Three Essays in the Dynamics of Political Behavior

Thesis by
Seo-young Silvia Kim

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended May 12, 2020

Seo-young Silvia Kim
ORCID: 0000-0002-8801-9210

ACKNOWLEDGEMENTS

First and foremost, I express my deepest gratitude to R. Michael Alvarez, Jonathan N. Katz, Michael Gibilisco, and Matthew Shum for their guidance and support. In particular, I am much indebted to Jonathan and Mike A, respectively my primary advisor and my co-advisor/dissertation chair, both of whom have demonstrated how interesting political science can be and firmly and patiently supported my professional and personal growth. Jonathan and Mike have been ideal advisors one can only dream of—they have had and continue to have my best interests at heart, and always encouraged confidence. In particular, both have great insights about my personal strengths and weaknesses and know how to turn even my weaknesses into strengths. I am where I am thanks to them.

I sincerely thank the John Randolph Haynes and Dora Haynes Foundation for providing funding to support the research reported in Chapter 1, and the Lance Davis Fellowship for supporting me in 2019–2020. I also thank Neal Kelley and his team at the Orange County Registrar of Voters (OCROV) office for their partnership on Chapter 1 and other papers not included in this dissertation. I especially thank Justin Berardino of the OCROV for answering many of my questions and facilitating the transfer of the data.

I also thank other faculty and fellow graduate students in and outside Caltech—among them, Laura Doval, Michael J. Ewens, Daniel R. Guth, Alex Hirsch, Byungkoo Kim, Jeongbin Kim, John S. Kuk, Shiro Kuriwaki, Inés Levin, Yimeng Li, Lucas Núñez, Thomas R. Palfrey, Alejandro Robinson-Cortés, Jean-Laurent Rosenthal, Bob Sherman, and Euncheol Shin.

A special shoutout to the staff of Caltech, including Laurel Auchampaugh, Kapauhi Stibbard, and Sabrina Heimeister. In particular, Daniel Yoder and Laura Flower Kim have helped me navigate the administrative difficulties faced at various points in time and deserve special love and the most sincere respect.

Finally, I am very grateful for my family, Myung-hwan, Chung Mee, and Ah-young, who have unconditionally supported everything I do from afar. Most of all, I thank YongKi Hong, fiancé, best friend, and colleague—he has made this all possible with a positive, humorous disposition and steadfast love.

ABSTRACT

In this thesis, I empirically assess the dynamics of political behavior. More specifically, I analyze what creates—or does not create—change in political participation, such as voting in elections and contributing to campaigns. Through this, I intend to show that paying close attention to dynamics can help answer fundamental questions of political behavior and offer important insights for real-life policies.

In Chapter 1, I focus on how non-political life events and election administration policy impact voter turnout. I analyze (1) the effect of moving on turnout over time and (2) how an election administration policy helps with the recovery of lowered turnout by lowering the re-registration burden of movers.

Moving depresses turnout by imposing various costs on voters. However, movers eventually settle down, and such detrimental effects can disappear over time. I analyze these dynamics using United States Postal Services (USPS) data and detailed voter panel data from Orange County, California. Using a generalized additive model, I show that previously registered voters who move close to the election are significantly less likely to vote (at most -16.2 percentage points), and it takes at least six months on average for turnout to recover. This dip-and-recovery is not observed for within-precinct moves, suggesting that costs of moving matter only when the voter's environment has changed much. I then evaluate an election administration policy that resolves their re-registration burden. This policy proactively tracks movers, updates their registration records for them, and notifies them by mailings. Using a natural experiment, I find that this policy is effective in boosting turnout (+5.9 percentage points). This success of a simple, pre-existing, and non-partisan safety net is promising, and I conclude by discussing policy implications.

Chapter 2 (published at *Election Law Journal*, doi: 10.1089/elj.2019.0593, coauthored with R. Michael Alvarez and Jonathan N. Katz) shows how the participation dynamics of political participation differ between two distinct classes of donors—hidden and visible (from data), based on their amount contributed. In campaign finance we find that there is something about the data generating process that is often overlooked, but which affects the interpretation of data greatly. This precedes Chapter 3 as it provides some important intuitions as to how the data should be filtered, wrangled, and interpreted for usage.

More specifically, inferences about individual campaign contributors are limited by

how the Federal Election Commission (FEC) collects and reports data. Only transactions that exceed a cycle-to-date total of $200 are individually disclosed, so that contribution histories of many donors are unobserved. We contrast visible donors and "hidden donors," or small donors who are invisible due to censoring and routinely ignored in existing research. I use the Sanders presidential campaign in 2016, whose unique campaign structure received money only through an intermediary (or conduit) committee. These are governed by stricter disclosure statutes, allowing us to study donors who are normally hidden. For the Sanders campaign, there were seven hidden donors for every visible donor, and altogether, hidden donors were responsible for 33.8% of Sanders' campaign funds. We show that hidden donors start giving relatively later, with contributions concentrated around early primaries. We suggest that as presidential campaign strategies change towards wooing smaller donors, more research on what motivates them is necessary.

In Chapter 3, I focus on how events in the election cycle affect political behavior—this time, campaign contributions. I show how the aggregate behavior of campaign contributors is *not* affected as a function of election cycle dynamics and events.

Using the 2016 campaign finance data from the FEC as a daily time-series, I test the hypothesis that if presidential donors are either instrumental or momentum-driven, they will be responsive to events that reveal new information about candidate viability, such as early victories or unexpected upsets in primaries. I employ the sequential segmentation spline method to detect structural breaks while providing smooth estimates between the jumps. I find that on the national level, daily aggregates for any candidate is a slow-moving, smooth process, without any particular critical events. Even when data is disaggregated by state, events expected to create shocks hardly ever do, such as the Iowa caucus or the New Hampshire primary. This is also observed for a preliminary analysis of the 2020 contribution data. I conclude that campaign contributing is, in aggregate, a smooth process, and that donors are neither uniformly instrumental nor momentum-driven.

In all these chapters, my methodological contribution is in taking advantage of extremely large administrative datasets and harnessing the power of the large sample size with nonparametric and semiparametric methods. The rich world of nonparametric and semiparametric methods remains largely untapped by political science studies. I hope to show through this thesis that they can answer new questions, answer old questions in new ways, and provide strong insight that the default linearity model cannot provide.

PUBLISHED CONTENT AND CONTRIBUTIONS

Alvarez, R. Michael, Jonathan N. Katz, and Seo-young Silvia Kim (2020, March). Hidden donors: The censoring problem in U.S. federal campaign finance data. *Election Law Journal: Rules, Politics, and Policy 19*(1), 1–18. doi: 10.1089/elj.2019.0593.

Personal contribution statement: S-Y. S. K. conceived the idea, contributed to the methodology, wrote the software/code, planned and carried out the analysis, designed figures and tables, and contributed to the writing of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

In this thesis, I empirically assess the dynamics of political behavior. More specifically, I analyze what creates—or does not create—change in political participation, such as voting in elections and contributing to campaigns. Through this, I intend to show that paying close attention to dynamics can help answer fundamental questions of political behavior and offer important insights for real-life policies.

Dynamics of individual-level political behavior span across many subjects and intellectual traditions since the so-called "behavioral revolution." In the political psychology tradition, they can speak about the dynamic effects of the media, with negative information or events perceived to be important potentially changing over time. In the political sociology tradition, they can speak to the effects of changing social contexts. In the political economy tradition, they can speak to how the tangible benefits and costs of participation changes over time, as perceived by the individual citizen.

Despite these important connections, the dynamics of individual-level, micro-political behavior have not received the interest they deserve within the field. This is perplexing, because analyzing dynamics is particularly important in political participation, a subtopic of political behavior. Analysis of dynamics complements the scholarship about a mechanism more frequently highlighted: habit formation. The literature has shown that many prominent political behaviors such as turnout are formed by habit—that is, that once a behavior is acted upon, it increases the probability of that behavior in the future. Some examples include Green and Shachar (2000), Gerber et al. (2003), Aldrich et al. (2011), and Coppock and Green (2016).

But it is awkward to answer the why-question in political participation only with the habit theory, because (1) it cannot explain the origin of the behavior, and (2) empirically, behavior often changes. First, habit theory as a standalone answer to participation would be the equivalent of answering the question of "Why are we acting this way?" with "Because we have always done so." There is no doubt that there are evidences showing the observed "habit" is not simply a residue of unexplained variance. But it still does not answer how the initial act came to be.

Moreover, empirically observed political participation is never completely static. This is unsurprising, as the environment in which these behaviors are shaped are constantly shifting. The ebb and flow of media/public interest in politics, changes in institutions or policies, information revelation from events such as elections—all contribute to creating a dynamics that guarantees that there is no fixed behavior at the micro-level that can be fully replicated in the next cycle of interest. For example, in every election cycle, there are strong election-specific characteristics that affect participation in all areas and levels differently.

The evidence for habit formation naturally calls for two separate but related questions: (1) how does the behavior form in the first place, and (2) if the behavior changes, what causes that change? This dissertation is comprised of three essays with independent research questions that come under the umbrella of the second question: I take advantage of aforementioned changes in the environment to empirically assess the dynamics in political behavior—especially participation—and to determine what changes them, if anything.

Again, although political behavior itself is a broad theme, dynamics have been of less interest, relative to more traditional research questions. For example, the role of individual determinants of participation such as education, race, gender, demographics, and economic constraints such as income have been much explored (Wolfinger and Rosenstone, 1980; Verba and Nie, 1987; Verba et al., 1995; Leighley and Vedlitz, 1999; Schlozman et al., 2012; Leighley and Nagler, 2013). Systematic, long-term determinants such as legal and institutional constraints have also been lengthily explored (Rosenstone and Wolfinger, 1978; Powell, 1986; Rosenstone and Hansen, 1993; Wolfinger et al., 2005; Geys, 2006; Burden et al., 2014). But determinants that are more short-term, local, contextual, and therefore less predictable have received less limelight. These include changes to the individual and systematic environment, such as a sudden change in an individual voter's life, a chance decision by local or state-level election administrators, a political candidate's unexpected victory in caucuses and primaries, and so on. This dissertation is meant to be an endeavor to fill some of this gap in the literature.

In Chapter 1, I focus on how non-political life events and election administration policy impact political behavior—specifically, voter turnout. I analyze (1) the effect of moving on turnout over time and (2) how an election administration policy helps with the recovery of lowered turnout by reducing the re-registration burden of movers.

Moving depresses turnout by imposing various costs on voters. However, movers eventually settle down, and such detrimental effects can disappear over time. I analyze these dynamics using USPS data and detailed voter panel data from Orange

County, California. Using a generalized additive model, I show that previously registered voters who move close to the election are significantly less likely to vote (at most -16.2 percentage points), and it takes at least six months on average for turnout to recover. This dip-and-recovery is not observed for within-precinct moves, suggesting that costs of moving matter only when the voter's environment has sufficiently changed. I then evaluate an election administration policy that resolves their re-registration burden. This policy proactively tracks movers, updates their registration records for them, and notifies them by mailings. Using a natural experiment, I find that it is extremely effective in boosting turnout (+5.9 percentage points). This success of a simple, pre-existing, and non-partisan safety net is promising, and I conclude by discussing policy implications.

Chapter 2 (published at *Election Law Journal*, doi: 10.1089/elj.2019.0593, coauthored with R. Michael Alvarez and Jonathan N. Katz) shows how the dynamics of political participation differ between two distinct classes of donors—hidden and visible (from data), based on their amount contributed. In campaign finance we find that an often-overlooked part of the data generating process has great implications about the interpretation of data. This precedes Chapter 3 as it provides some important intuitions as to how the data should be filtered, wrangled, and interpreted for usage.

More specifically, inferences about individual campaign contributors are limited by how the FEC collects and reports data. Only transactions that exceed a cycle-to-date total of $200 are individually disclosed, so that contribution histories of many donors are unobserved. We contrast visible donors and "hidden donors," or small donors who are invisible due to censoring and routinely ignored in existing research. I use the Sanders presidential campaign in 2016, whose unique campaign structure received money only through an intermediary (or conduit) committee. These are governed by stricter disclosure statutes, allowing us to study donors who are normally hidden. For the Sanders campaign, there were seven hidden donors for every visible donor, and altogether, hidden donors were responsible for 33.8% of Sanders' campaign funds. We show that hidden donors start giving relatively later, with contributions concentrated around early primaries. We suggest that as presidential campaign strategies change towards wooing smaller donors, more research on what motivates them is necessary.

In Chapter 3, I focus on how events in the election cycle affect political behavior—this time, campaign contributions. I show how the aggregate behavior of campaign contributors is *not* affected as a function of election cycle dynamics and events.

Using the 2016 campaign finance data from the FEC as a daily time-series, I test the hypothesis that if presidential donors are either instrumental or momentum-driven, they will be responsive to events that reveal new information about candidate viability, such as early victories or unexpected upsets in primaries. In addition, I provide a smooth, de-noised estimate of the underlying campaign dynamics by candidate. For this purpose, I employ the sequential segmentation spline method to detect structural breaks while providing smooth estimates between the jumps. I find that on the national level, daily aggregates for any candidate is a slow-moving, smooth process, without any particular critical events. Even when data is disaggregated by state, events expected to create shocks hardly ever do, such as the Iowa caucus or the New Hampshire primary. This is also observed for a preliminary analysis of the 2020 contribution data. I conclude that campaign contributing is, in aggregate, a smooth process, and that donors are neither uniformly instrumental nor momentum-driven.

In all these chapters, my methodological contribution is in taking advantage of extremely large administrative datasets and harnessing the power of the large sample size with nonparametric and semiparametric methods. The rich world of nonparametric and semiparametric methods remains largely untapped by political science studies, often because much of the literature has relied on surveys, which can fall short in the number of observations necessary to run more flexible regressions. I hope to show through this thesis that they can answer new questions, answer old questions in new ways, and provide strong insight that the default linearity model cannot provide.

Chapter 1

GETTING SETTLED IN YOUR NEW HOME: THE COSTS OF MOVING ON VOTER TURNOUT

1.1 Introduction

Americans are very mobile—every year, at least 10% of the total population moves (United States Census Bureau, 2018a,c). This is an internal migration rate that is almost twice as high as other developed countries' rates. In 2018, with 10.1% of 'mover' rate, more than 30 million people changed residences in the United States. And while moving in itself many not necessarily be a political life event, it has a large impact on people's political participation, particularly by reducing their turnout rate (Squire et al., 1987; Highton, 2000).

Many different types of costs obstruct movers' turnout. A voter whose residence changed has to re-register to vote with her new address and figure out where her new polling place is, which poses an administrative burden. She has to learn the names and the issue positions of her new political representatives if she crosses political district lines. In addition, she may no longer have friends and neighbors in her new community, which can break the social and contextual cues that motivate her to turn out.

On the other hand, a voter is rarely a mover for a long time. She eventually settles into her new home, transitions into a 'stayer,' and overcomes the detrimental shock of moving on political participation. Eventually, she will have more time to re-register, to learn about the new political districts, and to build social ties. Given this, what is the *dynamic* impact of moving on turnout? If there is a significantly negative effect of moving, can we offset the reduced turnout of movers by a policy intervention? I answer these questions using detailed voter panel data from California's Orange County between the 2016 and 2018 elections, appended with data from the United States Postal Services (USPS).

The existing studies have been somewhat limited, by either relying on settings where some types of costs are entirely alleviated by the institutional setting, or by using surveys with a small sample size and rough, self-reported measurements related to moving. This unique administrative dataset on the American electorate—large, accurate, and comprehensive—helps fill the gap in the literature by overcoming measurement constraints present in survey-based research.

Using a generalized additive model, I show that previously registered voters who move close to the election are significantly less likely to vote. Compared to a voter who has lived for a full two years at her new residence, the propensity to vote is at most 16.2 percentage points lower. The detrimental effect is largely transitory, but I find that it takes at least six months on average for turnout to recover. The nonlinear dip-and-recovery pattern is not seen for moves where information costs of voting are nonexistent or very low. This suggests that costs of moving matter only when the voter's environment has sufficiently changed.

Time does help a voter recover from moving. Yet, is there a way to quickly offset the lowered turnout other than simply waiting—that is, speed the convergence? I evaluate an election administration policy designed to retain movers by lifting their burden of having to re-register to vote. I exploit a natural experiment in which because of a policy implemented in California, which I call NCOA automatic voter registration, only some movers were proactively tracked, had their voter registration updated for them, and were notified of the automated change by an official mailing. I find those who received this mailing turned out 5.9 percentage points more. This is a highly effective get out the vote (GOTV) measure. Moreover, this is a simple, non-partisan, and pre-existing policy based on the National Voter Registration Act of 1993, which is very promising for a scale-up. I discuss the policy implications and suggest that restrictions placed on this policy should be lifted if election administrators want to increase turnout of voters who move close to the Election Day.

1.2 Literature

Who Are the Movers?

Annually, at least 10% of Americans move.[1] This is a decreased proportion compared to when Squire et al. (1987) performed their survey (30%), but still a formidable percentage. Table 1.1 shows geographic mobility for the last five years, estimated from the Current Population Survey (CPS), 2013-2018. It also provides more details into where movers are headed to. Two-thirds of internal migration within the United States is same-county moves. One-fifth of moves are same-state, cross-county moves, and about 15% of Americans cross state borders.

It should first be recognized that movers are nonrandom, self-selected group. What

[1] This estimate of 10% is, of course, pre-COVID-19 statistics, and the number is likely to decrease.

	2017-2018	2016-2017	2015-2016	2014-2015	2013-2014
As proportion of population,					
— Movers	10.1	11.0	11.2	11.6	11.5
— Non-movers	89.9	89.0	88.8	88.4	88.5
As proportion of within-country movers,					
— Same-county movers	63.7	64.2	63.8	66.1	67.8
— Same-state, different-county movers	20.8	19.8	22.2	19.2	18.7
— Different state movers	15.5	15.9	14.1	14.7	13.5

Table 1.1: Annual Geographical Mobility Rates, By Type of Movement: 2013-2018

does it mean for a voter to have changed residences?[2] They are likely to be younger and renters (Squire et al., 1987; McDonald, 2008). Some also document that they are more likely to be non-white (McDonald, 2008) and higher educated (Squire et al., 1987). But what makes them move?

	Reason for Moving	2017-2018	2016-2017	2015-2016	2014-2015	2013-2014
1	Wanted new or better home/apartment	16.4	16.0	17.4	15.3	15.8
2	To establish own household	12.6	11.5	12.2	11.0	11.1
3	Other family reason	11.1	11.3	10.5	14.3	13.4
4	New job or job transfer	10.3	9.9	10.8	10.6	9.7
5	Wanted cheaper housing	7.9	8.3	8.2	7.5	9.4
6	Wanted to own home, not rent	7.3	7.3	5.9	5.3	5.6
7	Other housing reason	6.7	7.6	6.7	14.4	12.8
8	To be closer to work/easier commute	5.6	5.5	6.0	4.9	6.2
9	Other reasons	5.0	5.0	4.4	1.5	1.0
10	Change in marital status	4.4	5.1	4.8	5.8	4.9
11	To attend or leave college	2.8	3.0	3.2	0.3	0.5
12	Wanted better neighborhood/less crime	2.6	2.8	3.1	2.9	3.0
13	Health reasons	1.8	1.9	1.8	0.3	0.4
14	To look for work or lost job	1.5	1.3	1.5	1.6	2.1
15	Retired	1.1	0.8	0.7	1.1	0.7
16	Other job-related reason	1.1	0.9	1.2	2.3	2.0
17	Foreclosure/eviction	0.7	1.1	0.9	0.7	1.3
18	Change of climate	0.7	0.5	0.8	0.2	0.1
19	Natural disaster	0.5	0.3	0.1	—	0.0

Table 1.2: Reasons for Moving, 2013-2018, The Census Bureau, Aligned in Descending Order Using 2017-2018 Responses

Table 1.2 displays the reasons cited for moving in percentages from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (United States Census Bureau, 2018b), 2013-2018. In 2017 to 2018, 40.9% of moves were

[2]I use residential mobility instead of geographic mobility to refer to any changes in residence, regardless of distance. According to Gillespie (2016), geographic mobility refers to "long-distance household migration across some administrative or geopolitical boundary," whereas residential mobility refers to "short-distance household mobility." Highton (2000) uses residential mobility vs. community mobility similarly to distinguish changes in residences from changes in communities.

housing-related,[3] 28.1% were family-related,[4] and 18.5% job-related.[5] The first and foremost reason seems to be that the mover wanted a new or better home (16.4%).

As can be seen, factors behind moving are diverse and not dominated by one particular reason. Age is clearly a key higher-level variable, as many events in a person's life cycle such as marriage or job seem to trigger residential mobility. Indeed, the Census Bureau has consistently reported that people in their mid to late twenties have the highest mobility rate, at 65.5% during a five-year interval (Ihrke and Faber, 2012).[6] While I recognize that movers are a self-selected group, the interest of this paper is analyzing the changes to turnout only within the group of movers. Hence, the upcoming analysis is entirely conditional on being a mover within the last two years.

The Costs of Turnout

The analysis of turnout for movers is part of the larger literature of what influences individual turnout, a key indicator and foundation of legitimate democracy. Previous research has shown that there are a variety of reasons why movers will face additional burdens when turning out to vote. Residential mobility can activate one or more of these barriers, as can other factors. While many different costs of voting exist, three types of costs are relevant: convenience costs, information costs, and social costs.

The key idea behind *convenience costs* is that turnout can deteriorate due to barriers between the voter and the polling place (or a vote-by-mail ballot). Because voter registration was not fully automatic anywhere in the U.S. until recently,[7] the major

[3] These are, per exact quote from the data, "wanted own home," "not rent," "wanted new or better home/apartment," "wanted better neighborhood/less crime," "wanted cheaper housing," "foreclosure/eviction," and "other housing reason."

[4] These are change in marital status, to establish own household, and other family reason. Other family reason is the third-largest factor in moving, but an ambiguous category. To supplement this loss in information, the Census Bureau has conducted an analysis of a write-in expansion (Ihrke, 2016). They determined that the common write-in responses for this particular category were such as moved with family member(s), pregnant/had a baby/adoption, assist or take care of family member(s), death of a family member, and move closer to family.

[5] These are new job or job transfer, to look for work or lost job, to be closer to work/easier commute, retired, and other job-related reason.

[6] In sociology, the life-cycle theory of household mobility argues that individuals relocate because they are dissatisfied with their current housing when there is a change in family size and household composition (Rossi, 1980; Gillespie, 2016). Highton and Wolfinger (2001) on the other hand concluded that early adult roles have inconsistent and sometimes negative association with turnout, while age significantly boosts turnout, independent of assuming social roles.

[7] The only exception is the state of Oregon, which first implemented automatic voter registration (AVR) in 2015. Of course, limited forms of AVR exist such as through the DMV or state-level benefits agencies. For a full list of states participating in various levels of AVR, see https://www.ncsl.org/research/elections-and-campaigns/automatic-voter-registration.aspx.

discussion has been focused on institutional barriers to registration (Rosenstone and Wolfinger, 1978; Wolfinger et al., 2005; Ansolabehere and Konisky, 2006; Nickerson, 2014; Street et al., 2015). For movers, convenience costs equate to the burden of re-registering to vote to reflect their recent-most address. Squire et al. (1987) show that the longer people live in their homes, the more likely that they will turn out, attributing it to having more time to newly register. Highton (2000) shows that changing residences accounts more for a drop in turnout than changing communities, suggesting that re-registration costs are high.

Information costs constitute the cost of (1) having to learn about the choices available on the ballot for the new political jurisdiction, and (2) having to learn where, if changed, a new polling place is. By the same vein, voters roll off for unfamiliar choices (Wattenberg et al., 2000) or when they are redistricted/reprecincted (Hayes and McKee, 2009; Brady and McNulty, 2011; Amos et al., 2017). Hansen (2016) finds that crossing municipality borders did not lower turnout, but along with Squire et al. (1987), finds that educated voters are less affected by moving.

Social costs usually indicates a decrease in social rewards from voting, or a decreasing D term. Because turnout is perceived as a desirable behavior, it results in intrinsic satisfaction from social networks (Rosenstone and Hansen, 1993). A higher degree of social "embeddedness" will give higher turnout, while its disruption result in a lower turnout, such as recent loss of a spouse (Hobbs et al., 2014). Moving can also cut social ties and lower turnout: Aldrich et al. (2011) discuss the disruption of voter's habit formation by moving; Gay (2012) shows that mobility experiments had a negative impact on poor voters' turnout whose social relations were severed; Hansen (2016) argues that when there are no convenience costs, absence of evidence for information costs translates into evidence of social costs on turnout.

All in all, moving hampers turnout in a variety of ways. But how does the voter *adapt* to the damage incurred by moving over time? This can depend on residential stability, or how long she has lived in her new residence. Previous literature has been mostly built on surveys, which have many covariates, but often lack power to answer these questions, as the number of survey respondents is limited. Movers are but a small part of the already strained survey sample, and it is plausible that movers are less likely to be contacted for or answer surveys. Highton and Wolfinger (2001) pooled six presidential elections' worth respondents for a sufficiently large sample size within the National Election Studies (NES) for a total of 9,435 respondents. The CPS's sample size is much larger, but it faces the same problem, and its technical documentation briefly touches upon the bias and sample variation from the exclusion of movers.[8] Moreover, the CPS breaks down residential stability into five uneven categories.[9] The only paper that uses administrative data is Hansen (2016), but here residential stability is again presented in uneven, arbitrary categories of 0-30 days, 31-90 days, and 90 days and beyond.

The detailed voter data that I use improves upon these measurement constraints. I have sufficient sample size to fully use the residential stability information without discretizing them into coarser categories. What is more, it is accurate and not reliant upon self-reported moving and turnout information, the latter of which especially can be subject to social desirability bias. To cap it all, the data contains granular information on their old and new residences and political districts, which I can geocode and use to identify different environments in which adaptations can occur heterogeneously.

In particular, I build upon the less-studied aspect of information costs of moving and show that turnout depends on not only how long you have lived there, but also how much your environment changed by moving. I show that dynamic adaptation occurs differently by what information barriers the movers is facing.

Finally, this paper provides a first-ever evaluation of an election administration policy aimed to retain movers, rooted in the National Voter Registration Act of 1993 (NVRA). The policy idea is in fact first suggested in Squire et al. (1987) in discussing how to increase turnout and also briefly discussed in Wolfinger and Highton (1995) and Highton and Wolfinger (1998). However, again due the low power of surveys, the authors were not able to estimate the partial effect of the policy.

1.3 Data and Context

This Section describes the data acquisition and the sample in detail. The data is provided by official election administrators, and it is a combination of official voter registration records and the change-of-address requests filed at the USPS, a rare chance to look in-depth at movers' political participation. In my final sample, out of

[8] In 16-4, *Quality Indicators of Nonsample Errors*, the authors write as follows:

> **Panel nonresponse**. (M)overs are not followed, but the new household members are interviewed ... Out-movers were more likely to be unemployed but more likely to respond compared with in-movers.

[9] CPS distinguishes the length by (1) less than 1 month, (2) 1-6 months, (3) 7-11 months, (4) 1-2 years, (4) 3-4 years, (5) 5 years or longer.

roughly 1.5 million registered voters, I have 102,425 movers in the data. For details on data wrangling and descriptive statistics, see Appendix A.2 and A.3.

Official Voter Database from Orange County, California

Data Acquisition. The voter database is provided by the Orange County Registrar of Voters (OCROV) in California. The acquisition is part of a long-term, larger project built on strong cooperation and trust with election administrators of Orange County, who are the leading public servants in terms of innovative administration practices. Including the 2016 general election snapshot, I have received 156 daily "snapshots" of the data for more than 1.5 million unique voters from April 26, 2018 to December 31, 2018, which cover 89% of business days within the period. This enables me to observe exactly when the voter data changes, and as I will illustrate in Section 1.3, can be used to find movers with details about when and where they moved. This is an unprecedented level of granular details provided for academic studies and provides high accuracy in capturing the dynamic aspect of the voter file.

Given the daily snapshots, I apply entity resolution between them to reverse-engineer transaction logs to the data. More simply, this allows me to look into what records are added, dropped, or changed, on a daily level. By closely observing how the data changes day to day, I am able to extract not only who moves, but when they re-register and through what means. These details contain insight into what the voter is doing, and what policies are affecting the re-registration or voting decisions. No other existing compilation of voter data provides such information, and none certainly have been augmented with the National Change of Address (NCOA) dataset. For more details on the data, refer to Kim et al. (2019).

Asides from being able to determine mover status, the voter data carries many useful covariates. It carries full street-level addresses of old and new residences which can be geocoded into specific latitude/longitudes and accompanying political districts. It also has date of birth, original place of birth, partisan affiliation, precinct assignment, political district assignments such as Congressional districts, first and most recent voter registration date, and the reasons for the last update of the registration. Most of all, it has accurate records of voting history that is not inflated by social desirability bias as in surveys. I also augment the data with imputed gender and race.[10]

[10] While there is a 'gender' entry in the Orange County dataset, most of the entries are missing, as it is not an official field in the voter registration document. R package `gender` (Mullen, 2018; Blevins and Mullen, 2015) of rOpenSci project helps infer gender by first names and the Social Security Administration's yearly dataset. If there is an entered gender or a prefix (e.g. 'Mr.'), it

Differentiating Movers. Movers are defined as those who have moved after the 2016 general Election Day and before the 2018 general Election Day, up to October 31, 2018—that is, those who have moved within two years of Election Day, as in Squire et al. (1987).[11] My sample is limited to in-county movers, which enables full visibility of their voting history for the dependent variable, which is the 2018 general election turnout. Note that in-county movers form the lion's share of movers at more than 63% as seen in Table 1.1, giving sufficient number of cases to analyze movers' political behavior.

I differentiate movers from stayers by closely monitoring the changes in the residential addresses of voters. This may seem initially odd because not all movers will voluntarily report their new address to the Registrar, who maintains the countywide voter database. Indeed, while a voter can voluntarily re-register to vote with their new address or visit the Department of Motor Vehicles and update their registration information there, not all voters will do either of these things. The OCROV writes in its website as follows:

> *Unfortunately most people that move notify their banks, car lenders, family and friends - even magazine subscriptions before they change their voter registration. You can help us to keep our voter lists up-to-date by taking a few minutes to notify us of changes in your life.*

In Orange County, it is possible to detect movers just by observing the Registrar's changing data due to a particular election administration practice called NCOA processing. While originally designed as voter list maintenance activity, the processing allows the Registrar to detect movers in advance, even when they do not voluntarily inform the election administrators. Section 1.3 illustrates this.

National Change-of-Address (NCOA) Data

The NVRA, while mainly about offering more opportunities to register to vote, also requires the states to maintain accurate, up-to-date database. To achieve this goal, states can use the permanent change-of-address requests submitted to the USPS. Individuals submit such requests to the USPS because then the agency will forward

---

overrides the inferred gender. For race/ethnicity, R package `wru` uses surname and geolocation to infer race (Khanna et al., 2017; Imai and Khanna, 2016) using Bayesian updating. The inference is primarily performed on the census block level.

[11] While ideally I would like to observe movers for a longer period, this is restricted because the data collection began in 2018.

mail from their old residence to the new one. USPS maintains the last 48 months' requests, called NCOA data, which approximates 160 million change-of-address (COA) records with accurate old and new residences, as well as when the individual moved and requested the data.

While not strictly required to use the NCOA dataset, California's Election Code requires the Secretary of State to match the statewide voter file to NCOA data (called NCOA processing). If existing voters have changed addresses, the Secretary will then transfer the data to relevant counties. For protection of voters' privacy, the change-of-address request dates are coarsened to the month of the move instead of the exact date, and the data is disseminated on a monthly basis. Within the Orange County data in question, there were two major NCOA processing, respectively, on July 26, 2018, and December 20, 2018, to monitor which voters moved each month. Again, note that while this is coarsened, this is still an unprecedented level of detailed data compared to the previous literature.

If the address change is within the same county, the voter file is *automatically* updated with a forwardable address confirmation mailing sent to the voter. If this update was a mistake because there was an error or the move is temporary, the voter can inform the Registrar using a prepaid postage or a phone call. For the full wording of the various legal statutes and California's NVRA guidelines, see Appendix A.1.

In sum, by examining the changing data, it is possible to detect between the 2016-2018 elections (1) all voters who have voluntarily reported their change of address to the Registrar *before* any NCOA processing, and (2) all voters who did not voluntarily report, but filed a change of address with the USPS, thereby being detected through NCOA processing, and ultimately had their address updated within the Registrar's database.[12] I classify them altogether as movers between the 2016-2018 elections. Again, this is a valuable addition by the NCOA dataset.

Validated Movers. The NCOA data enables the classification of movers, as the Registrar performed NCOA processing up to movers of November 2018. The only undetected movers would be those who did not voluntarily inform the Registrar, the DMV, nor the USPS, and did not vote in either the primary or general election with the updated address. While this is theoretically possible, I limit my sample of

[12]In Orange County, another source of third party address changes are consumer credit reporting agencies, which a county can use to verify a voter's residence per CA Elec Code § 2227 (2017). However, as we detail above, we do not use this class of movers, which form a very small percentage.

movers to those *who requested a change of address with the USPS*, independent of informing the Registrar.

This is a validation measure which ensures that the measured mobility is not a correction of incorrect data entries/typos. Suppose that a voter has lived at 110 N. California Boulevard, and the address changes to 1100 N. California Boulevard in the voter data. Or suppose that I see a voter's record change from 200 S. Main Street to 200 N. Main Street. Is this a real change in physical residences, or a modification in data with no entailing real-world change? By itself it is difficult to discern. However, even if the old and new addresses look similar, if I see that in the USPS data that the voter has requested a change of address, I can be assured that there is a true change in residences.

In addition, I am able to get an accurate measure of when a voter has moved, which is something not available in the voter database. For instance, if a voter voluntarily reports a new address to the Registrar in October 2018 but does not have any record in the NCOA database, it is incorrect to impute her moving date to October 2018 because it would be confounded by the fact that closer to the election, voters will remember to re-register more. As this paper's interest is in the dynamic effect of moving, ascertaining the timing of the move is vital.

While these two points are strong pros in limiting the sample of movers to those who officially requested the change of address, one point should be noted. There is no study to my knowledge about who chooses to request the change of address as opposed to those who do not. Therefore, it is unclear whether and if so, how inference will be affected by the decision to limit the analysis to validated movers. The data itself is certainly popular—for example, the Census Bureau has used NCOA data to supplement the tracking of migrations (Hogan, 2008). The younger electorate such as teenagers may be underrepresented, as they are likely to have little mail in their name. Not many more educated guesses are possible.[13]

The OCROV used a USPS-licensed vendor to provide me with the augmented data. Some 2 out of 3 movers that I had classified could be matched to the NCOA dataset, with the same set of old and new addresses as can be found in the voter file. In the end I have around 100,000 voters.

[13] Comparison within the available voter file is not a valid comparison, since the baseline population will be then those who voluntarily reports to the Registrar or those who vote without the change of address requests.

1.4 Dynamic Impact of Moving

In this Section, I describe how I estimate how voters adapt to moving—the impact of moving over time. In particular, I show how the impact varies by different hurdles of moving, defined by different informational environments.

Methods

The dependent variable used in this paper is turnout for 2018 general election. Residential stability, or months lived at the new residence, is the key explanatory variable. This is a continuous variable ranging from 1 to 24 months lived in the new residence before the 2018 General Election. Figure 1.1 shows the distribution of number of movers by each value. For example, those who have lived four months at their new home by November Election Day are essentially movers in July 2018. There is a seasonality as summer is the most popular time for moving, and January the least popular month.

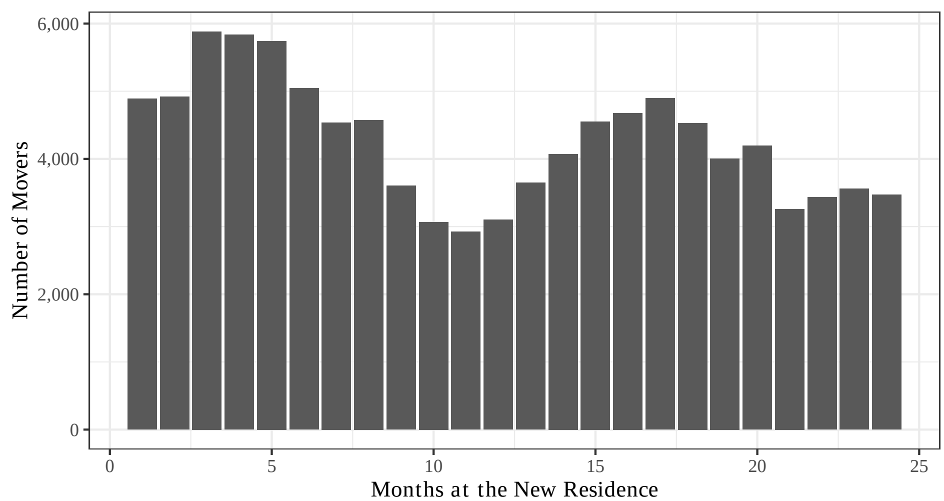


Figure 1.1: Distribution of Months at the New Residence

Ideally within a controlled environment, a researcher could randomize (1) the timing of voter's moves given a particular election date and (2) the various costs that she may potentially face. As this is not possible, the data is observational. An experimental research design is also difficult, since ignorability of treatment assignment is likely to be violated with voters who have moved in various points in time. However, the cross-sectional analyses presented in this Section, which show the conditional prediction of turnout given the values of the independent variables, still provide insight into a problem that practically does not have an experimental design. In addition, it is likely that voters do not self-select into choosing their timing of residential move based on the decision of whether to turn out to vote in a future election. Hence in this paper, I assume that the cross-sectional variation between the 24 months is conditional on the observables. For example, the seasonality of moving is age-dependent, hence I control for age.

