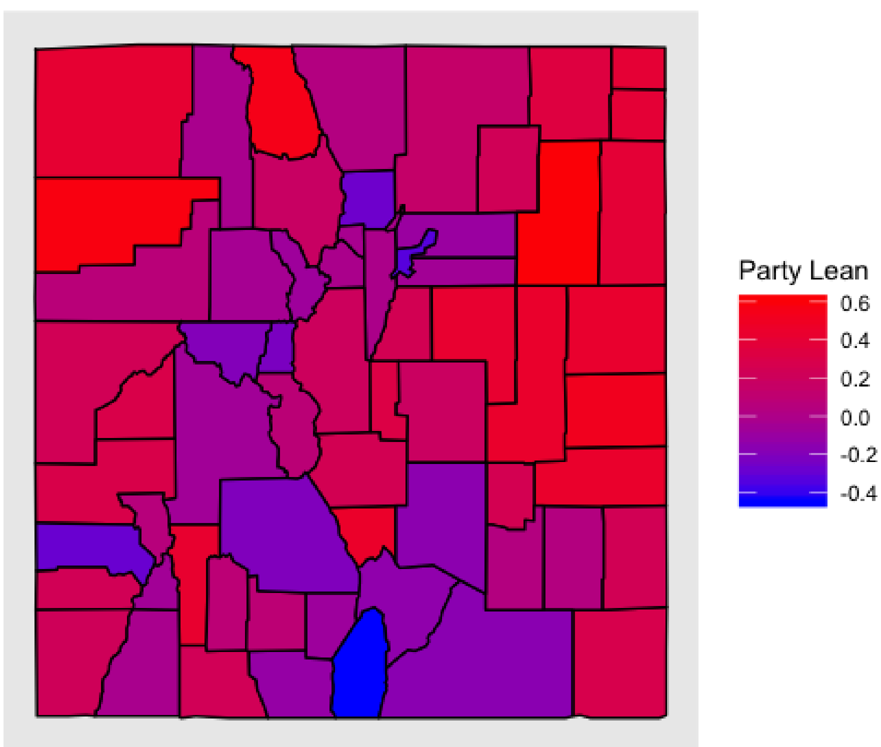

Figure 3.3: Democratic/Republican party lean per Colorado county

In the process of my research, I have acquired sufficient data to cover the second and third of these areas. I was unable to procure individual level data on demographic characteristics apart from gender, age, and party registration[2]. Reasonable conclusions can still be drawn from county or precinct aggregates.

3.2.1 Sources and first glance

I used two sources of data: Colorado voter records gathered by the Colorado Secretary of State's office, and demographic data from the 2010 US Census. In the process of procuring these data I was aided by a series of other researchers and professionals with experience in the field of elections administration. Andrew Menger, Postdoctoral Fellow at the Weidenbaum Center on the Economy, Government, and Public Policy at Washington University, was kind enough to give me access to data files for Colorado for the years of 2012-2016 that he had already collected for his research[3]. I directly obtained the 2017 data from the Colorado Secretary of State's office, with the help of Mr. Judd Choate, Director of Elections.

2010 US Census

The US Census is conducted country-wide every ten years, with the goal of procuring accurate data on the demographic characteristics of the population. The Census uses a combination of federal field workers conducting door-to-door canvassing and statistical methods for data aggregation. The 2010 Census allows public access to total population counts, characteristics on race, and rural/urban population counts for Colorado.

I use two datasets from the Census. For both, the unit of observation is one of the 64 counties of Colorado, and both include the same total population counts. One contains racial demographic characteristics and the other contain percentages of rural and urban populations in each county. The racial demographic dataset needed some wrangling work to extract a percentage of white residents in each county. Individuals were coded as "white" when they identified as exclusively white–this does not include mixed-race individuals reporting white ancestry.

Colorado Voter Files

As mandated by HAVA, Colorado maintains a statewide registry of all currently registered voters. This registry is typically under the purview of the Secretary of State. Voter Registration Files are constantly updated with new information on existing voters, new voters, or with the removal of inactive or otherwise ineligible voters. Therefore, this file will be different every time it is accessed or shared. Based on when this file is accessed, only a "snapshot" of the file can be obtained. Similarly

[2]Data this specific are commonly purchased from companies such as TargetSmart or Catalyst, obtained through record linkage, or inferred from characteristics such as last name (as accomplished with the `wru` *R* package).

[3]Doctor Menger's website with links to his research can be found at www.andrewmenger.com

with VRFs, a Voter History File is maintained and constantly updated by the state. This file is uniquely connected to its VRF: only voters showing up as registrants will have their histories included. I have both Voter Registration and History files for the years between 2012-2017, obtained with the help of Judd Choate and Andrew Menger.

In the Voter Registration files, the unit of observation is the individual voter, and all variables are initially coded as character strings. Each voter is assigned a unique voter ID, which serves as a point of reference between the two files. Broadly speaking, data in this file can be divided between three categories: first, personal identification information like address, ZIP code, or phone number; second, demographic information like age and gender; third, information pertinent to elections administration like congressional district, local elections for which the individual should receive a ballot, voter ID, and party registration. I will further elaborate on relevant variables in the wrangling section.

In the Voter History files, the unit of observation here is a single ballot cast, and all variables are initially coded as character strings. This means that for each voter registered–and so included in the VRF–the history file should contain an observation for each time they voted. This file includes two types of data: first, identifiers for the election like county, date, description, and type; second, identifiers for the individual vote including voter ID and voting method.

Samples of what a Voter Registration File (3.5) and Voter History File (3.6) can be found at the end of this chapter. For the purpose of privacy, I have randomly generated IDs and included random names in place of the original entries. I have also selected only specific fields to display, based on the data I use in this thesis.

3.2.2 Why Voter Registration Files?

Voter Registration and History files are suitable sources for my analysis because they contain all the data that is necessary for a first pass at testing my hypotheses: voting method, county, election level, active registration, registration dates, and a series of individual characteristics like party registration, age, or gender. These data, and the demographic data in particular, are also in the most basic unit of observation: the ballot level[4]. This means that I do not need to establish any process to infer individual characteristics from population-wide statistics.

In statistical science, sampling is the process by which individual units are selected from a population. The sample selected should be *representative* of the whole so that it can be used to infer characteristics of the general population. Despite a vast array of techniques to ensure that sampling is representative, there is always room for error. Voter Registration Files are an excellent source of data because they do not involve any sampling whatsoever; they include *all* currently registered voters and voter histories. This characteristic helps cut down on data-related errors and on the complexity of data extraction.

An additional characteristic of these data is that they are concentrated and

[4] A good heuristic for what "ballot" level means is a specific individual at the time when they cast a specific ballot. Between ballots all characteristics may change: age, gender, party registration etc., which is why the "ballot" level is distinct from the individual level.

relatively uniform. Data transfer errors still exist, and the wrangling process is never entirely straightforward. Still, the data are almost completely uniform in how variables are encoded. Over thirty five million observations are included in my final, cumulative voter history dataset, and all of them have, for example, the same types of entry for party registration (REP, DEM, UAF etc.). Admittedly, this may just be a characteristic of the Colorado files, since they are the only ones I used for my research.

A last benefit of using Voter Registration Files comes from the replicability they allow for. These files are generally public, with access to them including only data transfer and administrative costs. This makes peer-review and replication less complicated than if, say, I was using private survey data. It also allows for expansion that goes beyond the State of Colorado; my code can be adapted to fit different data, making future comparative studies more likely and less time-consuming than starting from scratch.

3.3 Wrangling the Data

The process of “wrangling” refers to manipulating the data into a form that can then be used for graphing, exploratory data analysis, modelling, or presentation. In this case, wrangling also included aggregating data across multiple sources and datasets. For this purpose, I made heavy use of the `tidyverse` *R* package, and in particular the `dplyr` package. In this section I will go through some of the key problems encountered during the wrangling of these data, and then discuss the final form each variable takes.

3.3.1 Initial Problems with the 2017 Voter File and Solution

In my initial research I intended to only use the 2017 snapshots of the Colorado Registration and History file. The major issue I encountered, which merits discussion in its own section, comes from the fact that the records I have access to are “snapshots”. What this means, is that for each person in each year of voter registration files, I have their corresponding history files for all ballots they have cast in Colorado, but not their own history of registration and migration. If, say, a voter moved from Boulder County to Summit County, I would have their votes in Boulder County show up in the voter history file, but them being registered in Summit. If you recall the turnout calculations specified earlier on, this implies an overestimation when looking back at elections that happened some time before the date of the “snapshot”. Additionally, “snapshots” of current voter files do not reflect voters dropping off the rolls for whatever reason (death, moving out of the state, long term inactivity, non-confirmable personal data etc.)

After going through turnout calculations with the 2017 files, a significant majority of counties appeared to have turnout exceeding 80%, particularly for years between 2000 and 2012. This was, to put it mildly, concerning. With the aforementioned help, I was given access to similar “snapshots” for each year between 2012-2016. After similar calculations, I returned figure 3.4 for the eight most populous counties as described above, including different shapes for election type, colors for county, and a vertical

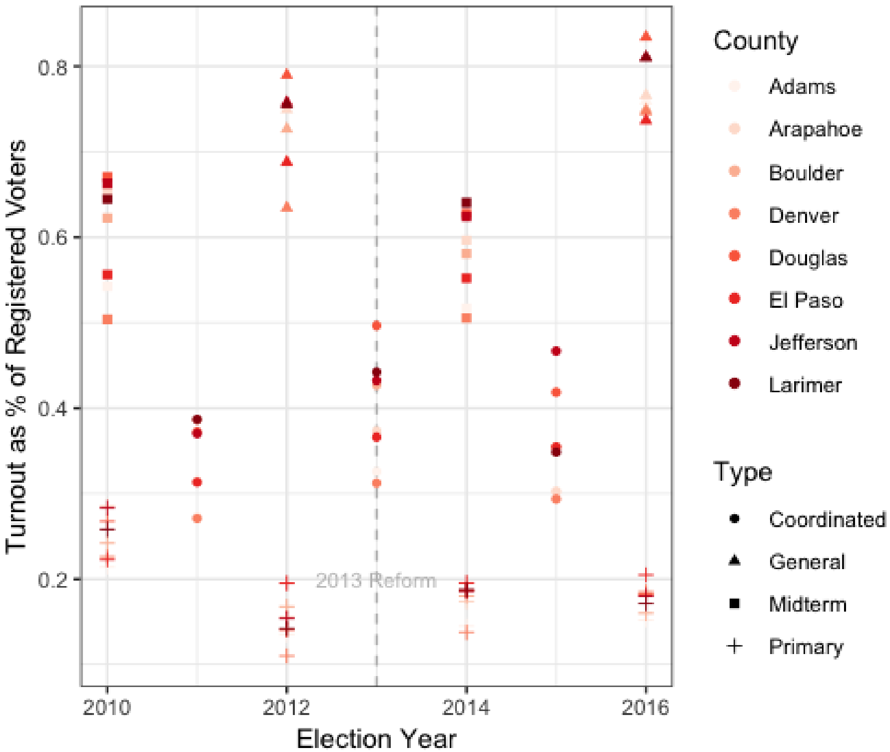

Figure 3.4: Turnout plot for eight largest Colorado counties, 2012-2016

line at 2013 to signify the latest major change in how Colorado administers elections.

