How I learned to stop worrying and Love Voter Registration Files

---

A Thesis
Presented to
The Division of Mathematics and Natural Sciences and History and Social Sciences
Reed College

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

Theodore Dounias

December 2018

Approved for the Division
(Mathematics)

Andrew Bray	Paul Gronke

Preface

This is an example of a thesis setup to use the reed thesis document class.

Table of Contents

List of Tables

List of Figures

Introduction

The democratic system is based on procedures as much as principles. The way that democracies chose to tally the will of the people is always a messy, controversial process. 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 is inherently coupled with the outcome. Underlying those decisions is a nebulous, inconclusively answered question: are elections fair, and how can we make them more so.

The passage of the Help America Vote Act—or HAVA–(Robert Nay, 2002), which mandated states to update and consolidate public voter registration files, and created the US Elections Assistance Commission that makes available county level data, innovated the way we use data based approaches to answer this question. HAVA offered political scientists and statisticians direct access to the voting population's voting patterns, political registration, age, geolocation and much more; information that up to then was only accessible by sampling through surveys. The immense leap here happens because true population data does away with the need for sampling techniques that are often biased and inaccurate. We can now not only get a complete picture of the data, but also link and merge with other sources of information such as US Census data on religion, race, education, or income—work that has been lucrative for firms such as Catalist or Target Smart. By posing Political Scientific questions, and trying to respond with rigorous statistics, 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).

Chapter 1

The State of the Literature

In this chapter I will go through the exsting literature on Vote-By-Mail (VBM). I will define what Vote-By-Mail is; I will then summarize the expectations that researchers have of the effects of VBM on turnout, based on existing theories of electoral participation. I will continue with a summary of previous quantitative research on the effects that VBM and similar policies have had on turnout. I will conclude with some more general comments on the available data, and literature concerning the most commonly used quantitative methods.

1.1 What is VBM?

Gronke (2007, 2008), RMStein (1998)

Vote-By-Mail 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 these ballots, ranging from dropping them off at pre-designated locations, to mailing them in, to bringing them to a polling place and voting conventionally. This varies across states that have implemented VBM. Some common forms of the VBM policy are:

- *Postal Voting*: All voters receive a ballot by mail, which can then be returned to a pre-designated location or mailed in to be counted. This is the current system in Oregon, is an option in Colorado, and is implimented by a number of counties in California, Utah, and Montana.

- *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 Washington, Kansas, and New Jersey.

- *Hybrid or Transitionary Systems*: In hybrid systems, voters receive a mail ballot but can choose to disregard it and vote conventionaly. This is the case in Colorado. Transitionary systems exist in states that have chosen to eventually conduct all elections by postal voting, but have given counties an adjustment period during which this shift is not mandatory, or mandatory only for certain elections. This is the case in California, Utah, and Montana.

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 timeframe. This means that Vote-By-Mail can be counted as a "convenience voting" reform. These are usually impliented 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.

1.2 The Calculus of Voting

Grimmer (2011), Burden (2013)

1.2.1 Why Turnout Matters

Geys (2006), Smets (2013) ++ book sources

Turnout is the most commonly used measure for electoral participation. It is important because it signifies the level of engagement of the population with the state, th elevel of incorporation of different subgroups of the population into democratic processes, and the legitimacy of elected officials. It is widely accepted that turnout shuld be maximized so that the democratic franchise represents the majority of citizens.

1.2.2 Theories of Voting

Aldrich (1993), Berinsky (2001, 2005), Edin (2007), Bendor (2003), Gerber and Green (2015), Matsusaka (1997), Fowler (2006)

1.2.3 How they Apply to VBM

Berisnky (2005), Banducci (2000), Gronke and Toffey (2008), and several applications of sources in above sections.

1.3 Previous Study Results

1.3.1 General Results

Arcenaux (2012), Bergman (2011), Burden (2014), Edelman + Pantheon Analytics (2018), Gerber (2013), Rhine (1995), Neihelsen (2012), Keele (2018), Richey (2008), RMStein (1997, 2007), Gronke (2007, 2008, 2012, 2017)

1.3.2 The Gerber Piece

Gerber(2017)

1.4 Voter Registration Files

1.4.1 Inaccuracy of Survey Data

Ansolabehere and Hersch (2012), Burden (2000), Deufel (2010)

1.4.2 The Importance of VRF

Books, mentioned later

1.5 Common Methods Used and Problems Encountered

1.5.1 Methods

- *Synthetic Control Group*: Abadie (2010), McClleland (2017), Gronke (2017)
- *Record Linkage*: Ansolabehere and Hersch (2017), Harvey (1994, 97), Koudas (2013)
- *GLM (Probit/Logit/Poisson)*: Barreto (2004), Dow (2004)
- *DID*: Bertrand et al. (2002)
- *E.I.*: King (2013), Burden (1998), Calvo (2003), Chao (2004), RMStein (2002)
- *Mixed-Effects*: Gelman and Hill (2007)
- *General EDA and Models*: James et al. (2013), Chapman and Hall (2017)

1.5.2 Issues

Grimmer (2015) {Not always best to do inferential models}, Ansolabehere and Hersch (2010) {Problems with Voter Reg Files}, other sources from the previous section

Chapter 2

Hypothesis, Data, and Methods

2.1 The Data

2.1.1 Source

2.1.2 Structure

2.1.3 Wrangling

2.2 Hypotheses

2.2.1 Description of Hypotheses

2.2.2 Criteria

2.2.3 Expected Results

2.3 Methodology

2.3.1 EDA

2.3.2 Description and Parametrization of Models

2.4 Gerber Replication

Chapter 3

Results

3.1 EDA

3.2 Model Output

3.3 Gerber Expansion Results

Conclusion

References