In the following Subsections, I show how turnout is depressed by moving by semiparametric regressions. More specifically, I use generalized additive models (GAMs) which can flexibly fit continuous variables. The functional form associated with costs that change over time is not known, and there is no reason a priori to assume that the relationship will be linear. Using GAMs can expand the analysis beyond the default linearity assumption while keeping the interpretability of an additive model. In addition, while parametric handling of nonlinearity such as log transformations or polynomial regressions will impose a global function, GAMs are local and more flexible (Beck and Jackman, 1998). The results indeed show that the relationship between residential stability and turnout is strongly nonlinear and need a local specification.

GAMs are fitted by the R package `mgcv` using thin plate regression splines (TPRS) and fitted by restricted maximum likelihood (REML) method. The smoothness is automatically selected using a penalized spline approach and are checked with standard diagnostics to see if the basis dimensions for smooth functions are adequate (Wood, 2017). TPRS have one basis function per data point, but the basis functions are reduced by eigendecomposition, hence fitting on a reduced problem (Wood, 2003).

Because I show how costs of moving depends on different environments, I further use factor-by-curves, which is an extension of simple GAMs, by allowing the smoothed continuous variable to 'interact' with a specified factor. In other words, the smoothing curves are fitted separately for each level of the factor, subject to a centering constraint. For more reference on GAMs, see Hastie (1992), Beck and Jackman (1998), Keele (2008), Wood (2011), and Wood (2017). All continuous variables are smoothed. The main independent variable is the months lived at the new residence, which is an interval variable, taking discrete values from 1 to 24. Given this, the usage of GAM provides a unique specification into the relationship between residential stability and turnout, further proving their usefulness in social science research.

Differentiating Information Costs

Not all moves are created equal—a move across the street is not equal to a move thirty miles away from your original home. As explored in the literature, these situations pose different costs on turnout. While it is impossible to fully net out convenience and social costs, it is possible to differentiate the information environments that voters are facing. For factor-by-curves, I use the following categorical proxies for information costs:

1. The voter moved but stayed at the same street address (e.g. only changed units within the same apartment complex), labeled *Same Address* (3.3%);
2. The voter moved within the same precinct and with the same polling place, labeled *Same Precinct* (5.7%);
3. The voter crossed precinct boundaries but within the same state-level or local political districts,[14] and with the same polling place, labeled *Same Subdistricts* (8.3%);
4. The voter crossed some local or state-level districts or her polling place changed, but within the same Congressional district, labeled *Same Congressional District* (46.3%);
5. The voter crossed Congressional district lines: *Different Congressional District* (36.5%).

The variable captures five different situations a mover can face, with progressively larger burdens. Note that convenience cost should theoretically apply to all movers, and the social costs are not independent with information costs. For example, if social costs are roughly proxied by distance moved, if a voter moves within the same apartment, there are certainly no information or social costs involved, because the distance moved is zero. If a voter moves a mile or more, it is highly likely that she crosses precinct boundaries. If she moves sufficiently far away, she is also likely to cross local, state, and federal district boundaries. Therefore, I simply term this the difference in 'environment,' or (generic) costs of moving.

The logistic GAM with factor-by-curves used in this paper is as follows, for $i = 1, \cdots, n$:

[14] These include state Senate districts, state Assembly districts, supervisorial districts, and ward divisions.

$$\Pr(y_i = 1|x_i, \mathbf{X}_i^1, \mathbf{X}_i^2, z_i) = \left(1 + \left(\exp\left(\beta_0 + f_z(x_i)z_i + \gamma_z z_i + \sum_{j=1}^{J} \beta_j X_{i,j}^1 + \sum_{k=1}^{K} g_k(X_{i,k}^2)\right)\right)^{-1}\right)^{-1} \tag{1.1}$$

y_i is turnout in the 2018 general election, x is residential stability (months lived at the new residence), and z is the information cost category. $\mathbf{X}^1$ is the set of variables that are linearly added without smoothing, such as dummies for race/ethnicity or gender. $\mathbf{X}^2$ is the set of variables that are additive but smoothed, such as age, and g_k for $k = 1, \cdots, K$ are nonparametric smooth functions fitted for each of these variables. γ is separately included with the indicator variables z_i for information cost categories $z = 1, \cdots, 5$, subject to $\sum_{z=1}^{5} z_i = 0$, because the estimation of smooth functions is subject to a centered constraint. Note that the model is essentially specifying that the residential stability information should not be pooled for different environments, while for other variables such as age, voters are exchangeable across environments.

The following covariates are controlled for: number of times moved in a 24-month period, straight Euclidean distance from voter's home to the designated polling place (Gimpel and Schuknecht, 2003; Dyck and Gimpel, 2005; McNulty et al., 2009), permanent absentee voting status (Gronke et al., 2007), age, inferred gender[15] and race, partisan affiliation, 2016 general turnout, census block group-level median household income of both their old and new residences, whether the voter was born abroad (e.g. a naturalized citizen), and the Congressional district of the new residence. The Congressional district was added to address the fact that there were some hotly contested House races in the 2018 general election, as opposed to other districts where landslides were predicted.

Results

Heterogeneity by Costs of Moving

Figure 1.2 shows the relationship between residential stability and turnout by each information cost category, holding other covariates fixed.[16] What is immediately

[15] 3.8% of voters have an ambiguous gender that cannot be inferred from the first name. In these cases, these are treated as unknown and as a baseline group, instead of dropping them from the sample.

[16] The dependent variable's values displayed in conditional plots usually use the median value for continuous variables and mode values for categorical variables, as each covariate has to have

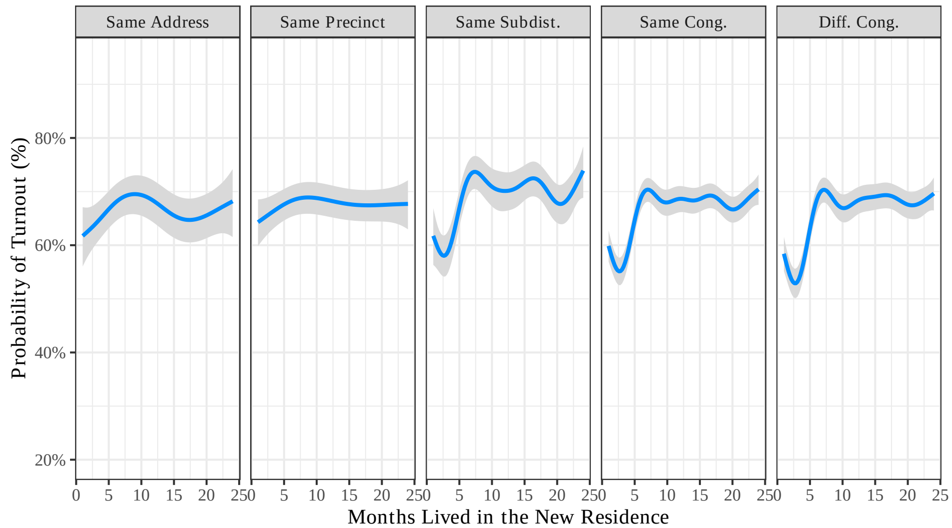


Figure 1.2: Fitted Smooth Functions for Residential Stability by Information Cost

striking is that there is a clear nonlinear trend for movers with medium to high information costs ($p < 0.001$). There is a severe dip in turnout for movers close to the election if they at least crossed precinct lines. By the projected propensity to vote, an average voter (regardless of information environment) who has lived fully two years at her new residence will vote 67.6% of the time. An average voter who has lived only two months will vote 51.4%, resulting in maximum difference of 16.2 percentage points. Turnout steadily climbs before it reaches a plateau with slight variance, displaying similar patterns for the last three panels.

For movers with sufficient change in their informational environment, it takes at least six months or more for their turnout to climb to the level of stayers (about 70%). The slight uptick of turnout for movers of October 2018 is puzzling, but it is likely because some of them were eligible to vote in their original precinct and polling place. By California election statutes, movers within 14 days of the Election Day can choose to vote either at their old or new residence's polling place.[17]

On the other hand, movers who have moved but stayed at the same street address ($p =$ 0.0924) or within the same precinct ($p = 0.3004$) do not show strong associations between their time at the new residence and turnout. Given that the sample size for

a fixed value. The predictions are not necessarily aligned with the average predicted value of the dependent variable. For more on visualizing the relationship between a single independent variable and the dependent variable, see (Breheny and Burchett, 2013). For the full set of values used for the displayed figures, see Appendix **??**.

[17] I do not interpret the bumps after turnout reaches a plateau—these are less likely to be 'interesting local features' (Beck and Jackman, 1998).

same-address movers is 3,000 or more voters, and the sample size for same-precinct movers at least 5,000, it seems to be the case that for those with low to no information cost, there is no strong evidence for turnout change over time. Crossing precinct lines or more seem to be a necessary condition to observe the dip and recovery.

		Imperfect Placebo			*Placebo Tests*		
	General 2018	General 2016	Primary 2016	General 2014	Primary 2014	General 2012	Primary 2012
A. Smooth terms (effective degrees of freedom / residual degrees of freedom)							
Res. Stability × Same Address	3.353	1.148*	2.698	2.655	1.030	1.003	1.002
	4.153	1.145	3.358	3.305	1.059	1.006	1.004
Res. Stability × Same Precinct	2.684	1.145	3.391	1.931	1.004	1.003	1.933
	3.341	1.276	4.206	2.418	1.007	1.007	2.421
Res. Stability × Same Subdist.	7.205***	1.034	2.305	1.015	2.007	3.971	3.473**
	8.232	1.067	2.878	1.030	1.792	4.899	4.305
Res. Stability × Same Cong.	8.446***	3.810***	3.910*	1.013	1.792*	2.296	1.002
	8.913	4.709	4.830	1.027	2.241	2.868	1.004
Res. Stability × Diff. Cong.	8.428***	3.177***	1.235	2.552	1.015	1.003*	1.002
	8.908	3.945	1.433	3.183	1.029	1.007	1.003
Distance Moved	1.139*	1.009*	2.828	2.270	1.006	1.005	1.002
	1.139	1.017	3.580	2.887	1.011	1.011	1.006
Age	7.991***	7.378***	7.314***	7.035***	7.603***	7.655***	8.115***
	8.655	8.227	8.200	7.998	8.473	8.365	8.752
Distance to Poll	4.947*	1.037	2.287	1.021	2.102*	1.011	1.814
	6.035	1.074	2.910	1.041	2.676	1.021	2.303
Old Residence's	8.623***	8.140***	5.252**	8.377***	8.060***	3.017***	7.732***
Neighborhood Income	8.955	8.786	6.318	8.882	8.754	3.808	8.573
New Residence's	8.357***	7.866***	1.009*	6.994**	3.103**	2.371*	1.003
Neighborhood Income	8.881	8.655	1.019	8.037	3.891	3.029	1.006
B. Parametric coefficients (estimate / standard error)							
Same Precinct	0.053	0.043	0.003	0.303***	0.163	0.134	0.094
	(0.050)	(0.058)	(0.053)	(0.064)	(0.084)	(0.069)	(0.080)
Same Subdist.	0.113*	0.156**	0.063	0.395***	0.170*	0.257***	0.064
	(0.048)	(0.054)	(0.050)	(0.060)	(0.079)	(0.064)	(0.075)
Same Cong.	−0.004	0.133**	0.019	0.312***	0.152*	0.201***	0.086
	(0.042)	(0.047)	(0.046)	(0.056)	(0.071)	(0.056)	(0.068)
Diff. Cong.	−0.026	0.104*	−0.018	0.325***	0.158*	0.183**	0.031
	(0.045)	(0.050)	(0.050)	(0.059)	(0.074)	(0.060)	(0.071)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	100,389	96,195	83,977	71,411	69,104	65,388	59,914
Adjusted R^2	0.159	0.052	0.090	0.134	0.165	0.067	0.164
Log Likelihood	−57,722.160	−45,323.170	−53,401.650	−41,960.780	−26,208.670	−34,083.580	−27,761.590
UBRE	57,859.880	45,412.880	53,480.550	42,045.840	26,277.130	34,150.990	27,832.550

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 1.3: Generalized Additive Model Results, Full Sample

Placebo Tests

For a robustness check, I perform placebo tests—that is, the same analysis on data where there was no 'intervention,' which should theoretically yield null results. An intuitive placebo test to check for sample self-selection is to not use turnout of 2018 general election but use turnout of previous elections. Because residential changes took place after the 2016 general election, theoretically, the imposed costs should not affect turnout for previous elections. Table 1.3 shows the results of the main regression and its placebo checks using primary and general elections of 2016, 2014, and 2012. The table contains both the parametric coefficients and the smooth terms, and covariates are excluded for brevity—the full table can be found in Appendix A.4.

The first noticeable results from placebo tests are the sample-selection issue of information cost categories. From Figure 1.2, comparative to movers within the same subdistricts, it appears that the mean of predicted turnout is lower for those with higher information costs. However, the placebo tests show that voters seem to be self-selecting into these different environments. A voter who moves within the same apartment complex can be a different voter than those who cross Congressional district lines, although what this means substantively is uncertain.

But having partialed out the difference in means by a centering constraint, do fitted smooth functions also pass placebo tests? Although not perfectly for every cost and past election combination, the results seem reasonable for elections *before* the 2016 general election. For the 2016 general, there is a strong nonlinear relationship documented for movers with high information costs, on a smaller scale (effective degrees of freedom ≈ 3) and in the opposite direction, if linearly fitted.

Why is this the case? The likely answer is that the 2016 general election is not a great placebo for this model, because costs may also incur *right before* the voter moves. Note that a voter who has lived in her new residence for twenty-four months before the 2018 election is essentially a voter who has moved in the thicket of the 2016 general election month. Take distraction costs. A voter may be more distracted *before* she moves, rather than *after*, because it would take at least a few months to search make a housing decision and to search for appropriate housing. In addition, a voter who will soon move has no objective benefit to reap from a community that she will soon leave. In that case, the information costs of learning about local issues can outweigh the benefits of voting (Dowding et al., 2012). Figure 1.3 shows the descriptive proportion of turnout for respectively the 2016 and 2018 elections by residential stability. The placebo tests excluding the 2016 general election tests generally seem to pass.

Disentangling Distance Moved

Although the distance moved is adjusted for in the result above, it should be still clarified—is it distance that is actually driving the dip and recovery of turnout for voters with high information costs?

I test whether there is still a dynamic relationship when there is little to no distance moved. To do this, I define a sufficiently small neighborhood by distance, and run the

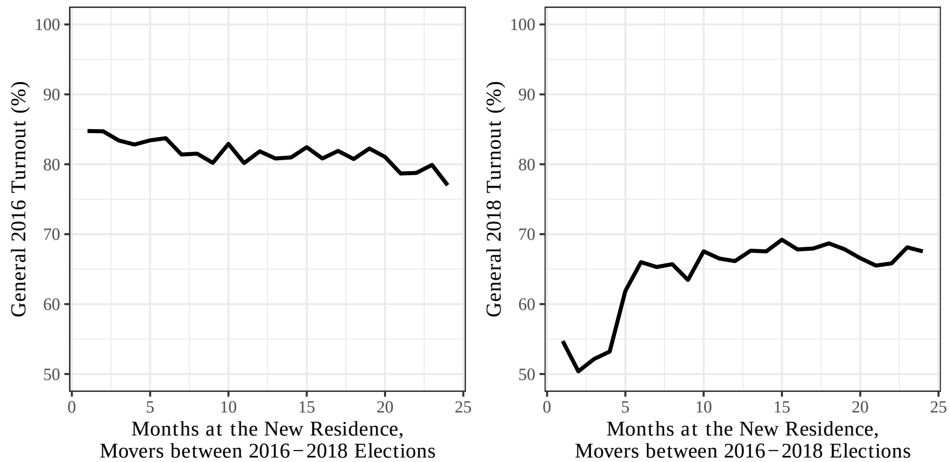


Figure 1.3: Turnout of Movers, 2016 and 2018 General Elections

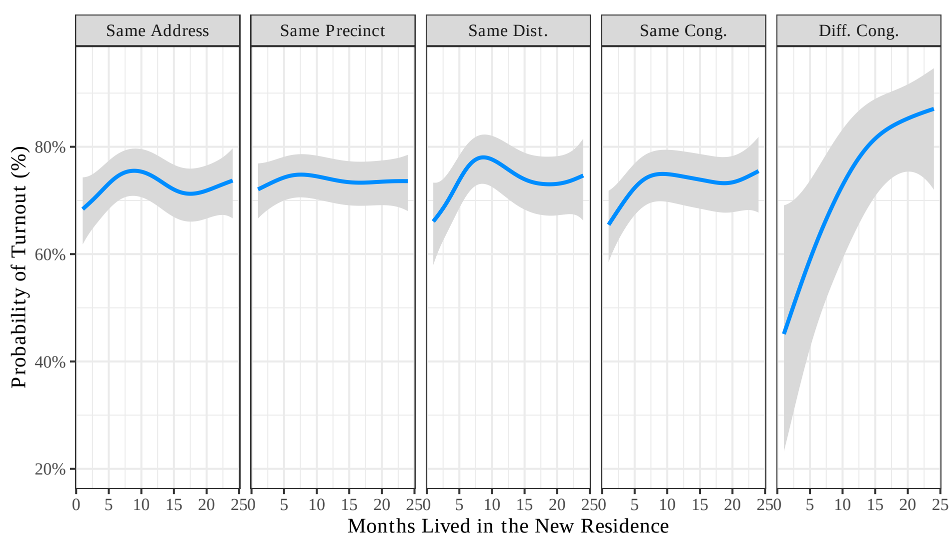


Figure 1.4: Fitted Smooth Functions for Residential Stability by Information Cost, Movers within Half-Mile

	General 2018	*Imperfect Placebo* General 2016	*Placebo Tests* Primary 2016	General 2014	Primary 2014	General 2012	Primary 2012
A. Smooth terms (effective degrees of freedom / residual degrees of freedom)							
Res. Stability × Same Address	3.310	1.708*	2.771	2.732	1.148	1.002	1.000
	4.102	1.004	3.448	3.400	1.281	1.003	1.001
Res. Stability × Same Precinct	2.559	1.004	3.169	1.930	3.181	1.541	3.155
	3.189	1.008	3.937	2.416	3.954	1.904	3.924
Res. Stability × Same Subdist.	3.613*	1.113*	2.019	4.075	1.001***	1.001	3.368**
	4.472	1.218	2.522	6.027	1.001	1.002	4.177
Res. Stability × Same Cong.	3.003*	1.001	1.438	1.001	1.001	1.504	1.001
	3.730	1.003	1.756	1.002	1.002	1.853	1.001
Res. Stability × Diff. Cong.	1.564**	1.003	1.002	1.002	1.808	1.003	1.084
	1.939	1.006	1.005	1.004	2.274	1.005	1.163
Age	6.009***	4.398***	4.492***	1.085***	4.788***	6.285***	3.961***
	7.102	5.392	5.510	1.166	5.856	7.297	4.907
Distance to Poll	1.016	1.003	1.004	1.003	1.001	1.000	1.001
	1.031	1.005	1.008	1.005	1.003	1.001	1.002
Old Residence's	5.223*	6.960**	3.146	3.645	1.002	1.000*	1.957
Neighborhood Income	6.297	7.998	3.944	4.534	1.003	1.001	2.484
B. Parametric coefficients (estimate / standard error)							
Same Precinct	0.057	0.053	−0.026	0.262***	0.120	0.118	0.031
	(0.058)	(0.068)	(0.061)	(0.073)	(0.098)	(0.080)	(0.094)
Same Subdist.	0.053	0.184*	0.109	0.253**	−0.093	0.238*	−0.089
	(0.077)	(0.089)	(0.079)	(0.093)	(0.130)	(0.104)	(0.122)
Same Cong.	0.009	0.305**	−0.006	0.266**	−0.024	0.312**	−0.041
	(0.080)	(0.093)	(0.083)	(0.097)	(0.134)	(0.109)	(0.127)
Diff. Cong.	0.001	−0.217	−0.172	−0.254	−1.032	0.086	0.333
	(0.232)	(0.241)	(0.239)	(0.296)	(0.569)	(0.286)	(0.326)
Distance Moved	−0.020	−0.106	0.185	0.407*	0.591*	−0.139	0.505
	(0.171)	(0.200)	(0.175)	(0.201)	(0.277)	(0.230)	(0.264)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	13,150	12,470	10,680	8,877	8,529	8,008	7,307
Adjusted R^2	0.157	0.051	0.090	0.137	0.173	0.068	0.186
Log Likelihood	−7,578.712	−5,909.561	−6,823.539	−5,199.565	−3,164.015	−4,214.901	−3,331.615
UBRE	7,619.864	5,942.088	6,855.060	5,224.926	3,178.778	4,238.004	3,348.890

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 1.4: Generalized Additive Model Results, Subsample of Movers within Half Mile

same generalized additive model with movers who have moved to a nearby house. Specifically, I use a walkable neighborhood of 0.5-mile distance, which is the mean walking value for those who have reported at least one walking trip daily (Yang and Diez-Roux, 2012). Assuming that a street block equals one-ninth of a mile, this equates to walking about four blocks. While other thresholds are possible, I use walking distance because (1) car ownership is not observed, and (2) larger travel distance by vehicles can be vary in their costs by time of the day and the available infrastructure surrounding a household.

The sample of movers who have moved less than half a mile is about 13.1% (13,150 voters). Given that now the distance is smaller, the sample is smaller, and that the distance moved has effective degrees of freedom[18] that does not exceed 2 in Table

[18] This is also called equivalent degrees of freedom (Beck and Jackman, 1998).

1.3, the distance moved is now treated as parametric (linear). I also exclude the neighborhood's median household income for the new household. This is because while in the full sample, the Pearson correlation coefficient was about 0.53, now in the small move sample the correlation coefficient is 0.98.

Figure 1.4 shows the fitted smooth functions by each information cost category. Again, while the smooth functions do not seem to have strong relationship for movers within the same apartment ($p = 0.089$) or within the same precinct ($p = 0.568$), those with higher information costs have clearer dynamic trends ($p < 0.05$). That is, if the voter crosses precinct lines or more, *even within the same county*, a recovery of turnout is documented. This shows that independent of distance moved, the detrimental effects of turnout apply heterogeneously by informational environments. Changing the cutoff from half-mile to values such as one mile or three miles produce the same results.

Therefore, the conclusion holds that sufficient changes in information seem to be a necessary condition to observe a dip and recovery pattern, regardless of distance moved. One plausible explanation is how voters subjectively perceive voting costs to be. For a small enough move, such as within the apartment or within the precinct, costs of moving may not matter so much, causing the voter to pay little attention to acclimating to the change.

For instance, for a within-apartment mover, a voter may be able to pick up her vote-by-mail ballot from a common mail room without necessarily having to re-register to have it forwarded. Or perhaps she could go to the same polling place, state her address up to the street-level and not the unit, and still be able to vote. Similarly within the precinct, the voter may simply be able to visit her polling place and state that she moved but is still within the precinct boundaries, or retrieve her ballot easily from her old home.[19]

Also note that in Figure 1.2 the nonlinear pattern in movers who cross precincts, local and state-level districts, and Congressional districts all show a very similar pattern. It seems to be that once the voter faces an environment in which some search cost must occur, how fast the voter adapts does not seem to be dependent much on varying degrees of information cost. All in all, the results show how it might not be entirely accurate to impose a global and additive functional form to different types of costs of moving.

[19] Ballots are not forwarded even with the change-of-address requests at the USPS.

1.5 Mitigating Re-registration Costs: Policy Evaluation

In Section 1.4 I have shown that turnout is significantly low for voters who move close to the election, especially if the movers are facing a big enough move in terms of changed informational environment. Unfortunately, in terms of policy interventions, it is not usually possible to manipulate the voters' change in environment, as election administrators have no voice in where the voters move to, what friends of neighbors they make, and so on. If movers stay six months or more at their new home, their turnout levels seem to recover almost to the turnout level of stayers. But of course, election administrators cannot manipulate how long a voter has lived at the new residence on Election Day.

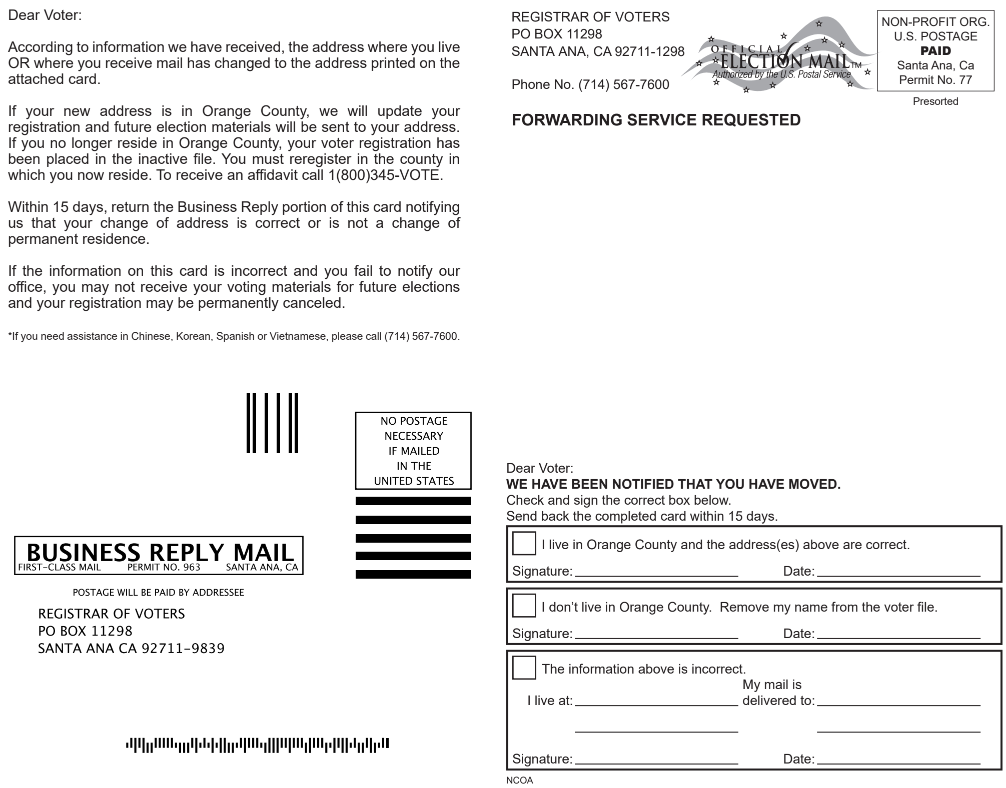
Is there any way to reliably help movers turn out, other than simply waiting? Very few papers have investigated how to boost turnout of movers. McDonald (2008) shows that portable statewide registration, which permits in-state movers to be registered and to be able to vote on Election Day, increases turnout by 2.4 percentage points. How about movers within-state, or even within the county? Here I evaluate an election administration policy first suggested in Squire et al. (1987) and subsequently implemented through the National Voter Registration Act of 1993, when evaluated, was found to greatly improve movers' turnout.

Background

The National Voter Registration Act of 1993 decrees that states can use the USPS's NCOA records to determine whether the voter still lives in the address or have moved. California actively uses this data to maintain its voter rolls up-to-date (California Secretary of State, 2019). In particular, when a residential move is detected via NCOA processing, election administrators have to send a *physical mail* to the voter to notify that their new address will be used for voting purposes, by CA Elec Code § 2225 (2017). This notification must be, substantially, in the following form (see Appendix A.1 for full statutes):

> "We have received notification that you have moved to a new residence address in California. You will be registered to vote at your new address unless you notify our office within 15 days that the address to which this card was mailed is not a change of your permanent residence. You must notify our office by either returning the attached postage-paid postcard, or by calling toll free. If this is not a permanent residence, and if you

do not notify us within 15 days, you may be required to provide proof of your residence address in order to vote at future elections."


Dear Voter:

According to information we have received, the address where you live OR where you receive mail has changed to the address printed on the attached card.

If your new address is in Orange County, we will update your registration and future election materials will be sent to your address. If you no longer reside in Orange County, your voter registration has been placed in the inactive file. You must reregister in the county in which you now reside. To receive an affidavit call 1(800)345-VOTE.

Within 15 days, return the Business Reply portion of this card notifying us that your change of address is correct or is not a change of permanent residence.

If the information on this card is incorrect and you fail to notify our office, you may not receive your voting materials for future elections and your registration may be permanently canceled.

*If you need assistance in Chinese, Korean, Spanish or Vietnamese, please call (714) 567-7600.

REGISTRAR OF VOTERS
PO BOX 11298
SANTA ANA, CA 92711-1298

Phone No. (714) 567-7600

OFFICIAL ELECTION MAIL™
Authorized by the U.S. Postal Service

NON-PROFIT ORG.
U.S. POSTAGE
PAID
Santa Ana, Ca
Permit No. 77

Presorted

FORWARDING SERVICE REQUESTED

NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY MAIL
FIRST-CLASS MAIL PERMIT NO. 963 SANTA ANA, CA

POSTAGE WILL BE PAID BY ADDRESSEE

REGISTRAR OF VOTERS
PO BOX 11298
SANTA ANA CA 92711-9839

Dear Voter:
WE HAVE BEEN NOTIFIED THAT YOU HAVE MOVED.
Check and sign the correct box below.
Send back the completed card within 15 days.

☐ I live in Orange County and the address(es) above are correct.
Signature: ____________ Date: ____________

☐ I don't live in Orange County. Remove my name from the voter file.
Signature: ____________ Date: ____________

☐ The information above is incorrect.
I live at: ____________ My mail is delivered to: ____________
Signature: ____________ Date: ____________

NCOA

Figure 1.5: NCOA Mailing of Orange County, California, Front and Back

This policy, henceforth *NCOA automatic voter registration* or NCOA treatment for simplicity,[20] serves two purposes. First, it rids the movers of their convenience costs of re-registering to vote. Second, it reminds them about the upcoming election, acting as a 'nudge.' Figure 1.5 shows the mailing sent out to the movers in Orange County, in its original form. Most importantly, if the voter actively does not deny moving to a new residence or notify the Registrar of a new mailing address, the voter file will reflect the new address. This is true even if the voter does not return the mailing checked with "I live in Orange County and the address(es) are correct," acknowledging the movement. Therefore, if there is no counteraction, the USPS information is treated as a true move. If the voter is a permanent absentee voter, the mail ballots will be sent to the new address. If the voter has crossed precinct

[20] Again, I would like to emphasize that this is a policy that needs both NCOA processing and automatic voter registration (AVR).

boundaries, the new polling place's roster will have her name printed, and not the old one.

This is an extraordinarily proactive measure by the Registrar and can potentially boost turnout. However, the efficacy of the policy has never been measured so far. How effective is this policy in stimulating movers to turn out? And how can we estimate it?

Natural Experiment

A natural experiment is available as follows. As explained in Subsection 1.3, no list maintenance is performed 90 days before the Election Day. In Orange County, the last NCOA automatic voter registration was performed on July 26th, 2018, up to the movers who moved before June 15, 2018, as the Secretary of State's office obtains and disseminates NCOA data in the middle of the month. This discontinuity creates an interesting quasi-experimental opportunity for policy evaluation, as those who have moved in the *latter half of June* and beyond did not get the NCOA mailings, as opposed to those who moved in the *early half of June*. This is a quasi- regression discontinuity design with NCOA mailing as an intervention. Although I cannot determine the exact date of the residential move, I can parse NCOA treatment from the voter file changes, which reveals whether the voter moved later than the cutoff.

The full sample for policy evaluation here is the set of June movers who filed a change of address. Note that the treatment group is the set of movers from June 1 to June 14 who *have not voluntarily updated their registration records until late July.* If the voter has already reported having changed residences to the Registrar so that the voter roll is already up-to-date, the mailing is not sent out. Therefore all others, including those who moved in early June and possibly disclosed it to the Registrar before July, are put to a control group.[21] If anything, this will estimate a lower bound of the effect of the policy, as those who voluntarily inform the Registrar are more likely to vote.

Results

The estimation of the policy effect is straightforward. The independent variable of interest is the NCOA treatment, which is binary, and I use the same set of covariates used in Section 1.4. Because the logistics regression result and the entailing average marginal effect is almost identical to the effect estimated by a linear probability

[21] Note that if the disclosure is voluntary, I cannot extract whether the voter has moved in early or late June.

model, I present the output from a linear probability model in Table 1.5 for a more direct interpretation.

The first column is the main regression with 2018 general turnout as the dependent variable. Using eligible voters in previous three general and primary elections, I also perform placebo tests, which all pass for the treatment. On average, the treated group is more likely to have voted in the 2018 general by 5.9 percentage points. Figure 1.6 shows the 95% confidence interval ($[0.0345, 0.0839]$) along with the intervals for placebo tests.

		Imperfect Placebo	*Placebo Tests*				
	General 2018	General 2016	Primary 2016	General 2014	Primary 2014	General 2012	Primary 2012
NCOA Treatment	0.059^{***}	-0.006	0.001	-0.004	-0.011	0.001	-0.010
	(0.013)	(0.011)	(0.015)	(0.015)	(0.012)	(0.015)	(0.015)
Same Address	-0.008	-0.051	-0.038	-0.046	0.026	-0.117^{**}	-0.041
	(0.040)	(0.036)	(0.049)	(0.050)	(0.039)	(0.049)	(0.047)
Same Precinct	0.014	-0.063^{*}	-0.032	-0.070	0.008	-0.079^{*}	-0.052
	(0.036)	(0.033)	(0.042)	(0.044)	(0.034)	(0.043)	(0.041)
Same Cong.	-0.040^{*}	-0.018	-0.015	-0.019	0.004	-0.039	-0.055^{**}
	(0.024)	(0.022)	(0.029)	(0.029)	(0.023)	(0.029)	(0.028)
Diff. Cong.	-0.051^{*}	-0.043^{*}	-0.022	-0.016	-0.013	-0.047	-0.076^{**}
	(0.027)	(0.024)	(0.032)	(0.033)	(0.026)	(0.032)	(0.031)
Distance Moved	-0.002^{*}	0.001	0.001	-0.002	0.0001	0.001	-0.0005
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	5,539	5,035	4,341	3,680	3,553	3,366	3,082
Adjusted R^2	0.126	0.043	0.081	0.135	0.136	0.057	0.136
Res. Std. Error	0.455 (df=5510)	0.392 (df=5007)	0.474 (df=4313)	0.448 (df=3652)	0.342 (df=3525)	0.419 (df=3338)	0.391 (df=3054)
F statistic	29.427^{***} (df=28; 5510)	9.334^{***} (df=27; 5007)	15.164^{***} (df=27; 4313)	22.291^{***} (df=27; 3652)	21.695^{***} (df=27; 3525)	8.587^{***} (df=27; 3338)	18.902^{***} (df=27; 3054)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 1.5: Effect of NCOA Automatic Voter Registration, Linear Probability Model

Effect Size. The estimated treatment effect is very large. In fact, if we could boost turnout of movers uniformly at 5.9%, much of turnout depression by moving would disappear. To put this in context, take the results of the landmark study in persuasive get-out-the-vote (GOTV) mailings in Gerber et al. (2008). The effect of showing households their own voting records and urging them to vote resulted in a 4.9 percentage point increase of turnout, and showing them both their own and the neighborhood voting records resulted in an 8.1 percentage point increase. Oftentimes, many GOTV devices are insignificant or have effect size lower than 5 percentage points (Schelker and Schneiter, 2017).

So why is the effect so large, comparative to the previous findings of the GOTV literature? There can be a few caveats. First, the width of the "window" of observations used for the regression discontinuity is two weeks. Two extra weeks may have helped voters overcome the detrimental effect of moving, thereby inflating the effect size.