To also further illustrate the in-county migration and dropped voter problem, I created a graph that includes logged total counts of registered voters calculated using the 2017 and the 2012-2016 files. The plot also includes a line at $y = x$. If in-Colorado migration and dropped voters are not an issue, most points on this graph should be at this line.

Two things should be clear from figure 3.5. First, there is significant deviation between the counts using just the 2017 file and all files across years. Specifically, the 2017 count consistently underestimates the total amount of registered voters–this is shown by the red linear model smoothing line. This consistent difference means that it is close to impossible to generate safe conclusions on my hypotheses using only the 2017 files and the methods I have outlined in Chapter 2. Second, counts get more accurate the closer to 2017 we get. This should be even more apparent in figure 3.6, which limits the scale to only some high registration counties, and adds a shape indicator for county.

Here the structure of the data becomes clear: for each county, there are a series of almost vertically distributed points, which get closer to the $y = x$ line the closer the counts get to 2017. Through this series of tests, it became clear that using multiple years of data was necessary in order to conduct an accurate test of my hypotheses. My selection was later vindicated, when looking at comparisons between reported

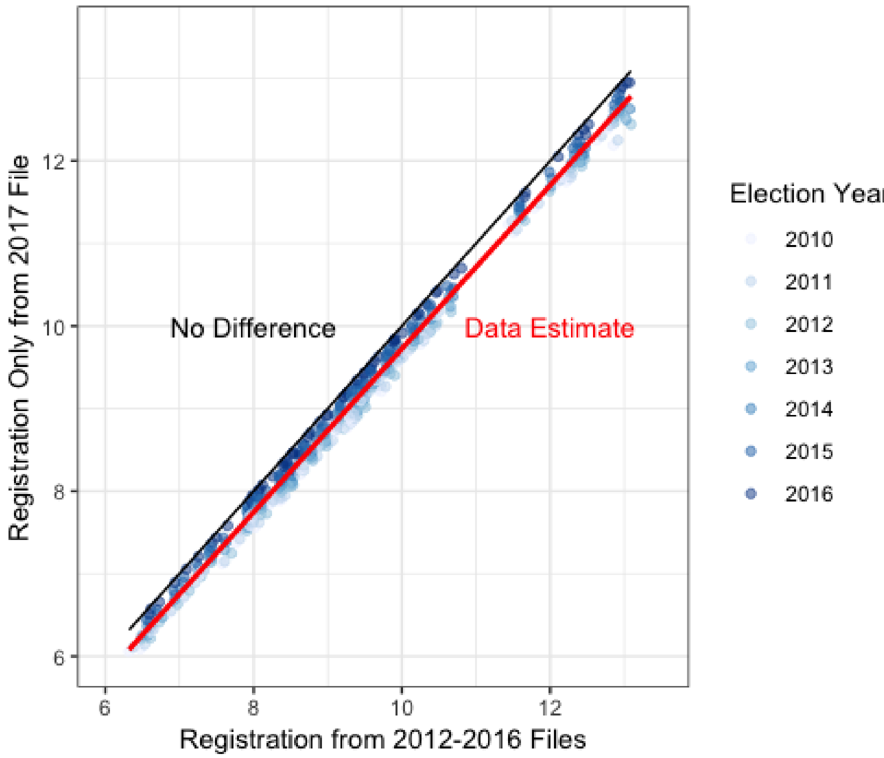

Figure 3.5: Comparison of registration count methods

rates of turnout[5] and turnout calculated through my dataset for the 2014 midterm election (see fig. 3.7).

The differences are insignificant. They exist because of “noise” added on because of errors in the data, misreporting, registration records redacted due to privacy concerns, voters dropped before the “snapshot” occurred, and other similar factors.

3.3.2 Other Wrangling Issues

Wrangling the data was the majority of the work that went into this thesis. As will become clear in this section, apart from accurately processing, diagnosing, and merging the data, the process of wrangling includes several non-trivial decisions about how to treat missing values and variable specification. Including a full account would probably read like the world’s most cliche crime novel: a series of elusive final datasets, a plucky yet occasionally naive young detective, two wisened mentors, clues, dead ends, frustration, compromise, and. . . spreadsheets. I will spare the reader the whole story, but I will include a non-comprehensive list of some of the difficulties associated with wrangling voter files, as it was a crucial part of the learning process I underwent while doing my research.

Missing Values: The decision on how to deal with missing values–or NAs–in a dataset is a lot more important than it may initially seem. A first, intuitive reaction

[5]Turnout is calculated over all registered voters

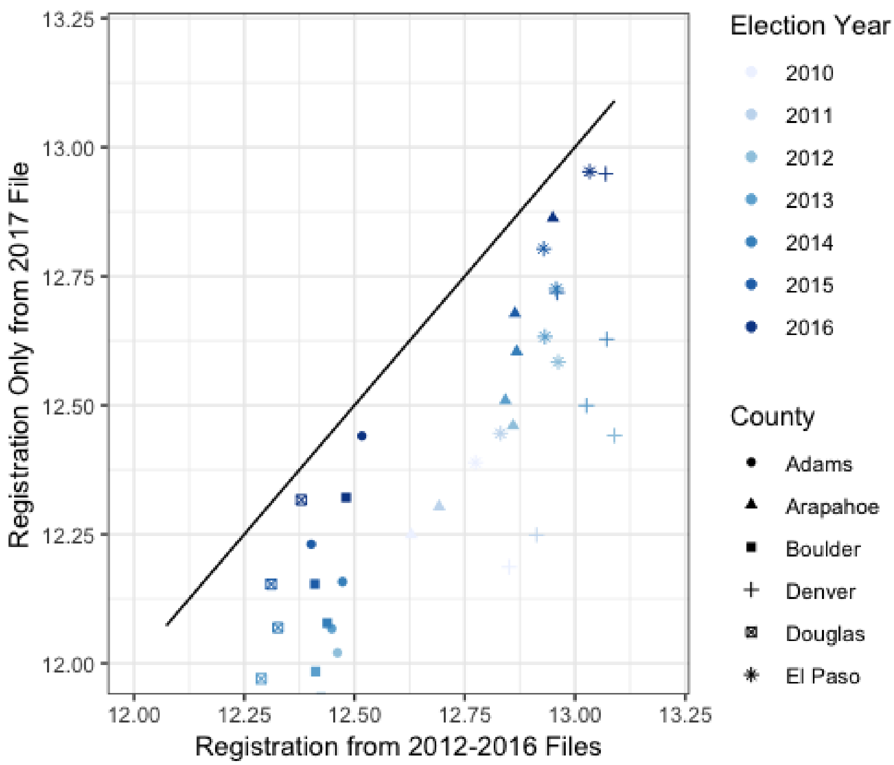

Figure 3.6: Comparison of registration count methods only for a few counties, 2012-2016

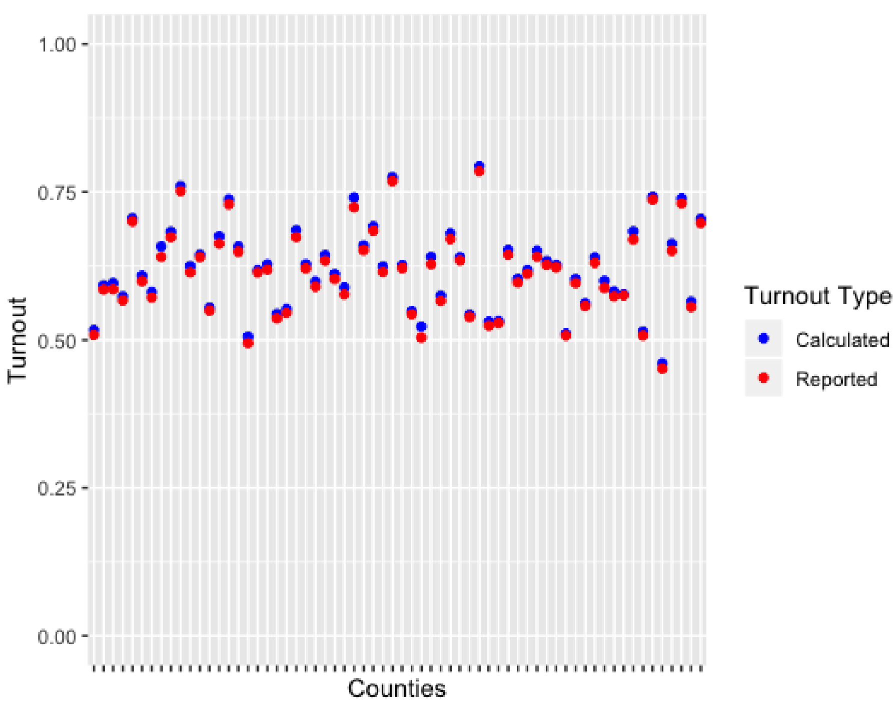

Figure 3.7: Comparison of reported and calculated turnout for 2014 midterms across county

might be to just disregard them; however this works under the assumption that there is no structure inherent to why these data are missing. To give just two examples, in the data I have collected the PARTY value for the 2015 voter registration file is missing. If I excluded all observations with missing PARTY values, I would be excluding a fifth of my data. Missing values were also present in the VOTING_METHOD variable of the voter history files. While this may have seemed troubling, after closer examination it was revealed that the vast majority of such missing values was concentrated in Jefferson County, and in elections prior to 2002. Therefore, these observations could be ignored, since they played no role in my final dataset. The conclusion should be that choices made on exclusion, inclusion, or estimation of missing data are very important, and should be taken with much care and consideration for the underlying structure of the data.

Data Input Errors: Is "QATAR" a political party in Colorado? State records say not. However, "QATAR" did show up as a value for the Party variable for my 2016 voter registration file snapshot. This may occur for a number of reasons, the most likely of which is the introduction of errors when transferring these data. The data have been read and written by multiple operating systems (iOS and Windows) and programming platforms (STATA, R); they have also been uploaded, downloaded, and written unto CDs, as well as transferred between County and Colorado Secretary of State's Office when they were created. Characters that would be normally read into one platform as line or value delimiters may have been misinterpreted by another platform, with no operator error involved. In my analysis I treated all values that seemed more likely than not to be errors as NAs. There were not many of these–less than .001% of my data–but they were a hassle to find, analyze, and then recode into some useful value.

Data Size: Nothing to write home about here, just an observation that multiple voter registration files can be *huge*, which puts considerable strain on a computer's processing power. This means that wrangling has to comprise of a series of careful, deliberate moves. Brute force should be discouraged, as a dead end means several hours of melodic computer fan panic.

Joining, Merging, Spreading, and the Multiplicity of Levels: For the data to end up in any functional shape, it eventually becomes necessary to start joining datasets. Thankfully, a clear division of modelling tasks between county and individual level models means that joining on COUNTY or VOTER_ID is ideal, and fairly straightforward. As will become clear in later sections, I also had to consider the variety of different units of observation, specifically: county, individual, ballot, election, county-by-election.

3.3.3 Final Variable Specifications

After the conclusion of the wrangling process, the resulting datasets included a series of discrete and continuous variables. I will briefly outline them here, along with their range and values.

- VOTER_ID: Discrete variable, unique value given to each individual voter.