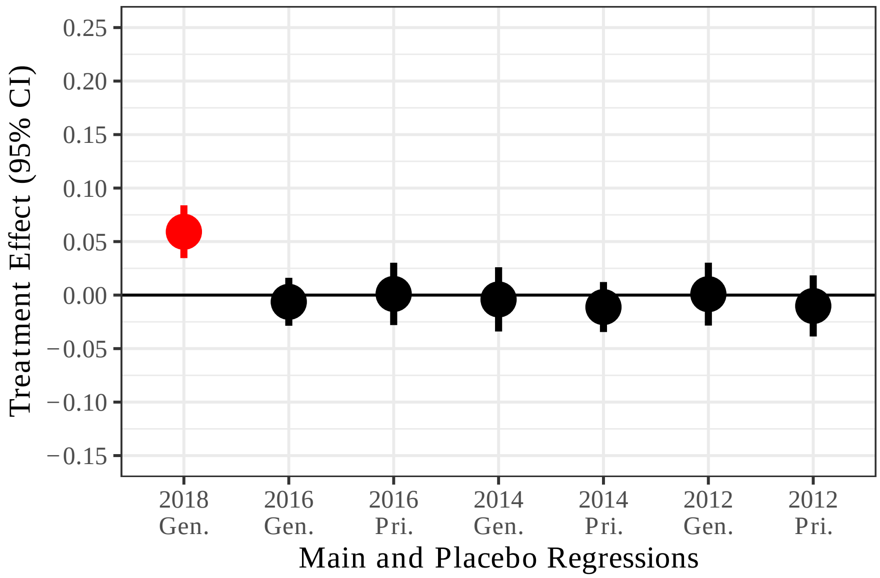


Figure 1.6: Effect of NCOA Automatic Voter Registration Treatment and Placebo Tests

Second, this was an official, pre-paid postage mailing from the election administrators, and not a GOTV mailing from third-party civic organizations. This may have made the difference. For example, Mann and Bryant (2019) show that even a simple postcard from official election administrators can encourage voter registration and turnout (2 percentage points increase), without any legal or administrative process changes—a 'nudge.' Malhotra et al. (2012) also show that while third-party organization's emails made no difference in turnout, emails from official sources increased turnout.

Third, Orange County in 2018 was a highly contentious area in which the entire county turned 'blue,' whereas it has been deemed a conservative bastion for very long. Some districts, which had a Republican representative for thirty-five years, now elected a Democratic House representative. Hence the voters may have been simply more responsive to any type of stimulus related to the general election. However, also note that Congressional districts of new residences are controlled for.

Lastly, this is not an average treatment effect (ATE) for all movers but the average treatment effect on the treated (ATT). Specifically, the treatment is on relatively *peripheral* voters who happened to move in the early weeks of June. The treated voters did not inform the Registrar in about two months of moving, which indicates that they are less interested in voting compared to those who do inform the Registrar. Also, treated group who moved within the first five days of June are likely not to have voted in the primary of 2018, which took place on June 5, 2018. If they voted with their new address, the database would have been updated, so that no treatment would have been necessary in July. In fact, if placebo tested with the dependent variable of 2018 primary, the treatment gives statistical negative effects on turnout. And peripheral voters are understood to be more responsive to stimulus than those who are already well-motivated to vote (Highton and Wolfinger, 1998).

However, there seems no denying that the NCOA automatic voter registration is effective, whatever the size may be. As can be seen with the placebo tests for 2016, 2014, and 2012 elections, the treated group is not so extremely peripheral to the degree that they have also voted less in previous cycles. In addition, as aforementioned, the effect is underestimated by including voluntary disclosers in the control group. Hence, it may well be the case that the estimated size is valid.

Overcoming Costs of Moving. In Section 1.4, I have concluded that dynamic effects of costs exist only when the voter's environment has sufficiently changed. Does the NCOA automatic voter registration still boost turnout of those with little to no change in environment?

Consistent with the results in Section 1.4, I find that the NCOA mailings do not have a turnout boosting effect on voters with 'small' moves. Table 1.6 shows that for movers within the same precinct or for movers within a half-mile of their original residence, the NCOA mailing has no effect. Again, this may indicate that while theoretically convenience/distraction costs should apply to all movers homogeneously, it only begins to be effective when the mover has moved sufficiently far away or crossed precinct lines while moving.

Policy Implications

In terms of election administration, the efficiency of an already existing NVRA policy is promising. It is designed for both higher turnout and for better voter list maintenance, it is relatively simple, and it is non-partisan—both in nature and in effect, as when estimated, no further mobilization of movers would have swayed any major election results, not even a state Senator or general Assembly.[22] Therefore, from a policy perspective, it is encouraging confirmation that the Registrar has "safety nets" that help movers adjust and participate in the political process.

[22]For Orange County 2018 general election results, see https://www.ocvote.com/fileadmin/live/gen2018/results.htm#c-1913.

	Dependent Variable: General 2018 Turnout			
	Low Info Cost	Distance Moved Less Than 0.5-mile	Distance Moved Less Than 1 Mile	Distance Moved Less Than 3 Miles
NCOA Treatment	0.044	0.050	0.060^{**}	0.076^{***}
	(0.043)	(0.034)	(0.027)	(0.018)
Same Address		0.035	0.015	−0.001
		(0.054)	(0.046)	(0.043)
Same Precinct	0.053	0.083	0.064	0.031
	(0.051)	(0.054)	(0.044)	(0.038)
Distance Moved	0.050			−0.009
	(0.111)			(0.012)
Same Cong.		0.012	0.015	−0.020
		(0.056)	(0.039)	(0.027)
Diff. Cong.		−0.149	0.103	−0.052
		(0.206)	(0.093)	(0.041)
Controls	Y	Y	Y	Y
Observations	464	724	1,143	2,548
R^2	0.176	0.179	0.145	0.137
Adjusted R^2	0.131	0.149	0.124	0.128
Residual Std. Error	0.441 (df = 439)	0.441 (df = 697)	0.446 (df = 1115)	0.450 (df = 2519)
F Statistic	3.898^{***} (df = 24; 439)	5.851^{***} (df = 26; 697)	6.989^{***} (df = 27; 1115)	14.323^{***} (df = 28; 2519)
Note:				$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 1.6: Effect of NCOA Automatic Voter Registration in Small Moves

However, the actual implementation of the policy can be difficult. List maintenance cannot take place less than 90 days before a federal election. Given a primary and a general election, this already equates to six months in the election year where NCOA mailings cannot be performed. If there is a special election or two, the moratorium period extends further. This means that the Registrar has to rely on a small window of time between those moratoriums to perform full NCOA processing and mailings.

When the usage of vote-by-mail is prevalent, or the election is conducted by all-mailing, this may pose a problem. This indicates that voters who move close to the election are relatively disenfranchised, policy-wide. The 90-days restriction is understandable in that it will lessen the load of election administrators too close to the election. If, however the election administrators have sufficient resources, loosening the 90-days restriction can improve turnout of movers. (Highton and Wolfinger, 1998) in particular expressed concern about the 90-days policy as follows:

> (A 90-day closing date) would be the first week in August for the general

> election ... Twelve percent of all adult citizens moved in the six months before November 1992, nearly three-quarters of everyone who moved during the entire year. Moving in the summer fits best with the school year, a rhythm that guides not only parents of school-age children but also significant groups such as teachers and university students. Purging and reregistration through the NCOA option misses all these people.

If the summer movers equate to younger and more transient electorate, easing the 90-days policy will benefit them in particular, who are already a low-turnout group.

Outside California, if NCOA treatment is mandated, this could greatly help voters. According to the National Conference of State Legislatures (NASS), at least thirty-six states authorize the usage of NCOA data to check whether voters' address changed (National Association of Secretaries of State, 2017). What happens *after* the voter data and NCOA data is matched is slightly unclear. Pursuant to 42 U.S.C. § 1973gg-6(c), if moves detected are within-jurisdiction moves, the election official is required to update the voter's registration and then send the notice.

How this is exercised is a little less straightforward—in-county movers' addresses are automatically updated only in Arizona, Arkansas, Colorado, California, Iowa, Kansas, Kentucky, New Jersey, New York, Oregon, Pennsylvania, Rhode Island, Virginia, Washington, and West Virginia.[23] In fact, according to NASS, in California, Florida, and Illinois, this applies to all in-state and not just in-county movers, which is an extraordinary feat. For some states such as Delaware, Indiana, Louisiana, Maine, Mississippi, Missouri, Montana, Nebraska, Nevada, New Mexico, Ohio, Oklahoma, South Dakota, or Texas, there is no mention of automatic updates *prior* to contact—some are contingent upon the voter actually returning the mailing notice, and some are silent on details. Although a follow-up study that includes all these states would be desirable, this would be extremely expensive to run, since it will require both a national voter file and NCOA processing of the entire national roster.

What a simple contact without re-registration would do is not entirely clear. Does NCOA automatic registration work mainly because it solves re-registration burden, or mainly because it is a reminder about the election from official sources? The answer is out of the scope of this paper. However, it is nonetheless clear that other

[23] In Connecticut and Wisconsin, only movers within the municipality are contacted. In Michigan similarly, only voters within the city or township.

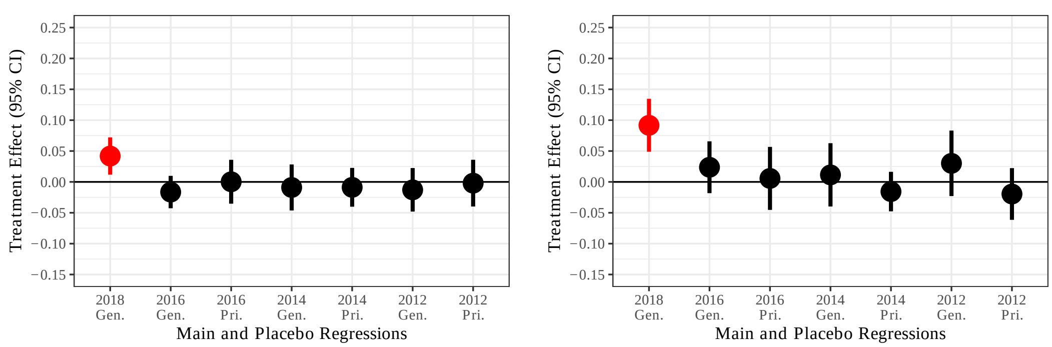
states should follow up on the usage of NCOA data, if they wish to best clean the voter data as well as maintain movers' turnout.

Subgroup Analyses

	Dependent Variable: General 2018 Turnout								
	Main	Subgroup Analyses							
		PAV		Race			Party		
	All	PAV	Not PAV	White	Hispanic	Asian	Dem	Rep	NPP/Third-Party
NCOA Treated	0.059*** (0.013)	0.042** (0.015)	0.092*** (0.022)	0.070*** (0.016)	0.065 (0.034)	0.137*** (0.040)	0.021 (0.021)	0.085*** (0.022)	0.074*** (0.022)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	5,539	3,632	1,907	3,255	850	573	1,793	1,815	1,931
R^2	0.130	0.099	0.196	0.119	0.175	0.181	0.127	0.124	0.129
Adjusted R^2	0.126	0.093	0.184	0.113	0.151	0.145	0.114	0.111	0.117
Residual Std. Error	0.455 (df = 5510)	0.457 (df = 3604)	0.447 (df = 1879)	0.446 (df = 3230)	0.460 (df = 825)	0.461 (df = 548)	0.438 (df = 1766)	0.454 (df = 1788)	0.469 (df = 1904)
F Statistic	29.427*** (df = 28 5510)	14.747*** (df = 27 3604)	16.921*** (df = 27 1879)	18.239*** (df = 24 3230)	7.287*** (df = 24 825)	5.055*** (df = 24 548)	9.869*** (df = 2 1766)	9.742*** (df = 26 1788)	10.843*** (df = 26 1904)

Note: *p<0.05; **p<0.01; ***p<0.001

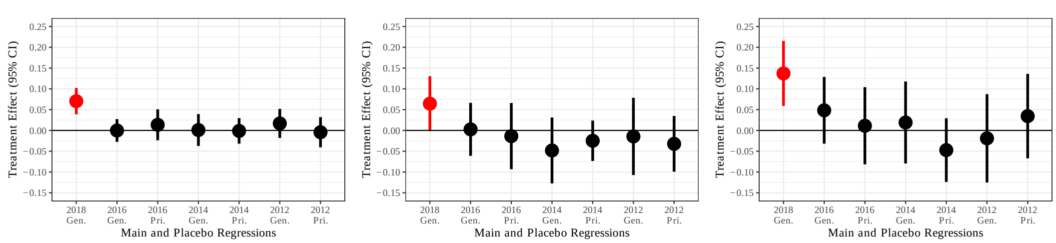
Table 1.7: Comparison of Main and Subgroup Analyses, Effect of NCOA Automatic Voter Registration



(a) Permanent Absentee Voter

(b) Not Permanent Absentee Voter

Figure 1.7: Effect of NCOA Automatic Voter Registration Treatment by Permanent Absentee Voter Status

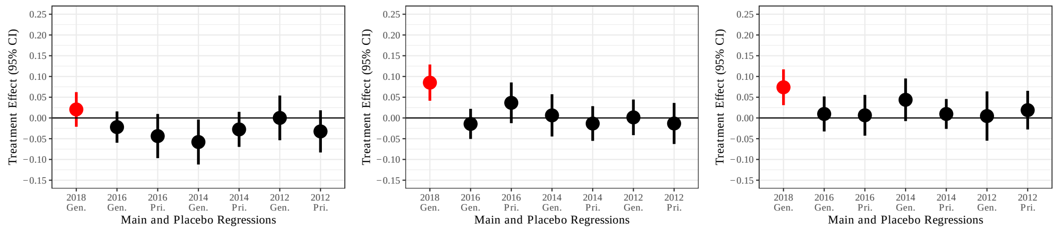


(a) White

(b) Hispanic

(c) Asian

Figure 1.8: Effect of NCOA Automatic Voter Registration Treatment by Race



(a) Democrat (b) Republican (c) No Party Preferences/Third Party

Figure 1.9: Effect of NCOA Automatic Voter Registration Treatment by Party

Here I discuss some important subgroup analyses policy effects, by (1) permanent absentee voter status (Figure 1.7), (2) race (Figure 1.8), and (3) party affiliation (Figure 1.9).[24] Table 1.7 shows these analyses in numbers. Note that as did for the full sample, all regressions pass the placebo tests. However, the results should still be taken with a grain of salt, as the subgroup estimates are not blocked or part of the randomization process.[25]

The subgroup analyses show that the policy has very different effect for subgroups. Per absentee voter status in Figure 1.7, the treatment effect is much stronger for those who are not permanent absentee voters, that is, those who are more likely to turn out physically to vote at a polling place. This is an important intuition in light of the recent COVID-19 pandemic of 2020, in which many of the states are hurriedly converting towards all-mail voting in the primaries and are actively discussing going all-mail in the general election as well. As election administrators are forced to operate with less polling places, this shows that NCOA automatic voter registration may aid voters who are not used to vote-by-mail.

Per racial category in Figure 1.8, the effect on Hispanic movers are somewhat muted, while the treatment has a strong effect on white movers. What is interesting is that the Asian movers are extremely responsive, with almost twice the coefficient of whites. What this signals for the Asian electorate as opposed to the Hispanic electorate is not immediately clear to me. One possibility is that the mailing was entirely in English, resulting in a language barrier.

As convenience voting mechanisms are often quoted as benefiting the Democrats,

[24] I exclude 'black' because there are too few voters who are black in Orange County, and 'others' due to their ambiguity.

[25] In the entire dataset, 59.4% are treated. The following is a simple descriptive of proportion treated per subgroup: PAV (55.8%), not PAV (66.3%), white (59.2%), Hispanic (63.6%), Asian (55.0%), Republican (59.1%), no party preference/third party (61.0%), and Democrat (58.2%).

it is very important to perform a subgroup analyses by party. This yields a very interesting result, shown in Figure 1.9. For Democrats, the NCOA automatic voter registration had almost no effect at all. For the rest of the movers, especially for Republicans, the effect was extremely strong. Although it remains to be seen whether we will see similar effects in other counties than this traditional conservative bastion, this is one evidence that convenience voting mechanisms are not uniformly beneficial to the Democratic party.

1.6 Discussion

In this paper, I first asked how the impact of moving on turnout changes over time. I show how voters who move close to the election are significantly less likely to vote, the difference being at most 16.2 percentage points. It took movers at least six months to recover to a level of turnout similar to stayers, and turnout was relatively steady for those who lived longer than six months at their new residence. This dip and recovery trend is not seen for small moves in which there were no or very little information costs, such as within-apartment movers. The time trend was only apparent for voters who at least cross precinct boundaries while moving, even when limited to movers within half-mile. Sufficient changes in information costs seem to be a necessary condition in activating dynamic costs of moving.

While election administrators cannot dictate when or where the voter moves to, I show that a simple, pre-existing, and non-partisan policy is effective in retaining movers and boosting their turnout. With this policy, election administrators can use the change-of-address data from the USPS to proactively track movers, automatically update the voter registration database's address for them, and notify them of this action. This resolves the convenience cost of voters or the re-registration burden from movers, and also reminds them about the upcoming election. Using a natural experiment that stems from legal requirements and how the policy was practiced, I evaluate the effect of the policy. I found that the propensity to vote was 5.9 percentage points larger for those who received the 'NCOA mailings.' I also discussed some related policy suggestions.

The importance of analyzing movers' turnout is even more important due to the changes in election administration practices and COVID-19 which has upended all elections since March. Many states were moving towards all-mail elections even before the pandemic, as vote-by-mail has been argued to increase turnout, especially for the peripheral voters in low-stimulus elections (Karp and Banducci, 2000; Southwell and Burchett, 2000; Gerber et al., 2013). As of 2018, twenty-two states had provisions allowing all-mail elections under some circumstances. Among them, three (Oregon, Washington, and Colorado) conduct *all* elections by all-mail, and California and Hawaii are also gradually making a transition. With the pandemic, many states have been forced to look for all-mail options, and it is uncertain whether in-person voting options can be as robust as before for the general election. To cap it all, voters themselves will choose to vote by mail more than ever, afraid of health hazards from voting. When a larger electorate receives ballots by physical mail, it is more vital than ever to analyze and remove barriers that movers may face when voting.

The detailed and large administrative data and the use of semiparametric regressions made this analysis possible. I would like to again emphasize that this study would not have been possible nor convincing with just survey data, as they lack both valid information about details of moving, and the power to perform such flexible analysis. In addition, the use of such large-N data ensures that we can explore beyond the linearity to reliably fit generalized additive models with splines (or any other smoothing functions). This paper shows that generalized additive models can open up exciting new analyses when combined with administrative datasets.

While I only use in-county movers, the costs of moving will only increase for out-county and out-state movers. Portable registration does not apply across county borders. In these cases, the lowered turnout is likely to be stretched across a longer period of time. These voters with geographic, community mobility will take much longer to adjust than six months—indeed, two years, the observation period of this study, may be insufficient for a full recovery. It may take more than three years, five years, or even close to a decade, a time frame as analyzed in Highton (2000). However, as a majority of movers are within-county, it is still important to recognize how in-county movers may face difficulties in voting, especially if they have moved close to the Election Day. In addition, multistate partnerships such as Electronic Registration Information Center (ERIC)'s data sharing may benefit from taking a leaf out of in-county NCOA automatic registration's book.

Chapter 2

HIDDEN DONORS: ANALYZING THE CENSORING PROBLEM IN U.S. FEDERAL CAMPAIGN FINANCE DATA

Alvarez, R. Michael, Jonathan N. Katz, and Seo-young Silvia Kim (2020, March). Hidden donors: The censoring problem in U.S. federal campaign finance data. *Election Law Journal: Rules, Politics, and Policy 19*(1), 1–18. doi: 10.1089/elj.2019.0593.

Personal contribution statement: S-Y. S. K. conceived the idea, contributed to the methodology, wrote the software/code, planned and carried out the analysis, designed figures and tables, and contributed to the writing of the manuscript.

The doi for this paper is at https://doi.org/10.1089/elj.2019.0593. I thank Mary Ann Liebert, Inc., publishers of Election Law Journal for granting the copyright permission to include elj.2019.0593 in my thesis.

2.1 Introduction

The United States has established one of the world's most comprehensive federal campaign disclosure processes. The Federal Election Commission (FEC), the federal entity which collects and disseminates these disclosures, produces data that are transparent, accessible, and up-to-date. Consequently, there has been an enormous amount of academic research on campaign finance, in particular on campaign expenditures, in the U.S. for the last few decades—see, for example, the recent review by Dawood (2015). Yet as we argue below, the FEC's data on campaign contributions is incomplete, and accordingly, our understanding of campaign contributions in the United States is incomplete as well.

One key feature of disclosure regulations in the U.S. is that currently, each federal campaign committee only has to report to the FEC contributions from individuals *who have already given $200 in aggregate to that campaign committee*, either within a year or a two-year election cycle according to the committee type. The $200 threshold has been in place for decades, and many donors' first few—in some cases all—contribution records are censored.[1] We call this the *censoring problem* in

[1] Prior to 1989, however, the data entry threshold of the FEC was set at $500. See the FEC's

campaign finance data, and the campaign contributors whose contribution amounts are below this threshold we call *hidden donors*.

Although a crucial part of the data generating process, the censoring problem and hidden donors have not been adequately addressed in the campaign finance literature (Key, 1964; Francia et al., 2003; Barber, 2016a; Magleby et al., 2018). There is, after all, little that could be done about this censoring, short of changing the law. However, in aggregate, hidden donors are an important force that politicians must cater to when deciding ideological positions and campaign strategies. Hidden donors are likely to become more important for researchers to study in the near future, as more candidates are wooing individual donors, and positioning themselves as not "buyable" by corporate PAC money.[2]

Fortunately, we can get an estimate of the censoring problem due to the unique fundraising approach taken by the 2016 presidential campaign of Bernard "Bernie" Sanders. For the first time, we can observe *all* of a major presidential candidate's individual contribution records, without any of them being censored. This is because the Sanders campaign chose to receive money only through an intermediary committee, and so these contributions are known as "earmarked contributions." These transactions are governed by a different, stricter set of regulations, resulting in full disclosure. Through the Sanders presidential campaign, we can compare datasets in which the censoring problem is and is not present. While the "law of available data" has driven us to investigate visible donors almost exclusively, this paper, using the valuable opportunity presented by the 2016 Sanders campaign, studies the seldom-explored world of "hidden" donors.

We show that hidden donors differ substantially from the visible donors. We do this using only the administrative data from the FEC, where we statistically impute contributor race/ethnicity and gender using only the contributor's name and address, which augments the thin selection of covariates available just from campaign reports. We find that hidden donors are more likely to be female, non-white, and younger. They may also have different political goals or interests, as they are more likely to donate later in the election cycle than visible donors, as their contributions are concentrated at points in the election cycle when the race is contested. Most importantly, there were seven times more hidden donors than visible donors for Sanders—suggesting that past research focusing only on visible donors may have

Thirty Year Report published in 2005. Also, because the $200 restriction is with nominal dollars, the data have to be filtered using inflation adjustments.

[2] See Appendix B.4 for additional information on the 2020 campaign and small donors.

only observed the tip of the iceberg in terms of individual donor behavior. We suggest that as presidential campaign fundraising methods are now shifting towards an individual and small-donor paradigm, more analysis of small donor behavior is necessary.

2.2 Past Research

Given the limitations of the FEC's campaign contributions data, most papers that study individual campaign contributors use survey data (Brown et al., 1995; Francia et al., 2003; Barber et al., 2017; Rhodes et al., 2018). However, such studies inevitably rely on the FEC records to sample respondents, meaning that they are restricted to visible donors. The comparisons that studies often make are between visible contributors and average citizens, or between different types of visible contributors. These studies have not examined how hidden contributors differ from those who are visible, nor do they compare hidden contributors and non-contributors.

It is not that scholars are altogether unaware of this censoring problem. Key (1964) noted that scholars do not know much about "little givers." Francia et al. (2003) recognized the threshold by labeling donations after the $200 threshold as "significant" donations, while Heerwig (2016) called these donors "elite" donors. Barber (2016a), while discussing how his survey does not include donors who gave less than $200 in a footnote, said that his picture of individual donors' motivations may be incomplete, if unitemized (hidden) donors have different motivations from those who are itemized (visible).

Although this censoring problem is well-known, there is no remedy short of changing federal campaign finance law, or asking campaigns to voluntarily report every small donation. Unfortunately, as contributing money to electoral campaigns is not a widespread form of political participation, the relative rarity of campaign donors in the population of adult Americans means that it is difficult to draw a representative sample of donors overall, and small or hidden donors specifically. There have been a few studies using the American National Election Studies (ANES) surveys (Panagopoulos and Bergan, 2006; Johnson, 2013), but the sample size that report contributing is very small—for instance, in the 2008 election, only 9.9% of respondents reported to have contributed money to a specific candidate.[3]

Despite these limitations, the literature has worked to generate stylized facts about

[3] 2008 Time Series Study, ANES, May 19, 2015 version, sub-sample of post-election survey respondents: question `V085033`.

the small, hidden donors. Most obviously, past research has argued that the hidden donors have less income than visible donors, because giving is related to disposable income. For example, Verba et al. (1995) show that the percentage of family income contributed to political campaigns increases with household income, sharply rising at $50,000 or more (See also Wilcox (2008) and Malbin (2013)). A series of papers such as Graf et al. (2006), Panagopoulos and Bergan (2007), and Malbin (2009) exploit surveys based on donor lists from the public matching funds program, in which presidential candidates can apply for government funding up to the first $250 of each contribution, but only after full disclosure. Johnson (2010) and Culberson et al. (2018) tackle the problem by using aggregate amounts reported at the campaign level, concluding that small donors are linked to more ideologically extreme candidates, although there is some mixed evidence with this argument (Malbin, 2013).

The most recent work that has systematically compared visible and hidden campaign contributors is Magleby et al. (2018). They were able to cooperate with major presidential candidates in 2008 and 2012 to receive a random sample of contributors who gave donations of all sizes. They generally do not find important ideological differences between visible and hidden donors.[4] In terms of their demographic profiles, they found that visible donors are older and wealthier than hidden donors. Hidden donors, meanwhile, were more likely to have been solicited to contribute online. Magleby et al. (2018) makes it clear that campaigns prioritize large donors over the small, especially due to changes introduced by the Bipartisan Campaign Reform Act of 2002. However, appeals to small donors are becoming more prevalent in American campaigns, especially in recent elections as the Internet is becoming central to campaign fundraising (Malbin, 2013; Karpf, 2013). In that case, it is crucial to study who these small donors are, and what their donation strategies may be, so that we can understand how they influence campaign strategy.

This paper contributes to the literature on small donors by analyzing an entire donor population for a presidential candidate using complete individual-level data. While our data do not have the more in-depth questions and breadth that are available in surveys like that of Magleby et al. (2018) or Graf et al. (2006), we are not limited to survey respondents—we can bring the entire population of visible and hidden Sanders contributors to study, which can complement the mostly survey-based existing literature. This allows us to provide a set of analyses regarding the

[4] They refer to this comparison as itemized vs. small donors.

hidden donors in 2016, to document their contribution behavior, and to establish a baseline for future research on small and hidden donors.

2.3 The Censoring Problem in Campaign Finance Data

In this Section, we discuss the censoring problem that is the consequence of federal regulations on individual campaign contributions—how the censoring problem may manifest itself, and how we can use the data from the 2016 Sanders campaign to estimate the extent of campaign contribution censoring.

Code of Federal Regulations, Title 11

The FEC administers federal campaign finance law, under Title 11 of the Code of Federal Regulations. A federal campaign committee that meets the conditions will be registered with the FEC and will regularly file reports that disclose funds that are raised and spent. 11 CFR 104.3 *Contents of Reports* dictates this, and how the information on receipts is censored is stated in 11 CFR 104.3(4)(i):[5]

> (4)(i) Each person, other than any political committee, who makes a contribution to the reporting political committee during the reporting period, whose contribution or contributions aggregate in excess of $200 per calendar year (or per election cycle in the case of an authorized committee), together with the date of receipt and amount of any such contributions, except that the reporting political committee may elect to report such information for contributors of lesser amount(s) on a separate schedule;

When a contribution pushes the sum of aggregated contributions over this threshold, this contribution is “itemized.” Unitemized contributions are aggregated into a lump-sum and reported as a single number, and hence no other details are reported for unitemized contributions. Also, donors who do not meet the threshold are entirely absent from receipt reports that campaigns file with the FEC.

As aforementioned, we utilize intermediary committees to investigate the censoring effect on the data. 11 CFR 110.6 *Earmarked Contributions* provide details as to what

[5] We immediately see that aside from the arbitrary $200, there are two additional problems: One is that the $200 is in nominal dollars, unadjusted for inflation, and the other is that whether a campaign committee is authorized or not—which is, in the campaign finance jargon, another name for candidate-affiliated committee, especially the principal campaign committees. For an exact definition, see 11 CFR 9032.1. We largely avoid the first problem as we only use a single cycle’s observation, and we will for the moment ignore the second problem.

intermediary committees are, and what they should disclose. While earmarked contributions are money designated to a clearly identified candidate/committee, intermediary committees (also called conduits) are "anyone who receives and forwards an earmarked contribution to a candidate or candidate's authorized committee."[6] The following excerpt from the FEC summarizes the special disclosure requirements:

> A political committee that serves as a conduit of an earmarked contribution must disclose the earmarked contribution, regardless of amount, on two separate reports: the committee's next regularly scheduled FEC report, and a special transmittal report sent to the recipient authorized committee. 110.6(c)(1).

As we can see, intermediary committees have stricter disclosure requirements than other campaign committees. They must report all transactions, not just those that cross the $200 threshold—see Appendix B.1 for details. The issue, which we explain with detailed hypothetical examples in Appendix B.2, is that for campaigns that undergo the typical process, we will not observe campaign contributors who donate less than $200.

The 2016 Sanders Campaign

The 2016 Sanders campaign was unique in many ways, including their approach to campaign finance. In 2016, one of the important issues for Sanders was his campaign's stance against the influence of "big" money and special interests. He claimed early on that his campaign would not be allied with any super PACs (Lee, 2016; Qiu, 2015). In addition, the Sanders campaign was reported by the media in late 2015 to have organized only seven traditional fundraisers, while the Clinton campaign had by then organized more than 110 (Associated Press, 2015). In March 2016 it was further reported that the campaign had only two more traditional fundraisers (Gaudiano, 2016).

This distaste for conventional fundraising meant that most of Sanders' donations were digitally processed. His website noted on April 30, 2016 that 94% of its

[6] While conduit or intermediary committees are interchangeably used and the FEC seems to prefer the former, we use the terminology 'intermediary committees.'

contributions were made online.[7,8] His fundraising appeals were mostly digital, using emails, texts, and social media (Corasaniti, 2016), and his website pointed to the ActBlue contributing page. The campaign used no other online platforms and had no offline fundraising staff. Hence, almost all individual donors who desired to contribute to Sanders would have had to donate through ActBlue, regardless of their wealth, connections, or intentions.

Because ActBlue is an intermediary committee, we can track all Sanders donors, regardless of the size of their donations to his campaign, which as far as we know is a first for federal campaigns. Regardless of whether they gave fifty cents or maxed out individual contribution limits, they would be captured in ActBlue reports. This means that the censoring problem disappears with respect to Sanders' contributors, giving us an unprecedented opportunity to study both visible and hidden donors to a major presidential campaign. See Appendix B.3 for details.

It is true that the data are restricted to a subset of presidential donors, and that Sanders is an unusual candidate, as he is a very progressive Independent and the first presidential candidate to rely exclusively on individual contributions, mostly raised online.[9] However, there are unmistakable advantages in utilizing the Sanders data. We may never have another major candidate whose contributions are so transparently presented to the public, and whose donors are exposed, both small and large.[10] This is also the first step into understanding the unique data generating

[7]The website address is `https://berniesanders.com/revolution/`. Our final estimate of how much Sanders campaign received online is 98.1% of his total individual contributions, which is greater than the 94% reported in the website. The discrepancy seems to arise because the campaign's estimate was calculated before mid-May 2016—and we believe that many of the donations made after that point in the campaign were made online. That the 94% is calculated before mid-May can be inferred from the Wayback Machine's snapshot of the website, the first snapshot of which is at May 21, 2016, and shows the 94% claim.

[8]5.1% of donors' year-to-date contribution records do not match the actual sum of records, (1) potentially because they have not been sufficiently record linked due to the conservative linking of contribution records that constitute the same individual, (2) the committee inaccurately calculated the contributor year-to-date by mixing up different contributors, or (3) because these donors have given in the few traditional fundraisers we discussed earlier. The third possibility seems viable, since we find that 68.7% of those with record irregularities are visible. The second explanation may also have merit, because we often find that intermediate records of contribution year-to-date sums are grossly inaccurate. This speaks to the difficulty of working with campaign finance data.

[9]As discussed earlier, there is some debate and mixed evidence in the literature about whether small donors are polarizing—that is, whether ideologically extreme candidates attract more small donors (Wilcox, 2008; Bonica, 2011; Malbin, 2013; Johnson, 2010; Culberson et al., 2018; Magleby et al., 2018). Unfortunately, although this is a very important and interesting question, due to the nature of our data, we cannot test this claim in this paper.

[10]While the 2020 Democratic primary contenders are also shunning corporate PACs, their fundraising also incorporates the traditional individual fundraising, such as private, invited events.

process presented by intermediary committees.

Table 2.1: Proportion of Unitemized Contributions in Individual Contributions, Dollar Amounts, the 2016 Election

Office/Party	Unitemized Individual $/ (Unitemized $ + Itemized $) (%)
House Democrats	22.4
House Republicans	13.9
Senate Democrats	25.0
Senate Republicans	18.7
Hillary Clinton (Dem.)	25.8
Bernie Sanders (Dem.)	58.1
Donald Trump (Rep.)	64.9
Ted Cruz (Rep.)	38.5

Table 2.1 shows the proportion of unitemized contribution amounts relative to all individual contributions for the four major presidential candidates of 2016, as well as House and Senate Democrats and Republicans in 2016. We see in Table 2.1 that 58.1% of Sanders' contribution amounts were unitemized, and a large proportion of contribution amounts going to Trump, Clinton, and Cruz were also unitemized. The same is the case for House and Senate campaigns in 2016. Note that the total number of unitemized contributions is not the same as the sum of funds contributed by hidden donors, since the first $200 of visible donors are still labeled as unitemized, and not corrected in retrospect. Again, this demonstrates that there are many hidden donors, and that they play a substantial role in elections, especially in presidential races.

2.4 Data

The data we use were obtained from the FEC. One can use the FEC "Download Bulk Data" page (`https://www.fec.gov/data/browse-data/?tab=bulk-data`) to download the individual contributions data or build a database using the OpenFEC API (`https://api.open.fec.gov/developers/`). While the bulk data are easy to download and do not need deduplication, they contain only visible donors meeting the $200 threshold, and they also lack the donor's address. Hence we built our database using raw FEC data using the OpenFEC API, and downloaded all of the 2016 cycle's raw individual contribution data.[11]

[11]The R scripts used to build the database will be publicly available upon publication. This script downloads the data as text files in batches of 100 records, resulting in more than 1,133,000 files and

We then augmented the records with geographic identifiers such as Congressional district information after geocoding each record.[12] We also used record linkage to connect the donations that came from the same individual contributor. We use exact matching on first name, last name, and street address, and allow matches if there is variation in these variables but the employer and occupation is exactly the same for consecutive contribution records. We use exact matching to take advantage of the sequential nature of repeating contributions—if donor A has been a teacher for the first part of 2016, and has a different address starting mid-2016, we would still link her records together given that the names and occupation/employer stay consistent. To prevent false matches, we only match within the state.

Next, we focused specifically on contributions to *Bernie 2016*, the Sanders presidential committee, and to *ActBlue*, creating a union of contribution records to Sanders while filtering out duplicate contributions. This is a very important and time-consuming step, because there is always a separate receipt for donor giving to the intermediary committee, the intermediary committee giving to the final destination campaign, and the acknowledgement of a donor giving to the final destination campaign. The OpenFEC API recognizes this and offers guidance as to how to exclude some entries as duplicates in its Receipts description. Every intermediary contribution must be checked against the destination committee's reports to eliminate duplicate entries, which will have the same contribution amount, date, and personal details. This way, we can uncover donations ranging from a dollar to the $2,700 individual limit. The campaign reports do not contain gender or racial/ethnic information per se, and only carry name, address, the money's destination, date, occupation, and employer, the minimal requirements by the federal regulations.

Finally, we supplemented the contributions data, using names and geo-locations to infer gender and race/ethnicity. To infer gender for each individual contributor, we use the R package `gender` (Mullen, 2018; Blevins and Mullen, 2015) of rOpenSci project, which uses first names and the Social Security Administration's yearly dataset of first names to infer the gender of the donor. If the donor has voluntarily entered a prefix such as a Mr., we override the gender inference with the gender inferred from the self-reported prefix. For race/ethnicity, we use the R package `wru` (Khanna et al., 2017; Imai and Khanna, 2016). That is, we utilize the Census Bureau's surname list and Bayes' rule to infer the race. In particular, we use a Census

over 100 million records of individual transactions.

[12] Given that each record had a limited number of fields, this database has a single table, and can be linked to other tables in future research (by contributor name and address, for example).

block level inference, after geocoding each address using the Census Geocoder and obtaining the latitude and longitude of the address. The Census block being the smallest unit in which detailed surname distributions are provided by the 2010 Census, this enables the most accurate approximation of race. For the empirical validation of the method, see Imai and Khanna (2016).

2.5 Who Are the Hidden Donors?

We define a Sanders contributor as anyone who donated to Sanders' presidential campaign committee during the 2016 presidential election cycle. The committee launched on April 30, 2015 and the cycle ended on December 31, 2016.[13] The total estimated number of Sanders contributors is over 2 million, including both visible and hidden donors, a large base of donors. However, among them, only 12.4% (250,352 out of 2,017,638) would have been visible if the Sanders campaign did not receive individual contributions only through intermediary committees. In other words, there were seven times more contributors when hidden donors were included. This simple fact shows the potential magnitude of the censoring problem, and is similar to the numbers reported in Magleby et al. (2018), where they estimated that 82% of Romney donors and 88% of Obama donors in 2012 were small donors.[14]

Figure 2.1 shows the donor base compared to the population estimated by the 2016 American Community Survey, geographically. Not surprisingly, we can see that the Sanders' donor population is large in Vermont, with the town of Newfane with the largest proportion of Sanders donors (32% of residents). 74% of the top 100 Zip Codes in Sanders-donor population relative to their total population are in Vermont, with California and Massachusetts trailing respectively at 12% and 5%.

Table 2.2 describes the demographic difference between visible and hidden Sanders contributors. A hidden donor is more likely to be female and a racial or ethnic minority, whether black, Hispanic, or Asian. Note that the differences in gender are consistent with previous literature (Graf et al., 2006; Magleby et al., 2018).

While the percentages of black or Asian contributors do not differ greatly between visible and hidden Sanders donors, a hidden contributor is much more likely to be Hispanic than a visible one.[15] This is an interesting observation and could reflect the

[13] Statement of Organization, Bernie 2016. Available at `http://docquery.fec.gov/pdf/537/15031422537/15031422537.pdf`, last accessed November 19, 2018.

[14] One thing that should be noted is that both Obama and Romney were candidates that won the primaries, while Sanders did not make it to the general election. Had he won the primaries, these descriptives may have changed. See Appendix B.6 for a brief illustration.

[15] All differences are statistically significant. We do not offer p- or t-statistics separately for any

Figure 2.1: Proportion of Sanders Donors By Estimated Population, Zip Code Level

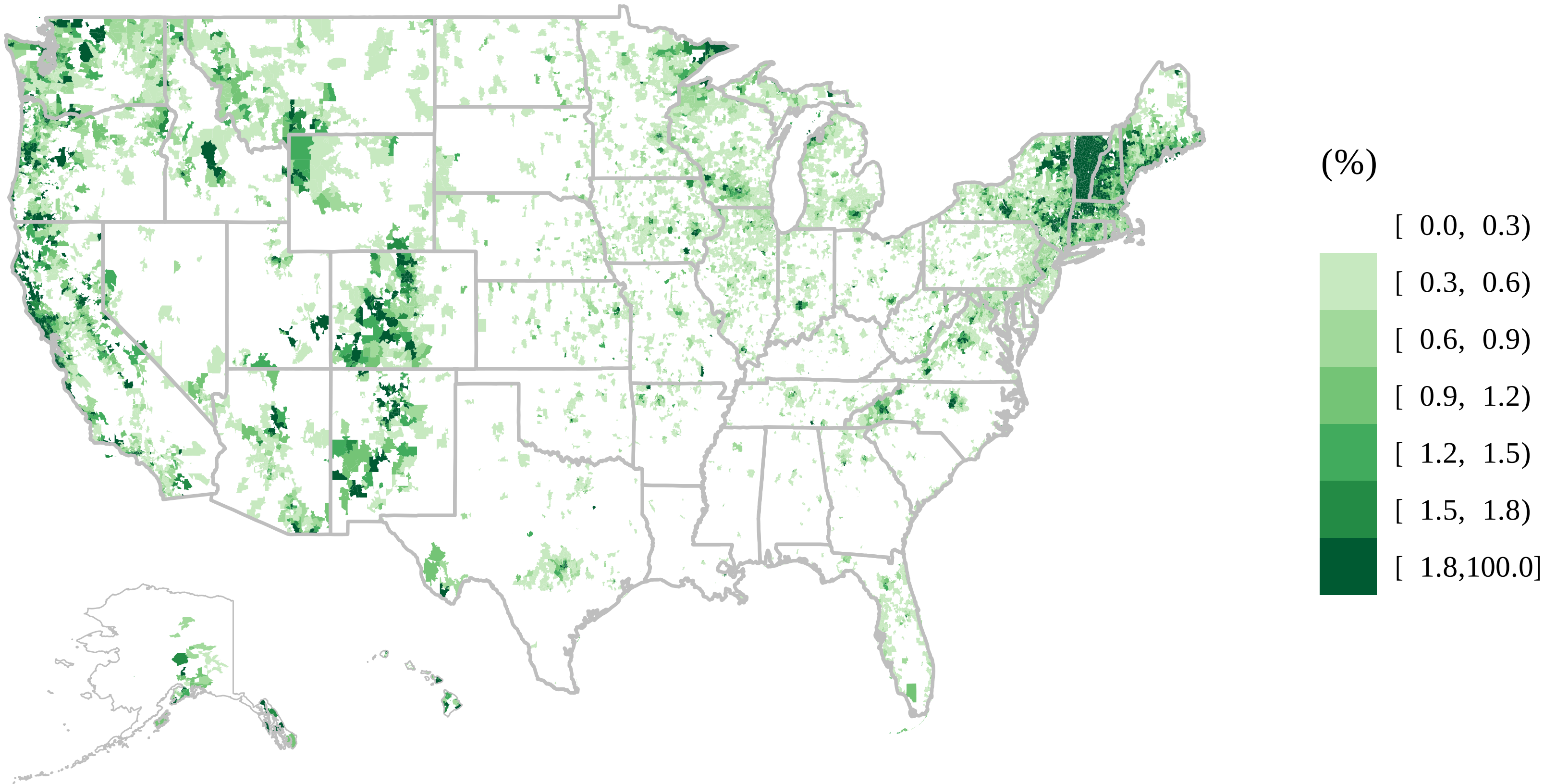


Table 2.2: Demographics and Occupations of Visible and Hidden Contributors, 2016 Sanders Campaign

	Visible	Hidden
% of Men	57.2%	53.2%
% of Whites	90.5%	86.7%
% of Blacks	2.8%	3.0%
% of Hispanics	4.2%	7.4%
% of Asians	2.5%	2.9%
% of Unemployed	26.7%	25.7%
% of Engineers	4.8%	3.1%
% of Teachers	4.1%	4.9%
% of Retired	3.9%	1.3%
% of Attorneys	2.3%	1.1%
% of Professors	2.1%	1.2%
% of Physicians	2.1%	0.8%
% of Consultants	1.9%	1.3%
% of Students	1.2%	4.0%
% of Homemakers	0.4%	0.3%

(Occupations are sorted by the visible donor percentage for a better contrast.)

fact that the Hispanic electorate is generally younger than other racial/ethnic groups (Patten, 2016), and perhaps unable to afford donating more than $200.

Table 2.2 also shows the top 10 types of employment of all 2016 contributors and their percentage for the subset of Sanders supporters. Noticeable is the presence of students in the hidden donor population, while the likelihood of being retired is much higher for visible donors. This signals a difference in age groups: that the hidden donors are likely to be younger in age. In addition, while visible donors are more likely to report being attorneys and physicians, hidden donors are slightly more likely to be teachers. Considering that the average annual salary of teachers in public schools was $56,383 in 2012[16] while a physician's lowest pay was $189,000[17], this hints that income and wealth might explain why some donors are hidden while others are not, which is as expected. However, it is also interesting to note that the proportion of visible donors reporting unemployment is slightly higher.

of our descriptive statistics—because the sample size is so large, *every* difference that we present in this paper is statistically significant ($p < 0.001$).

[16]National Center for Education Statistics, Digest of Education Statistics: 2013

[17]Forbes, The Best- And Worst-Paying Jobs For Doctors, July 20, 2012.

2.6 Differences in Contribution Patterns

Given the demographics differences that we have documented earlier, here we investigate whether hidden and visible donors vary in the amount, frequency, and timing of their donations.

Amount and Number

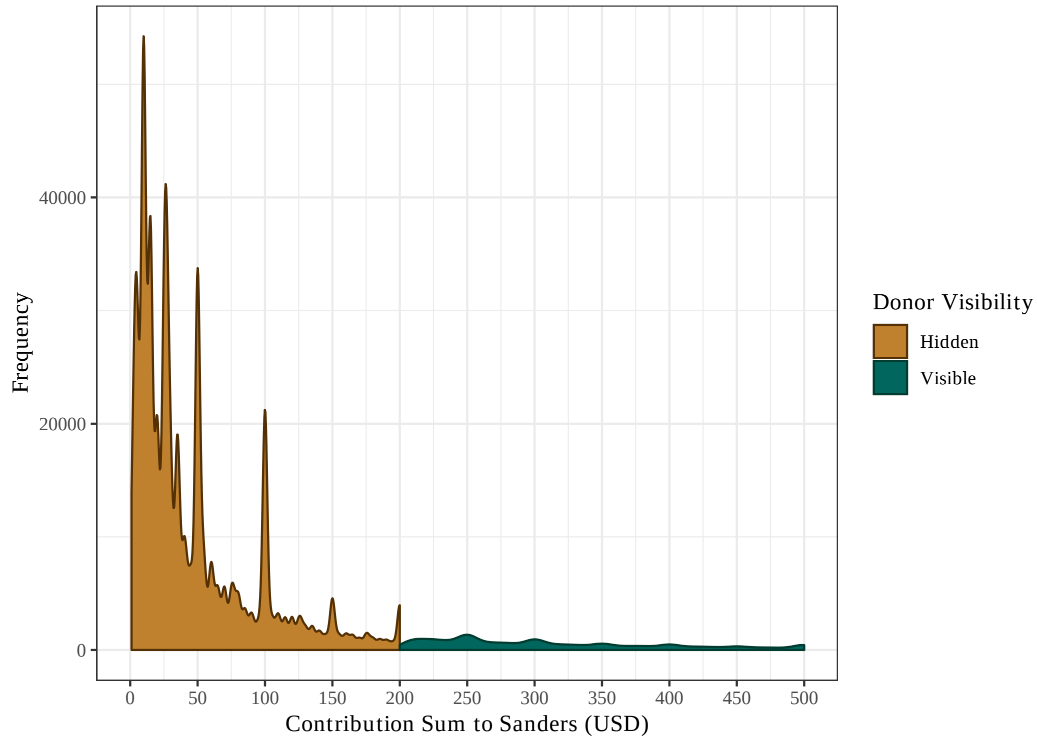
Table 2.3 displays the basic descriptive statistics for Sanders contributions in the 2016 cycle. Here we see the important differences between visible and hidden contributors. Figure 2.2 shows the density plot of the sum of Sanders contributions, capped at $500 to clearly highlight the difference. We see that the distribution of the contribution sum is extreme valued, much more pronounced with the hidden donors' data. Also, the distribution of hidden donors' contributions exhibits peaks at various salient contribution amounts such as $10 (12.0% of hidden donors), $15 (8.5%), $50 (7.6%), and $27 (6.1%). There are no such extreme peaks for visible donors' sum of contributions. The mode for visible donors is $250 (4.8%), followed by $300 (3.1%).

Table 2.3: Sanders Contributions, Visible and Hidden Contributors, 2016 Sanders Campaign

	Visible (12.4%)	Hidden (87.6%)
Sum of contributions:		
—— Mean	$632.0	$45.7
—— Q1	$270.0	$15.0
—— Q2	$400.0	$27.0
—— Q3	$700.0	$60.0
Number of contributions:		
—— Mean	13.2	2.5
—— Q1	5	1
—— Q2	10	1
—— Q3	17	3
Median first contribution date	10/04/2015	01/29/2016
Median last contribution date	05/18/2016	03/09/2016
% of donors who gave to Sanders, after		
—— Super Tuesday	83.5%	44.9%
—— Acela primaries (Super Tuesday III)	63.3%	16.9%
—— Clinton became presumptive nominee	29.5%	4.6%
—— Sanders endorsed Clinton	5.5%	0.8%
—— End of Democratic convention	0.1%	0.0%

The $27 contribution amount is particularly interesting. This is an amount that the Sanders campaign deliberately made salient mid-election, emphasizing that the

Figure 2.2: Distribution of Contribution Sum to Sanders, Visible and Hidden Contributors, Truncated at $500



average donation size to the campaign was $27, and thus the campaign relied on small donors and not on the fat-cats of Wall Street.[18] This emphasis has been frequently covered by the media (Bump, 2016; Foran, 2016; Mehta et al., 2016), and was further echoed in Sanders's later campaign strategies, as they added a $27 donation option in their suggested contribution amounts. And indeed, the campaign had successfully mobilized small donors enough that its median donor was a one-time giver of $27. However, $27-giving is more indicative of a visible donor compared to the benchmark distribution. While 20% of Sanders donors have given $27 at least once, conditional on a $27 giving record, a donor is 26.5% likely to be visible. This may suggest that such compliance with the campaign message—enough to choose such a non-salient amount—is indicative of donor loyalty. That is, who ends up contributing enough to earn the visibility.

Aside from the obvious differences in contribution amounts, Table 2.3 also shows important differences by the number of contributions. Note that visible donors have

[18]Note that the individual contribution limit of 2016 was $2,700, which could have resulted in this value.

given multiple times—indeed, while the median hidden donor has given only once, the median visible donor has given ten times to Sanders. This shows that not all visible donors are one-time donors of hundreds of dollars.

Timing

The timing of campaign contributions has not been widely analyzed, though there are a few recent papers on the subject (Christenson and Smidt, 2011; Smidt and Christenson, 2012; Mitchell et al., 2015; Magleby et al., 2018). These studies aggregate the number of donations on a daily level to study donation dynamics. Magleby et al. (2018) aggregated the data separately for small and large donors, and concluded that the same events motivate large and small donors alike, and thus they seem to be driven by the same dynamics.

However, the timing of giving disaggregated by first and last date of contributions may provide additional insight into donor motivations. If we view contributing to campaigns as a rational act in which a donor weighs her marginal benefits and costs, whether her gift is expressively or instrumentally motivated, the first contribution date reveals when the marginal benefits first exceed marginal costs of the person. If donations are driven by solicitation, timing may show how an equilibrium is struck between demand and supply.

Table 2.3 also shows various summaries of when the donors choose to give. Notice that visible contributors on average enter earlier in the race with their first contribution. While visible donors give long before the primaries, hidden donors enter right before the primaries start. Visible donors also leave later than hidden contributors, again by a difference of two to three months, a substantial amount of time. Note how half of the hidden donors started to leave after March 9th, although that was the day after the Sanders' surprise win in Michigan. In fact, by April 26 (the time of the so-called Acela primaries), less than one in six Sanders hidden donors remained (16.9%). For visible donors, a majority of them stopped only after May 18.

Table 2.3 shows the percentage of donors who gave to Sanders or other committees *after* some key dates in which Sanders' prospect declined. We include Super Tuesday (March 15), the Acela primaries (April 26), when Clinton became a presumptive nominee (June 6), when Sanders officially endorsed Clinton (July 12th), and when the Democratic Convention officially closed (July 28). Note that the percentage of hidden donors who give to Sanders rapidly declines as the election progresses. This contrasts with visible donors. By the time that Clinton is the presumptive nominee, almost 30% of visible donors still opened their wallets for further campaign contributions to Sanders, while only 5% of hidden donors did.[19] This suggests that "hidden" status was achieved because they were not attracted to give until the race was contested, and they were also quicker to lose interest. That hidden donors are late entrants to the market signals that they may not have been strong supporters to begin with, and that they may have been persuaded to give only after Sanders gained sufficient momentum.[20] Indeed, it is surprising to see that so many visible donors contributed to Sanders after he had effectively lost his quest for the Democratic nomination. Even if we account for the possibility of campaigns mislabeling donations with different dates then they actually received, the percentage seems substantial.

The difference in enthusiasm could be attributed to many factors, such as ideological differences (Ensley, 2009), candidate valence, information costs, or contributor's budgets. For example, given that donors are instrumental, the ones who are further away from Sanders ideologically may have been attracted to give by his rising prospects, but as his viability fell, they withdrew as their marginal benefit of giving fell below the costs. It could be that these are donors who are generally not sufficiently interested in politics to consider giving, and only when Sanders became a serious figure, were attracted just enough to give once or twice. There could be a donor class with limited budget for political donations, for instance due to a lower family income, who thus cannot afford to give expressive donations such as those made after Sanders loses the primary race. With only administrative data we cannot immediately conclude whether either of these two explanations hold true, but this is worthy of further investigation.

Differentiating Visible Donors

Given the results above, there is a strong possibility that visible donors become visible simply because they have patiently given many times, and eventually gave more than the $200 threshold. These donors, who we call *eventually visible* donors, may be different from the ones who have given over $200 in their first contribution, who

[19]Overall, 22.2% of donors started to give before the Iowa Caucuses but also stopped before the Iowa Caucuses, while 35.2% of donors gave before and after the Iowa Caucuses. Similarly, 35.9% started to give before Super Tuesday but not after, while 39.1% of donors gave before and after Super Tuesday.

[20]Note that this difference in timing is best highlighted also because Sanders had a lengthy campaign, as he did not leave the race until the Democratic convention. If a less-viable candidate had dropped out earlier or his/her prospects dwindled much more quickly, while the difference will still hold, it would be more subtle.

we will call *immediately visible* donors. If for instance a donor gives $50 monthly starting January 2016 and opts out during or before April, she will be hidden, while if she remains to give once more until May, she becomes visible. Distinguishing the different types of visible donors may be important in understanding differences in their political behavior from hidden donors.

Table 2.4 replicates Table 2.3 with these classifications. The immediately visible donors constitute only 2.7% of the donor population, while the eventually visible donors are 9.7% of the donor population. Note that the number and sum of contributions are correlated, as there is a $2,700 cap on the maximum amount of individual contributions—a donor who gives $500 regularly cannot give more than 5 times. Altogether, the immediately visible donors (2.7% of donors) give 20% of the individual contributions amount, the eventually visible donors (9.7%) give 46.2%, and the hidden donors (87.6%) give 33.8%. Again, this now excludes the first $200 from visible donors, and hence the proportion is reduced from what is presented in Table 2.1. Each class is a formidable force in the financial electorate, but it is the eventually visible donors who filled the campaign chest of the presidential candidate. Table 2.2's equivalent with these classifications is available in Appendix B.5.

Table 2.4: Sanders Contributions, Immediately Visible, Eventually Visible, and Hidden Donors, 2016 Sanders Campaign

	Immediately Visible (2.7%)	Eventually Visible (9.7%)	Hidden (87.6%)
Sum of contributions:			
—— Mean	$876.0	$564.0	$45.7
—— Q1	$277.0	$269.0	$15.0
—— Q2	$500.0	$389.0	$27.0
—— Q3	$1,000.0	$638.0	$60.0
Number of contributions:			
—— Mean	6.2	15.1	2.5
—— Q1	1	7	1
—— Q2	3	11	1
—— Q3	7	19	3
Median first contribution date	01/25/2016	09/23/2015	01/29/2016
Median date when voter became visible	01/25/2016	03/04/2016	—
Median last contribution date	04/03/2015	05/25/2016	03/09/2016
% of donors who gave to Sanders, after			
—— Super Tuesday	61.5%	89.6%	44.9%
—— Acela primaries (Super Tuesday III)	38.7%	70.1%	16.9%
—— Clinton became presumptive nominee	15.7%	33.3%	4.6%
—— Sanders endorsed Clinton	3.0%	6.2%	0.8%
—— End of Democratic convention	0.1%	0.0%	0.0%

On average, the eventually visible donors start early, but give frequently and are late to leave the donation game. For other campaigns than the Sanders campaign, we would not have been able to observe their first few donations, and hence would not know exactly when they entered. Instead, we would have observed only the date that they crossed the $200 threshold and became visible, the median of which is March 3, 2016, and would have only known that they would have entered before that. Because the best predictor of giving is their past giving history, it makes sense for campaigns to target early givers and attempt to solicit more money from them, given that they have not maxed out. It appears that donors who become eventually visible are the most loyal—or persuadable—set of donors. Indeed, conditional on having first given before June 1, 2015 (within two months of launching the campaign), 27.6% will become eventually visible.

One interesting point is that the median date of first donation for hidden donors and for immediately visible donors is similar—only four days apart, a week to go before the Iowa caucuses. The first and third quartiles are also similar: respectively September 12, 2015 vs. October 1, 2015, and March 6, 2016 vs. March 9, 2016. If first donation dates are indication of interest/persuasion, this suggests that the hidden and immediately visible donors may be quite similar in their motivations for contributing, and perhaps that income or wealth is the primary difference between hidden and immediately visible donors. It is the repeat donors—who are also mostly early donors—who stand out, and they might be the donors who most affect the campaign's policy positions.

Again, we are looking at the data from a single campaign. Depending on the nature of different campaigns, contextual factors, and primary election dynamics, the profile of the hidden donor population might vary across election cycles and candidates. If the hidden donors—who provided one-third of contribution dollars to Sanders in 2016—can be further mobilized with certain issue positions, a campaign may choose to do just that. Hidden donors have, after all, the potential to give more and perhaps become visible. If they are more economically representative, the campaign may find it more beneficial to focus on small donors and retaining policy positions closer to the median voter. However, if the campaign's donor base cannot be mobilized this way, the campaign may choose to diverge from the median voter to attract more rich donors.

Are the three classes of donors actually any different in policy preferences? We cannot conclude that they are with only administrative data, and much less with just Sanders' 2016 campaign. However, we believe it is worth further investigation. Uncovering hidden donors clearly shows that we need to better understand who the presidential campaign perceives to be major financial electorate, and whether catering improves a campaign's chances of primary and general election success.

2.7 Conclusion

How is our understanding of campaign finance restricted by the FEC's data reporting process? Because only transactions that exceed cycle-to-date aggregate of $200 are disclosed, many donors' contribution records are censored, either entirely or in part. We have shown that small donors, who are usually "hidden" due to censoring (hence the term *hidden donors*), are distinct from large donors. Unfortunately, it has been only large donors who were usually visible to campaign outsiders and researchers (*visible donors*). Hence, we argue that our knowledge of individual campaign contributors has been incomplete.

Hidden donors to Sanders 2016 campaign far outnumbered visible donors, and for every visible donor to his campaign, there were seven who were hidden. Also, these hidden donors, relative to visible ones, were more likely to be students, females, and racial/ethnic minorities. This hints, consistent with the existing literature, that we usually only see a limited picture of much wealthier, privileged donors' political behavior when using censored campaign contributions data.

Hidden donors on average contribute later in the election than visible donors, and their giving generally occurred when the race was competitive. The median hidden donor started giving to Sanders a week before the primaries started, and did not give more than once or twice—whether it be a lack of enthusiasm, Sanders' diminished viability, or the donor's constricted campaign budget, is difficult to identify with our data. Whichever is the case, the hidden donors largely cleared out around March 2016, while a substantial portion of visible donors continued to give even with their candidate's dwindled prospects. This observation cannot be made with survey-based methods, with daily aggregation over all donations, or with campaign-level data aggregation.

The hidden donors, in aggregate, provided up one-third of the contributions to Sanders from individuals. While this proportion may be overestimated due to the uniquely popular, small-donor powered nature of the Sanders campaign, the hidden donors are still a significant subset of the financial constituents that a politician must be responsive to. If hidden donors are providing significant financial support to a campaign, it is likely that they have a more important influence on campaign platforms than previously acknowledged.

We also demonstrated the importance of donors who were hidden in the beginning but *eventually* became visible, who provided about half of Sanders' funds from individuals. These eventually visible donors were more likely to be early givers than the hidden donors—something that was not observed with traditional FEC data. For this class of donors, campaigns are likely to target solicit money from them until they have maxed out their contribution limits. Overall, looking into hidden donors' contributions and eventually visible donors' first contributions demonstrate the need to better understand how campaigns perceive and tailor their messages to all their contributors.

Fortunately for campaign finance scholars, the reliance on intermediary committees is ever increasing. Not only is the usage of ActBlue to solicit online donations now widely prevalent in Democratic campaigns, but the GOP is also trying to attract small donors by building a similar intermediary committee called WinRed, which will also be subject to the same strict set of disclosure regulations as ActBlue. While we may never be as lucky have such rich contributions data from a major presidential candidate, this ever-increasing reliance on intermediaries/conduits, and that the $200 requirement stands unadjusted by inflation, will help us study small donors in the future.

Chapter 3

DONATION DYNAMICS: DO CRITICAL CAMPAIGN EVENTS INFLUENCE CONTRIBUTIONS?

3.1 Introduction

Individual donors play a key role in presidential elections. Their importance is ever growing (Goff, 2005; Magleby, 2008) such that other sources of money—such as PACs, party committees, candidate self-financing, public funds, and transfers from affiliated committees—pale against direct donations from individuals. Indeed, from 2012 on, no major presidential candidate has accepted public matching funds, due to the spending limitations that follow the money. The vacancy of public funds has been quickly filled by individuals tapped by campaign operations.

Yet we know surprisingly little about *why* these individuals give. What motivates individual campaign contributions? The classical literature of donor motivations are built on surveys (Brown et al., 1995; Francia et al., 2003; Barber, 2016b), and while surveys are useful, they do not quite take into account the rich information available in the campaign finance disclosure data. This paper complements the surprising lack of literature using this large administrative dataset and contributes as proof that there is much to explore in administrative data alone—especially about the *timing* of giving.

Each transaction in a campaign disclosure report is tagged with the date that the money was received by the campaign. Given a single transaction, I must the ask the following: why did this individual give on this particular date, and not on some other dates? The timing of individual contributions is a critical component to the why-question, as it may help better establish the empirical model of the "purposeful, instrumentally motivated citizen"—or its alternative. Yet, the timing and the dynamics of giving have been much under-explored in the literature.

Hence in this paper, I approach this *why* question by looking into *when* these individual contributions to the presidential race take place, using both nonparametric and semiparametric methods. I use the Federal Election Commission (FEC)'s 2016 election data, cast as a daily time-series of individual contributions in (1) sums and (2) counts. With this data, I test the hypothesis that if presidential donors are either instrumental or momentum-driven, they will be responsive to events that reveal new information about candidate viability, such as early victories or unexpected upsets in caucuses and primaries. That is, I look for *critical events* that significantly motivate or demotivate individual giving, while mapping out a smooth and succinct trend elsewhere to see how the overall giving takes place over time. For this purpose, I employ the sequential segmentation splines in Ratkovic and Eng (2010) to harness the power of smoothing splines while searching for critical breaks.

I find that on the national level, daily aggregates for any candidate are a slow-moving, smooth process, without any particular critical events. The overall smooth estimates show that there is certainly a dynamic element to the underlying political interest during an election cycle. However, there are no sudden changes—everything is more or less gradual. This is true even when the algorithm is explicitly modeled with known events, such as key primary races, dates that presumptive nominees emerge from each party, surprise, upset wins by candidates, or scandals.

Even when I disaggregate the data by state, events that I expect to create shocks hardly ever do. These include events such as the Iowa caucus or the New Hampshire primary, and the few breaks that I do detect run contrary to existing theory. I conclude that overall, donors seem neither uniformly strategic nor momentum-driven, and provide some preliminary interpretations of the exceptions and also of the structural breaks at other unexpected, seemingly uneventful dates.

3.2 Literature

Instrumental and Expressive Giving

Money is extremely important in presidential races, or any other election in the United States—one legislator called it "the mother's milk of politics,"[1] and election campaigns stop when they run out of funds. Thus political giving is another form of democratic political participation, just as is voting, volunteering in campaigns, contacting public officials, political discussions, grassroots local activities, and so on (Rosenstone and Hansen, 1993; Verba et al., 1995). Therefore, campaign contribution literature has relied on similar theoretical frameworks used to model and analyze turnout.

Naturally, contributing has been subject to the same scrutiny posed in the voting literature—that is, of the "failure of rational choice theory" or the "puzzle of participation" (Verba et al., 1995). Just as one voter's vote is not likely to be pivotal, a

[1]https://www.nytimes.com/1987/08/06/obituaries/jesse-unruh-a-california-political-power-dies.html

single campaign contributor's money to a candidate, given the existing legal caps,[2] are not likely to be enough to influence the outcome of the election. This is true of all federal races, including both Congressional and presidential races, and obviously more pronounced in the latter. Yet 23.4%[3] of respondents to the Cooperative Congressional Election Study claimed to have contributed in the 2016 election, among 80.2% of which have given to a presidential candidate. And among these American presidential givers, about 55.9% of them have given *only* to the presidential race. If one's money cannot influence the election outcome, why would anyone give?

Ever since Riker and Ordeshook (1968) attempted to rescue the instrumental calculus of voting in Downs (1957), voting has been regarded as a decision based on both instrumental and expressive components (Fiorina, 1976). By instrumental motivation, I refer to the Downsian utility in which a donor's money has value only when her dollar is pivotal in helping the more ideologically proximate candidate win. By expressive motivation, I refer to the utility that results from the act of contributing itself, such as psychological satisfactions in fulfilling the civic duty or expressing solidarity with the supporting candidate or cause. Which of these two prevail in individual campaign contributors?

One literature[4] has argued more for the instrumental and strategic motivations of campaign contributions, such as (Bartels, 1988), Mutz (1995), Ovtchinnikov and Pantaleoni (2012), and Hill and Huber (2017). This is especially because contributors are thought to be more politically involved, knowledgeable, and sophisticated (Mutz, 1995). Others have argued that political giving is expressive, especially

[2]Currently, an individual's contribution to a federal candidate is capped at $2,700 USD per election, counting the primary and the general elections separately.

[3]CCES 2016 cross tabulations without adjusting with survey weights. While the proportion of presidential givers among all campaign contributors are stably around four-fifth, the proportion of campaign contributors in presidential election years have been declining from 2008 (36.5%) and 2012 (31.5%).

[4]I do not mention some other prominent campaign finance literature, because they are not relevant. Within the thin individual contributor literature,[5] most of the past literature has focused on Congressional contributors. This is simply because their data is more useful in comparing donor types, with all the variability in candidates and states that can be exploited. Unfortunately, presidential donors and Congressional donors often do not overlap: for instance, Francia et al. (2003) have classified Congressional contributors into three types: investors, ideologues, and intimates, who are respectively donors who give expecting some quid pro quo benefit from the legislator, donors who give because of ideological similarity, and donors who give simply out the pleasure of associating with the high social circle created by the legislator. However, this frame is not informative when observing presidential contributors, who overlap little with presidential donors. For instance, in the CCES 2016, while 75% of Congressional donors are also presidential donors, only 26.4% of presidential donors are also Congressional donors. In addition, it is much more difficult to imagine that many investor-type or intimate-type contributors to the presidential candidates.

considering the limits of one individual's money being able to sway the election. Ansolabehere et al. (2003) describe political giving as akin to donating to the Red Cross.

The most prominent of the research explicitly dealing with presidential donor motivation, Brown et al. (1995), classify motivations into three: purposive, solidary, and material. The first is instrumental, and the latter two are expressive, as 'material' contributors in their book indicate those who answered that they give because it is something "expected of someone in my position," if I set aside donors with "business and employment reasons."[6] That is, they recognize that there are both types of donors. Or both motivations behind an individual, with perhaps different weights given between them.

Suppose a donor is strategic and cares about the outcome of the election. How would her contributing behavior change? First, if she is sufficiently forward looking, she is better off giving early in the campaign. Within the election cycle, if she receives new, important information on candidates, she will rationally update her priors on the viability of the candidate, which will increase or decrease the utility from giving. Suppose there is a positive shock to a candidate that is observable to most donors. Then on the aggregate level, it is likely that I will see a financial "bump", a reflection of donors revising their decision to invest in the candidate in question.

Now suppose that a donor is expressive. The Riker and Ordeshook approach, which sets the expressive utility as a constant, gives no information on when this person is likely to give. Given that there are dynamics in the campaign, it is difficult to imagine that the expressive utility from giving is constant. In Section 3.2 I present one branch of theory as to how expressive utility may look like.

Momentum in Elections

Another thread of the campaign finance literature deals with momentum (Mayer, 1996; Mutz, 1997; Steger et al., 2004; Norrander, 2006; Donovan and Hunsaker, 2009; Butler, 2009), which is something distinct from Bayesian updating and "investing" to increase the return. Butler (2009) summarizes this as follows:

> Victory creates enthusiasm among supporters, who increase their cam-

[6] Brown et al. (1995)'s donors were documented only after they have given \$200 or more. This is equivalent to more than \$400 in 2016 dollars. I expect that within the donors observable from FEC transactions, the proportion of donors with material motivations is much lower in recent elections than in past elections. For the discussion of hidden donors by reporting threshold, see Chapter 2.

paign contributions. ... The reverse effect happens for losing candidates.

Momentum in primary victories is an important factor well-recognized in political science, especially due to the sequential nature of the United States primaries (Knight and Schiff, 2010; Redlawsk et al., 2011; Collingwood et al., 2012). Candidates spend a disproportionate amount of money and time in the first few primary sites such as Iowa, which is one of prominent reasons why candidates now shun public matching funds—federal funds require that candidates spend proportionately to the state's population, a high disadvantage in the early primaries. It is also well-known that the media coverage is biased by the sequence in the primaries. Norrander (2015) states that Iowa and New Hampshire *each* receive five to ten times more coverage than the average state's presidential contest.

Why would there be any momentum, an overestimation or an underestimation of the underlying candidate viability? One reason could be that people like backing a winning horse, and they look to each other to catch signals on candidate viability—donating to those with a positive signal will then result in a herding behavior. Another idea is that there is a certain utility from "suspenses" or higher variance in beliefs, and surprises in themselves (Ely et al., 2015).

What is interesting is that whether a donor is instrumental or expressive-momentum driven, both of them will be responsive to the same critical events in the campaign. These include initial primary victories or upset, surprise wins in which the candidate exceeds expectations. Assuming that information on candidate viability is not quite accurate, and donors are generally instrumental or momentum-driven, I can then hypothesize that I will see sharp boosts, apart from a smooth, slow trend, at these particular event dates. In this paper I do not aim to disentangle the two but to test it jointly.

3.3 Data and Methodology

Data

I use the data from the FEC, which makes all contribution records from individuals to presidential committees available, including donors' names, addresses, and occupations. Using this, I can use entity resolution techniques to build a panel data, if there are multiple donations from the same individual. The data is open source but extremely large, complex, and growing exponentially. This has resulted in a low utilization of this data despite the magnitude of information that it provides to many important questions in political giving.

In this paper I limit the analysis to the 2016 election, and its six most prominent presidential candidates—Hillary Clinton (D) and Donald Trump (R), the general election candidates, and Bernie Sanders (D), Ted Cruz (R), Marco Rubio (R), and John Kasich (R), who were unsuccessful in their respective party's presidential primaries.[7] I choose to use only 2016, because it is the first election in which (1) there is no incumbent president running for reelection, and (2) there are no major party candidates that participated in the matching fund system, in either primaries or the general election.[8] This puts the candidates' solicitation strategies on a relatively equal footing. Moreover, 2016 was a highly interesting election. Neither the Democrats nor the Republicans were able to determine a general election candidate during the invisible primary period. By the first week of January 2016, FiveThirtyEight reported that Sanders had narrowed the margins between himself and Clinton to its lowest up-to-date, and Cruz and Trump were neck-and-neck.[9]

Individual contributions to election committees are, by the Code of Federal Regulations, only made public when the individual in question gives more than 200 USD to the committee in question, and only from that moment on.[10] This makes small and early donations invisible to the researcher, and the remaining contribution records sparse. In Table 3.1, I provide key campaign finance statistics, especially sources of their money, by candidate. Note that high reliance on small contributors relative to large means that the pattern of giving may seem more volatile to the eye of the researcher, which is something to watch out for with Sanders and Trump data.

Due to this and the fact that most contributors give only once or twice in the campaign, individual campaign contribution records are not very informative as a panel data, if the researcher wishes to observe the dynamics of a single election campaign. Not surprisingly, most research has to do with tracking donors over different years of elections. In this paper, I collapse the panel into a time-series

[7]For the Republican party, I have chosen candidates who have lasted longest. I could have included Ben Carson, whose popularity peaked and then dwindled from late 2015, or Jeb Bush, who was most prominent at the beginning of the election cycle but lost popularity—here I chose not to, because their observed numbers of individual donations were too small.