Useful for merging.
- COUNTY: Discrete variable, the 64 counties of Colorado.
- REGISTRATION_DATE: Discrete variable, date of registration for each registrant. Useful to get total registrants on election day.
- TURNOUT: Continuous variable, in the range [0,1]. The response variable for my county-level models.
- ELECTION_TYPE: Discrete variable, the four types of elections: Primary, Coordinated, Midterm, Presidential.
- ELECTION_DATE: Discrete variable, self-explanatory.
- VBM_PCT: Continuous variable, in the range [0,1]. This is the focus of my analysis, as it counts the percentage of total ballots that were mail ballots.
- PCT_WHITE: Continuous variable, in the range [0,1]. Percentage of white residents per county.
- PCT_URBAN: Continuous variable, in the range [0,1]. Percentage of urban residents per county.
- PARTY: Discrete variable. For each voter, the party they are registered with. Can be: Republican, Democrat, Other, or Unaffiliated.
- GENDER: Discrete binary variable, Male or Female.
- AGE: The age of the individual registrant.
- VOTING_METHOD: The method used by an individual voter to cast their ballot. Is coded as either VBM or In Person, according to Table 3.4.

Table 3.4: Voting methods coding table

Voting Method	Description of Method	Designation
Absentee Carry	Voters who carried an absentee ballot with them from an early voting location or county office	VBM
Absentee Mail	Voters who were sent an absentee ballot, and mailed it in	VBM
Early Voting	Voters who physically went to an Early Voting location and voted	In Person
In Person	Voters who physically went to a polling place and voted on paper	In Person
Mail Ballot	Vote By Mail	VBM
Polling Place	Traditional polling place voting, discontinued in 2013	In Person
Vote Center	Voters who cast their ballots at Vote Centers	In Person

	VOTER_ID	COUNTY	LAST_NAME	RESIDENTIAL_CITY	VOTER_STATUS	PARTY	GENDER
1	318090	Denver	Jefferies	DENVER	Active	DEM	Male
2	1224557	La Plata	Simmons	BAYFIELD	Inactive	REP	Female
3	833164	Arapahoe	Jenkins	CENTENNIAL	Active	UAF	Female
4	654070	Boulder	Giroux	LAFAYETTE	Active	UAF	Female
5	506568	Alamosa	Provorov	ALAMOSA	Inactive	UAF	Male
6	1467732	Mesa	Hoskins	GRAND JUNCTION	Active	DEM	Female
7	1061624	Boulder	Nola	BOULDER	Active	DEM	Male
8	1896894	Denver	Arrieta	DENVER	Active	DEM	Male
9	649632	Arapahoe	Butler	LITTLETON	Inactive	DEM	Female
10	1144545	Montrose	Clement	OLATHE	Active	DEM	Male

Table 3.5: Sample of a Voter Registration File

	VOTER_ID	ELECTION_TYPE	ELECTION_DATE	VOTING_METHOD	PARTY	COUNTY_NAME
1	318090	General	11/03/1998	Absentee Mail	REP	Gunnison
2	1224557	Coordinated	11/05/2013	Absentee Mail	DEM	El Paso
3	833164	General	11/07/2000	Absentee Mail	UAF	Arapahoe
4	654070	General	11/08/2016	Mail Ballot	CDP	Arapahoe
5	506568	Coordinated	11/02/1999	Absentee Mail	REP	Arapahoe
6	1467732	General	11/08/2016	Mail Ballot	REP	Pueblo
7	1061624	Coordinated	11/07/1995	Absentee Carry	REP	Larimer
8	1896894	General	11/03/1998	Polling Place	DEM	Larimer
9	649632	General	11/03/1992	Polling Place	UAF	Arapahoe
10	1144545	Coordinated	11/03/2015	Mail Ballot	DEM	Jefferson

Table 3.6: Sample of a Voter History File

Chapter 4

Model Specification and Results

The goal of this chapter is to apply inferential statistical modeling to the data. This task can be divided into three steps: specifying the models mathematically, fitting the models, and interpreting the results[1]. However, before jumping into this process, it is worth going into some key problems with the models produced. As a consequence of these issues, some of the models are not estimated to the standards of convergence that are commonly set.

4.1 Modelling Issues

4.1.1 Lack of variability

To put it very simply, it's not enough to have hundreds of thousands of observations if they are all almost identical to each other. If, for example, my data included a thousand people in Jefferson county, and 63 in all other counties of Colorado combined (one in each remaining county), then I would not be able to leverage my data to draw conclusions on county-level effects.

As previously stated, the data available includes registration files going back to 2012. From these files, I have extracted data for elections going back to 2010.[2] In order to make inferences on VBM and turnout effects it is necessary to have extensive and varied data. Specifically, it is necessary to have data that include a large enough sample of the voters in Colorado, with a substantial portion of them using different voting methods, from different counties, or in different election years etc.

The data are extensive (over 35 million observations at the individual level) but substantially lack variance in voting method. Put simply, the vast majority of registrants in Colorado from 2010 onward either did not vote at all, or voted by mail. If you recall the changes in Colorado election law, in 2008 counties were allowed to

[1]In theory there is also the step of translating the models from their mathematical specification into some sort of algorithmic process that produces estimates of coefficients and error terms. This process is arduous and long, so it is not included in this chapter. Appendix A deals with some of the techniques involved with model estimation

[2]See section 3.3.1; the extracted data is limited to this time period to avoid accuracy issues with migration and removal of inactive/unavailable voters.

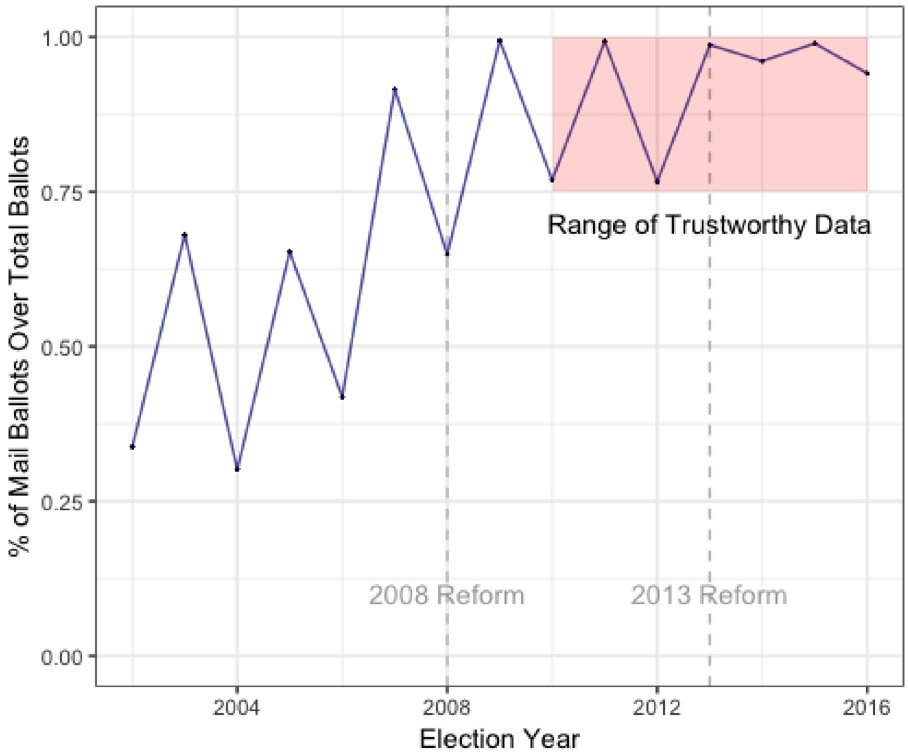

Figure 4.1: Percentage of mail ballots over total ballots by year

conduct all mail elections, and no-excuse permanent absentee voting was implemented state-wide; then in 2013 Colorado transitioned to full VBM for all elections. This means that few people were still using traditional polling places or vote centers to cast their ballots. Figure 4.1 shows how after 2013, and even before that in 2011 over 95% of ballots cast were mail ballots. Only in the general elections of 2010 and 2012 is there some variance, but mail ballots account for well over two thirds of total votes.

This issue is not completely fatal for county level models. There is still variance between counties that have 100% mail ballots and those that are around the 75-85% margin. For individual level models, where I am estimating voting probability, VBM will be an almost perfect predictor for voting, and therefore will not present me with any substantial analytical result on how it affects voting probability. There are some ways to compensate for this issue, which I outline; due to time or data constraints, not all of these will be implemented in this thesis:

- *More (Diverse) Data*: It would be very useful to get snapshot data of Colorado voter files from 2004 to today, because it would allow for an extensive study on how the 2008 and 2013 election laws re-shaped voting decisions in the state. It would be useful, but also expensive and very time consuming, involving several purchases of data from the Secretary of State of Colorado. Voter registration files also tend to get messier the further back one goes, which means that the process of cleaning up the data would get substantially harder. It would also require more processing power to handle more observations. My research here does not do this, as the scope of a senior thesis is limited; such an overarching study would probably be conducted by multiple researchers with several assistants. I do however present several replicable materials for such a study, through the creation of an *R* package I include on my GitHub page along with the final results of this thesis. This thesis does not go that far, but it may help similar studies in the future.

- *Localized, Natural Experiment Studies*: A natural experiment is when, due to policy changes and circumstances, a "control" and "treatment" group for a policy are created in the same approximate geographical area. This happens when, for example, only some of the counties in a state enact a specific change. Several such studies exist already, with some even tackling VBM in Colorado (Keele & Titiunik, 2017), or how turnout rates are affected by new, restrictive registration laws (Burden & Neiheisel, 2013). This method allows for more accuracy in both the individual and county level models, and through the existence of a treatment and control group guarantees the variability that I currently lack.

- *Synthetic Control Group*: The synthetic control group method is a way of creating a control group when no such group exists. It involves gathering a set of characteristics from the treatment group members and then using statistical methods to combine them into making the appropriate control (McClelland & Gault, 2017). I will not go into the particulars of this method (the sources cited should provide a decent introduction), but this method has been successful in assessing policy effects such as anti-smoking laws (Barr et al., 2012), or even motor voter laws in Oregon(Gronke, McGhee, Romero, & Griffin, 2017).

4.1.2 Computational Considerations

The process of computing estimates for model coefficients can often be very computationally intensive. This issue is particularly present during estimation of individual level models, which alongside complex hierarchical structure also draw on a huge dataset of 35 million observations. This computationally intensive procedure requires more processing power than I currently have available. For now, I compensated for this problem by using stratified sampling to sample a subset of my observations[3].

The form of stratified sampling I am using is very simple; based on county, mail vote, and electoral participation, I use `dplyr` in *R* to draw a sample that contains equal proportions of every combination of values of these variables to those in the original dataset. If, for example, the original dataset had 2% of entries being voters from Jefferson county that participated using a mail ballot, the sampled dataset would have a proportion that is approximately equal to 2% (Chihara & Hesterberg, 2011). In this way I draw a sample of around 370,000 observations from my initial ballot dataset, on which I run all my individual models. After checking the variable ratios in sampled and population datasets, I found that the differences between ratios had a

[3]A long term solution to this issue could be the use of a more powerful local RStudio server, or Amazon Web Services (AWS).

mean and standard deviation of less than a hundredth of a percentile. Therefore this sampled dataset serves as a decent approximation of my population.