[8]https://www.washingtonpost.com/posteverything/wp/2016/02/09/public-campaign-funding-is-so-broken-that-candidates-turned-down-292-million-in-free-money/?utm_term=.2cd0a21dc7cd. In 2000, Bush rejected public matching funds for primaries. In 2008, Obama rejected the public matching funds for the general election, a first for a major party candidate. In 2016, every major candidate rejected the public money for both races, not counting O'Malley.

[9]FiveThirtyEight, 2016 Primary Forecasts, Iowa Democratic and Republican Caucuses, polls-only forecast. I do not use the poll-plus forecast because this takes the fundraising state into account as well.

[10]11 CFR 104.8 - Uniform reporting of receipts.

Contribution Type	Clinton	Sanders	Trump	Cruz	Rubio	Kasich
Individual, Large	52.67%	41.62%	14.01%	60.99%	66.37%	75.98%
Individual, Small	18.53%	57.70%	25.93%	38.16%	24.34%	22.04%
PACs	0.31%	0%	0.04%	0.10%	2.08%	1.2%
Public Funds	0%	0%	0%	0%	0%	0%
Self-finance	0.25%	0%	19.77%	0%	0%	0%
Other Transfers	28.23%	0.68%	40.24%	0.75%	7.21%	0.78%
Number of Donors Observed	519,714	250,352	205,847	99,554	35,136	13,795
Number of Donations Observed	3,007,977	2,089,355	402,917	509,474	93,381	24,845

Table 3.1: Key Campaign Finance Statistics by Candidate

observed daily, with two quantities of interest—daily contribution sum, and daily contribution counts. Although I am primarily interested in the former, the latter is a more robust measure to individual-level heterogeneity in wealth, and I intend to look for structural breaks in both quantities.

Using campaign finance data as a time-series is a relatively unexplored approach in the campaign finance literature. The only such paper to our knowledge is Christenson and Smidt (2011), which explored the daily fundraising for the 2008 primaries for both major parties using Kalman filtering to account for the noise created by days of the week and FEC report deadlines. The paper provides an excellent argument as to why the daily observations of fundraising measures may not equal the underlying contribution dynamics. However, their interpretation of candidates' latent fundraising strengths relies heavily on visually interpreting the figures. In this paper, I offer more concrete, objective measures that determine whether an event is "critical" enough to create a structural break, and yet is simultaneously able to smooth the estimates, as will be explained in Section 3.3.

Sequential Segmentation Smooth Splines

In this paper I have two purposes: to test existing hypotheses of giving by looking for structural breaks at certain dates, as well as estimating a stable, smooth estimate of campaign finance dynamics. For this purpose, I employ the sequential smooth spline method of Ratkovic and Eng (2010), a modification of the smooth spline method designed to look for location and number of unknown jumps.

This nonparametric method is a modification to partial splines (Wahba, 1990; Gu and Wahba, 1993; Kim and Gu, 2004; Gu, 2013) designed to account for sudden breaks in the data. Inspired by how the smooth splines systematically overestimates and underestimates George W. Bush's popularity, right before and after the

9/11 Attacks in 2001 respectively, the method looks for the existence, the number of, and the locations of unknown jumps in the data. After potential breaks are identified, several models are fitted, sequentially adding breaks. Then the modified Bayesian Information Criterion (BIC) is calculated for each model and the algorithm automatically determines if jumps should be added, and if so, how many.

More specifically, recursive procedure is as follows: firstly, the algorithm fits a cubic smoothing spline to the data. That is, it assumes the target y_i is linear in a smooth function of x_i, so that $y_i = f(x_i) + \epsilon_i$ with $E(\epsilon_i) = 0$, $\mathrm{var}(\epsilon_i) = \sigma_\epsilon^2$. This smooth function is estimated with an unpenalized linear trend and penalized interpolation:

$$\hat{f}_{SS} = \min_f \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int \{f''\}^2 \, dt \tag{3.1}$$

I then remove the linear trend from the residuals, $\hat{e}_1, \cdots, \hat{e}_n$, and search the residuals for a breakpoint, by running a modified binary segmentation method.[11] Next, I add the most likely break found to the partial spline's unpenalized space as follows:

$$\hat{f}_{SS} = \min_{f,\beta} \sum_{i=1}^{n} (y_i - f(x_i) - \beta I\{x > \text{breakpoint}\})^2 + \lambda \int \{f''\}^2 \, dt \tag{3.2}$$

This is repeated a number of times, sequentially finding the next most likely break and augmenting it to the existing breaks. After a series of models are generated—model with no jumps, model with only the first break, model with the first and the second breaks, and so on—I calculate the modified BIC for each of these models. The model with the lowest statistic is chosen as the final model.

The original paper is focused on comparing the sequential segmentation splines with existing smoothing functions, including the canonical smooth splines, Kalman filters, and LOESS, and shows the strength of the method in several different ways, such as the behavior of the residuals. In addition, the method warrants comparisons with existing literature on detecting breakpoints, such as Bai (1994, 1997); Bai and Perron (2003), and Bai and Perron (1998). The classical structural break tests are essentially "tests against changes in the coefficients of a linear regression." (Zeileis et al., 2003). Here I am not fitting linear regressions but taking a nonparametric

[11] For details, see Ratkovic and Eng (2010) and Sen and Srivastava (1975).

approach to accommodate both (1) a flexible underlying dynamics in individual campaign contributions, and (2) automatic selection of breaks with principled model selection. Hence Ratkovic and Eng (2010) is better than classical tests in this regard. I briefly go over how each method performs with observational data in Appendix C.1.

Practical Extensions

In this paper, I make some extensions to (Ratkovic and Eng, 2010). To test for known events, I not only run the spline method as is—that is, assuming that all breaks are unknown—but also run the method a second time, in which the jump is directly modeled into the unpenalized space of the partial splines. This is a form of semiparametrics, and a measure to recognize that I do indeed *know* the potential location of jumps, although these may not be the first and foremost choices in structural breaks. If I intentionally fit known jumps into the first of the sequence of breaks, the algorithm will compare the BIC statistics between the different models and decide whether adding this known jump—and a series of other jumps—will benefit than having only smooth splines. However, as the original paper warns, this is essentially "assuming the answer to the question we are asking."

I apply the following two principles to guard against false positives. First, after the initial estimation, I bootstrap the residuals and re-estimate the splines—and jumps, if any—from the generated fake data. This is to generate a confidence interval for our estimates, and especially the standard errors for the jump sizes, as suggested by the original paper.[12] However, one caveat is that when the signals are not strong enough in the data, the jump may disappear when the spline method is rerun on the newly generated fake data. The more this takes place, the more it is likely that the initial estimate with the jump is likely to be a false positive, especially when I forcibly model known jumps into the splines. I declare a jump as a true positive only when the jump is rediscovered 90% or more times in its bootstrapped estimates. Second, when the jump size is very small, the rediscovered jumps sometimes display different signs. Inconsistency in the sign of jumps estimates is also an indication that the jump is a false positive. For jumps that are not 90% times or more positive or negative, I assess them as false positives.

[12] After performing an augmented Dickey-Fuller test, I find that the residuals are not quite stationary. This indicates that it would be better to block bootstrap the residuals, by either non-overlapping block bootstraps (Carlstein et al., 1986) or moving block bootstraps (Kunsch, 1989; Liu et al., 1992).

The set thresholds for both principles are arbitrary, and I leave a formal assessment such as simulations for future work. However, these are very stringent criteria.[13] Figure 3.1 is a case in which no jumps are found without a model, but when a known jump is specified in the spline, there is a jump which is rediscovered 70% or more times when bootstrapped. Figure 3.3 is a case in which a jump is found even without a model. When bootstrapped, the jump is found 99.6% of the time.

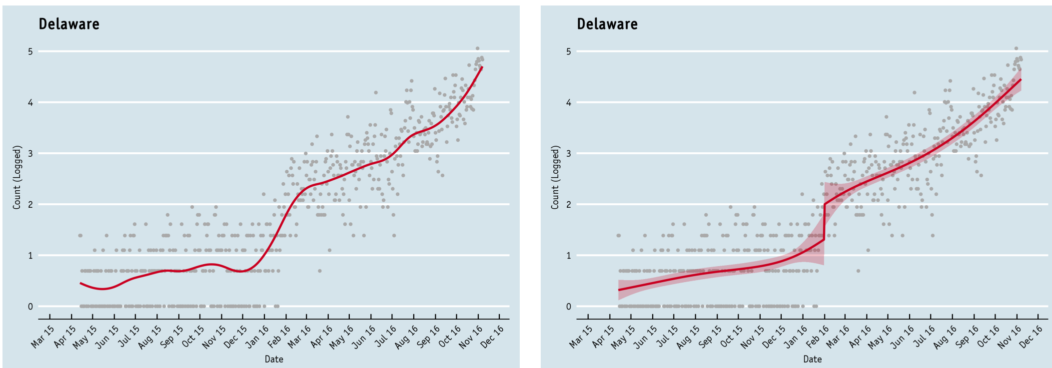


Figure 3.1: Comparison of Estimates with and Without a Model, Jump Rediscovery Rate of 70.1%

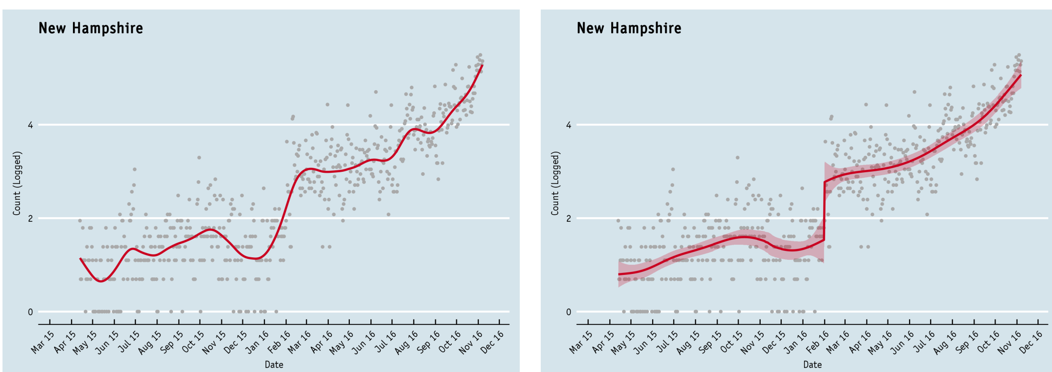


Figure 3.2: Comparison of Estimates with and Without a Model, Jump Rediscovery Rate of 88.8%

Note that I use logged measures of our original quantities of interest. This is due to the fact that as the election intensifies, the variance of the data increases. I show an example of the raw and logged measures in Figure 3.4. The residuals, as can be seen in Figure 3.5, is not well-distributed around zero, and cannot be bootstrapped

[13]For instance, even the Bush poll data used in Ratkovic and Eng (2010) will fail the first criteria—the large and immediate jump, which is apparent to the eye, will only be estimated 70% of the time when bootstrapped. The jump at the Iraq invasion will be rediscovered even less.

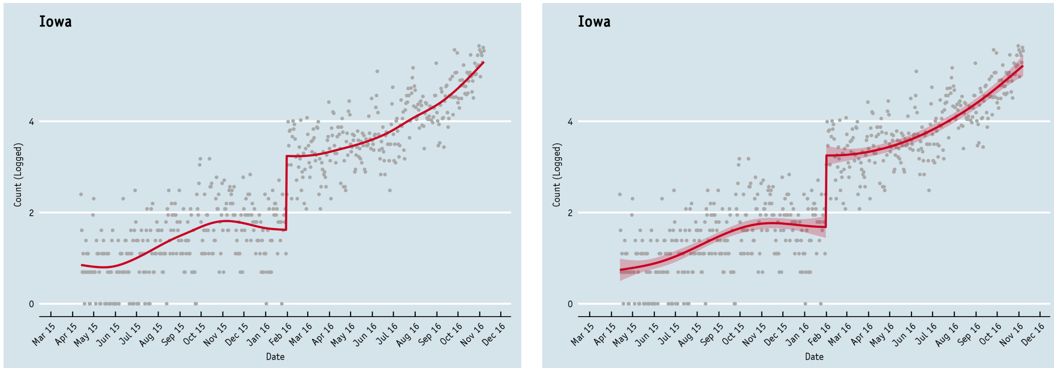
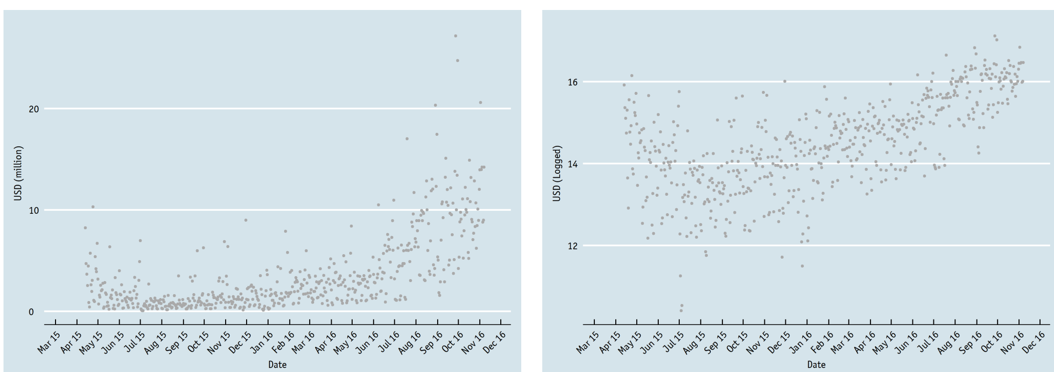


Figure 3.3: Comparison of Estimates with and Without a Model, Jump Rediscovery Rate of 99.6%

without violating the underlying distributional assumptions. The logged measure does much better in terms of residual distributions. Note that this changes the interpretation to responsiveness in *elasticities*.



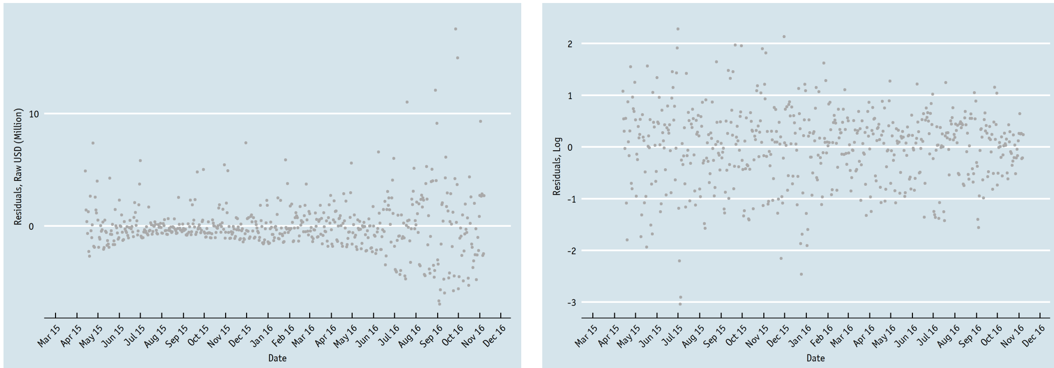
(a) Raw USD

(b) Logged USD

Figure 3.4: Example of Raw and Logged Contribution Sum, Clinton, National Aggregate

Lastly, I further validate the jump by running the spline method not only on the logged quantity of interest, but also on the residuals of the said quantity after netting out the effects of weekends, FEC deadlines, and end-of-month dates. This is the minimal de-seasonalization of campaign finance, some mentioned already in Christenson and Smidt (2011). These variables indeed have a strong linear correlation with campaign contributions, and I accept only the jumps that are also found when the algorithm is run on the residuals. That is, jumps that persist when I have excluded the effects of the three covariates.[14]

[14] I can also directly model those covariates into the unpenalized space of the splines, but this is



(a) Raw USD (b) Logged USD

Figure 3.5: Example of Residuals, Raw and Logged Contribution Sum, Clinton, National Aggregate

3.4 National Jumps

I first ask the following question: do any known critical events create a national jump in campaign finance? If so, which ones are more important than others? Figure 3.6 shows national trends estimated by the sequential smoothing spline method for all six candidates.

I see that the underlying process estimated is surprisingly smooth on the national level—there are no jumps anywhere, for any candidate, and for any of the two quantities of interest. Table 3.2 presents the BIC statistics for these estimates. It is clear that the minimum BIC statistics from sequentially augmented models are nowhere close to the BIC without breaks, in all twelve cases.

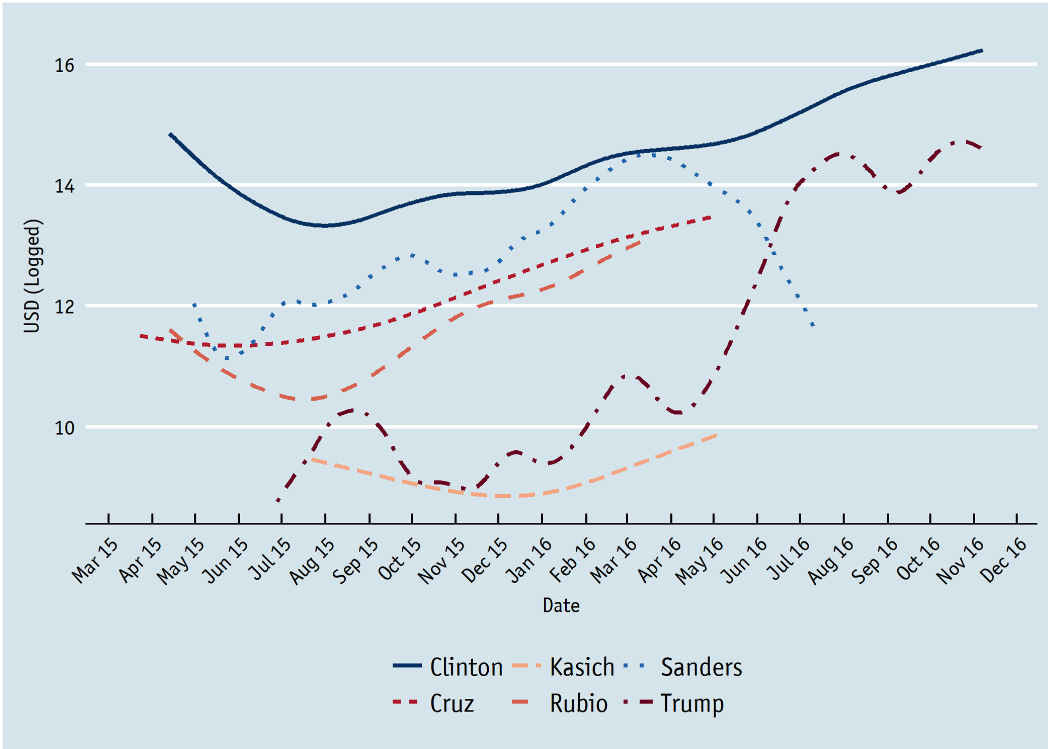
Although no events were detected, I can still extract useful information from the smooth estimates. First of all, note how the candidate strength measured by campaign finance does not correspond to forecasts from national primary polling, shown in Figure 3.7.[15] For instance, in the Democratic party, although the campaign finance began falling during the month of April 2016, polls for Sanders was still rising. In the Republican party, note how Kasich garnered popularity as Rubio faltered and withdrew, but elasticity[16] of giving did not follow the popularity increases. The comparison remains similar when compared to log of popularity to make the quantities comparable in interpretation.

Save for Trump, all candidates start off their campaign with large contributions

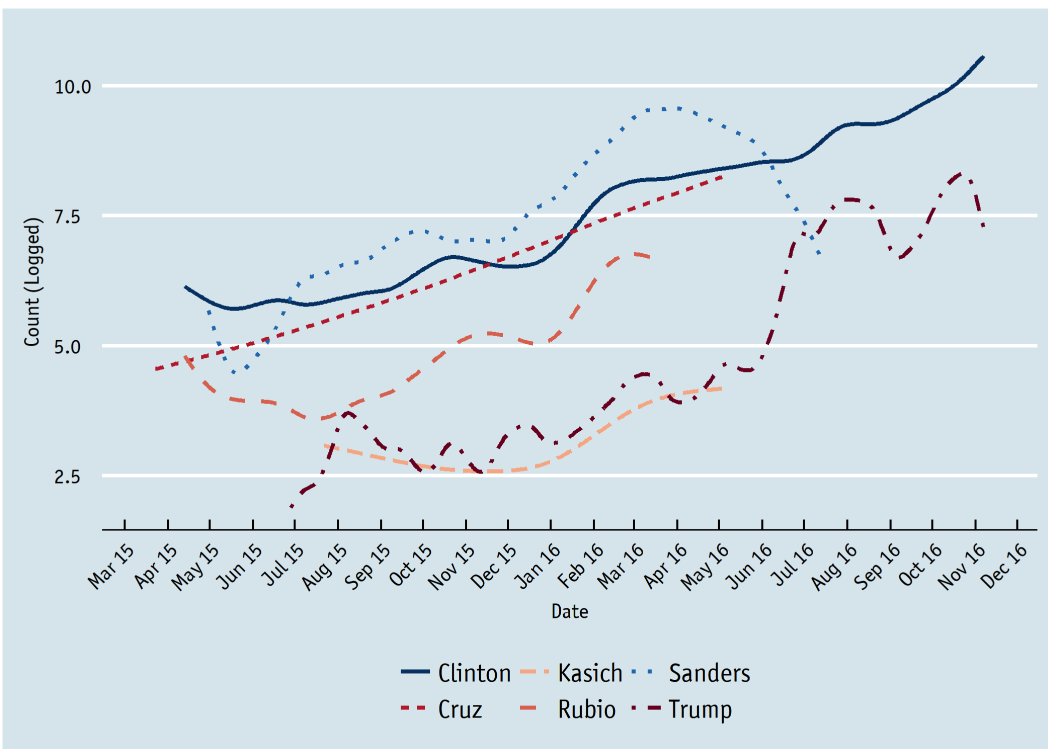
outside the scope of this paper.

[15]For all major and minor candidates, see Appendix C.2.

[16]Note again that these are logged measures.



(a) Contribution Sum



(b) Contribution Counts

Figure 3.6: National Estimate by Candidate

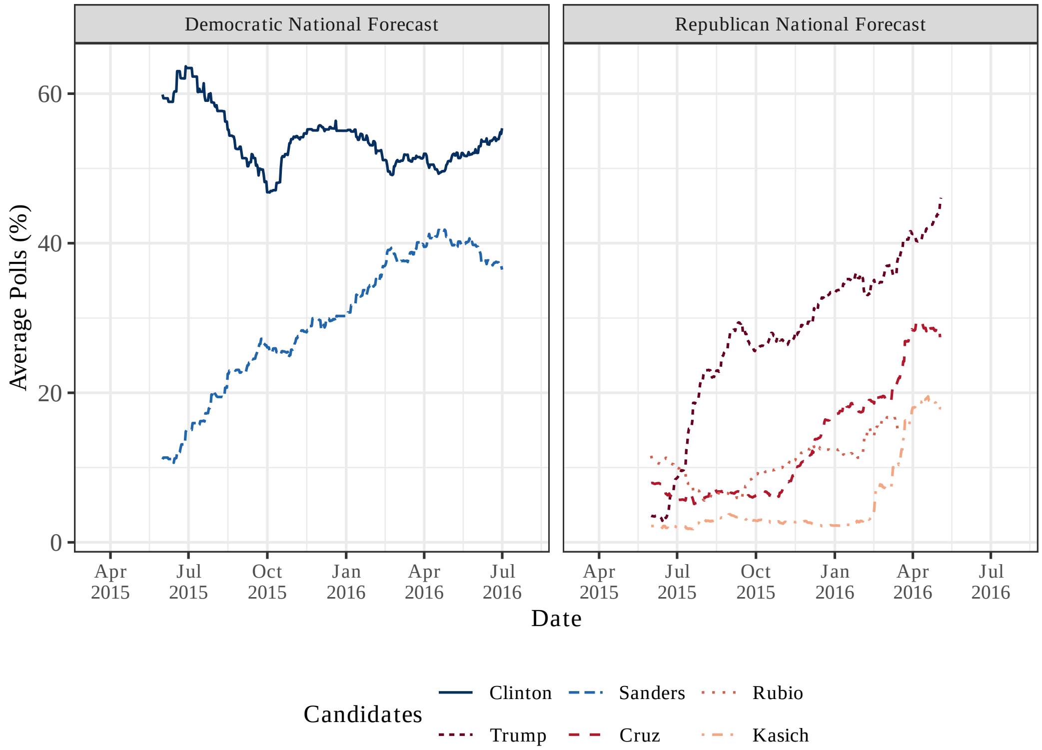


Figure 3.7: National Primary Polls by Party, FiveThirtyEight, 2016 (Selected Candidates)

arriving on the first few days. The excitement cools over the first half of year 2015, then starts climbing again at various points by the candidate. The exception, Trump's war chest, shows large fluctuations—I believe this is mainly because Trump relied much less on individual contributors than any other candidate. Plus, I can clearly see that Clinton's war chest overwhelms any other candidates' at any given point in time during the election cycle. Sanders donors give many more times though, suggesting that each transaction from a Clinton supporter is much larger than one from a Sanders supporter, which is not very surprising.

The local maxima of these smoothed estimates are worth noting. Sanders' global maximum is at Super Tuesday (March 1, 2016), and Trump's local maxima in 2016 are each at March 3, July 30, and October 23, which are respectively quite close to Super Tuesday, Republican convention, and the last general election debate. Naturally, I must ask whether critical events will be detected if I aid the algorithm by specifying where the known events occurred. That is, I can try modeling a set of breaks that I have reasonable prior expectations to be influential on campaign finance, directly into the unpenalized component of the partial splines.

Candidate	Quantity	Selected Date	BIC	Candidate	Quantity	Selected Date	BIC
Clinton	Sum	Splines Only	11.81	Sanders	Sum	Splines Only	18.02
		2015-07-02	17.83			2015-12-07	23.87
		2015-07-06	19.98			2015-05-25	28.69
		2015-05-08	18.16			2015-06-27	33.67
		2015-07-18	20.48			2015-06-25	35.32
		2016-01-25	24.58			2016-10-16	38.87
	Counts	Splines Only	20.40		Counts	Splines Only	21.35
		2015-07-01	26.81			2015-05-25	25.73
		2015-07-18	29.21			2015-06-25	25.56
		2015-05-12	33.15			2015-12-12	29.53
		2016-06-10	37.86			2016-06-27	35.59
		2016-03-17	42.19			2016-07-02	38.34
Trump	Sum	Splines Only	19.09	Cruz	Sum	Splines Only	6.322
		2016-07-01	24.75			2015-10-26	9.107
		2015-12-23	27.76			2016-01-25	11.77
		2016-09-26	30.51			2015-06-15	17.42
		2015-07-30	33.77			2015-09-14	21.13
		2015-10-12	38.20			2015-07-22	25.87
	Counts	Splines Only	34.63		Counts	Splines Only	5.281
		2016-07-01	36.82			2016-01-18	9.451
		2015-07-30	39.10			2015-04-11	13.08
		2015-10-15	41.60			2015-10-26	17.31
		2016-06-20	43.62			2016-06-15	19.62
		2015-12-23	47.17			2015-09-14	25.55
Rubio	Sum	Splines Only	7.175	Kasich	Sum	Splines Only	4.527
		2015-07-01	13.88			2016-02-16	7.501
		2015-12-24	17.62			2016-01-05	11.25
		2015-04-17	15.33			2015-12-18	15.00
		2015-05-01	19.16			2015-11-16	18.75
		2016-01-25	24.26			2016-01-01	22.50
	Counts	Splines Only	11.67		Counts	Splines Only	6.148
		2015-07-01	16.09			2016-02-08	7.501
		2016-01-25	16.45			2016-01-11	11.25
		2015-04-17	18.27			2015-12-18	15.00
		2015-12-17	20.26			2015-11-16	18.75
		2015-05-01	23.57			2016-03-07	22.67

Table 3.2: First Five Events Selected, Per Candidate and Quantity of Interest, National Level

I divide the election cycle into three periods in which the dynamics differ: the pre-primary period, the primary period, and the post-primary period. For each of the periods, I use the following events while searching for possible breaks: (1) Republican and Democratic primary debate dates for the pre-primary period,[17] (2)

[17] Norrander (2015) points out that debates are high-stake events during the invisible primary and the earlier primary season, and that voters are influenced by the candidates' performance during

Iowa caucus, New Hampshire primary, Super Tuesday, Super Tuesday II,[18] and the Acela Primaries[19], dates in which presumptive nominees emerged[20], and end-of-convention dates for the primary period, and (3) general election debate dates and October 7, 2016, for the post-primary period, in which *Access Hollywood* video tape with Trump appeared in the Washington Post and WikiLeaks released Podesta emails, a rare date in which big scandals for both party presidential candidates took place. While the choice of these events is not based on any existing measures, it spans key events in which the media's horse-race coverage were focused on.

Again, I find no breaks—modified BIC statistics are lower when there are no jumps but just splines. I conclude that the national level campaign finance is too smooth to have any jumps in the elasticity of giving. Henceforth, I now look for jumps by disaggregating the data into money from each state, in an attempt to prevent the data from canceling out meaningful variations.

3.5 Initial Primary Victories and Surprise Victories

I now ask our main question: do donors respond to either (1) "surprises" in viability or (2) initial primary wins and losses? In order to ask this question, I must first calculate at which races—if any—candidates exceeded expectations. I utilize FiveThirtyEight, a website that analyzes opinion polls since 2008, and which uses a model to weight pollsters by date and their accuracy, and provide a daily forecast of by-candidate vote share.[21]

The races in question for the two separate questions overlap, especially for the Republican party. Rubio did very well in the Iowa caucus, 6.7% more (23.1%) than was predicted (16.4%) right before the caucus. Cruz also did better (27.6%) than predicted (23.9%), while Trump fared worse (24.3%) than predicted (28.6%). In the New Hampshire primary, while Trump did much better (35.2%) than pro-

these dates. More than fifteen million watched the first Democratic debate on October 13, 2015, broadcast on CNN. The Republican counterpart, first on August 6, 2015 and then on September 16, 2015, boasted more than twenty-three million viewers separately for each event.

[18]This is March 15, 2016, in which there were races in five states, Florida, Illinois, Missouri, North Carolina, and Ohio

[19]This is April 26, 2016, in which there were races in five states, Connecticut, Delaware, Maryland, Pennsylvania and Rhodes Island, also referred to as Super Tuesday III.

[20]These are respectively June 6, 2016, and May 4, 2016, for Democratic and Republican party. Super Tuesday IV, in which there were races in California, Montana, New Jersey, New Mexico, South Dakota, was not a consideration for this reason.

[21]Unfortunately, there are states in which no or very few polls are conducted due to low interest in its outcome, resulting in only around half of the fifty states to have reasonable forecasts before the race. However, this also means that the race is not of much import, so I can safely set aside these states for my purposes.

jected (29.8%), it was Kasich who surprised, coming in second rather than third behind Rubio as was predicted. After that, there were no notable upset victories or surprises,[22]

For the Democratic party, Iowa and New Hampshire races did not deliver much surprises compared with the polls. However, Sanders delivered upset victories in Michigan and Indiana, also confirmed by news reports at that time. In Michigan race, March 8, 2016, Sanders was projected to get a vote share of 36.9% against Clinton who would win 58.2%, but the results came in 49.8% against 48.3%, Sanders winning the state. Similarly, in Indiana race, May 3, 2016, Sanders scored 52.5% of vote share against 47.5%, beating his expectations at 42.2% against 49.3%. Given this knowledge about the two major parties, are there any jumps from each candidate-state time-series on the dates in question?

Estimating Events as Unknown Jumps. Again, when the sequential smoothing spline method is run without any specified jumps, there are no particular breaks detected on the day of the primaries, the day before, or the day after. One exception is Kasich's donors in five states: Connecticut, Florida, New York, Texas, and Virginia, in all of which there is a positive jump on February 8th.[23] Again, this means that there was a surge *starting from* the day of the New Hampshire primary itself.

Although the trend is not strong enough for the jumps to persist when bootstrapped (see Figures C.7 and C.8), this is an interesting phenomenon documented: Rubio and Kasich donors had some states with breaks on dates close to the Iowa caucus and New Hampshire primary, respectively. Note that for both candidates, these races are the ones in which they have exceeded expectations. I find that in contribution counts, from January 21 to January 31, 2016, twenty-one states had positive jumps for Rubio.[24] I also find that there was a jump from Kasich supporters in Kansas on

[22] This can be also seen from Google New searches with relevant keywords during the primary period.

[23] There is also a case in which sum of contributions from Trump's supporters in Illinois showed a slight jump on April 19, 2016, the date of the New York primary in which Trump had a large win in his home state. However, given that the first three dates selected and recognized as breaks in this particular time-series is November 5, 2015, April 19, 2016, December 23, 2015, these jumps may have been purely accidental due to Trump having so few individual over-$200 donors. Corresponding BIC statistics are 23.17, 26.89, and 16.10, and the splines-only BIC is 16.49. Surely enough, when bootstrapped, the jump disappears.

[24] These are, in the order of the date of the jump, Rhodes Island (January 21), Alaska and Oregon (January 24), Iowa and New Mexico (January 25), Colorado, Minnesota, Oklahoma, South Carolina, South Dakota, and Utah (January 26), Arkansas, Louisiana, Wisconsin, and Wyoming (January 27),

February 7. Note that these are all *before* the races and not after, contrary to our expectations—this is shown graphically in Section 3.7. Again, although only one—Rubio's Arkansas, in contribution counts—survives our stringent bootstrapping criterion (See Figures C.9 and C.10), this is an interesting pattern in the data, even with such few sample size comparative to other candidates.

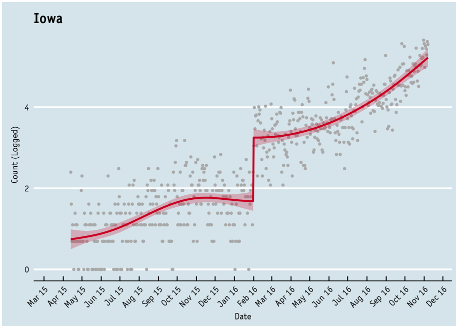
What the pattern documents is not very clear-cut at this point. If candidates that are—by January 2016—not the top two primary candidates experience jumps before key races in which they have done well, why would that be? I could speculate on several possibilities. First is that the donors are rationally investing in the upcoming battle, in the hopes that they can increase the marginal probability of winning for their favorite candidate. That is, it is forward-looking agent's—either on the demand or supply side, or both—attempt to back up the campaign's war chest to influence the key races. Second is that the campaign finance is simply a real-time reflection of the surging popularity of these candidates—that is, there is an endogeneity, but somehow happened to create a break. Third is a possible winner-take-all type of forecasting utility, a satisfaction from backing the winning horse, which makes agents overstate their private signal in the Ottaviani and Sørensen (2006) sense, which may be why these jumps are observed only in second-tier candidates. Distinguishing between these behaviors will require much more data, and an analysis of other election years as well.

Estimating Events as Known Jumps. When the four races of interest—Iowa, New Hampshire, Michigan, and Indiana—are directly modeled and tested for jumps, and the same bootstrapping principles applied, only a handful of Clinton's and Trump's data survives, and only the jumps at either the Iowa caucus or the New Hampshire primary. These are presented in Figure 3.8, and Figure C.11 plots how the splines would have been estimated without direct modeling of the events.

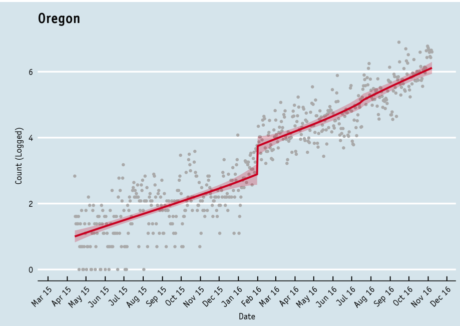
The final results defy expectations somewhat. Clinton's Iowa race, although Clinton won, was such a narrow win, that some caucuses resorted to such extreme measures as coin tosses to decide a winner.[25] In addition, the vote share was below her original expectations. Hence it is not the case that an unexpected victory brought her a positive boost in campaign money. However, there *was* a boost. The case of

Connecticut and Indiana (January 28), Mississippi and Nebraska (January 30), and Kentucky and North Dakota (January 31).