4.2 Variable Specification

I will not go through each individual variable in this section, but will briefly describe my notation for the following models. I will include more comments whenever they seem necessary under each model. In this thesis I include predictors on a series of variables that can be divided into five categories based on unit of observation: county, election, individual, local result, and ballot. The last two are functions of other units: local result units are equal to the product of elections and counties, while ballot units are equal to the number of unique individuals multiplied by the number of elections each of them was registered in. For notation, I follow this set of rules:

1. If the variable is a response, it is coded y.
2. If the variable is a predictor, it is coded x
3. The variable's superscript will provide information on what it represents, else it will be explained.
4. All variables represent a single value (scalar) of that variable unless stated otherwise.
5. Unit of observation will also be specified in subscript, according to the indices described in Table 4.1. These indices are also used in sum notation.
6. All Greek characters represent coefficients to be calculated.
7. By $k[j]$ I represent the k-th value corresponding to the j-th observation. In this case, this would be the county that an individual is registered in.
8. Note that for Local Result level variables, I use k, l as an index. This is because there are very few variables at this level, it is a direct Cartesian product of two other units, and this notation avoids confusion with even more index types.

Table 4.1: Variable indices per unit of observation

Units	Index
Ballot	i
Individual	j
County	k
Election	l
General Index	v

4.3 County Level Models

4.3.1 Specifications

In this section I will go through a step-by step creation of models at the county level. County level models use a series of variables at the election, county, and local result levels. The response variable is always turnout in one county after a particular election. With no other information, this model could be thought of as an assignment of voting tendencies across counties; each county independent of election has a unique range of turnout results. In this way it is possible to build a naive, baseline model of turnout as follows:

$$Y_{k,l}^{turnout} = \beta_0 + (\sum_{k=1}^{64} \beta_k x_k^{county}) + \epsilon,\ \epsilon \sim N(0, \sigma^2) \quad \text{(Model 1)}$$

where x_k^{county} is a series of 64 dummy variables for each county of Colorado. Here differences between elections come from normally distributed error terms, rather than predictors. I name this **Model 1**, and it does not fit the data particularly well. First off, this model includes the assumption that counties are independent of one another, which is probably false; just consider that these counties are areas of the same state, in the same country, with populations moving between them at regular intervals, and many of them covering the same metropolitan area or congressional district. Additionally, the model matrix here is rank deficient: there are two county coefficients that are perfect linear combinations of other coefficients. This means they will be dropped by *R* when the model is called in the `lm()` function.

A way to fix both these issues is to use a multilevel model with mixed effects for county. By constraining coefficients at the county level to a set distribution, this model does away with the assumption of independence. The other county level predictors help to explain some of the unexplained group level variation, which reduces the standard deviation of county coefficients and helps provide more exact estimates (Gelman & Hill, 2006). I call this **Model 2**, which can be written as:

$$Y_{k,l}^{turnout} = a_k + \beta_1 x_k^{\%white} + \beta_2 x_k^{\%urban} + \epsilon, \quad \text{(Model 2)}$$

$$a_k \sim N(\gamma_0, \sigma_\alpha^2)$$

$$\epsilon \sim N(0, \sigma^2)$$

This model provides a more reasonable set of estimates for each county, but still fails to provide any information as to secular trends, time-specific effects, election type effects, or mail voting. I will amend this by adding a set of variables at the election and local result levels: election type and an interaction term between election type and mail voting. This variable should reflect whether turnout effects of mail voting are more pronounced in a specific type of election. I call this **Model 3** and it can be specified as follows:

$$Y_{k,l}^{turnout} = a_k + \beta_1 x_k^{\%white} + \beta_2 x_k^{\%urban} + \overbrace{(\sum_{v=3}^{6} \beta_v x_v^{electiontype} x_{k,l}^{\%mail\ vote})}^{\text{Interaction Effect with Type}} + \overbrace{(\sum_{v=7}^{10} \beta_v x_v^{electiontype})}^{\text{Main Effect of Election Type}} + \epsilon, \quad \text{(Model 3)}$$

$$a_k \sim N(\gamma_0, \sigma_\alpha^2)$$
$$\epsilon \sim N(0, \sigma^2)$$

where $x_v^{electiontype}$ is a series of four dummy variables for each type of election (General, Primary, Coordinated, Midterm). This model reflects nearly all the information I have available, apart from election date. For the incorporation of election dates there are two possible alternatives. First, I can simply add a dummy variable for each year. This would assume independence between each year, as it would specify different, independent "slopes" for the seven years I have data for–this is like calculating seven different models, one for each year. This is not particularly elegant as a solution nor does it reflect the fact that years actually are interconnected; of course there can be massive shifts in national or regional political climates, but those shifts happened *from some baseline*, which is reflected in previous years.

These elections can be thought of as systems for which prior condition affects future outcomes, and therefore time cannot be modeled as a series of independent effects. The solution here is adding a spline function for time, using a general additive multilevel model. I use a natural cubic spline function for the purposes of this model (S. N. Wood, 2006). More on the subject of splines can be found in the Wood (2006) textbook. The model, which I call **Model 4** can be written as follows:

$$Y_{k,l}^{turnout} = a_k + \beta_1 x_k^{\%white} + \beta_2 x_k^{\%urban} + \overbrace{(\sum_{v=3}^{6} \beta_v x_v^{electiontype} x_{k,l}^{\%mail\ vote})}^{\text{Interaction Effect with Type}} + \overbrace{(\sum_{v=7}^{10} \beta_v x_v^{electiontype})}^{\text{Main Effect of Election Type}} + s(x_l^{year}) + \epsilon, \quad \text{(Model 4)}$$

$$a_k \sim N(\gamma_0, \sigma_\alpha^2)$$
$$\epsilon \sim N(0, \sigma^2)$$

where $s()$ is a natural cubic regression spline function with seven knots–equal to the number of years.[4] A summary of these four models is provided in the following table:

[4]I used the `gam.check()` function that is present in the `mgcv` *R* package, whose call determined

Table 4.2: County level model descriptions

Model No	Model Description
Model 1	Baseline model with only county specific effects
Model 2	Multilevel model; added county level predictors
Model 3	Multilevel model; added VBM, interaction terms, and election fixed effects
Model 4	Multilevel General Additive model; added spline function for election year

4.3.2 Results

The table in this section presents coefficients and standard errors for all four county level models. This table does not include any metrics for county, either mixed or fixed effects. I have chosen to omit these because they firstly are not very relevant to my hypotheses, and secondly because they are very extensive (64 coefficients for each of the four models). I have also not included any metric for time, here measured in years and used only in the fourth model. Both the mixed effects for county and the measure for time should be considered as controls: the first controls for county-specific trends while still restricting these to allow for non-independence, and the second makes sure that my results are indicative of a secular trend, accounting for any shifts along time.

In terms of goodness-of-fit, I use 5-fold cross-validated Mean Squared Error (MSE) for all of the models. There is a significant drop-off in MSE between Models 1, 2 and Models 3, 4 of around .35, which shows that the variables introduced in the later models substantially increase how well the models explain variability in the data. There is also a small increase of CV MSE between models 3 and 4, but the numbers are very comparable[5].

Table 4.3: Estimated county level coefficients (**Significant at the .01 level *at the .05 level)

Variables	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.369	0.492**	0.455	0.470**
	(0.60)	(0.045)	(0.078)**	(0.072)
Pct_white		0.034	0.033	0.031
		(0.053)	(0.050)	(0.050)
Pct_urban		-0.118**	-0.117**	-0.119**
		(0.022)	(0.021)	(0.021)

that the number of knots here may be too low. However, given the data available to me, I was limited to the inclusion of seven years and as such cannot increase the number of knots any further. Setting the number of knots to seven also gave the lowest CV MSE.

[5]If I was trying to make a model for predictive purposes I would probably choose Model 3; however, there is still value in comparing Models 3 and 4, even if the later doesn't fit the data better than 3. The difference in coefficient values, after controlling for time, is a particularly interesting result.

Variables	Model 1	Model 2	Model 3	Model 4
typeGeneral			0.190**	0.254**
			(0.070)	(0.065)
typeMidterm			0.252**	0.070
			(0.068)	(0.063)
typePrimary			-0.071	-0.170**
			(0.069)	(0.062)
typeCoordinated*VBM			-0.001	0.002
			(0.067)	(0.058)
typeGeneral*VBM			0.151*	0.087*
			(0.073)	(0.037)
typeMidterm*VBM			-0.058	0.109*
			(0.026)	(0.030)
typePrimary*VBM			-0.089	-0.003
			(0.028)	(0.027)
CV MSE	0.041	0.040	0.004	0.006
Obs	704	704	704	704
Groups	64	64	64	64

Given that, the first observable result is that the percentage of white population and the percentage of urban population are fairly stable indicators of slightly higher and lower levels of turnout respectively, although only urban population reaches statistical significance at the .05 level. The lack of variability between models is not surprising; these represent a county-level, time-independent demographic statistic, and there would be no reason to assume that part of their effect would be subsumed by other variables in Models 3 and 4.

Moving on to election type, the coefficients for the different election types should be read as differences from the "baseline" that is typeCoordinated. First surprising result here is that the coefficient for general presidential elections is substantially lower than that of midterms. Rather, this would be surprising if we did not notice the interaction terms with VBM, which indicate that after allowing for VBM effects, presidential elections do actually have higher turnout in my model than midterms do[6]. Other than this, coefficients in Model 3 and Model 4 make sense, in the assumed ordering of turnout in such elections: presidential, then midterm, then coordinated and primary.

Next, taking election type and all interaction terms into consideration, let's examine what happens when the spline function for time is introduced between Models 3 and 4. Most coefficients shift dramatically, with the exception of the interaction between coordinated elections and VBM. This dramatic shift–between 5 and 15(!) percentage points–indicates that several of the effects that the third model estimated are actually time-specific trends, and that there is a significant difference if we account for them. In the fourth model, the coefficients for election type on their own are still indicative

[6]Remember here that due to Figure 4.1 most counties will have a proportion of mail ballots close to .9

of a common assumption for turnout in such elections[7]. As for interaction terms with VBM, the effect of VBM on primary election turnout is almost wiped out entirely, the interaction with general election turnout is depleted but still present at around 8%, and coordinated election VBM effects remain statistically insignificant. Interestingly, the effect of VBM on midterm turnout switches sign from a negative effect of 5% to a positive effect of around 11%.

Taking my hypotheses one by one, these models present evidence in favor of H1. Mail voting does seem to affect turnout in a way consistent across time–see the coefficients for VBM effects on general elections–but this effect is not particularly more strong than the percentage of urban population in each county. Conversely, my second and third hypotheses can be convincingly rejected at the county level. After controlling for time, the effect that VBM has on coordinated or primary elections is not statistically significant, compared to significant, consistent effects on midterm and general elections. The one point in favor of H3 here is that the effect of VBM on midterm elections is slightly higher–about 2%–than the effect on presidential elections in model 4. However, this difference is not enough to rule in favor of H3; if this difference was caused by the lack of presence of national effects, it would be more pronounced in primary and coordinated elections as well.