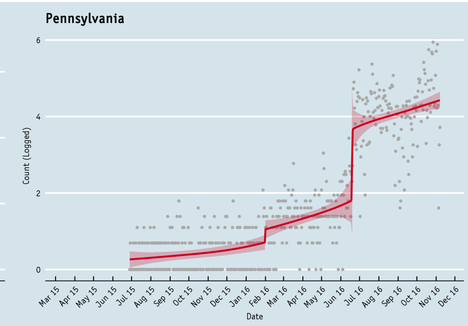
[25]https://www.washingtonpost.com/news/the-fix/wp/2016/02/02/heres-just-how-unlikely-hillary-clintons-6-for-6-coin-toss-victories-were/?utm_term=.85d8f65dcc84



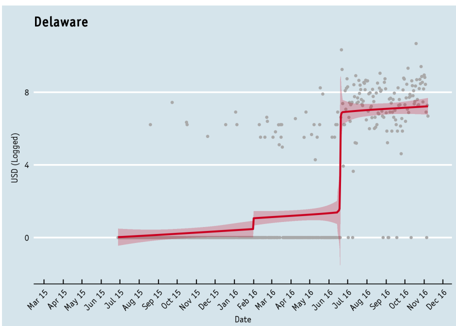
(a) Clinton, Iowa, Jump at the IA Caucus, Jump Size: Mean 1.6, Std. Err. 0.2



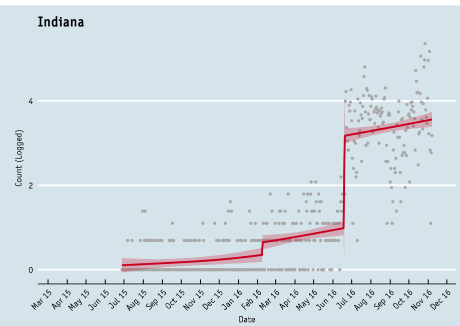
(b) Clinton, Oregon, Jump at the IA Caucus, Jump Size: Mean 0.9, Std. Err. 0.1



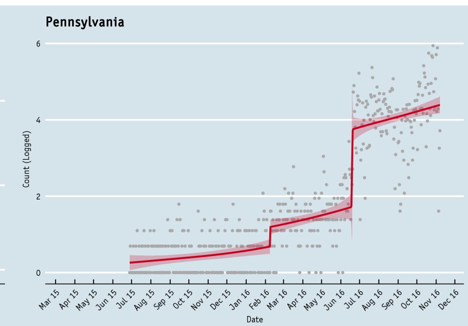
(c) Trump, Pennsylvania, Jump at the IA Caucus Jump Size: Mean 0.4, Std. Err. 0.1



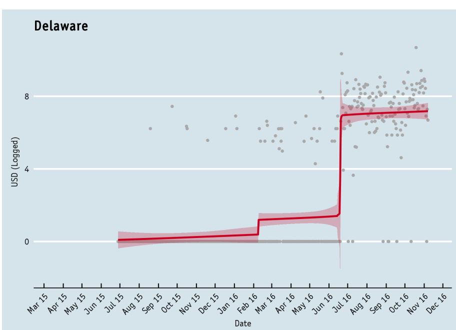
(d) Trump, Delaware, Jump at the IA Caucus, Jump Size: Mean 0.6, Std. Err. 0.4



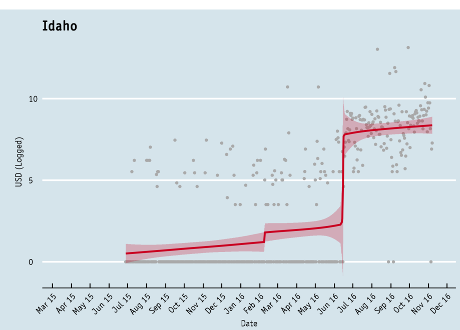
(e) Trump, Indiana, Jump at NH Primary, Jump Size: Mean 0.3, Std. Err. 0.1



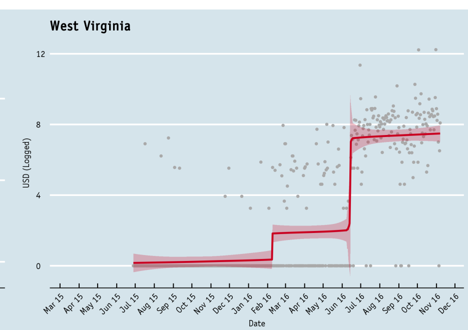
(f) Trump, Pennsylvania, Jump at the New Hampshire Primary, Jump Size: Mean 0.5, Std. Err. 0.1



(g) Trump, Delaware, Jump at NH Primary, Jump Size: Mean 0.8, Std. Err. 0.3



(h) Trump, Idaho, Jump at NH Primary, Jump Size: Mean 0.6, Std. Err. 0.5



(i) Trump, West Virginia, Jump at NH Primary, Jump Size: Mean 1.5, Std. Err. 0.4

Figure 3.8: Cases with Structural Breaks at Initial Primaries

Clinton's contributions in Oregon is also rather peculiar. Clinton was always leading in the few opinion polls led in Oregon[26] but Sanders declared victory in Oregon by twelve percentage-points. Why Oregon of all places should be responsive to the Iowa caucus, albeit its jump size smaller than in Iowa, it not clear.

The various results for Trump are also difficult to interpret—these states are not easily clusterable by covariates such as size of the delegates, the timing of the primary, popularity of Trump, and so on. However, I am wary of reading too much into jumps from Trump's data, as he had much less donation counts than Clinton, and jumps may be products of a small sample size. Note also that Pennsylvania and Delaware will display the Iowa caucus as a jump or the New Hampshire primary as a jump, depending on what known jump it is model with. A model that includes both is rejected by the BIC statistics. This indicates that I must further have means to choose one of the known jumps if two jumps are close to each other and incompatible.

3.6 Local Events: In-state Wins or Losses

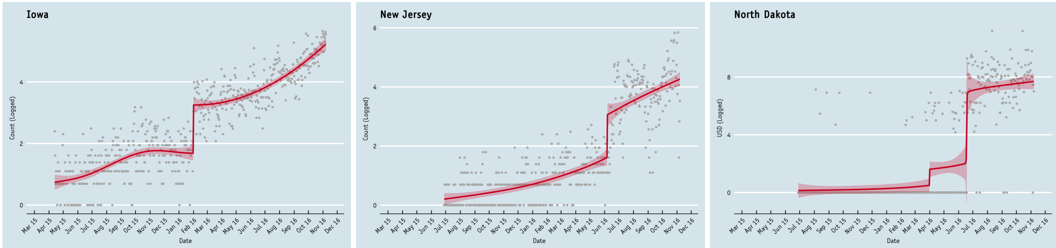
I ask one more question: do donors respond to local events? Here I limit local events to in-state primaries.[27] This is not a question directly related to the theory that we presented in Section 3.2. However, in light of finding a significant jump in contributions from Clinton-supporting Iowans on the Iowa caucus date, I wish to make sure that I can rule out the effect of local races. Theoretically, a local race's influence may be justified per the social network literature, notably (Sinclair, 2012). Sinclair showed that in the 2008 election, people were influenced by how their friends were giving.

For each state, I model a jump at their primary or caucus date and test if the jump survives (1) the modified Bayesian Information Criteria, (2) de-seasonalizing weekend effects, FEC deadline effects, and end-of-month effects, (3) bootstrapping residuals and again detecting for jumps. If I only take jumps that survive 90%[28] or more times and have consistent jump sign during bootstrapping, I end up with only two cases aside from the aforementioned Clinton-Iowa, one from Trump-supporters of New Jersey and one from North Dakota. The following figures show the two jumps:

[26] http://www.oregonlive.com/mapes/index.ssf/2015/08/oregon_presidential_poll_hilla.html#incart_river https://www.opb.org/news/series/election-2016/bernie-sanders-hillary-clinton-donald-trump-oregon-poll/

[27] It would be interesting to track where the candidates rally, geocode the location, and see if these boost local contributions, but I leave this for future research.

[28] If I restrict this to 95%, only the New Jersey case of Trump supporters survives.



(a) Clinton, Iowa, Jump at the Iowa Caucus (Feb 1, 2016), Jump Size: Mean 1.6, Std. Err. 0.2

(b) Trump, New Jersey, Jump at the New Jersey Primary (Jun 7, 2016), Jump Size: Mean 1.5, Std. Err. 0.2

(c) Trump, North Dakota, Jump at the North Dakota Caucus (Apr 1, 2016), Jump Size: Mean 1.2, Std. Err. 0.3

Figure 3.9: Cases with Structural Breaks at In-state Primaries

On top of finding very few jumps with this local event specification, the few results have nothing in common. They are neither similar in primary form, dates, whether candidate exceeded expectations, and so on.[29] Indeed, the only thing common about these three outcomes is that they are for candidates who became the general election candidates, the same conclusion I have drawn in Section 3.5. In addition, if I do not specify these events, the sequential smoothing spline method will conclude that these are not jumps, except for the first case. For now, I leave set aside these findings as exceptions and conclude that local events per se have no effects on campaign finance.

3.7 Events Unknown A Priori

Finally, I ask the following: are there any critical events in campaign finance that are unknown a priori? This is an important question, and the methodology itself was designed to find location and number of unknown breakpoints.

I run the algorithm by each candidate and state, without specifying any known events in the partial splines. In Figures 3.10 and 3.11 I plot these break dates against the break sizes. The point sizes in the Figure are each state's population size, which is a measure to see if these breakpoints are results of small potential donor population.

I see that there are certain clusters of jumps by state per candidate. Setting aside the aforementioned jumps of Rubio and Kasich donations near Iowa and New Hampshire races, first of all, note that there is a big jump in Trump's war chest around mid-June in both contribution sum and counts. In Figure 3.10 it starts with Wisconsin on May 30. Twenty-one states experience jumps starting from May 30 to June 21, dispersed

[29]New Jersey race was long after Trump became the presumptive nominee, and North Dakota has no meaningful enough set of in-state primary polls to have a forecast.

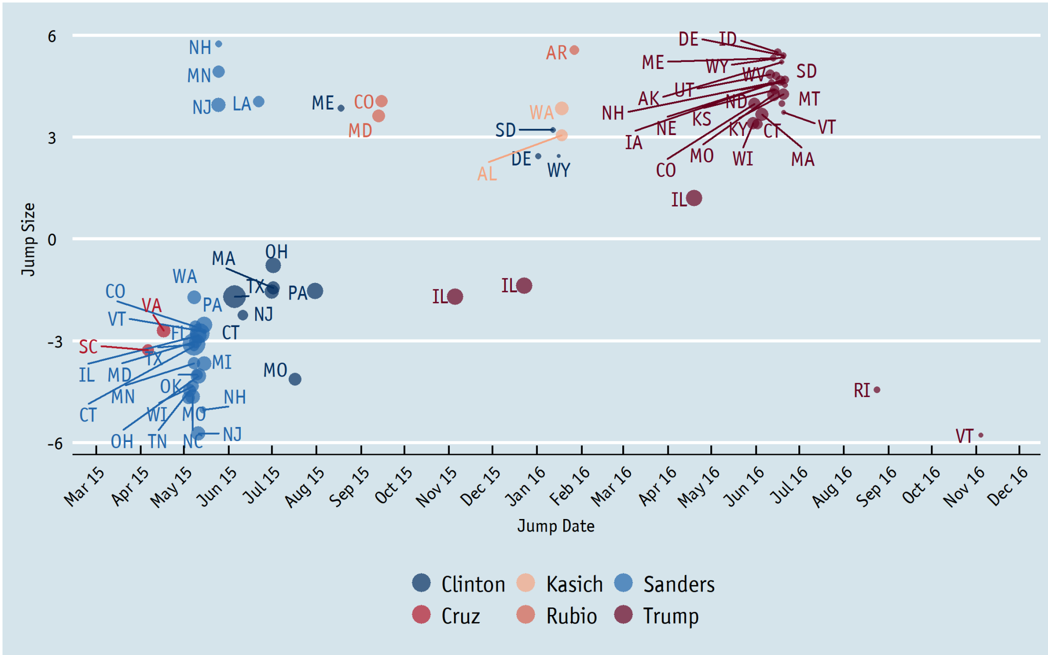


Figure 3.10: All Detected Breaks by Candidate and State, Contribution Sum

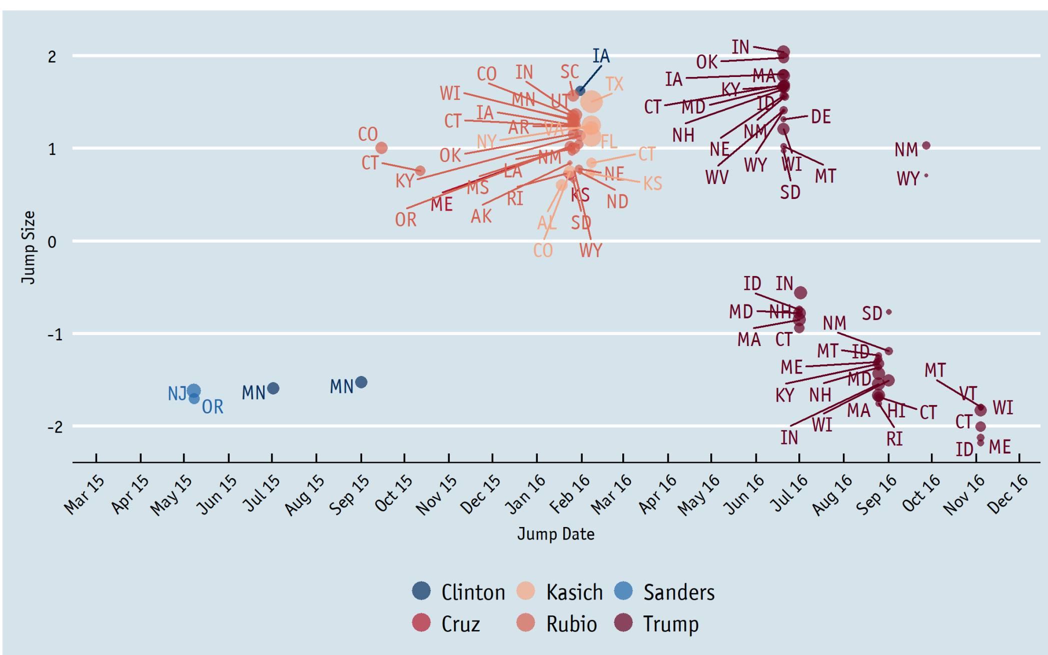


Figure 3.11: All Detected Breaks by Candidate and State, Contribution Counts

over the three weeks but most concentrated on June 20. In Figure 3.11 there are fifteen states that had a jump on June 20.

What happened on this day with the Trump campaign? Trump fired Corey Lewandowski, his campaign manager, on this date, but it is hard to believe that this had such a big, positive impact on donor behavior. I interpret this to be not an external shock but an internal change within the campaign operations. June 20 is a major FEC deadline mid-year in which the campaign must put its full effort into raising money, to display its viability to the public and convince voters. The Trump campaign had done quite poorly on raising money before the month of June, at which they only had 1.3 million cash on hand. The Trump campaign launched extensive solicitations during June to compensate, and was able to gather 20.2 million cash on hand by the end of June.[30] However, why only some states were responsive, and others not, is not immediately clear—if the intensity of solicitations notably differed by state, they do not align with more sensible swing state choices, or anything reported from the media.

Secondly, for the Trump campaign's contribution counts, I also observe three clusters of dips, each respectively on July 1 (six states, ± 1 day), August 25 (eleven states, ± 1 day), and November 4 (six states). Again, there seem to be no particular external events that could have hurt the Trump campaign, and this time it is neither the case that these dates are linked to FEC deadlines. The latest of the dips, November 4, may be understandable if donors calculate that further giving on a date so close to the campaign will not be very effective, but it is then questionable why I do not see the same pattern for Clinton supporters, or for Trump supporters from other states. Overall, there are these inexplicable breaks if approached from information update point of view, as they are relatively uneventful dates.

Finally, as shown in Figure 3.10, there are the dips in the contribution sum from Sanders donors from May 4 to May 15, in nineteen different states. Considering that the campaign was announced on April 30 and Sanders' major rally only began on May 26 in Burlington, Vermont,[31] it is likely that this is a relic of bunched reporting from the campaign, and nothing of substantial import.

[30] See FEC filing here. A Fortune article on July 21, 2016 summarizes this well.
[31] http://time.com/3895770/bernie-sanders-hillary-clinton-vermont/

3.8 An Early Look into the 2020 Data

So far, I have explored the 2016 election, where there were only a handful of presidential candidates in each party. One may argue that the smooth processes seen in Figure 3.6 is because there was not much uncertainty in winnowing of the candidate pools, perhaps except for Trump's surprise strength in the early primaries.

In this 2020 election cycle, the situation was dramatically different. With an incumbent president, there were only four candidates who threw their hats in the Republican party, but in the Democratic party, there were twenty-eight named Democrats vying for the president (Burns et al., 2020), maximum twenty-four at any given point in time (Lee, 2020). With Sanders ending his presidential bid on April 15, 2020, the Democratic field has finally narrowed down to Biden. But Biden was not a strong primary candidate in the pre-primary period according to the polls, with many strong candidates such as Sanders, Buttigieg, Warren, Klobuchar threatening his candidacy. What is more, his numbers fell rapidly with the beginning of the Iowa Caucus as he experienced devastating losses in the early states, before drastically shooting up starting from Super Tuesday. This can be seen in Figure 3.12, which is Figure 3.7's equivalent. For a figure with all candidates, see Appendix C.2.

With such sharp turn of events and much uncertainty about candidate viability much more even than the 2016 Republican primary, the 2020 Democratic primary is a race best suited for testing whether donors are actually instrumental or momentum-driven.

Moreover, on September 24, 2019, the House speaker, Nancy Pelosi, formally announced that the committees of the House would begin an impeachment inquiry, with the actual impeachment starting mid-December and concluding early February. While there is hardly any literature on what effects impeachment of an existing president has on political participation, either turnout or campaign contributions, the effect will be seen for the data gathered—if any.

Figure 3.13 shows Figure 3.6's equivalent for the 2020 cycle, with data collected up to March 1, 2020. Figure 3.14 shows the same jump estimates with the underlying data points, as well as smooth estimates for each candidate for more information.

As can be seen, in either contribution sum or in counts, there are no structural breaks—save for one for Biden in counts, when there is a large jump in the very beginning of his campaign at April 24, 2019. This is hardly any surprise, as he officially announced his campaign by video on April 25, 2019. That is, the few

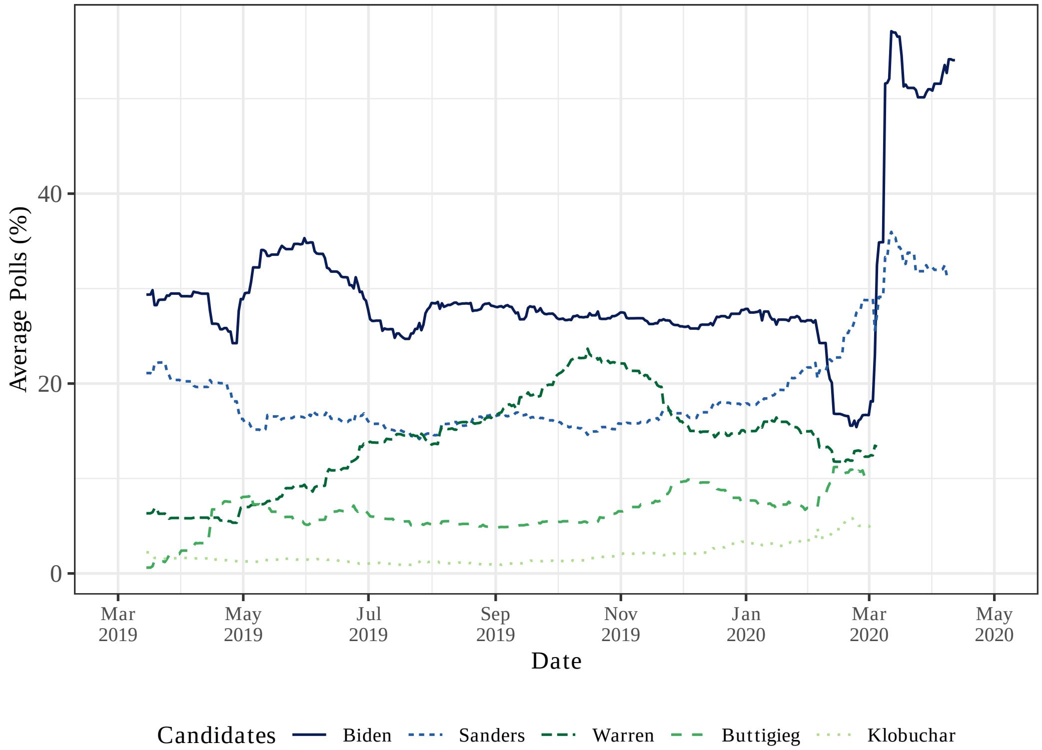
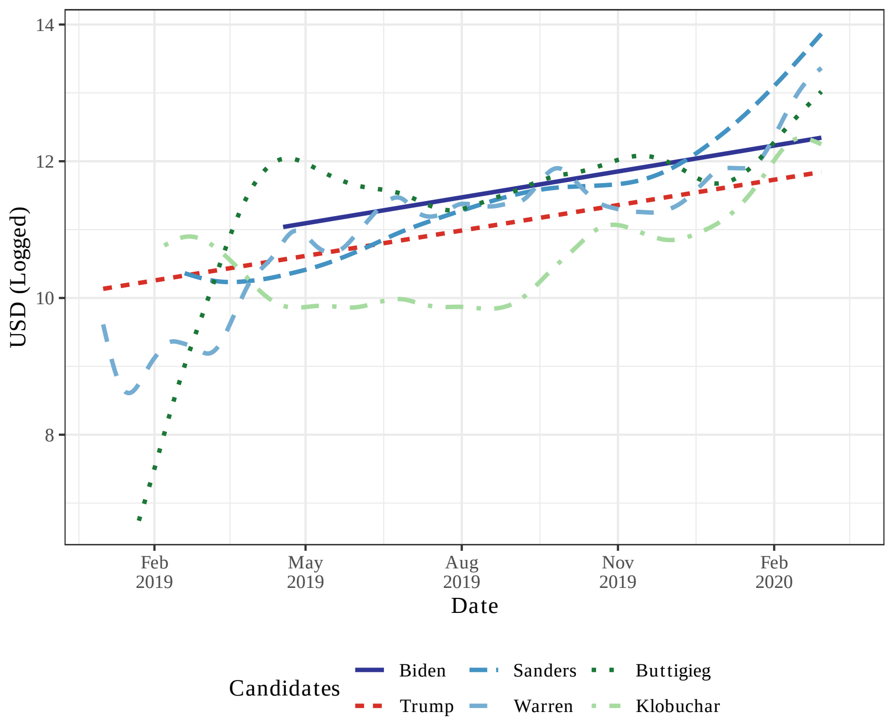


Figure 3.12: National Primary Polls by Party, FiveThirtyEight, 2020 (Selected Candidates)

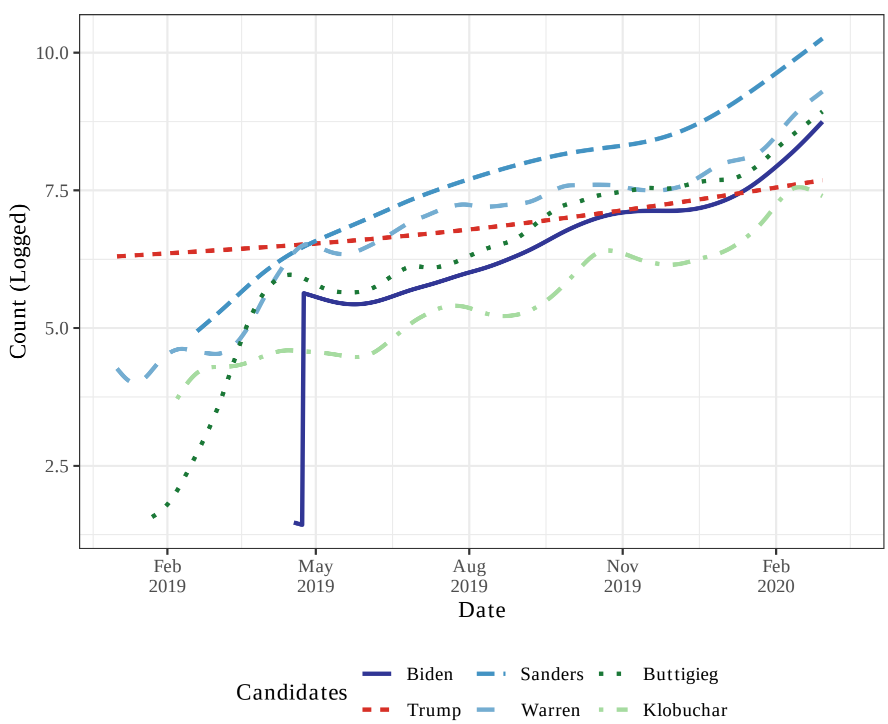
observations prior to April 25, 2019 are likely from very political sophisticated donors with inside contact or enthusiasm enough to start donating even before the official announcement. Hence the jump, although robust, is hardly of much interest.

Note in Figure 3.14 that Trump did have several days around the day that impeachment was announced and around the day that impeachment began with very high contribution sum. In fact, the day his campaign hauled in the most in both sum and counts was December 18, 2019, when impeachment formally began (more than 900,000 USD). But overall, his cash influx from individuals was quite consistent over the observed election cycle, resulting in almost a flat line when smoothed.

Now, since this dataset does not include observations after Super Tuesday and after the Democratic field has winnowed down to a single candidate, we may observe jumps on the national level later in the election cycle. In addition, the COVID-19 pandemic that is raging across the United States and the world will surely have significant impact on individual contributions, as citizens suffer economically (supply effect) and face-to-face campaign activities are virtually impossible (demand

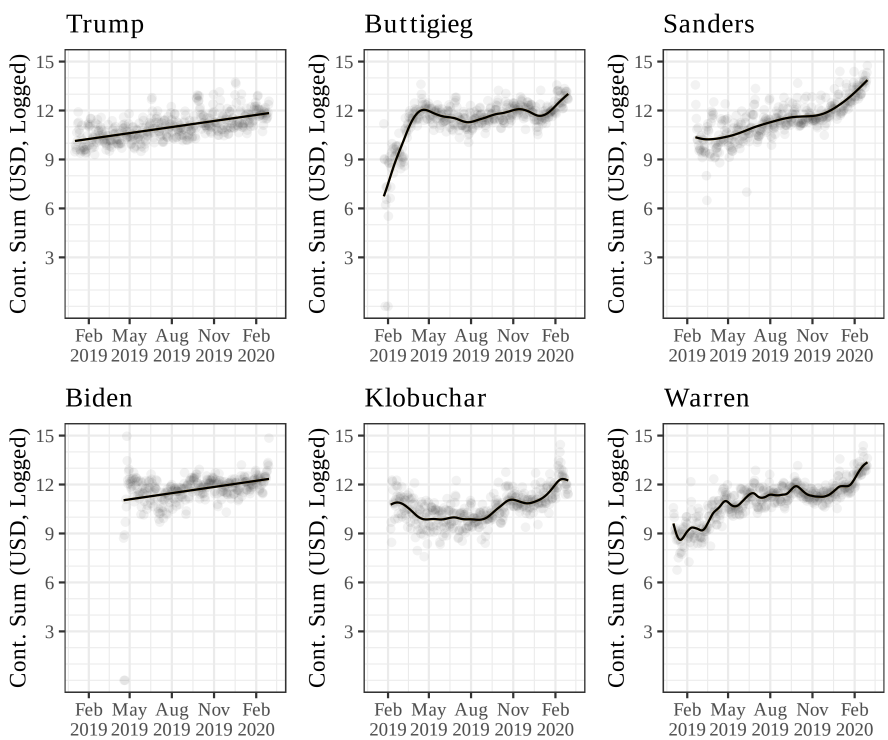


(a) Contribution Sum

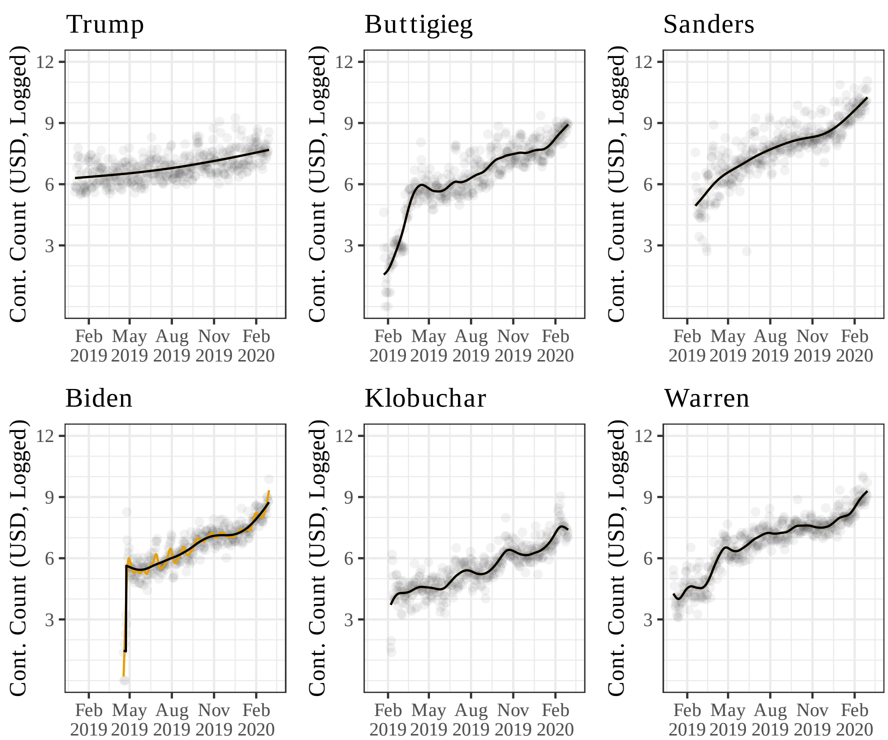


(b) Contribution Counts

Figure 3.13: National Estimate by Candidate, the 2020 Election Cycle (Up to March 1, 2020)



(a) Contribution Sum



(b) Contribution Counts

Figure 3.14: National Estimate by Candidate, the 2020 Election Cycle (Up to March 1, 2020), with Underlying Data Points and Smoothing Spline Estimates

effect).

While I have not performed state-by-state analysis for the 2020 cycle, nor modelling for known jumps, the preliminary analysis of the 2020 cycle on the national level seems to be consistent with what I have shown earlier—that campaign contributions is a slow-moving but a smooth process, and there are no significant structural breaks that would be observed if donors are largely instrumental or momentum-driven.

3.9 Conclusion

In this paper, I have attempted to answer "why do campaign contributors give" with "*when* do campaign contributors give", using presidential donors to the 2016 race. I hypothesized that if either donors are instrumentally motivated, or are expressively motivated but primarily momentum-driven, I will observe structural breaks at key campaign events, such as initial primary victories or surprise, upset wins. I also purported to pin down the slow-moving process underlying campaign finance data.

Instrumental donors will Bayesian update with new information about the true state of candidate viability unless most donors have already accurately inferred the true state of the world. With momentum-driven expressive donors I hypothesized as such because by definition momentum indicates how early candidate performances boosts or hurts the candidates, whatever the reason behind it is, such as the desire to back a winning horse. I tested this hypothesis using the sequential segmentation spline method taken from Ratkovic and Eng (2010), which detects the number and location of unknown and known jumps while providing smooth estimates to the underlying dynamics elsewhere.

I found that when contributions are aggregated nationally, there were no events detected whatsoever, either with a blank slate or when key events were modeled into the unpenalized space of the partial splines. Only when I have disaggregated the data by contributor's state, I find jumps—but not where I have expected them to be. If anything, there is a surge in contributions *before* the candidates exceed expectations, not *after*, for some second-tier candidates. In addition, there were a few sporadic responses to the first two primaries—Iowa and New Hampshire—from the general election candidate's money, but without a readily discernible pattern. All in all, known events do not seem to be creating immediate and lingering shocks for the algorithm to detect as a structural break. Although they ran contrary to our hypothesis, these provide interesting new pictures that contribute to the campaign finance literature.

I asked two additional questions. Do donors respond to local events—here more specifically, in-state caucuses and primaries? Are there any other critical events in campaign finance that are not considered important a priori? I find that generally, in-state primaries do not cause structural breaks. I also find some dates that emerge as important in many states for each candidate, but these are dates with no particularly interesting event, such as August 25th for the Trump campaign in which donations sharply dipped from twelve states. I interpret these to be demand effects from the campaigns themselves, and not the supply effect—that is, the campaign may fire up or cease operations at times that are not particularly linked to any external event, but makes sense internally, such as an FEC deadline.

Lastly, I performed the same check on the national-level daily aggregated data of the 2020 presidential election cycle, using data collected up to March 1, 2020. I again see virtually no events detected either on contribution sum or counts. It remains to be seen whether there will be any structural breaks later in the cycle post Super Tuesday, or in the mayhem of the COVID-19 pandemic.

BIBLIOGRAPHY

Aldrich, John H., Jacob M. Montgomery, and Wendy Wood (2011). Turnout as a habit. *Political Behavior 33*(4), 535–563.

Amos, Brian, Daniel A. Smith, and Casey Ste Claire (2017). Reprecincting and voting behavior. *Political Behavior 39*(1), 133–156.

Ansolabehere, Stephen, John M. De Figueiredo, and James M. Snyder Jr. (2003). Why is there so little money in US politics? *Journal of Economic Perspectives 17*(1), 105–130.

Ansolabehere, Stephen and David M. Konisky (2006). The introduction of voter registration and its effect on turnout. *Political Analysis 14*(1), 83–100.

Associated Press (2015). Bernie Sanders pulls in nearly as much cash this quarter as Hillary Clinton. *Chicago Tribune*.

Bai, Jushan (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis 15*(5), 453–472.

Bai, Jushan (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*(3), 315–352.

Bai, Jushan and Pierre Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Bai, Jushan and Pierre Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Barber, Michael J. (2016a). Donation motivations: Testing theories of access and ideology. *Political Research Quarterly 69*(1), 148–159.

Barber, Michael J. (2016b). Ideological donors, contribution limits, and the polarization of american legislatures. *The Journal of Politics 78*(1), 296–310.

Barber, Michael J., Brandice Canes-Wrone, and Sharece Thrower (2017). Ideologically sophisticated donors: Which candidates do individual contributors finance? *American Journal of Political Science 61*(2), 271–288.

Bartels, Larry M. (1988). *Presidential Primaries and the Dynamics of Public Choice*. Princeton University Press.

Beck, Nathaniel and Simon Jackman (1998). Beyond linearity by default: Generalized additive models. *American Journal of Political Science 42*(2), 596.

Blevins, Cameron and Lincoln Mullen (2015). Jane, John... Leslie? A historical method for algorithmic gender prediction. *DHQ: Digital Humanities Quarterly 9*(3).

Bonica, Adam (2011). Small donors and polarization. *Boston Review*.

Brady, Henry E. and John E. McNulty (2011). Turning out to vote: The costs of finding and getting to the polling place. *American Political Science Review 105*(1), 115–134.

Breheny, Patrick and Woodrow Burchett (2013). Visualization of regression models using visreg. *R package*, 1–15.

Brown, Clifford W., Lynda W. Powell, and Clyde Wilcox (1995). *Serious Money: Fundraising and Contributing in Presidential Nomination Campaigns*. Cambridge University Press.

Bump, Philip (2016). Bernie Sanders keeps saying his average donation is $27, but his own numbers contradict that. *Washington Post*.

Burden, Barry C., David T. Canon, Kenneth R. Mayer, and Donald P. Moynihan (2014). Election laws, mobilization, and turnout: The unanticipated consequences of election reform. *American Journal of Political Science 58*(1), 95–109.

Burns, Alexander, Matt Flegenheimer, Jasmine C. Lee, Lisa Lerer, and Jonathan Martin (2020). Who's running for president in 2020? *The New York Times*.

Butler, R. Lawrence (2009). Momentum in the 2008 presidential contests. *Polity 41*(3), 331–344.

California Secretary of State (2019). California national voter registration act manual. https://elections.cdn.sos.ca.gov/nvra/nvra-manual/complete.pdf. Last accessed June 29, 2019.

Carlstein, Edward et al. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics 14*(3), 1171–1179.

Christenson, Dino P. and Corwin D. Smidt (2011). Riding the waves of money: Contribution dynamics in the 2008 presidential nomination campaign. *Journal of Political Marketing 10*(1-2), 4–26.

Collingwood, Loren, Matt A. Barreto, and Todd Donovan (2012). Early primaries, viability and changing preferences for presidential candidates. *Presidential Studies Quarterly 42*(2), 231–255.

Coppock, Alexander and Donald P. Green (2016). Is voting habit forming? New evidence from experiments and regression discontinuities. *American Journal of Political Science 60*(4), 1044–1062.

Corasaniti, Nick (2016). Bernie Sanders campaign showed how to turn viral moments into money. *The New York Times*.

Culberson, Tyler, Michael P. McDonald, and Suzanne M. Robbins (2018). Small donors in congressional elections. *American Politics Research*, 1532673X18763918.

Dawood, Yasmin (2015). Campaign finance and american democracy. *Annual Review of Political Science 18*, 329–348.

Donovan, Todd and Rob Hunsaker (2009). Beyond expectations: Effects of early elections in us presidential nomination contests. *PS: Political Science & Politics 42*(1), 45–52.