4.4 Individual Level Models

4.4.1 Specifications

For the rest of this section, assume the following:

$$y_i \sim \text{Bernoulli}(p)$$

Where $y_i \in \{0, 1\}$ is the probability that the i-th ballot was completed. The goal of such an individual level model is to estimate p as a function of variables measured at four different level of observation:

1. Ballot
2. Individual
3. County
4. Election

For this section models are built based on the following concept: assume that an election worker is trying to asses whether a ballot in their hands is completed or not, without opening the envelope it is included in. The possible outcomes are either that the person it corresponds to decided not to complete it (an outcome that includes the individual not going to a polling place at all), or that they decided to fill it in and vote. Assume that the ballot has some information written on the cover, such as the

[7] Also see Figure 3.4

county it is from, or the election date, or whether it is a mail ballot. The models in this section are built for different combinations of such information.

As a preliminary baseline model I would predict the probability that an individual voted in a particular election would be equal to turnout, as calculated through all other ballots. Therefore:

$$\hat{\mathbb{P}}(y_i = 1) = \frac{\#\text{votes cast}}{\#\text{ballots}}$$

4.4.2 Estimation with only one type of data

County Level

Given that the ballot I am assessing has county of origin written on it, there are two ways to predict $\mathbb{P}(y_i = 1)$. First, assume that each different county has a different, independent $\mathbb{P}(y_i = 1)$, then:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}\left(\sum_{k=1}^{64} x_{k[i]}^{county} \beta_k\right)$$

Where k counts over the 64 counties of Colorado, and x_k is an indicator variable for each county. Without assuming independence between counties, I could also fit a mixed effects model. This is a reasonable step since these counties are in the same state and country, and also often share borders.

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(a_{k[i]}), \qquad \text{(Model 1)}$$

$$a_k \sim \text{N}(\gamma_0, \sigma_\alpha^2)$$

Where $\alpha_{k[i]}$ varies by county, constrained by its standard deviation and γ_0, an intercept coefficient. I name this **Model 1**.

Assume that along with the one ballot, I was given a short list of $n^{\text{county vars}}$ other county-level variables, be they discrete, continuous, or indicators. With this additional information, the models can be updated to the following form:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}\left(\sum_{k=1}^{64} x_{k[i]}^{county} \beta_k + \sum_{v=1}^{n^{\text{county vars}}} x_{k[i],v} \beta_{v+64}\right)$$

Where $x_{k[i]}$ is the k-th value of the v-th variable. If, as before, I do not assume independence, the model can be written as:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(a_{k[i]}), \qquad \text{(Model 2)}$$

$$a_k \sim \text{N}\left(\gamma_0 + \sum_{v=1}^{n^{\text{county vars}}} x_{k[i],v} \gamma_v, \sigma_\alpha^2\right)$$

In the case of my specific data, for the time being I have county-level data for white population and urban population, so $n^{\text{county vars}} = 2$. Both of these variables are ratios over the total county population, taking values in $[0, 1]$. I name this **Model 2**.

Individual Level

Assuming that I know the voter ID of the individual that cast their ballot, I can treat this piece of information in about the same way that I did for county as described above. This means that the following is mostly an exercise in maintaining consistent notation. For these purposes, let n^{ID} be the number of total unique voter IDs, or individuals, that I have data on, and j an index that sums over all individuals. Also let x_j^{ID} be an indicator variable for each individual. Then:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\sum_{j=1}^{n^{ID}} x_{j[i]}^{ID} \beta_j)$$

And the second model, not assuming independence, would be:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\delta_{j[i]}),$$
$$\delta_j \sim \text{N}(\zeta_0, \sigma_\delta^2)$$

Again, in a similar way to county level data, there are variables at an individual level, thus making it relatively easy to build further models. Let's say now that along with the one ballot, I was given a short list of $n^{\text{indiv vars}}$ other individual-level variables, be they discrete, continuous, or indicators. The two models would then look like:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\sum_{j=1}^{n^{ID}} x_{j[i]}^{ID} \beta_j + \sum_{v=1}^{n^{\text{indiv vars}}} x_{j[i],v} \beta_{v+n^{ID}})$$

Where $z_{j[i]}$ is the j-th value of the v-th variable, where j is the individual corresponding to ballot i. If, as before, I do not assume independence, the model can be written as:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\delta_{j[i]})$$

$$\delta_j \sim \text{N}(\zeta_0 + \sum_{v=1}^{n^{\text{indiv vars}}} x_{j[i],v} \delta_v, \sigma_\delta^2)$$

In the case of my specific data, for the time being I have individual-level data for gender, so $n^{\text{indiv vars}} = 1$. I name the combination of this model and Model 2: **Model 3**. Model 3 can be written as follows:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\delta_{j[i]} + a_{k[i]}), \tag{Model 3}$$

$$a_k \sim \text{N}(\gamma_0 + \sum_{v=1}^{n^{\text{county vars}}} x_{k[i],v} \gamma_v, \sigma_\alpha^2)$$

$$\delta_j \sim \mathrm{N}(\zeta_0 + \sum_{v=1}^{n^{\text{indiv vars}}} x_{j[i],v}\delta_v, \sigma_\delta^2)$$

Election Level

Additional models result from the inclusion of election-level data. The first assumes I only knew what specific election the ballot comes from. Let x_l^{elect} be an indicator variable for each election and n^{elect} the number of elections. The model assuming independence, with x_l being indicator variables for each election, is:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\sum_{l=1}^{n^{elect}} x_{l[i]}^{elect}\beta_l)$$

In a similar way to county- and individual-level data, I add in variables at an election-level. Let's say now that along with the one ballot, I was given a short list of $n^{\text{election vars}}$ other election-level variables, be they discrete, continuous, or indicators. The two models would then look like:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\sum_{v=1}^{n^{\text{election vars}}} x_{l[i],v}\beta_{v+n^{elect}} + ns(x^{\text{year}})) \qquad \text{(Model 4)}$$

Where $x_{l[i],v}$ is the l-th value of the v-th variable, where l is the election corresponding to ballot i. For the time being I have two different variables that describe individual elections: date and type. I choose to fit a GAM with a natural cubic smoothing spline function for year. This would also include four distinct indicators for election type. I name this **Model 4**. Model 4 is not a mixed effects model, since all the variability between elections is incorporated in election type and election year; with those two variables I can fully describe each election[8].

Ballot Level

In this section I assume that the ballot has some key features written on it, like the voting method, age, or party registration of the person that filled it out. A mixed effects model here would make no sense, since all the data is at the same unit of observation. Therefore, when adding ballot level variables, the model would look like:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\beta_0 + \sum_{v=1}^{n^{\text{ballot vars}}} x_{i,v}\beta_v) \qquad \text{(Model 5)}$$

Where $x_{i,v}$ is the i-th value of the v-th variable, and $n^{\text{ballot vars}}$ is the number of ballot level variables. For now, I have data on voting method, age, and party. Voting method is coded as a binary variable with value one if the method was a Mail Vote.

[8] It is much safer to assume election types to be independent when it comes to turnout, than to make the same assumption for individuals or counties. The reason is that different election types historically have different levels of turnout. Any dependence can safely be estimated by the inclusion of a trend over time.

Party includes four distinct indicators for REP, DEM, Other, and Unaffiliated. A linear term is used for age. I name this **Model 5**.

4.4.3 Estimation with the full dataset

I now proceed to include variables from all units of observation into one model. The first model, assuming independence, is:

$$\hat{\mathbb{P}}(y_i = 1) \sim \text{logit}^{-1}(\sum_{k=1}^{64} x_k^{county}\beta_* + \sum_{v=1}^{n^{\text{county vars}}} x_{k[i],v}\beta_* + \sum_{j=1}^{n^{ID}} x_j^{ID}\beta_* + \sum_{v=1}^{n^{\text{indiv vars}}} x_{j[i],v}\beta_* + \sum_{v=1}^{n^{\text{election vars}}} x_{l[i],v}\beta_* + \sum_{v=1}^{n^{\text{ballot vars}}} x_{i,v}\beta_*)$$

You will notice that I have omitted the subscript for all beta coefficients. This is because after two or three parameters the subscript becomes increasingly large. For simplicity, assume increasing indexes for different beta coefficients from left to right in this expression.

The mixed effects model will again operate on two "levels" of hierarchy, but the second level will now include two distinct regressions. Caveats for variables like age and date should be noted from previous sections. This, the most complex model, will be **Model 6**

$$p\_\hat{v}ote \sim \text{logit}^{-1}(\sum_{v=1}^{n^{\text{ballot vars}}} x_{i,v}\beta_v + \sum_{v=1}^{n^{\text{election vars}}} x_{l[i],v}\beta_{v+n^{ballot}} + ns(x^{\text{year}}) + \delta_{j[i]} + \alpha_{k[i]}), \tag{Model 6}$$

$$\alpha_k \sim \text{N}(\gamma_0 + \sum_{v=1}^{n^{\text{county vars}}} x_{k[i],v}\gamma_v, \sigma_\alpha^2)$$

$$\delta_j \sim \text{N}(\zeta_0 + \sum_{v=1}^{n^{\text{indiv vars}}} x_{j[i],v}\delta_v, \sigma_\delta^2)$$

In summary, Table 4.3 includes all noteworthy models from the previous section. I add a few models which should be easily understood based on the specifications given above.

Table 4.4: Individual level model descriptions

Model No	Model Description
Model 1	Naive model with only county mixed effects
Model 2	Multilevel model; added county level predictors
Model 3	Multilevel model; individual- and county-level mixed effects and predictors

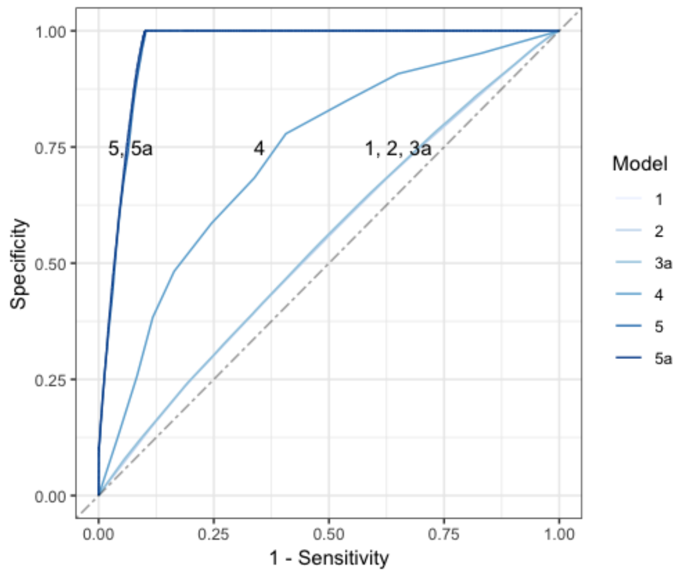

Figure 4.2: ROC Curve for all individual models

Model No	Model Description
Model 3a	Multilevel model; same as 3 without individual-level mixed effects
Model 4	General Additive model; election predictors and time smoothing splines
Model 5	Ballot-level predictors fixed effects model
Model 5a	Multilevel model; ballot predictors with county mixed effects
Model 6	Multilevel General Additive model; year splines; individual, county mixed effects and all predictors

4.4.4 Results

While the aforementioned models are sound in their construction, the leap from theory to implementation hit a few roadblocks as a direct result of the first section in this chapter and the problems outlined within. However, there are still valid reasons to include whatever results were possible to estimate. The results indicate that the data I have on their own *can* be used to build and run an individual model of turnout regardless of if that model is useful in responding to my hypotheses on VBM.