Dowding, Keith, Peter John, and Daniel Rubenson (2012). Geographic mobility, social connections and voter turnout. *Journal of Elections, Public Opinion and Parties 22*(2), 109–122.

Dyck, Joshua J. and James G. Gimpel (2005). Distance, turnout, and the convenience of voting. *Social Science Quarterly 86*(3), 531–548.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica (2015). Suspense and surprise. *Journal of Political Economy 123*(1), 215–260.

Ensley, Michael J. (2009). Individual campaign contributions and candidate ideology. *Public Choice 138*(1-2), 221–238.

Fiorina, Morris P. (1976). The voting decision: instrumental and expressive aspects. *The Journal of Politics 38*(2), 390–413.

Foran, Clare (2016). Bernie Sanders's big money. *Washington Post*.

Francia, Peter L., John C. Green, Paul S. Herrnson, Clyde Wilcox, Lynda W. Powell, et al. (2003). *The Financiers of Congressional Elections: Investors, Ideologues, and Intimates*. Columbia University Press.

Gaudiano, Nicole (2016). Revolution messaging helps drive Sanders' 'political revolution'. *USA Today*.

Gay, Claudine (2012). Moving to opportunity: The political effects of a housing mobility experiment. *Urban Affairs Review 48*(2), 147–179.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review 102*(1), 33–48.

Gerber, Alan S., Donald P. Green, and Ron Shachar (2003). Voting may be habit-forming: Evidence from a randomize field experiment. *American Journal of Political Science 47*(3), 11.

Gerber, Alan S., Gregory A. Huber, and Seth J. Hill (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.

Geys, Benny (2006). Explaining voter turnout: A review of aggregate-level research. *Electoral Studies 25*(4), 637–663.

Gillespie, Brian Joseph (2016). *Household Mobility in America: Patterns, Processes, and Outcomes*. Springer.

Gimpel, James G. and Jason E. Schuknecht (2003). Political participation and the accessibility of the ballot box. *Political Geography 22*(5), 471–488.

Goff, Michael J. (2005). *The Money Primary: The New Politics of the Early Presidential Nomination Process*. Rowman & Littlefield.

Graf, Joseph, Michael J. Malbin, and Costas Panagopoulos (2006). *Small Donors and Online Giving: A study of Donors to the 2004 Presidential Campaigns*. Institute for Politics, Democracy & the Internet, Graduate School of Political Management, George Washington University.

Green, Donald P. and Ron Shachar (2000). Habit formation and political behaviour: Evidence of consuetude in voter turnout. *British Journal of Political Science 30*, 561–573.

Gronke, Paul, Eva Galanes-Rosenbaum, and Peter A. Miller (2007). Early voting and turnout. *PS: Political Science & Politics 40*(4), 639–645.

Gu, Chong (2013). *Smoothing Spline ANOVA Models*, Volume 297. Springer Science & Business Media.

Gu, Chong and Grace Wahba (1993). Smoothing spline anova with component-wise bayesian "confidence intervals". *Journal of Computational and Graphical Statistics 2*(1), 97–117.

Hansen, Jonas Hedegaard (2016). Residential mobility and turnout: The relevance of social costs, timing and education. *Political Behavior 38*(4), 769–791.

Hastie, Trevor J. (1992). Generalized additive models. In *Statistical Models in S*, pp. 249–307. Routledge.

Hayes, Danny and Seth C. McKee (2009). The participatory effects of redistricting. *American Journal of Political Science 53*(4), 1006–1023.

Heerwig, Jennifer A. (2016). Donations and dependence: Individual contributor strategies in house elections. *Social Science Research 60*, 181–198.

Highton, Benjamin (2000). Residential mobility, community mobility, and electoral participation. *Political Behavior 22*(2), 109–120.

Highton, Benjamin and Raymond E. Wolfinger (1998). Estimating the effects of the national voter registration act of 1993. *Political Behavior 20*(2), 79–104.

Highton, Benjamin and Raymond E. Wolfinger (2001). The first seven years of the political life cycle. *American Journal of Political Science*, 202–209.

Hill, Seth J. and Gregory A. Huber (2017). Representativeness and motivations of the contemporary donorate: Results from merged survey and administrative records. *Political Behavior 39*(1), 3–29.

Hobbs, William R., Nicholas A. Christakis, and James H. Fowler (2014). Widowhood effects in voter participation. *American Journal of Political Science 58*(1), 1–16.

Hogan, Howard (2008). Measuring population change using the american community survey. In *Applied Demography in the 21st Century*, pp. 13–30. Springer.

Ihrke, David K. (2016). *Why Did You Move?: An Overview and Analysis of the Annual Social and Economic Supplement's Reason for Move Write-In Expansion*. United States Census Bureau.

Ihrke, David K. and Carol S. Faber (2012). *Geographical Mobility: 2005 to 2010*. Current Population Reports, P20-567. U.S. Census Bureau, Washington, DC.

Imai, Kosuke and Kabir Khanna (2016). Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Analysis 24*(2), 263–272.

Johnson, Bertram (2010). Individual contributions: A fundraising advantage for the ideologically extreme? *American Politics Research 38*(5), 890–908.

Johnson, Bertram (2013). *Political Giving: Making Sense of Individual Campaign Contributions*. FirstForumPress.

Karp, Jeffrey A. and Susan A. Banducci (2000). Going postal: How all-mail elections influence turnout. *Political Behavior 22*(3), 223–239.

Karpf, David (2013). The Internet and American political campaigns. In *The Forum*, Volume 11, pp. 413–428. De Gruyter.

Keele, Luke (2008). *Semiparametric Regression for the Social Sciences*. Wiley Online Library.

Key, Valdimer Orlando (1964). *Politics, Parties, and Pressure Groups* (5 ed.). Crowell.

Khanna, Kabir, Kosuke Imai, and Hubert Jin (2017). *wru: Who are You? Bayesian Prediction of Racial Category Using Surname and Geolocation*. R package version 0.1-7.

Kim, Seo-young Silvia, Spencer Schneider, and R. Michael Alvarez (2019). Evaluating the quality of changes in voter registration databases. *American Politics Research*.

Kim, Young-Ju and Chong Gu (2004). Smoothing spline gaussian regression: more scalable computation via efficient approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(2), 337–356.

Knight, Brian and Nathan Schiff (2010). Momentum and social learning in presidential primaries. *Journal of Political Economy 118*(6), 1110–1150.

Kunsch, Hans R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 1217–1241.

Lee, Jasmine C. (2020). How the democratic presidential field has narrowed. *The New York Times*.

Lee, Michelle Ye Hee (2016). Fact checker: Sanders's claim that he 'does not have a super PAC'. *Washington Post*.

Leighley, Jan E. and Jonathan Nagler (2013). *Who Votes now?: Demographics, Issues, Inequality, and Turnout in the United States*. Princeton University Press.

Leighley, Jan E. and Arnold Vedlitz (1999). Race, ethnicity, and political participation: Competing models and contrasting explanations. *The Journal of Politics 61*(4), 1092–1114.

Liu, Regina Y., Kesar Singh, et al. (1992). Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the Limits of Bootstrap 225*, 248.

Magleby, David B. (2008). Rolling in the dough: The continued surge in individual contributions to presidential candidates and party committees. *The Forum 6*(1).

Magleby, David B., Jay Goodliffe, and Joseph A. Olsen (2018). *Who Donates in Campaigns?: The Importance of Message, Messenger, Medium, and Structure*. Cambridge University Press.

Malbin, Michael J. (2009). Small donors, large donors and the Internet: The case for public financing after Obama. *Unpublished manuscript*.

Malbin, Michael J. (2013). Small donors: Incentives, economies of scale, and effects. *The Forum 11*, 385–411.

Malhotra, Neil, Melissa R. Michelson, Ali Adam Valenzuela, et al. (2012). Emails from official sources can increase turnout. *Quarterly Journal of Political Science 7*(3), 321–332.

Mann, Christopher B. and Lisa A. Bryant (2019). If you ask, they will come (to register and vote): Field experiments with state election agencies on encouraging voter registration. *Electoral Studies*.

Mayer, William G. (1996). Comment: Of money and momentum. *Political Research Quarterly 49*(4), 719–726.

McDonald, Michael P. (2008). Portable voter registration. *Political Behavior 30*(4), 491–501.

McNulty, John E., Conor M. Dowling, and Margaret H. Ariotti (2009). Driving saints to sin: How increasing the difficulty of voting dissuades even the most motivated voters. *Political Analysis 17*(4), 435–455.

Mehta, Seema, Anthony Pesce, Maloy Moore, and Christine Zhang (2016). Who gives money to Bernie Sanders? *Los Angeles Times*.

Mitchell, Joshua L., Karen Sebold, Andrew J. Dowdle, Scott Limbocker, and Patrick A. Stewart (2015). *The Political Geography of Campaign Finance: Fundraising and Contribution Patterns in Presidential Elections, 2004–2012*. Springer.

Mullen, Lincoln (2018). *gender: Predict Gender from Names Using Historical Data*. R package version 0.5.2.

Mutz, Diana C. (1995). Effects of horse-race coverage on campaign coffers: Strategic contributing in presidential primaries. *The Journal of Politics 57*(4), 1015–1042.

Mutz, Diana C. (1997). Mechanisms of momentum: Does thinking make it so? *The Journal of Politics 59*(1), 104–125.

National Association of Secretaries of State (2017). NASS report: Maintenance of state voter registration list. https://www.nass.org/sites/default/files/reports/nass-report-voter-reg-maintenance-final-dec17.pdf. Last accessed June 29, 2019.

Nickerson, David W. (2014). Do voter registration drives increase participation? for whom and when? *The Journal of Politics 77*(1), 88–101.

Norrander, Barbara (2006). The attrition game: Initial resources, initial contests and the exit of candidates during the us presidential primary season. *British Journal of Political Science 36*(3), 487–507.

Norrander, Barbara (2015). *Super Tuesday: Regional Politics and Presidential Primaries*. University Press of Kentucky.

Ottaviani, Marco and Peter Norman Sørensen (2006). The strategy of professional forecasting. *Journal of Financial Economics 81*(2), 441–466.

Ovtchinnikov, Alexei V and Eva Pantaleoni (2012). Individual political contributions and firm performance. *Journal of Financial Economics 105*(2), 367–392.

Panagopoulos, Costas and Daniel Bergan (2006). Contributions and contributors in the 2004 presidential election cycle. *Presidential Studies Quarterly 36*(2), 155–171.

Panagopoulos, Costas and Daniel Bergan (2007). Online fund-raising and contributors in the 2004 presidential campaign. *Social Science Computer Review 25*(4), 484–493.

Patten, Eileen (2016). The nation's Latino population is defined by its youth. *Pew Research Center*.

Powell, G Bingham (1986). American voter turnout in comparative perspective. *American Political Science Review 80*(1), 17–43.

Qiu, Linda (2015). Is Bernie Sanders the only presidential candidate without a super PAC? *Politifact*.

Ratkovic, Marc T. and Kevin H. Eng (2010). Finding jumps in otherwise smooth curves: Identifying critical events in political processes. *Political Analysis 18*(1), 57–77.

Redlawsk, David P., Caroline J. Tolbert, and Todd Donovan (2011). *Why Iowa?: How Caucuses and Sequential Elections Improve the Presidential Nominating Process*. University of Chicago Press.

Rhodes, Jesse H., Brian F. Schaffner, and Raymond J. La Raja (2018). Detecting and understanding donor strategies in midterm elections. *Political Research Quarterly*, 1065912917749323.

Riker, William H. and Peter C. Ordeshook (1968). A theory of the calculus of voting. *American Political Science Review 62*(1), 25–42.

Rosenstone, Steven J. and John Hansen (1993). *Mobilization, Participation, and Democracy in America*. Macmillan Publishing Company.

Rosenstone, Steven J. and Raymond E. Wolfinger (1978). The effect of registration laws on voter turnout. *American Political Science Review 72*(1), 22–45.

Rossi, Peter Henry (1980). *Why Families Move*. Sage Publications, Inc.

Schelker, Mark and Marco Schneiter (2017). The elasticity of voter turnout: Investing 85 cents per voter to increase voter turnout by 4 percent. *Electoral Studies 49*, 65–74.

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

Sen, Ashish and Muni S. Srivastava (1975). On tests for detecting change in mean. *The Annals of Statistics*, 98–108.

Sinclair, Betsy (2012). *The social citizen: Peer networks and political behavior*. University of Chicago Press.

Smidt, Corwin and Dino P. Christenson (2012). More bang for the buck: Campaign spending and fundraising success. *American Politics Research 40*(6), 949–975.

Southwell, Priscilla L. and Justin I. Burchett (2000). The effect of all-mail elections on voter turnout. *American Politics Quarterly 28*(1), 72–79.

Squire, Peverill, Raymond E. Wolfinger, and David P. Glass (1987). Residential mobility and voter turnout. *American Political Science Review 81*(1), 45–65.

Steger, Wayne P., Andrew J. Dowdle, and Randall E. Adkins (2004). The New Hampshire effect in presidential nominations. *Political Research Quarterly 57*(3), 375–390.

Stratmann, Thomas (1992). Are contributors rational? untangling strategies of political action committees. *Journal of Political Economy 100*(3), 647–664.

Stratmann, Thomas (1998). The market for congressional votes: Is timing of contributions everything? *The Journal of Law and Economics 41*(1), 85–114.

Stratmann, Thomas (2005). Some talk: Money in politics. a (partial) review of the literature. In *Policy Challenges and Political Responses*, pp. 135–156. Springer.

Street, Alex, Thomas A. Murray, John Blitzer, and Rajan S. Patel (2015). Estimating voter registration deadline effects with web search data. *Political Analysis 23*(2), 225–241.

The United States Department of Justice, Civil Rights Division (2017). The National Voter Registration Act of 1993 (NVRA): Questions and answers. https://www.justice.gov/crt/national-voter-registration-act-1993-nvra. Last accessed June 29, 2019.

United States Census Bureau (2018a). CPS historical geographical mobility/migration graphs. https://www.census.gov/library/visualizations/time-series/demo/historic.html. Last accessed June 29, 2019.

United States Census Bureau (2018b). CPS historical migration/geographic mobility tables. https://www.census.gov/data/tables/time-series/demo/geographic-mobility/historic.html. Last accessed June 29, 2019.

United States Census Bureau (2018c). Geographical mobility: 2017 to 2018. https://www.census.gov/data/tables/2018/demo/geographic-mobility/cps-2018.html. Last accessed June 29, 2019.

Verba, Sidney and Norman H. Nie (1987). *Participation in America: Political Democracy and Social Equality*. University of Chicago Press.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady (1995). *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press.

Wahba, Grace (1990). *Spline Models for Observational Data*, Volume 59. Siam.

Wattenberg, Martin P., Ian McAllister, and Anthony Salvanto (2000). How voting is like taking an SAT test: An analysis of american voter rolloff. *American Politics Quarterly 28*(2), 234–250.

Wilcox, Clyde (2008). Internet fundraising in 2008: A new model? In *The Forum*, Volume 6. De Gruyter.

Willis, Derek (2014). How ActBlue became a powerful force in fund-raising. *The New York Times*.

Wolfinger, Raymond E. and Benjamin Highton (1995). Can more efficient purging boost turnout? *Unpublished manuscript*.

Wolfinger, Raymond E., Benjamin Highton, and Megan Mullin (2005). How postregistration laws affect the turnout of citizens registered to vote. *State Politics & Policy Quarterly 5*(1), 1–23.

Wolfinger, Raymond E. and Steven J. Rosenstone (1980). *Who Votes?* Yale University Press.

Wood, Simon N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(1), 95–114.

Wood, Simon N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B) 73*(1), 3–36.

Wood, Simon N. (2017). *Generalized Additive Models: an Introduction with R* (2 ed.). Chapman and Hall/CRC.

Yang, Yong and Ana V. Diez-Roux (2012). Walking distance by trip purpose and population subgroups. *American Journal of Preventive Medicine 43*(1), 11–19.

Zeileis, Achim, Christian Kleiber, Walter Krämer, and Kurt Hornik (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis 44*(1-2), 109–123.

Appendix A

APPENDIX FOR CHAPTER 1

A.1 Legislation Related to Voter Registration

The National Voter Registration Act of 1993

The National Voter Registration Act of 1993 (NVRA), also known as the "motor voter law," is about increasing the opportunities of voter registration through various means. Its various Sections decree that the States[1] offer citizens the chance to register to vote through motor vehicle agencies, by mail-in applications, and by public assistance and disability offices.

Section 8, *Administration of Voter Registration*, requires States to maintain accurate and up-to-date data. Specifically, the it mandates that the States conduct "a general voter registration list maintenance program that makes a reasonable effort to remove ineligible persons from the voter rolls by reason of the person's death, or a change in the residence of the registrant outside of the jurisdiction, in accordance with procedures set forth in the NVRA" (The United States Department of Justice, Civil Rights Division, 2017).

The National Change of Address Program (NCOA)

For the second requirement, the NVRA offers one example. The States can use the permanent change-of-address records submitted to the USPS by voters. A United States resident can fill out a change of address form on the Official USPS Change of Address website or physically at a local post office, to have her mail forwarded to the new residence or a PO Box. While from a voter's point of view this serves primarily not to lose any mail while moving for a price of 1.05 USD,[2] the accumulated data contains movers' old and new addresses, including the date they requested the service to start.

This data can be used by the States to check their voter data and to discern voters who have moved. By distinguishing those who have moved away from the jurisdiction, the election officials can remove the names of some voters. This is an important

[1]The States here indicate 44 States and the District of Columbia, with Idaho, Minnesota, New Hampshire, North Dakota, Wisconsin, and Wyoming as exceptions.

[2]This is for ID verification.

step in list maintenance, because it can reduce the cost of direct mail operations by creating a cleaner list with fewer undeliverables and mistakes in delivery.

The only legal requirement is that this removal is performed 90 days prior to the date of the federal election. To see the actual screenshots of change-of-address applications and how it prompts voter registration data update, see Appendix A.1. Note that the usage of the NCOA data is not mandatory.[3] The NCOA processing is just one example of a potential list maintenance activity that can be performed by the States. A majority of the states do implement the NCOA processing (National Association of Secretaries of State, 2017), California being one prominent state that mandates NCOA processing by its own election laws.

California Elections Code

In California, counties can opt in to integrate NCOA processing into their list maintenance, as an alternative to a residency confirmation postcard (CA Elec Code § 2222 (2017); 52 U.S.C. § 20507(c)(1)(A)). This is classified as third party address changes, as opposed to first party address changes. Because there is a statewide voter registration system in California, it is the Secretary of State that is in charge of obtaining and disseminating the NCOA data (California Secretary of State, 2019).

Pursuant to California Code of Regulations § 20108.50 National Change of Address Processing, the Secretary of State must NCOA process the statewide voter list, and send any records of registrants that seem to have changed their address to the relevant county officials.

The Justice Department's Summary

The following is the 36th question posted in the Justice Department's questions and answers over the NVRA (The United States Department of Justice, Civil Rights Division, 2017). It details the role that NCOA processing plays in voter list mainte-

[3] Highton and Wolfinger (1998) wrote as follows (page 92):

> The NVRA provides one alternative to this daunting list-cleaning procedure: States may identify movers with the Postal Service's computer file of address-change information, known as the National Change of Address (NCOA) program. About 40 million permanent change-of-address notices are filed each year with the Postal Service. The NCOA file is updated daily and each change is kept for three years. This information can be bought from two dozen licensed vendors who distribute customized NCOA data sets. The NVRA requires that people purged by NCOA who move inside the same county (about 60% of all movers) be automatically re-registered at their new address.

nance.

> 36. Do States have to use the NCOA process to initiate the notice process?
>
> No. States do not have to use the NCOA process. Under the NVRA, States must have a general program that makes a reasonable effort to identify and remove the names of voters who have become ineligible to vote by means of a change of address. The program has to be uniform, non-discriminatory, in compliance with the Voting Rights Act and must be completed 90 days before a federal election. States otherwise have discretion under the NVRA and HAVA in how they design their general program, and States
>
> For example, some general programs involve a State undertaking a uniform mailing of a voter registration card, sample ballot, or other election mailing to all voters in a jurisdiction, and then using information obtained from returned non-deliverable mail as the basis for correcting voter registration records (for apparent moves within a jurisdiction) or for initiating the notice process (for apparent moves outside a jurisdiction or non-deliverable mail with no forwarding address noted).
>
> Another example involves general programs where States initiate the notice process based on information showing that a voter has not voted in elections nor made contact with a registrar over some period of time. This is not prohibited by the NVRA and its bar on removing voters from the list solely for failure to vote, since it relies on the NVRA notice process, and thus utilizes both a notice and a waiting period of two federal general elections.

The following is the 38th question from the same source, detailing how the within-county movers can vote.

> 38. Are there any protections in the NVRA for those eligible registered voters who have changed address to another location within a registrar's jurisdiction, or are otherwise on an inactive voter list, but have not notified the registrar prior to the date of a federal election?
>
> Yes. The NVRA contains fail-safe provisions to enable such persons who show up to vote on a federal election day to update their registration

and to vote in that election even though they have not notified the registrar of the address change:

1. An eligible registered voter who has moved to an address in an area covered by the same polling place as his or her previous address is permitted to vote at that same polling place upon oral or written affirmation by the registrant of the change of address at the polling place;
2. An eligible registered voter who has moved to an address in an area covered by a different polling place from the polling place for his or her previous address, but within the same registrar's jurisdiction and the same congressional district, at the option of the registrant:
   a) shall be permitted to correct the voting records and vote at the old polling place upon oral or written affirmation by the registrant of the new address before an election official at that polling place; or
   b) shall be permitted to correct the voting records and vote at a designated central location within the same registrar's jurisdiction, upon written affirmation by the registrant of the new address on a standard form provided by the registrar; or
   c) shall be permitted to correct the voting records for purposes of future elections at the new polling place, and shall be permitted vote in the current election at that polling place if allowed under State law, upon confirmation by the registrant of the new address by such means as are required by law.

A central voting location need not be made available by the registrar if State law allows the person to vote at either the old or new polling place in the current election upon oral or written affirmation of the address change.

The failsafe provisions of Section 8 draw a distinction between the registrant's need for "affirmation" or "confirmation" of a new address, depending upon the circumstances in which the failsafe voting occurs.

California Elections Code Section 2225

The following is the full text of CA Elec Code § 2225 (2017).

> (a) Based on change-of-address data received from the United States Postal Service or its licensees, the county elections official shall send a forwardable notice, including a postage-paid and preaddressed return form, to enable the voter to verify or correct address information.
>
> Notification received through NCOA or Operation Mail that a voter has moved and has given no forwarding address shall not require the mailing of a forwardable notice to that voter.
>
> (b) If postal service change-of-address data indicates that the voter has moved to a new residence address in California, the forwardable notice shall be in substantially the following form:
>
> "We have received notification that you have moved to a new residence address in California. You will be registered to vote at your new address unless you notify our office within 15 days that the address to which this card was mailed is not a change of your permanent residence. You must notify our office by either returning the attached postage-paid postcard, or by calling toll free. If this is not a permanent residence, and if you do not notify us within 15 days, you may be required to provide proof of your residence address in order to vote at future elections."
>
> (c) If postal service change-of-address data received from a nonforwardable mailing indicates that a voter has moved and left no forwarding address, a forwardable notice shall be sent in substantially the following form:
>
> "We are attempting to verify postal notification that the voter to whom this card is addressed has moved and left no forwarding address. If the person receiving this card is the addressed voter, please confirm your continued residence or provide current residence information on the attached postage-paid postcard within 15 days. If you do not return this card and continue to reside in California, you may be required to provide proof of your residence address in order to vote at future elections and, if you do not offer to vote at any election in the period between the date of this notice and the second federal general election following this

notice, your voter registration will be cancelled and you will have to reregister in order to vote."

(d) The use of a toll-free number to confirm the old residence address is optional. Any change to the voter address must be received in writing.

(Amended by Stats. 2015, Ch. 728, Sec. 68. (AB 1020) Effective January 1, 2016. Operative September 26, 2016, when the Secretary of State issued the certification prescribed by Stats. 2015, Ch. 728, Sec. 88.)

California Code of Regulations, Title 2: Administration

The following is § 20108.50. National Change of Address Processing in Division 7. Secretary of State, Chapter 2. Statewide Voter Registration Database.

> Except during the 90 days prior to a Federal election, the Secretary of State shall conduct monthly voter registration list maintenance using a change of address service or services based on the United States Postal Service National Change of Address (NCOA) database to identify address changes for registered voters. For records showing a change of address, the Secretary of State shall automatically transmit a change of address notice to the elections official in the county from or within which a voter has moved. Within five (5) business days of receipt of a change of address notice from the Secretary of State the elections official shall process the change of address notice pursuant to California Elections Code Section 2225 and Section 2226 and submit any changes in the registration record to Calvoter in accordance with Section 20108.15 and Section 20108.40.

United States Postal Services NCOALink

Privacy Act Statement. The following is the privacy act statement that accompanies the web-based USPS change of address as of June 30, 2018. The emphasis is added by the author.

> Your information will be used to provide you with mail forwarding and change of address services. Collection is authorized by 39 U.S.C. 401, 403, and 404. Providing the information is voluntary, but if not provided we will not be able to process your request. **We do not**

disclose your information to third parties without your consent, except to facilitate the transaction, to act on your behalf or request, or as legally required. This includes the following limited circumstances: to a congressional office on your behalf; to financial entities regarding financial transaction issues; to a U.S. Postal Service (USPS) auditor; to entities, including law enforcement, as required by law or in legal proceedings; to contractors and other entities aiding us to fulfill the service (service providers); to federal, state, local or foreign government agencies regarding personnel matters or for the performance of its duties; for the service of legal process; **for voter registration purposes**; for jury service duties; to a disaster relief organization if the address has been impacted by a disaster or manmade hazard; to individuals or companies already in possession of your name and old mailing address, as an address correction service. Information will also be provided to licensed service providers of the USPS to perform mailing list correction service of lists containing your name and old address. A list of these licensed service providers can be obtained at the following URL: `https://postalpro.usps.com/mailing-and-shipping-services/NCOALink`. For more information regarding our privacy policies visit `www.usps.com/privacypolicy`.

A.2 Data Wrangling

Re-processing the Database with NCOA

While the classification of movers can be performed by just monitoring the changes to the voter data, I have re-processed the database with NCOA with the help of Orange County election officials.[4] This is to detect the moving dates of the first class of voters, who disclosed their new address prior to having detected via NCOA. This final step augments the USPS data to the voter file and determine movers' their residential stability—that is, the months spent at the new residence.

Data Filtering

A couple caveats should be noted. I have excluded voters whose age was observed to be more than a 100. This decision accounts for the fact that for some voters, the date of birth is either entered wrongly (e.g. January 1, 1900), or the dead voters

[4]The NCOA processing is formally named the NCOALink Product. According to the Postal Service, the NCOALink Product is only provided to a selection of companies licensed by the Service. The OCROV processed the data through a vendor of their choice at my request.

have not been fully accounted for.

A.3 Descriptive Statistics

Here I present some descriptive statistics of data, as well as the values used in the main text's conditional plots.

Movers vs. Stayers

Table A.1 shows some comparisons between movers between the 2016-2018 elections and stayers in the final sample.

Variables	Movers	Stayers
% of General 2018 Turnout	63.6	70.1
% of Female, If Classified	51.8	50.6
Median Age	42	50
% of Republicans	33.2	31.4
% of Democrats	33.4	33.4
% of Voters Born Abroad	24.1	28.0
% of General 2016 Turnout, if Eligible	80.3	81.9

Table A.1: Movers vs. Stayers, Sample Data

Who Requests Change of Address?

Table A.2 shows comparisons between those who request change of address and those who do not, given the initial classification of movers. Note that this comparison is not a universal comparison.

Variables	Change-of-Address Requesters	Non-Requesters
% of General 2018 Turnout	63.6	70.1
% of Female, If Classified	51.8	50.6
Median Age	42	50
% of Republicans	33.2	31.4
% of Democrats	33.4	33.4
% of Voters Born Abroad	24.1	28.0
% of General 2016 Turnout, if Eligible	80.3	81.9

Table A.2: Movers with Change-of-Address Requests and Those Without, Sample Data

A.4 Regression Results in Main Text, Full Table

This Section shows the summary of the generalized additive models and the linear probability models in their full form, including all the coefficients from control variables.

		Imperfect Placebo	Placebo Tests				
	General 2018	General 2016	Primary 2016	General 2014	Primary 2014	General 2012	Primary 2012
A. Smooth terms (effective degrees of freedom / residual degrees of freedom)							
Res. Stability × Same Address	3.353	1.148*	2.698	2.655	1.030	1.003	1.002
	4.153	1.145	3.358	3.305	1.059	1.006	1.004
Res. Stability × Same Precinct	2.684	1.145	3.391	1.931	1.004	1.003	1.933
	3.341	1.276	4.206	2.418	1.007	1.007	2.421
Res. Stability × Same Subdist.	7.205***	1.034	2.305	1.015	2.007	3.971	3.473**
	8.232	1.067	2.878	1.030	1.792	4.899	4.305
Res. Stability × Same Cong.	8.446***	3.810***	3.910*	1.013	1.792*	2.296	1.002
	8.913	4.709	4.830	1.027	2.241	2.868	1.004
Res. Stability × Diff. Cong.	8.428***	3.177***	1.235	2.552	1.015	1.003*	1.002
	8.908	3.945	1.433	3.183	1.029	1.007	1.003
Distance Moved	1.139*	1.009*	2.828	2.270	1.006	1.005	1.002
	1.139	1.017	3.580	2.887	1.011	1.011	1.006
Age	7.991***	7.378***	7.314***	7.035***	7.603***	7.655***	8.115***
	8.655	8.227	8.200	7.998	8.473	8.365	8.752
Distance to Poll	4.947*	1.037	2.287	1.021	2.102*	1.011	1.814
	6.035	1.074	2.910	1.041	2.676	1.021	2.303
Old Residence's	8.623***	8.140***	5.252**	8.377***	8.060***	3.017***	7.732***
Neighborhood Income	8.955	8.786	6.318	8.882	8.754	3.808	8.573
New Residence's	8.357***	7.866***	1.009*	6.994**	3.103**	2.371*	1.003
Neighborhood Income	8.881	8.655	1.019	8.037	3.891	3.029	1.006
B. Parametric coefficients (estimate / standard error)							
Same Precinct	0.053	0.043	0.003	0.303***	0.163	0.134	0.094
	(0.050)	(0.058)	(0.053)	(0.064)	(0.084)	(0.069)	(0.080)
Same Subdist.	0.113*	0.156**	0.063	0.395***	0.170*	0.257***	0.064
	(0.048)	(0.054)	(0.050)	(0.060)	(0.079)	(0.064)	(0.075)
Same Cong.	−0.004	0.133**	0.019	0.312***	0.152*	0.201***	0.086
	(0.042)	(0.047)	(0.046)	(0.056)	(0.071)	(0.056)	(0.068)
Diff. Cong.	−0.026	0.104*	−0.018	0.325***	0.158*	0.183**	0.031
	(0.045)	(0.050)	(0.050)	(0.059)	(0.074)	(0.060)	(0.071)
2016 Turnout	1.467***						
	(0.017)						
Times Moved	−0.229***	−0.442***	−0.254***	−0.302***	−0.282***	−0.117***	−0.238***
	(0.025)	(0.027)	(0.026)	(0.031)	(0.045)	(0.032)	(0.042)
PAV	0.291***	0.358***	0.376***	0.390***	0.831***	0.064**	0.681***
	(0.015)	(0.017)	(0.016)	(0.018)	(0.026)	(0.020)	(0.024)
Female	0.172***	0.230***	0.105*	0.136**	0.042	0.136**	0.166*
	(0.038)	(0.042)	(0.041)	(0.049)	(0.066)	(0.052)	(0.067)
Male	0.215***	0.114**	0.113**	0.305***	0.235***	0.096*	0.273***
	(0.038)	(0.042)	(0.041)	(0.049)	(0.066)	(0.052)	(0.067)
Black	0.187*	0.253**	0.142*	−0.096	−0.271*	0.316**	−0.354**
	(0.073)	(0.083)	(0.079)	(0.093)	(0.133)	(0.102)	(0.132)
Hispanic	−0.106***	0.190***	0.108***	−0.315***	−0.570***	0.093*	−0.462***
	(0.028)	(0.031)	(0.030)	(0.036)	(0.051)	(0.037)	(0.049)
Others (Race)	0.125***	0.232***	0.113***	−0.162***	−0.331***	0.158***	−0.171***
	(0.030)	(0.033)	(0.032)	(0.038)	(0.050)	(0.040)	(0.049)
White	0.176***	0.418***	0.227***	0.059*	−0.130**	0.334***	0.041
	(0.025)	(0.027)	(0.027)	(0.031)	(0.040)	(0.033)	(0.040)
Independent/Third-Party	−0.135***	−0.354***	−0.416***	−0.491***	−0.507***	−0.523***	−0.616***
	(0.018)	(0.021)	(0.019)	(0.021)	(0.030)	(0.024)	(0.028)
Democrat	0.316***	0.144***	0.601***	−0.193***	−0.176***	−0.145***	−0.327***
	(0.019)	(0.022)	(0.018)	(0.020)	(0.027)	(0.024)	(0.025)
Born Abroad	−0.263***	−0.251***	−0.196***	−0.203***	−0.192***	−0.275***	−0.265***
	(0.018)	(0.021)	(0.020)	(0.024)	(0.033)	(0.027)	(0.032)
39th Cong. District	0.115	0.200	0.279	−0.019	0.202	0.134	0.236
	(0.155)	(0.170)	(0.171)	(0.186)	(0.268)	(0.196)	(0.247)
45th Cong. District	0.231	0.278*	0.325*	0.012	0.248	0.169	0.151
	(0.155)	(0.169)	(0.170)	(0.186)	(0.267)	(0.195)	(0.246)
46th Cong. District	−0.022	0.178	0.334*	−0.025	0.169	0.079	0.169
	(0.156)	(0.170)	(0.171)	(0.187)	(0.269)	(0.197)	(0.248)
47th Cong. District	0.015	0.147	0.243	−0.105	0.208	0.071	0.128
	(0.156)	(0.171)	(0.172)	(0.188)	(0.270)	(0.198)	(0.249)
48th Cong. District	0.230	0.267	0.257	−0.021	0.182	0.154	0.055
	(0.154)	(0.169)	(0.170)	(0.185)	(0.267)	(0.195)	(0.246)
49th Cong. District	0.289*	0.390**	0.254	−0.006	0.220	0.262	0.081
	(0.157)	(0.172)	(0.171)	(0.188)	(0.269)	(0.198)	(0.248)
Constant	−0.907***	1.031***	−0.901***	−0.777***	−2.195***	0.880***	−1.597***
	(0.167)	(0.183)	(0.184)	(0.204)	(0.289)	(0.214)	(0.269)
Observations	100,389	96,195	83,977	71,411	69,104	65,388	59,914
Adjusted R^2	0.159	0.052	0.090	0.134	0.165	0.067	0.164
Log Likelihood	−57,722.160	−45,323.170	−53,401.650	−41,960.780	−26,208.670	−34,083.580	−27,761.590
UBRE	57,859.880	45,412.880	53,480.550	42,045.840	26,277.130	34,150.990	27,832.550

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.3: Generalized Additive Model Results, Full Sample