I am confident in the results of models 1, 3a, and 4, somewhat confident in model 5, and less so for models 2, 3, and 5a. Models 2, 3 and 5a "failed to converge"; this means that the numeric approximation process by which *R* implements maximum likelihood estimation[9] for coefficients doesn't give stable results, within certain conditions. While model 5 did converge, it suffers from lack of variance in the predictor for VBM, as explained in the beginning of this chapter; this is the reason why the coefficient for

[9]Estimation of maximum likelihood here uses Adaptive Gaussian-Hermitian Quadrature (AGQ) to estimate coefficients (Handayani, Notodiputro, Sadik, & Kurnia, 2017)

mail vote is so disproportionately large and variable. Model 6 simply did not run, even on a sub-sampled dataset.

While I can't really derive any conclusions from this fact, there is a distinct possibility that this either occurred due to a lack of processing power, or lack of sufficient data for the model estimation to even reach close to convergence. It is also important to point out that model non-convergence is not a fatal issue in and of itself, but becomes so if outputted coefficients wildly differ between calls of the model. Running code to output 5-fold CV AUC involves estimating each model 5 times on different "folds", or sub-samples, of the dataset. During this process the coefficients and AUC values remained fairly stable between runs of the model, despite non-convergence. This leads me to assume that the problem is significant, but not fatal for my models.

In terms of model fit, the models neatly fall into three groups based on their cross-validated AUC. The first group, consisting of models 1, 2, and 3a has an AUC of around .544, making them only slightly distinguishable from a coin-toss. This is fairly reasonable, since I am building a model to make predictions at the ballot level while only using county data or gender[10]. The second group based on AUC includes model 4, the only non-multilevel individual model, with an AUC of around .733, significantly outperforming the first three models. Again, this is reasonable considering how wildly different turnout is between election types; it is only natural that these election-level variables would be so informative. The third group, with the highest AUC of around .96 are models 5 and 5a. This is a direct result of the lack of variability in my data: Mail Voting is an almost perfect predictor of the probability of voting.
I also include another metric under AUC, which I name "Baseline". This is the 5-fold CV difference in predictive power between the model fit and a baseline model that predicts everyone voted. For example, Model 1 has a percentage of true positive and true negative predictions that is 17.8% higher the baseline that everyone voted. The values of the "Baseline" metric fully corroborate what I mention above for AUC. I include it here because it is common practice in elections sciences to compare classification models of participation to the baseline that everyone participated, rather than the baseline of an uninformed "coin-toss" used by the ROC.

There are two conclusions that can be reached from these results. None of these conclusions are, sadly, related to my hypotheses on VBM. The first is that the lack of variance in the data and a lack of processing power are direct causes of my inability to estimate these models. This is apparent in how model 6 does not run, other models do not converge, and the coefficient for VBM is very large, since it doesn't vary enough even after stratified sampling to account for any variance between mail vote and conventional ballots. The second conclusion here is that, despite these issues, there are some confirmable results on turnout in general that are common between individual and county models. For example, across models 2, 3, 3a the urban population of a county is a substantial, negative factor in probability of voting, while the white population is a very small, positive effect[11]. Similar conclusions can be drawn for

[10]Few counties wildly differ in their turnout percentages, and that the coefficient for male gender results in only around a 2.5% decrease in voting probability

[11]This did, however, fail to reach statistical significance in both county and individual level models.

male gender, which is a very small negative effect in voting probability as compared to female gender. These effects being stable across several models means that they are independent of the additions to those models; for example, gender, urban population, and white population have effects that are not accounted for when adding individual level mixed effects. Despite not being able to assess VBM as a factor of turnout probability, these models at least show that the data have substantial use for modelling at the individual level.

Table 4.5: Estimated individual level coefficients (**Significant at the .01 level *at the .05 level)

Predictor	Model 1	Model 2	Model 3	Model 3a	Model 4	Model 5	Model 5a
(Intercept)	-0.175	-0.042	0.001	0.001	-0.541**	-2.478**	-1.888**
	(0.030)	(0.083)	(0.060)	(0.076)	(0.009)	(0.015)	(0.238)
Pct_urban		-0.423**	-0.436**	-0.424**			-0.538**
		(0.055)	(0.059)	(0.062)			(0.114)
Pct_white		0.067	0.075	0.073			-0.151
		(0.102)	(0.073)	(0.094)			(0.281)
genderMale			-0.097**	-0.094**			0.094**
			(0.007)	(0.007)			(0.017)
Republican						0.233**	0.208**
						(0.021)	(0.073)
Other						-0.085	-0.124
						(0.073)	(0.073)
UAF						-0.308**	0.325**
						(0.021)	(0.021)
VBM						23.764**	26.502**
						(45.255)	(285.774)
Age						0.093**	0.086**
						(0.009)	(0.009)
typeGeneral					1.537**		
					(0.011)		
typeMidterm					0.829**		
					(0.011)		
typePrimary					-0.880**		
					(0.010)		
CV AUC	0.543	0.543		0.545	0.733	0.961	0.963
Baseline	0.178	0.178		0.179	0.291	0.547	0.548
Obs	370,586	370,586	370,586	370,586	370,586	370,586	370,586
Groups	64	64	64	64	64	64	64

This means that the small, positive effect is not that distinguishable from no effect at all.

Conclusion

This thesis started off with the prospect of testing three hypotheses, on the impact of VBM on turnout (H1) and how that impact may differ based on how prominent national effects are during election time (H2), specifically using the metric of election type between Midterm, General, Primary, and Coordinated elections (H3). These hypotheses derived from a theory of voting "at the margins", first developed by Aldrich (1993), which stated that the decision on whether to turn out or not is based on a rational calculus made by individual voters. This calculus is affected by a wide range of effects, from small local ones to large, national variables. In my first Chapter I identified how this theory fits in the general literature on voter behavior, and made a series of observations on how each theory views the impact of VBM. The hypotheses were then tested on Voter Files from Colorado, which described registration and voter history for elections ranging from 2010-2016. The process of data wrangling, modeling, and estimations revealed that the data I used was not ideal for the purposes I intended, particularly for fitting individual-level models.

At the end of my thesis I am able to confirm H1, but reject H2 and H3. These results come from my County level models. Despite non-convergence my individual level models have some inferential potential, but they are still not particularly useful at testing my hypotheses on VBM. This is true because of the variability concerns outlined early in Chapter 4. I confirm H1 because mail voting seems to be an incremental, comparably small effect on turnout. I reject H2/H3 because that effect is not more pronounced for elections that have less strong national influences. In fact, my county-level models showed the exact opposite: VBM had a strong, significant positive effect for presidential elections, but no discernible effect for primary, coordinated races, or midterm races.

This is evidence against voting "at the margins", since my hypotheses that were most connected to Aldrich's model (H2, H3) were convincingly rejected. I find evidence against a hypothesis of habitual voting as well, since in my county models Mail Voting has a significant, positive effect. This means that the absence of election-day effects or the strong presence of habitual voting did not override the effect on turnout that mail voting has. The same conclusion can be drawn for a social theory of voting, which predicts no effect; despite my prediction in Table 2.1 for this theory being a confirmation of H1 and rejection of H2/3, the fact that I find a consistent, positive effect on turnout in general elections is enough to rule out a social theory of voting effect.

Interpreting my results through a resources and organizational lens seems on the surface more convincing, since this theory predicts a consistent, positive effect of VBM on turnout because of the increase in capacity it gives to individuals. However the effect is not comparativelly large, as demonstrated by the fact that I confirm H1. It should also be noted that this effect should be present in *all* elections to be convincing evidence for the resources electoral participation paradigm. This is not the case. One explanation may be that my models were sensitive to the vastly greater turnout that occurs during presidential and midterm years, particularly since the differences in mail voting percentage between counties was at most around 20%. This, again, leads to the conclusion that the lack of variability in my data does not permit any more precise analysis.

The first step towards future research, directly resulting from this problem, would be to get more data. This is possible, but a task I was unable to accomplish due to time constraints related to this project. More data would allow for a comparative study of Colorado elections before and after all major legislative electoral changes of this century (2008, 2013). Another way to expand on my research would be to look at even lower level elections, like municipal elections, school board elections, or recall elections; voter history files contain data on all of these races. A study with more data could use such races as well for testing the hypotheses I set up in this thesis. Third, it is possible to replicate my research here for other states like Oregon and Washington, or for the specific counties in Arizona, Utah, or California that have all-mail elections. Fourth, elections thankfully do not stop happening; my research here can always be updated with data from the 2017 Coordinated Colorado election, or the 2018 Midterm that just took place. I would caution that the same issue with data variability would still exist in this case; there is a much greater need for data going back, rather than new all-mail elections.

Apart from expanding my research in terms of data, another path would be to implement some of the methods I outline in the beginning of Chapter 4, like the Synthetic Control Group. Such methods would allow for inferences to be made despite the data issues I encountered. Lastly, it should be noted that VBM is just one of a series of electoral reforms like Automatic Voter Registration (AVR), Early Voting, or Voter ID Restrictions, all of which can potentially be tested using multilevel models and data wrangling methods that I have applied in this thesis.

Another contribution of my thesis is my analysis of data wrangling, the construction of multilevel general additive models for turnout, and the accompanying *R* package. To take these one by one, I have provided arguments in favor of preferring multiple snapshots of registration files rather than just the latest iteration of the record. I have analyzed the pitfalls that exist in such documents, and given specific examples on how this can be dealt with for the Colorado data files. I have also provided a set of variable specification that can be useful as indicators of the content of these data, or the potential uses of voter registration files in other future studies. Finally, I have presented potential future solutions to issues these data have with variability, and ways to circumvent processing power limitations.

Additionally, I have meticulously gone through the creation of multilevel general additive models of individual and county level turnout. While due to data and processing power limitations I am unable to run all these models to typical standards of convergence. This does not mean that they present no value to future research. Quite the contrary, future researchers just have to go through the data clean-up stage, and then implement my models without having to construct them from scratch. In particular, mixed effects and general additive models are not widely used in such studies, making their presentation and specification rather unique regardless of their application in this piece of research.

Lastly, I provide an extensive library of code used to create this document and the research I conduct. I have made an *R* package–which I named `riggd`–that includes more than a dozen different functions that serve data wrangling and presentation purposes. These functions are made for use on Colorado files, but require relatively small amounts of changes to be applied to voter files from around the US. I also provide code for all tables and graphics that are included in this thesis on gitHub, which is a testament to the reproducibility and future value of the research I conducted.

I recognize that, despite many obstacles in terms of data or computing power, the outcome of this thesis being more constructive rather than conclusive is to some extent my fault. There were many problems in this thesis that I should have been aware of earlier in the process, which may have allowed me to present more concrete results rather than a series of tools and methods. However, in the combination of my existing conclusions and the materials I have created through this process, it is my belief that this thesis does in fact present a step forward in the literature, and that it adds to existing quantitative elections studies works. This is a small step, but it helps in our understanding of how voters behave, what the actual results of election policy are, and how to expand participation.

Appendix A

MCMC Estimation Processes for Multilevel Models

In statistical science a Markov Chain is a sequence of random variables whose value depends on the value of the exact previous random variable. In mathematical terms, this would be a sequence $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, ..., \theta^{(t)}$ where $\mathbb{P}(\theta^{(t)} = y|\theta^{(n)}, n < t) = \mathbb{P}(\theta^{(t)} = y|\theta^{(t-1)})$. A Markov Chain Monte Carlo simulation uses Bayesian estimation to update each sequential estimate of θ, leading it to converge to the true value being estimated (Gelman & Hill, 2006; Jackman, 2009).

Multilevel models can be estimated using MCMC sampling. Indicatively, this appendix presents the construction and coding of two types of MCMC samplers based on the Gibbs algorithm. The code and mathematical derivations are adapted to my models from Gelman and Hill (2006).

A.1 Gibbs Sampler for the County Models

The Gibbs algorithm works as follows:

1. Choose a number of parallel simulation runs (chains). This number should be relatively low. In this example it is set to 3.

2. For each chain do the following:

   (a) Initialize vector of parameters $\Theta^{(0)} = \{\theta_1^{(0)}\, \theta_2^{(0)}\, ..., \theta_n^{(0)}\}$

   (b) Choose a number of iterations. For each iteration update every parameter in vector $\Theta^{(n_{iteration})}$, based on the values of vector $\Theta^{(n_{iteration}-1)}$.

3. Evaluate convergence between the chains.

If convergence is poor, repeat for more iterations, or follow diagnostic procedures. These are not specified here, but Gelman and Hill provide a good overview (Gelman & Hill, 2006; Gelman, Carlin, Stern, & Rubin, 2003).

A.1.1 County Model 1 (Only Random County Effects)

A basic multilevel model with only group-level intercept mixed effects can be written as follows:

$$y_i \sim N(a_{j[i]}, \sigma_y^2),\ i \in [1, n] a_j \sim N(\mu_\alpha, \sigma_\alpha^2),\ j \in [1, J]$$

This specification is slightly different from that presented in Chapter 2. Here $\alpha_{j[i]}$ is the coefficient for the group j that individual i belongs to, σ_y, σ_α the variances of the individual and group level distributions respectively, and μ_α the mean of the group-level distribution. In the case of the most basic county-level model estimated in my thesis (County Model 1), $n = 704$ and $J = 64$. Using Maximum Likelihood Estimation, and given that:

$$\alpha_j | y, \mu_\alpha, \sigma_y, \sigma_\alpha \sim N(\hat{\alpha_j}, V_j) \tag{A.1}$$

we can obtain estimates:

$$\hat{\alpha_j} = \frac{\frac{n_{[j]}}{\sigma_y^2}\bar{y}_{[j]} + \frac{1}{\sigma^2 \alpha}}{\frac{n_{[j]}}{\sigma_y^2} + \frac{1}{\sigma^2 \alpha}}, \qquad V_j = \frac{1}{\frac{n_{[j]}}{\sigma_y^2} + \frac{1}{\sigma^2 \alpha}}, \tag{A.2}$$

where $n_{[j]}$ is the number of observations for group j, and $\bar{y}_{[j]}$ is the mean response for group j. Using these estimates and the common MLE estimates for variance and mean in a normal distribution, it is possible to construct a Gibbs sampler for model coefficients and errors. Step 2(b) in the Gibbs sampler would then be:

1. Estimate $a_j,\ j \in [1, J]$ using equations (1), (2).
2. Estimate μ_α by drawing from $N(\frac{1}{J}\sum_1^J \alpha_j, \sigma_\alpha^2/J)$ using the previous values estimated in step 1.
3. Estimate σ_y^2 as $\frac{\frac{1}{n}\sum_1^n (y_i - \alpha_{j[i]})^2}{X_{n-1}^2}$ where X_{n-1}^2 is a draw from a χ^2 distribution with $n-1$ degrees of freedom.
4. Estimate σ_α^2 as $\frac{\frac{1}{J}\sum_1^J (\alpha_j - \mu_\alpha)^2}{X_{J-1}^2}$ where X_{n-1}^2 is a draw from a χ^2 distribution with $J-1$ degrees of freedom.

While each step here seems relatively intuitive, the derivations behind some of the details (like the chi-squared distribution) are complex MLE processes and beyond the scope of this thesis. The R code for this algorithm is as follows:

```
## Gibbs sampler in R
a.update <- function(){
  a.new <- rep (NA, J)
  for (j in 1:J){
    n.j <- sum (model_dt$county==cnt_vec[j])
    y.bar.j <- mean (model_dt$turnout[model_dt$county==cnt_vec[j]])
```

```
    a.hat.j <- ((n.j/sigma.y^2)*y.bar.j + (1/sigma.a^2)*mu.a)/
                 (n.j/sigma.y^2 + 1/sigma.a^2)
    V.a.j <- 1/(n.j/sigma.y^2 + 1/sigma.a^2)
    a.new[j] <- rnorm (1, a.hat.j, sqrt(V.a.j))
  }
  return (a.new)
}
mu.a.update <- function(){
  mu.a.new <- rnorm (1, mean(a), sigma.a/sqrt(J))
  return (mu.a.new)
}
sigma.y.update <- function(){
  sigma.y.new <- sqrt(sum((model_dt$turnout-
                              a[model_dt$county])^2)/rchisq(1,703))
  return (sigma.y.new)
}
sigma.a.update <- function(){
  sigma.a.new <- sqrt(sum((a-mu.a)^2)/rchisq(1,J-1))
  return (sigma.a.new)
}

J <- 64
n.chains <- 3
n.iter <- 1000
sims <- array (NA, c(n.iter, n.chains, J+3))
dimnames (sims) <- list (NULL, NULL,
                         c (paste ("a[", 1:J, "]", sep=""), "mu.a",
   "sigma.y", "sigma.a"))

for (m in 1:n.chains){
  mu.a <- rnorm (1, mean(model_dt$turnout), sd(model_dt$turnout))
  sigma.y <- runif (1, 0, sd(model_dt$turnout))
  sigma.a <- runif (1, 0, sd(model_dt$turnout))
  for (t in 1:n.iter){
    a <- a.update ()
    mu.a <- mu.a.update ()
    sigma.y <- sigma.y.update ()
    sigma.a <- sigma.a.update ()
    sims[t,m,] <- c (a, mu.a, sigma.y, sigma.a)
  }
}
```

Table A.1: Gibbs sampler results for County Model 1

Calculated from...	mu.a	sigma.y	sigma.a
Sampler	0.4688	0.2004	0.0385
Model	0.469	0.199	0.039

As is obvious from Table, the Gibbs sampler produces values very similar to the ones given by an *R* call of Model 1.

A.1.2 County Model 2 (Random County Effects and County-Level Predictors)

With slight changes from the previous model the following is the mathematical expression for a mixed effects model with group-level predictors:

$$y_i \ \sim \ N(a_{j[i]}, \sigma_y^2), \ i \in [1,n] a_j \ \sim \ N(U_j \gamma, \sigma_\alpha^2), \ j \in [1, J],$$

where U_j is a vector of predictor values for group j, and γ a vector of group-level coefficients, with the rest of the parameters having the same designation as previously. Bear in mind that the second of the previous expressions can also be written as:

$$\alpha_j = U_j \gamma + \eta_j, \quad \eta_j \ \sim \ N(0, \sigma_\alpha^2) \tag{A.3}$$

Updating the estimates used previously, it is again possible to construct a Gibbs sampler for model coefficients and errors. Step 2(b) in the Gibbs sampler in this case is:

1. Estimate $a_j, \ j \in [1, J]$. Start by calculating $y_i^{temp} = y_i - U_{j[i]}\gamma$. Then calculate an estimate $\hat{\eta}_j$ and variance matrix V_j from equations (1), (2), by replacing $\hat{\alpha}_j$ with $\hat{\eta}_j$ and y with y^{temp}. Use $\eta_j \ \sim \ N(\hat{\eta}_j, V_j)$ to draw errors η_j and then use (3) to estimate α_j for $j \in [1, J]$.

2. Estimate γ by first regressing α by predictor matrix U to obtain $\hat{\gamma}$ and variance matrix V_γ. Then use distribution $\gamma_j \ \sim \ N(\hat{\gamma}_j, V_j)$ to obtain estimates for vector γ.

3. Estimate σ_y^2 as $\frac{\frac{1}{n}\sum_1^n (y_i - \alpha_{j[i]})^2}{X_{n-1}^2}$ where X_{n-1}^2 is a draw from a χ^2 distribution with $n-1$ degrees of freedom.

4. Estimate σ_α^2 as $\frac{\frac{1}{J}\sum_1^J (\alpha_j - U_j\gamma)^2}{X_{J-1}^2}$ where X_{n-1}^2 is a draw from a χ^2 distribution with $J-1$ degrees of freedom.

County Model 2, as presented in Chapter 4, includes two county-level predictors: percentage of white residents and percentage of urban population; this means that $U = \{x^{\%white}, x^{\%urban}\}$. Keeping this in mind the following code estimates the coefficients and standard errors for Model 2:

```
## Gibbs sampler for a multilevel model with county predictors


a.update <- function(){
  y.temp <- model_dt$turnout -
    (model_dt$pct_urban*g[1] + model_dt$pct_white*g[2])
  eta.new <- rep (NA, J)
  for (j in 1:J){
    n.j <- sum (model_dt$county==cnt_vec[j])
    y.bar.j <- mean (y.temp[model_dt$county==cnt_vec[j]])
    eta.hat.j <- ((n.j/sigma.y^2)*y.bar.j/
                    (n.j/sigma.y^2 + 1/sigma.a^2))
    V.eta.j <- 1/(n.j/sigma.y^2 + 1/sigma.a^2)
    eta.new[j] <- rnorm (1, eta.hat.j, sqrt(V.eta.j))
  }
  a.new <- (U$urban*g[1] + U$white*g[2]) + eta.new
  return (a.new)
}

g.update <- function(){
  lm.0 <- lm (a ~ U$urban + U$white)
  g.new <- coef (lm.0)[2:3]
  return (g.new)
}

sigma.y.update <- function(){
  sigma.y.new <- sqrt(sum((model_dt$turnout-
                  a[model_dt$county])^2)/rchisq(1,703))
  return (sigma.y.new)
}

sigma.a.update <- function(){
  sigma.a.new <- sqrt(sum((a-(model_dt$pct_urban*g[1] +
                  model_dt$pct_white*g[2]))^2)/rchisq(1,J-1))
  return (sigma.a.new)
}

J <- 64
n.chains <- 3
n.iter <- 2000
sims <- array (NA, c(n.iter, n.chains, J+4))
dimnames (sims) <- list (NULL, NULL, c (paste ("a[", 1:J, "]", sep=""),
   c("coef.urban", "coef.white"),
   "sigma.y", "sigma.a"))
```

```
for (m in 1:n.chains){
  g <- rnorm (2)
  sigma.y <- runif (1, 0, sd(model_dt$turnout))
  sigma.a <- runif (1, 0, sd(model_dt$turnout))
  for (t in 1:n.iter){
    a <- a.update ()
    g <- g.update ()
    sigma.y <- sigma.y.update ()
    sigma.a <- sigma.a.update ()
    sims[t,m,] <- c (a, g, sigma.y, sigma.a)
  }
}
```

Table A.2: Gibbs sampler results for County Model 2

Calculated from…	coef_urban	coef_white	sigma.y	sigma.a
Sampler	-0.1182	0.0337	0.1997	1.684
Model	-0.118	0.034	0.199	2.631

As previously the values outputted by the Gibbs sampler are very close to those estimated by the model, apart from the group level standard deviation. Some variability in how closely the sampler approximates the model call is to be expected, due to the difference in how the model is estimated in R (much more precise Bayesian processes).

References

Absentee and Early Voting. (2018, October). *National Council of State Legislatures.* Retrieved from http://www.ncsl.org/research/elections-and-campaigns/absentee-and-early-voting.aspx#a

Aldrich, J. H. (1993). Rational Choice and Turnout. *American Journal of Political Science*, *37*(1), 246–278. http://doi.org/10.2307/2111531

Ansolabehere, S., & Hersh, E. (2010). The Quality of Voter Registration Records: A State-by-State Analysis. *Institute for Quantitative Social Science and Caltech/MIT Voting Technology Project Working Paper.* Retrieved from https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/18550

Ansolabehere, S., & Hersh, E. D. (2017). ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name. *Statistics and Public Policy*, *4*(1), 1–10. http://doi.org/10.1080/2330443X.2017.1389620

Barr, C. D., Diez, D. M., Wang, Y., Dominici, F., & Samet, J. M. (2012). Comprehensive Smoking Bans and Acute Myocardial Infarction Among Medicare Enrollees in 387 US Counties: 1999–2008. *American Journal of Epidemiology*, *176*(7), 642–648. http://doi.org/10.1093/aje/kws267

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1). Retrieved from http://arxiv.org/abs/1406.5823

Bergman, E., & Yates, P. A. (2011). Changing Election Methods: How Does Mandated Vote-By-Mail Affect Individual Registrants? *Election Law Journal: Rules, Politics, and Policy*, *10*(2), 115–127. http://doi.org/10.1089/elj.2010.0079

Berinsky, A. J. (2005). The Perverse Consequences of Electoral Reform in the United States. *American Politics Research*, *33*(4), 471–491. http://doi.org/10.1177/1532673X04269419

Berinsky, A. J. (2016, February). Making Voting Easier Doesn't Increase Turnout. *Stanford Social Innovation Review.* Retrieved from https://ssir.org/articles/entry/making_voting_easier_doesnt_increase_turnout

Burden, B. C., & Neiheisel, J. R. (2013). Election Administration and the Pure Effect of Voter Registration on Turnout. *Political Research Quarterly*, *66*(1), 77–90.

http://doi.org/10.1177/1065912911430671

Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2014). Election Laws, Mobilization, and Turnout: The Unanticipated Consequences of Election Reform. *American Journal of Political Science*, *58*(1), 95–109. http://doi.org/10.1111/ajps.12063

Campbell, A. L. (2002). Self-Interest, Social Security, and the Distinctive Participation Patterns of Senior Citizens. *American Political Science Review*, *96*(3), 565–574. http://doi.org/10.1017/S0003055402000333

Chen, J. (2013). Voter Partisanship and the Effect of Distributive Spending on Political Participation. *American Journal of Political Science*, *57*(1), 200–217. http://doi.org/10.1111/j.1540-5907.2012.00613.x

Chihara, L. M., & Hesterberg, T. C. (2011). *Mathematical Statistics with Resampling and R* (1 edition). Hoboken, N.J: Wiley.

Cronin, T. E., & Loevy, R. D. (2012). *Colorado Politics and Policy: Governing a Purple State*. Lincoln: University of Nebraska Press. Retrieved from http://ebookcentral.proquest.com/lib/reed/detail.action?docID=1034959

Edelman, G., & Glastris, P. (2018). Letting people vote at home increases voter turnout. *Washington Post*. Retrieved from https://www.washingtonpost.com/outlook/letting-people-vote-at-home-increases-voter-turnout-heres-proof/2018/01/26/d637b9d2-017a-11e8-bb03-722769454f82_story.html

Edlin, A., Gelman, A., & Kaplan, N. (2007). Voting as a Rational Choice: Why and How People Vote To Improve the Well-Being of Others. *Rationality and Society*, *19*(3), 293–314. http://doi.org/10.1177/1043463107077384

Ewald, A. C. (2009). *The Way We Vote: The Local Dimension of American Suffrage*. Nashville: Vanderbilt University Press.

Fortier, J. C. (2006). *Absentee and early voting: Trends, promises, and perils*. Washington, DC: AEI Press.

Fowler, J. H. (2006). Habitual Voting and Behavioral Turnout. *Journal of Politics*, *68*(2), 335–344. http://doi.org/10.1111/j.1468-2508.2006.00410.x

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (1st Edition). Cambridge, NY: Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis, Second Edition* (2nd Edition). Boca Raton, Fla: Chapman Hall.

Gerber, A. S., Huber, G. A., & Hill, S. J. (2013). Identifying the Effect of All-Mail Elections on Turnout: Staggered Reform in the Evergreen State. *Political Science*

Research and Methods, *1*(1), 91–116. http://doi.org/10.1017/psrm.2013.5

Geys, B. (2006). Explaining voter turnout: A review of aggregate-level research. *Electoral Studies*, *25*(4), 637–663. http://doi.org/10.1016/j.electstud.2005.09.002

Griffin, R., Gronke, P., Wang, T., & Kennedy, L. (2017). *Who Votes With Automatic Voter Registration?* Center for American Progress. Retrieved from https://www.americanprogress.org/issues/democracy/reports/2017/06/07/433677/votes-automatic-voter-registration/

Gronke, P., & Miller, P. (2012). Voting by Mail and Turnout in Oregon: Revisiting Southwell and Burchett. *American Politics Research*, *40*(6), 976–997. http://doi.org/10.1177/1532673X12457809

Gronke, P., Galanes-Rosenbaum, E., Miller, P. A., & Toffey, D. (2008). Convenience Voting. *Annual Review of Political Science*, *11*(1), 437–455. http://doi.org/10.1146/annurev.polisci.11.053006.190912

Gronke, P., McGhee, E., Romero, M., & Griffin, R. (2017). Voter Registration and Turnout under "Oregon Motor Voter": A Second Look. In. Portland. Retrieved from https://dl.dropboxusercontent.com/nativeprint?file=https%3A%2F%2Fwww.dropbox.com%2Fs%2Fo63a0zi7j8plhgl%2FOMV_and_Turnout_McGhee_Gronke_Romero_Griffin_July2017.pdf%3Fdisable_range%3D1%26from_native_print%3D1%26preview%3D1

Hamm, K. (2017). How Colorado has voted in presidential elections (and how its politics have changed) since 1980. *The Denver Post*. Retrieved from https://www.denverpost.com/2017/12/22/how-colorado-votes/

Handayani, D., Notodiputro, K. A., Sadik, K., & Kurnia, A. (2017). A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (GLMM). In. Jawa Barat, Indonesia. http://doi.org/10.1063/1.4979449

Hersh, E. D. (2015). *Hacking the Electorate: How Campaigns Perceive Voters*. New York, NY: Cambridge University Press.

Hullinghorst, D. L., & Pabon, D. (2013, May). Voter Access and Moderninzed Elections Act.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences* (1st edition). Chichester, U.K: Wiley.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R* (1st ed. 2013, Corr. 7th printing 2017 edition). New York: Springer.

Keele, L., & Titiunik, R. (2017). Geographic Natural Experiments with Interference:

The Effect of All-Mail Voting on Turnout in Colorado.

Martin, C. (1962). *Colorado politics* (2nd ed.). Denver, Colorado: Big Mountain Press. Retrieved from http://hdl.handle.net/2027/mdp.39015024371158

Matsusaka, J. G., & Palda, F. (1999). Voter turnout: How much can we explain? *Public Choice*, *98*(3-4), 431–446. http://doi.org/10.1023/A:1018328621580

McClelland, R., & Gault, S. (2017). The Synthetic Control Method as a Tool to Understand State Policy. *Washington, DC: Urban-Brookings Tax Policy Center*.

McDonald, M. P. (2007). The True Electorate: A Cross-Validation of Voter Registration Files and Election Survey Demographics. *Public Opinion Quarterly*, *71*(4), 588–602. http://doi.org/10.1093/poq/nfm046

Mettler, S., & Stonecash, J. M. (2008). Government Program Usage and Political Voice. *Social Science Quarterly*, *89*(2), 273–293. http://doi.org/10.1111/j.1540-6237.2008.00532.x

Neiheisel, J. R., & Burden, B. C. (2012). The Impact of Election Day Registration on Voter Turnout and Election Outcomes. *American Politics Research*, *40*(4), 636–664. http://doi.org/10.1177/1532673X11432470

Plutzer, E. (2002). Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood. *The American Political Science Review*, *96*(1), 41–56. Retrieved from https://www.jstor.org/stable/3117809

Richey Sean. (2008). Voting by Mail: Turnout and Institutional Reform in Oregon. *Social Science Quarterly*, *89*(4), 902–915. http://doi.org/10.1111/j.1540-6237.2008.00590.x

Rosenstone, S. J. (2003). *Mobilization, participation, and democracy in America*. New York: Longman.

Saltman, R. (2009). *The History and Politics of Voting Technology: In Quest of Integrity and Public Confidence* (2006 edition). Gordonsville: Palgrave Macmillan.

Schneider, A., & Ingram, H. (1990). Behavioral Assumptions of Policy Tools. *The Journal of Politics*, *52*(2), 510–529. http://doi.org/10.2307/2131904

Smets, K., & Ham, C. van. (2013). The embarrassment of riches? A meta-analysis of individual-level research on voter turnout. *Electoral Studies*, *32*(2), 344–359. http://doi.org/10.1016/j.electstud.2012.12.006

Staff, L. C. (2009, July). TABOR and Referendum C. *Colorado General Assembly*. Retrieved from https://leg.colorado.gov/publications/tabor-and-referendum-c-2009

Stein, R. M., & Vonnahme, G. (2008). Engaging the Unengaged Voter: Vote Centers and Voter Turnout. *The Journal of Politics*, *70*(2), 487–497. http://doi.org/

10.1017/S0022381608080456

Thompson, J. (2016). The first Sagebrush Rebellion: What sparked it and how it ended. Retrieved from https://www.hcn.org/articles/a-look-back-at-the-first-sagebrush-rebellion

US Census Bureau QuickFacts: Colorado. (2010). *US Census Bureau Quickfacts*. Retrieved from https://www.census.gov/quickfacts/co

Verba, S., & Nie, N. H. (1972). *Participation in America: Political democracy and social equality*. New York: Harper & Row.

Wand, J. N., Shotts, K. W., Sekhon, J. S., Mebane, W. R., Herron, M. C., & Brady, H. E. (2001). The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida. *American Political Science Review*, *95*(4), 793–810. http://doi.org/10.1017/S000305540040002X

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R* (1st edition). Boca Raton, FL: Chapman; Hall/CRC.

Wood, S., & Scheipl, F. (2017, July). Gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4'. Retrieved from https://CRAN.R-project.org/package=gamm4