		Imperfect Placebo	*Placebo Tests*				
	General 2018	General 2016	Primary 2016	General 2014	Primary 2014	General 2012	Primary 2012
A. Smooth terms (effective degrees of freedom / residual degrees of freedom)							
Res. Stability × Same Address	3.310	1.708*	2.771	2.732	1.148	1.002	1.000
	4.102	1.004	3.448	3.400	1.281	1.003	1.001
Res. Stability × Same Precinct	2.559	1.004	3.169	1.930	3.181	1.541	3.155
	3.189	1.008	3.937	2.416	3.954	1.904	3.924
Res. Stability × Same Subdist.	3.613*	1.113*	2.019	4.075	1.001***	1.001	3.368**
	4.472	1.218	2.522	6.027	1.001	1.002	4.177
Res. Stability × Same Cong.	3.003*	1.001	1.438	1.001	1.001	1.504	1.001
	3.730	1.003	1.756	1.002	1.002	1.853	1.001
Res. Stability × Diff. Cong.	1.564**	1.003	1.002	1.002	1.808	1.003	1.084
	1.939	1.006	1.005	1.004	2.274	1.005	1.163
Age	6.009***	4.398***	4.492***	1.085***	4.788***	6.285***	3.961***
	7.102	5.392	5.510	1.166	5.856	7.297	4.907
Distance to Poll	1.016	1.003	1.004	1.003	1.001	1.000	1.001
	1.031	1.005	1.008	1.005	1.003	1.001	1.002
Old Residence's	5.223*	6.960**	3.146	3.645	1.002	1.000*	1.957
Neighborhood Income	6.297	7.998	3.944	4.534	1.003	1.001	2.484
B. Parametric coefficients (estimate / standard error)							
Same Precinct	0.057	0.053	−0.026	0.262***	0.120	0.118	0.031
	(0.058)	(0.068)	(0.061)	(0.073)	(0.098)	(0.080)	(0.094)
Same Subdist.	0.053	0.184*	0.109	0.253**	−0.093	0.238*	−0.089
	(0.077)	(0.089)	(0.079)	(0.093)	(0.130)	(0.104)	(0.122)
Same Cong.	0.009	0.305**	−0.006	0.266**	−0.024	0.312**	−0.041
	(0.080)	(0.093)	(0.083)	(0.097)	(0.134)	(0.109)	(0.127)
Diff. Cong.	0.001	−0.217	−0.172	−0.254	−1.032	0.086	0.333
	(0.232)	(0.241)	(0.239)	(0.296)	(0.569)	(0.286)	(0.326)
Distance Moved	−0.020	−0.106	0.185	0.407*	0.591*	−0.139	0.505
	(0.171)	(0.200)	(0.175)	(0.201)	(0.277)	(0.230)	(0.264)
Turnout 2016	1.484***						
	(0.046)						
Moved Times	−0.222***	−0.612***	−0.276***	−0.287***	−0.307**	−0.166	−0.226*
	(0.066)	(0.072)	(0.070)	(0.085)	(0.119)	(0.088)	(0.111)
PAV	0.163***	0.346***	0.251***	0.335***	0.850***	0.087	0.806***
	(0.042)	(0.048)	(0.044)	(0.051)	(0.078)	(0.056)	(0.070)
Female	0.189	0.090	0.264*	0.294*	−0.195	0.317*	0.331
	(0.102)	(0.116)	(0.115)	(0.141)	(0.179)	(0.140)	(0.198)
Male	0.238*	0.031	0.251*	0.463**	0.062	0.251	0.510*
	(0.102)	(0.116)	(0.116)	(0.141)	(0.179)	(0.140)	(0.198)
Black	0.206	0.083	0.075	−0.095	0.008	0.742*	−0.828
	(0.209)	(0.237)	(0.233)	(0.284)	(0.373)	(0.321)	(0.465)
Hispanic	0.083	0.112	0.122	−0.203	−0.623***	−0.053	−0.597***
	(0.079)	(0.087)	(0.087)	(0.105)	(0.148)	(0.109)	(0.146)
Others (Race)	0.305***	0.194*	0.159	−0.101	−0.336*	0.166	−0.134
	(0.084)	(0.094)	(0.093)	(0.110)	(0.145)	(0.119)	(0.143)
White	0.332***	0.330***	0.245**	0.043	−0.163	0.260*	0.020
	(0.072)	(0.082)	(0.080)	(0.094)	(0.122)	(0.102)	(0.122)
Independent/Third-Party	−0.226***	−0.361***	−0.398***	−0.471***	−0.392***	−0.472***	−0.472***
	(0.051)	(0.059)	(0.054)	(0.061)	(0.085)	(0.069)	(0.080)
Democrat	0.270***	0.164**	0.587***	−0.166**	−0.217**	−0.122	−0.260***
	(0.052)	(0.062)	(0.051)	(0.058)	(0.078)	(0.070)	(0.074)
foreign_born	−0.223***	−0.309***	−0.251***	−0.324***	−0.284**	−0.207**	−0.366***
	(0.049)	(0.057)	(0.056)	(0.068)	(0.093)	(0.076)	(0.091)
39th Cong. District	0.364	0.730	0.636	−0.791	−0.559	−0.777	−1.640**
	(0.395)	(0.403)	(0.475)	(0.508)	(0.612)	(0.777)	(0.569)
45th Cong. District	0.644	0.807*	0.542	−0.733	−0.581	−0.691	−1.877***
	(0.393)	(0.401)	(0.473)	(0.505)	(0.608)	(0.775)	(0.566)
46th Cong. District	0.157	0.698	0.634	−0.746	−0.581	−0.876	−1.704**
	(0.397)	(0.405)	(0.477)	(0.510)	(0.614)	(0.778)	(0.572)
47th Cong. District	0.210	0.614	0.735	−0.839	−0.717	−0.828	−1.944***
	(0.400)	(0.408)	(0.480)	(0.513)	(0.620)	(0.781)	(0.578)
48th Cong. District	0.654	0.900*	0.568	−0.708	−0.704	−0.744	−1.926***
	(0.393)	(0.400)	(0.473)	(0.505)	(0.608)	(0.775)	(0.565)
49th Cong. District	0.610	1.062**	0.461	−0.793	−0.776	−0.502	−2.182***
	(0.400)	(0.412)	(0.478)	(0.511)	(0.617)	(0.781)	(0.575)
Constant	−1.313**	0.877*	−1.250*	−0.141	−1.092	1.730*	0.085
	(0.419)	(0.430)	(0.498)	(0.538)	(0.655)	(0.797)	(0.619)
Observations	13,150	12,470	10,680	8,877	8,529	8,008	7,307
Adjusted R^2	0.157	0.051	0.090	0.137	0.173	0.068	0.186
Log Likelihood	−7,578.712	−5,909.561	−6,823.539	−5,199.565	−3,164.015	−4,214.901	−3,331.615
UBRE	7,619.864	5,942.088	6,855.060	5,224.926	3,178.778	4,238.004	3,348.890

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.4: Generalized Additive Model Results, Subsample of Movers within Half Mile

		Imperfect Placebo	*Placebo Tests*				
	General 2018	General 2016	Primary 2016	General 2014	Primary 2014	General 2012	Primary 2012
NCOA Treatment	0.059***	−0.006	0.001	−0.004	−0.011	0.001	−0.010
	(0.013)	(0.011)	(0.015)	(0.015)	(0.012)	(0.015)	(0.015)
Same Address	−0.008	−0.051	−0.038	−0.046	0.026	−0.117**	−0.041
	(0.040)	(0.036)	(0.049)	(0.050)	(0.039)	(0.049)	(0.047)
Same Precinct	0.014	−0.063*	−0.032	−0.070	0.008	−0.079*	−0.052
	(0.036)	(0.033)	(0.042)	(0.044)	(0.034)	(0.043)	(0.041)
Same Cong.	−0.040*	−0.018	−0.015	−0.019	0.004	−0.039	−0.055**
	(0.024)	(0.022)	(0.029)	(0.029)	(0.023)	(0.029)	(0.028)
Diff. Cong.	−0.051*	−0.043*	−0.022	−0.016	−0.013	−0.047	−0.076**
	(0.027)	(0.024)	(0.032)	(0.033)	(0.026)	(0.032)	(0.031)
Distance Moved	−0.002*	0.001	0.001	−0.002	0.0001	0.001	−0.0005
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
2016 Turnout	0.275***						
	(0.014)						
Times Moved	−0.057***	−0.072***	−0.070***	−0.071***	−0.037***	−0.018	−0.042**
	(0.015)	(0.013)	(0.017)	(0.017)	(0.014)	(0.017)	(0.017)
Distance to Poll	0.013	−0.012	−0.030	0.007	−0.004	−0.006	−0.005
	(0.017)	(0.016)	(0.020)	(0.021)	(0.016)	(0.020)	(0.020)
PAV†	0.054***	0.051***	0.078***	0.090***	0.098***	0.026*	0.124***
	(0.013)	(0.012)	(0.015)	(0.016)	(0.012)	(0.015)	(0.015)
Age	0.002***	0.003***	0.005***	0.009***	0.007***	0.005***	0.008***
	(0.0004)	(0.0004)	(0.0005)	(0.001)	(0.0004)	(0.0005)	(0.0005)
Female‡	0.017	−0.006	−0.020	0.028	−0.020	0.038	0.015
	(0.034)	(0.032)	(0.042)	(0.043)	(0.034)	(0.044)	(0.044)
Male‡	0.062*	−0.026	0.0001	0.093**	0.028	0.050	0.055
	(0.034)	(0.032)	(0.042)	(0.043)	(0.034)	(0.044)	(0.044)
White	0.040*	0.094***	0.045*	0.010	0.007	0.051*	−0.009
	(0.022)	(0.020)	(0.027)	(0.028)	(0.022)	(0.027)	(0.027)
Black	0.245***	0.113*	0.192**	0.111	0.015	0.141	0.163*
	(0.064)	(0.060)	(0.084)	(0.086)	(0.068)	(0.088)	(0.086)
Hispanic	−0.030	0.063***	0.041	−0.066**	−0.035	0.037	−0.061*
	(0.026)	(0.023)	(0.031)	(0.032)	(0.025)	(0.032)	(0.032)
Others (Race)‡‡	−0.005	0.073***	0.029	−0.034	0.020	0.020	−0.019
	(0.026)	(0.024)	(0.031)	(0.033)	(0.025)	(0.032)	(0.032)
Independent/Third-Party	−0.016	−0.055***	−0.081***	−0.102***	−0.038***	−0.088***	−0.068***
	(0.015)	(0.014)	(0.018)	(0.019)	(0.014)	(0.018)	(0.018)
Democrat	0.104***	0.021	0.142***	−0.036*	−0.0004	−0.021	−0.042**
	(0.016)	(0.014)	(0.018)	(0.019)	(0.015)	(0.018)	(0.018)
Old Residence's	0.001***	0.0003	−0.001*	−0.001*	−0.0001	0.0003	−0.0002
Neighborhood Income	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
New Residence's	0.001*	0.0005*	0.0003	0.00003	−0.0002	0.001	−0.00003
Neighborhood Income	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0003)
Born Abroad	−0.044***	−0.019	−0.055***	−0.047**	−0.055***	−0.044**	−0.080***
	(0.015)	(0.014)	(0.020)	(0.021)	(0.017)	(0.021)	(0.021)
39th Cong. District	0.113	0.229*	0.116	−0.145	−0.005	0.292**	−0.173
	(0.139)	(0.120)	(0.153)	(0.145)	(0.117)	(0.143)	(0.142)
45th Cong. District	0.156	0.252**	0.141	−0.148	−0.012	0.305**	−0.156
	(0.138)	(0.120)	(0.152)	(0.144)	(0.116)	(0.142)	(0.141)
46th Cong. District	0.108	0.234*	0.132	−0.152	−0.008	0.253*	−0.155
	(0.139)	(0.120)	(0.153)	(0.145)	(0.117)	(0.143)	(0.142)
47th Cong. District	0.125	0.208*	0.164	−0.153	−0.006	0.325**	−0.111
	(0.141)	(0.122)	(0.155)	(0.147)	(0.118)	(0.145)	(0.144)
48th Cong. District	0.157	0.239**	0.142	−0.156	−0.007	0.273*	−0.176
	(0.139)	(0.120)	(0.152)	(0.144)	(0.116)	(0.142)	(0.141)
49th Cong. District	0.150	0.246**	0.056	−0.147	−0.026	0.305**	−0.213
	(0.140)	(0.121)	(0.153)	(0.146)	(0.117)	(0.144)	(0.142)
Constant	0.011	0.410***	0.100	0.195	−0.146	0.142	0.099
	(0.151)	(0.132)	(0.167)	(0.161)	(0.129)	(0.159)	(0.157)
Observations	5,539	5,035	4,341	3,680	3,553	3,366	3,082
Adjusted R^2	0.126	0.043	0.081	0.135	0.136	0.057	0.136
Res. Std. Error	0.455 (df=5510)	0.392 (df=5007)	0.474 (df=4313)	0.448 (df=3652)	0.342 (df=3525)	0.419 (df=3338)	0.391 (df=3054)
F statistic	29.427*** (df=28; 5510)	9.334*** (df=27; 5007)	15.164*** (df=27; 4313)	22.291*** (df=27; 3652)	21.695*** (df=27; 3525)	8.587*** (df=27; 3338)	18.902*** (df=27; 3054)

Note: *p<0.05; **p<0.01; ***p<0.001
†: Permanent absentee voter.
‡: Voters with no clear gender classification are included as a baseline.
‡‡: Asian Americans are baseline race/ethnicity.

Table A.5: Effect of NCOA Automatic Voter Registration, Linear Probability Model

	Dependent Variable: General 2018 Turnout			
	Low Info Cost	Distance Moved Less Than 0.5-mile	Distance Moved Less Than 1 Mile	Distance Moved Less Than 3 Miles
NCOA Treatment	0.044	0.050	0.060**	0.076***
	(0.043)	(0.034)	(0.027)	(0.018)
Same Address		0.035	0.015	−0.001
		(0.054)	(0.046)	(0.043)
Same Precinct	0.053	0.083	0.064	0.031
	(0.051)	(0.054)	(0.044)	(0.038)
Distance Moved	0.050			−0.009
	(0.111)			(0.012)
Same Cong.		0.012	0.015	−0.020
		(0.056)	(0.039)	(0.027)
Diff. Cong.		−0.149	0.103	−0.052
		(0.206)	(0.093)	(0.041)
2016 Turnout	0.299***	0.313***	0.286***	0.272***
	(0.047)	(0.038)	(0.031)	(0.021)
Times Moved	0.024	−0.027	−0.028	−0.041*
	(0.056)	(0.038)	(0.031)	(0.022)
Distance to Poll	−0.074	−0.002	−0.040	0.004
	(0.063)	(0.049)	(0.040)	(0.026)
PAV	0.045	0.006	0.021	0.068***
	(0.045)	(0.035)	(0.028)	(0.019)
Age	0.004***	0.002**	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.024	−0.093	0.017	0.015
	(0.113)	(0.090)	(0.077)	(0.048)
Male	0.097	−0.037	0.051	0.057
	(0.114)	(0.091)	(0.077)	(0.048)
Black	0.488**	0.481***	0.397***	0.333***
	(0.194)	(0.158)	(0.140)	(0.101)
Hispanic	0.053	0.082	0.003	−0.050
	(0.098)	(0.077)	(0.059)	(0.037)
Others (Race)	0.065	0.130	0.084	−0.034
	(0.101)	(0.080)	(0.060)	(0.037)
White	0.048	0.142**	0.061	0.024
	(0.093)	(0.072)	(0.054)	(0.032)
Independent/Third-Party	−0.022	−0.085**	−0.044	−0.033
	(0.056)	(0.043)	(0.034)	(0.023)
Democrat	0.136**	0.102**	0.103***	0.078***
	(0.055)	(0.043)	(0.035)	(0.023)
Old Residence's	0.022	−0.002	0.0002	0.0001
Neighborhood Income	(0.022)	(0.004)	(0.002)	(0.001)
New Residence's	−0.019	0.003	0.002	0.001*
Neighborhood Income	(0.022)	(0.004)	(0.002)	(0.001)
Born Abroad	−0.013	0.009	−0.013	−0.036
	(0.052)	(0.041)	(0.033)	(0.022)
39th Cong. District			0.042	0.224
			(0.346)	(0.148)
45th Cong. District	−0.047	−0.011	0.037	0.238
	(0.074)	(0.059)	(0.345)	(0.147)
46th Cong. District	−0.057	−0.006	−0.028	0.154
	(0.086)	(0.071)	(0.346)	(0.148)
47th Cong. District	0.111	0.115	0.067	0.210
	(0.116)	(0.090)	(0.350)	(0.151)
48th Cong. District	0.066	0.079	0.090	0.245*
	(0.074)	(0.058)	(0.345)	(0.147)
49th Cong. District	−0.111	−0.055	−0.062	0.187
	(0.107)	(0.086)	(0.348)	(0.150)
Constant	−0.158	0.033	0.012	−0.060
	(0.206)	(0.167)	(0.369)	(0.169)
Observations	464	724	1,143	2,548
R^2	0.176	0.179	0.145	0.137
Adjusted R^2	0.131	0.149	0.124	0.128
Res. Std. Error	0.441 (df = 439)	0.441 (df = 697)	0.446 (df = 1115)	0.450 (df = 2519)
F Statistic	3.898*** (df = 24; 439)	5.851*** (df = 26; 697)	6.989*** (df = 27; 1115)	14.323*** (df = 28; 2519)

Note: *p<0.05; **p<0.01; ***p<0.001
†: Permanent absentee voter.
‡: Voters with no clear gender classification are included as a baseline.
‡‡: Asian Americans are baseline race/ethnicity.

Table A.6: Effect of NCOA Automatic Voter Registration by Information/Social Costs

Appendix B

APPENDIX FOR CHAPTER 2

B.1 11 CFR 110.6 Earmarked Contributions

The following is the full excerpt of the United States Code, the material in italics has been copied verbatim.

§ *110.6 Earmarked contributions 52 U.S.C. 30116(a)(8)).*

(a) General. All contributions by a person made on behalf of or to a candidate, including contributions which are in any way earmarked or otherwise directed to the candidate through an intermediary or conduit, are contributions from the person to the candidate.

(b) Definitions.

(1) For purposes of this section, earmarked means a designation, instruction, or encumbrance, whether direct or indirect, express or implied, oral or written, which results in all or any part of a contribution or expenditure being made to, or expended on behalf of, a clearly identified candidate or a candidate's authorized committee.

(2) For purposes of this section, conduit or intermediary means any person who receives and forwards an earmarked contribution to a candidate or a candidate's authorized committee, except as provided in paragraph (b)(2)(i) of this section.

(i) For purposes of this section, the following persons shall not be considered to be conduits or intermediaries:

(A) An individual who is an employee or a full-time volunteer working for the candidate's authorized committee, provided that the individual is not acting in his or her capacity as a representative of an entity prohibited from making contributions;

(B) A fundraising representative conducting joint fundraising with the candidate's authorized committee pursuant to 11 CFR 102.17 or 9034.8;

(C) An affiliated committee, as defined in 11 CFR 100.5(g);

(D) A commercial fundraising firm retained by the candidate or the candidate's authorized committee to assist in fundraising; and

(E) An individual who is expressly authorized by the candidate or the candidate's authorized committee to engage in fundraising, and who occupies a significant position within the candidate's campaign organization, provided that the individual is not acting in his or her capacity as a representative of an entity prohibited from making contributions.

(ii) Any person who is prohibited from making contributions or expenditures in connection with an election for Federal office shall be prohibited from acting as a conduit for contributions earmarked to candidates or their authorized committees. The provisions of this section shall not restrict the ability of an organization or committee to serve as a collecting agent for a separate segregated fund pursuant to 11 CFR 102.6.

(iii) Any person who receives an earmarked contribution shall forward such earmarked contribution to the candidate or authorized committee in accordance with 11 CFR 102.8, except that -

(A) A fundraising representative shall follow the joint fundraising procedures set forth at 11 CFR 102.17.

(B) A person who is prohibited from acting as a conduit pursuant to paragraph (b)(2)(ii) of this section shall return the earmarked contribution to the contributor.

(c) Reporting of earmarked contributions -

(1) Reports by conduits and intermediaries.

(i) The intermediary or conduit of the earmarked contribution shall report the original source and the recipient

candidate or authorized committee to the Commission or the Secretary of the Senate, as appropriate (see 11 CFR part 105), and to the recipient candidate or authorized committee.

(ii) The report to the Commission or Secretary shall be included in the conduit's or intermediary's report for the reporting period in which the earmarked contribution was received, or, if the conduit or intermediary is not required to report under 11 CFR part 104, by letter to the Commission within thirty days after forwarding the earmarked contribution.

(iii) The report to the recipient candidate or authorized committee shall be made when the earmarked contribution is forwarded to the recipient candidate or authorized committee pursuant to 11 CFR 102.8.

(iv) The report by the conduit or intermediary shall contain the following information:

(A) The name and mailing address of each contributor and, for each earmarked contribution in excess of $200, the contributor's occupation and the name of his or her employer;

(B) The amount of each earmarked contribution, the date received by the conduit, and the intended recipient as designated by the contributor; and

(C) The date each earmarked contribution was forwarded to the recipient candidate or authorized committee and whether the earmarked contribution was forwarded in cash or by the contributor's check or by the conduit's check.

(v) For each earmarked contribution passed through the conduit's or intermediary's account, the information specified in paragraph (c)(1)(iv) (A) through (C) of this section shall be itemized on the appropriate schedules of receipts and disbursements attached to the conduit's or intermediary's report, or shall be disclosed by letter, as appropri-

ate. For each earmarked contribution forwarded in the form of the contributor's check or other written instrument, the information specified in paragraph (c)(1)(iv) (A) through (C) of this section shall be disclosed as a memo entry on the appropriate schedules of receipts and disbursements attached to the conduit's or intermediary's report, or shall be disclosed by letter, as appropriate.:

(2) Reports by recipient candidates and authorized committees.

(i) The recipient candidate or authorized committee shall report each conduit or intermediary who forwards one or more earmarked contributions which in the aggregate exceed $200 in any election cycle.

(ii) The report by the recipient candidate or authorized committee shall contain the following information:

(A) The identification of the conduit or intermediary, as defined in 11 CFR 100.12;

(B) The total amount of earmarked contributions received from the conduit or intermediary and the date of receipt; and

(C) The information required under 11 CFR 104.3(a) (3) and (4) for each earmarked contribution which in the aggregate exceeds $200 in any election cycle.

(iii) The information specified in paragraph (c)(2)(ii) (A) through (C) of this section shall be itemized on Schedule A attached to the report for the reporting period in which the earmarked contribution is received.

(d) Direction or control.

(1) A conduit's or intermediary's contribution limits are not affected by the forwarding of an earmarked contribution except where the conduit or intermediary exercises any direction or control over the choice of the recipient candidate.

(2) If a conduit or intermediary exercises any direction or control over the choice of the recipient candidate, the earmarked contribution shall be considered a contribution by both the original contributor and the conduit or intermediary. If the

> *conduit or intermediary exercises any direction or control over the choice of the recipient candidate, the report filed by the conduit or intermediary and the report filed by the recipient candidate or authorized committee shall indicate that the earmarked contribution is made by both the original contributor and the conduit or intermediary, and that the entire amount of the contribution is attributed to each.*

[54 FR 34113, Aug. 17, 1989 and 54 FR 48580, Nov. 24, 1989; 61 FR 3550, Feb. 1, 1996; 81 FR 94240, Dec. 23, 2016]

B.2 Hypothetical Examples of Campaign Contributions

An example best illustrates how some donors become hidden. We also show how that affects summary statistics that might be computed using these data. Suppose that the following four hypothetical individual contributors donate during the 2016 election cycle (January 1, 2015 to December 31, 2016). We assume that a committee does not voluntarily report any contributions not required by the Federal Code of Regulations.

1. Anne gives \$500 as a one-time donation to the Democratic National Committee.
2. Diana gives to her junior Senator \$60 every month.
3. Ruby gives \$5 to her House Representative every month.
4. Gilbert gives \$5 to the same Representative every month, but via an intermediary committee.

Because Anne's single transaction well-exceeds the threshold, she is free of any kind of censoring. The rest, however, may be subject to censoring.

1. For Diana, given the \$200 threshold, only her fourth and later contributions will be disclosed, since \$180 < \$200 < \$240. Her first three months' contributions will not be recorded.
2. Ruby's contributions altogether only add up to \$120, well below the disclosure threshold, and none of her contributions will be made public. Her \$120 will be, however, added to a total sum of 'unitemized contributions' of the committee.

3. Although Gilbert's contribution pattern is identical to Ruby, because he gives through an intermediary committee, every single one of his records will be itemized.

Given these, what we really see in the FEC database are only the following itemized contributions:

1. Anne has a single transaction of $500. Her year-to-date aggregate is recorded as $500.
2. Diana has a twenty-one transactions of $60. The first transaction records her year-to-date aggregate as $240, and later ones correspondingly $300, $360, ...
3. Gilbert has twenty-four transactions each of $5 with corresponding year-to-date aggregates, clearly marked as earmarked contributions, *in the intermediary committee's report.* Gilbert's true intended recipient, his representative, need not disclose any contributions from him on her own reports.

If the researcher utilizes only data from committees that are final recipients, both Gilbert and Ruby are wiped from view. If she cares to investigate intermediary committees' reports as well, Gilbert's actions are fully observable, while Ruby, identical to Gilbert except for intermediaries, becomes completely hidden.

We can immediately spot several problems with descriptive statistics. With only observed transactions, it is difficult to infer the true state of the world. For example, the true mean of donor's total contribution sum is $545. The mean estimated using only itemized contribution would be little more than $626—an upwardly biased estimate. The mean frequency estimate may also be biased. From a researcher's point of view, how many times Diana has actually given is a mystery. She may have given just one more time. For instance, she may have given $180 in her first contribution, then started giving in $60 increments. However, we also cannot completely rule out the possibility that she has given ten, twenty, fifty, or even a hundred different times.

Another concern is how Gilbert is visible while Ruby is not, and what researchers may infer from seeing just the censored sample. For example, two major parties have different utilization levels of intermediary committees, with the Democratic party much more reliant on them than the Republican party. Suppose we pool every transaction observable, regardless of which report it comes from. Then the data will be over-representative of small Democratic donors, and if not careful, one may falsely conclude that Democratic committees are much more "grassroots" than they truly. In truth, Republicans are giving just as similarly but just not through intermediary committees, hence the small Republican donors hidden. To compare committees on an equal footing, there must be substantial filtering that accounts for the data generating process.

B.3 ActBlue

ActBlue is a non-profit that offers "fundraising technology for the left."[1] First established in 2004, it contracts with Democratic campaigns to serve as a conduit, with a flat rate of 3.95% fee for every contribution it processes. Having grown into the central "online clearinghouse for Democratic action,"[2] more than 400 federal committees used ActBlue to raise money in 2014 (Willis, 2014). In 2016, more than 1,400 federal campaign committees have used ActBlue as an intermediary committee. In the 2018 election cycle, it raised 85 million as of May 20, 2017, and more than 1 billion as of October 17, 2018—the top political action committee in the amount raised, almost four times the Republican National Committee, which is next in line in amounts raised from a political action committee.

Unlike Emily's List, however, ActBlue's website does not recommend candidates or causes to donate to. It simply facilitates receipts for political operations and forwards contributions to its final destination. In this sense, it is more like a credit card company for political fundraising, but for mostly Democratic causes. This is a different practice from bundling, notably used by Emily's List.[3]

As of 2017, ActBlue reported to have processed \$522,705,365 USD with \$31.95 average contribution size, and had 7,892 groups raising money,[4] ranging from the Democratic Senate Congressional Committee to Alexandria Ocasio-Cortez, who

[1] ActBlue, About Us, https://secure.actblue.com/about, retrieved April 28, 2017.

[2] ActBlue Summary, Center for Responsive Politics, https://www.opensecrets.org/orgs/summary.php?id=D000021806. Retrieved April 28, 2017.

[3] The following receipt of Emily's List report shows this clearly. All three contributions here are not earmarked, but they have an 'X' mark in their Memo Item where it is written "Tammy Duckworth Contributions." This means that all three individuals have directly contributed to Emily's List, which will bundle these receipts and contribute to Tammy Duckworth's principal campaign committee as an organizational contribution. If we only investigate individual contributions, it may make sense to include these observations as well. However, I do not yet include such observations, as some of them are inconsistent in how they are forwarded and there is room for miscounting one duplicate contribution as two. See the following receipt of Emily's List and another receipt of Hillary for America, in which there was no mention of an earmark in the former but is marked as an earmark in the destination committee's report.

[4] ActBlue, 2017 in Review, https://report.actblue.com/. Retrieved November 27, 2018.

continues Sanders' legacy by offering students and activists the option to donate $27 on her ActBlue pages.

B.4 2020 Democratic Campaign Quotes

The following are some quotes from the Democratic primary candidates' donation websites, simultaneously captured on June 23, 2019, trying to sway individuals and especially potentially smaller donors:

- *Our campaign finance system is broken and I want to fix it. We need to end Citizens United and get dark money out of politics. That's why we're rejecting donations from corporate PACs and federal lobbyists. Let's build this campaign from the ground up. Are you in?* (Cory Booker)
- *"We're powering our movement one donation at a time. No corporate PACs. No federal lobbyists. No individual super PAC. Just us—together."* (Kirsten Gillibrand)
- *"We're not taking a dime from corporate PACs, so your donation is critical to our success."* (Kamala Harris)
- *"Donate $5 today to show the strength of our people-powered campaign."* (Amy Klobuchar)
- *"Your contribution doesn't just ensure that we have the resources we need to run and to win the White House: it ensures that our democracy is once again powered by people, and only people. Not PACs, not special interests, not corporations."* (Beto O'Rourke)
- *"We made it to the first debates! Keep the momentum going by donating $1 today!"* (Eric Swalwell)
- *Warren for President does not accept contributions from PACs of any kind or federally registered lobbyists.* (Elizabeth Warren)

B.5 Demographics and Occupations of Donors

Table B.1: Demographics and Occupations of Immediately Visible, Eventually Visible, and Hidden Donors, 2016 Sanders Campaign

	Immediately Visible	Eventually Visible	Hidden
% of Men	60.7%	56.2%	53.2%
% of Whites	92.0%	90.0%	86.7%
% of Blacks	2.3%	2.9%	3.0%
% of Hispanics	2.9%	4.5%	7.4%
% of Asians	2.8%	2.4%	2.9%
% of Unemployed	23.7%	27.5%	25.7%
% of Engineers	4.9%	4.8%	3.1%
% of Teachers	2.8%	4.4%	4.9%
% of Retired	7.4%	2.9%	1.3%
% of Attorneys	3.6%	1.9%	1.1%
% of Professors	2.4%	2.0%	1.2%
% of Physicians	3.5%	1.7%	0.8%
% of Consultants	2.1%	1.8%	1.3%
% of Students	0.8%	1.3%	4.0%
% of Homemakers	0.5%	0.3%	0.3%

B.6 Comparisons with Primary Winners

One thing that should be noted is that Sanders is a candidate who did not make it past the primaries in 2016. Therefore, the descriptive statistics that we see may differ had he won the primaries. Although we do not have individual-level details for other candidates, in this Section we present some aggregate trends to compare the trends. Figure 1 shows the proportion of unitemized contributions against total individual contributions reported in a given period—monthly for the even year and quarterly for the odd year—for 2012 and 2016 candidates. Except for Sanders, all others are general election candidates.

Notice that the trends are quite heterogeneous by candidate, and there is no fixed trend for general election candidates. In case of Romney's 2012 race and Clinton in 2016, the proportion of unitemized money slowly seems to climb, while it drops for Obama 2012 or Trump 2016. In addition, note that Figure 1 is not equal to showing how small and large donors give, since the unitemized givers in January of 2012/2016 can turn into itemized/large donors at a later month.

For Sanders, it is possible that had he won the primary he would have changed so-

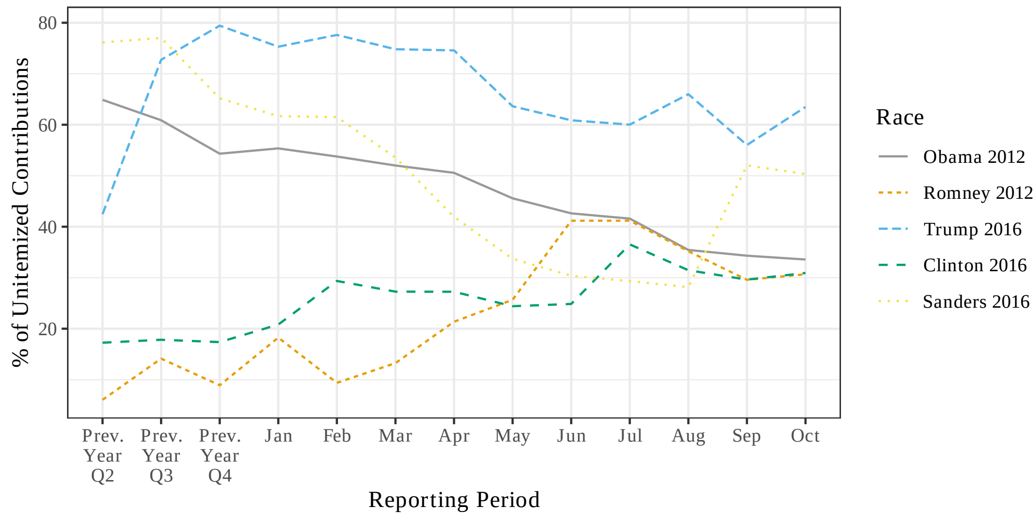


Figure B.1: Proportion of Unitemized Contributions in Total Individual Contributions, by FEC Regular Reporting Period

B.1

licitation strategies, since his campaign was uniquely leaning towards small donors. In that hypothetical case, itemized donations would increase. However, it is also possible that the trend from Obama 2012 continues, and even more unitemized contributions emerge, since the campaign's small donor strategy was a strongly pushed rhetoric, especially against PACs, corruption, and lobbyists.

Appendix C

APPENDIX FOR CHAPTER 3

C.1 Classical Calculations of Breakpoints

In this section, we briefly compare the sequential segmentation smooth splines with the classical approaches to detecting breakpoints. For this, we use the `strucchange` package in `R`, written and maintained by Achim Zeileis et al. The package description is as follows: "testing, monitoring, and dating structural changes in (linear) regression models." It introduces two approaches, one of which is F tests (Chow tests) designed for testing known breaks, and generalized fluctuation tests which minimizes the residual sum of squares of segmented linear regressions. Again, these are a different approach with a different philosophy from the nonparametric, sequential segmentation splines method.

One canonical data that the two classical approaches are applied to is the Nile data, or the 102 years' worth of the Nile's annual flow at Aswan. There is a known break for this data which is the completion of the Aswan dam in 1898, from which point the flow level decreases. From both the OLS-based CUSUM tests and the F statistics this single break is detected, and a two-segment model declared optimal. Figure C.1 shows the final model fit.

When the spline method is applied to the Nile data, it gives the following estimate as in Figure C.1. No breaks are detected. This is because when smooth splines are default, it smooths over the Nile flow over a long interval as downwards.

The results do not change even when we model the 1898 as a known break, as the BIC statistics are as follows for the sequential candidates of breaks: 6.443 for 1898, 10.024 for 1965 augmented, 14.482 for 1890 augmented, and so on. Splines only BIC is 5.621, hence there are no breaks.

Is this a false negative? It depends on the prior belief of the researcher about the data. If the researcher expects the data to be more or less stationary around a linear trend, and wishes to detect for level changes, the classical approaches fit better. If trends are expected to be nonlinear and we are still interested in sudden breaks, the spline method is our friend. On a more positive note, note that the first of the breaks chosen by the spline method is also 1898.

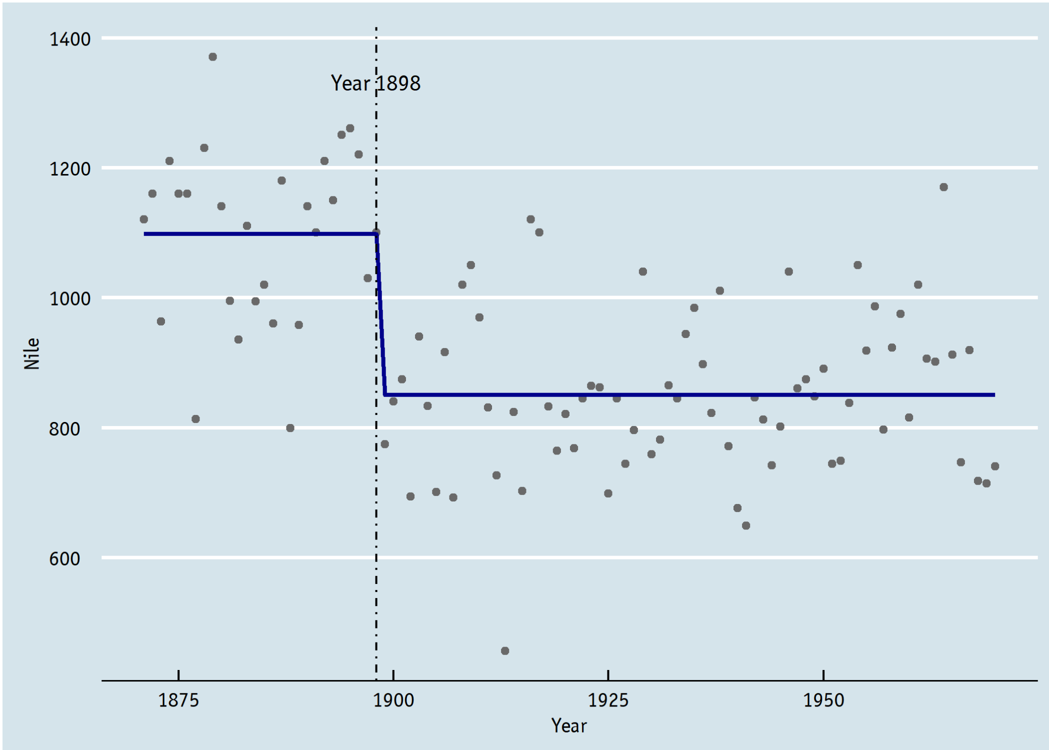


Figure C.1: Optimal Fitted Model for the Nile Data, F Statistics Method

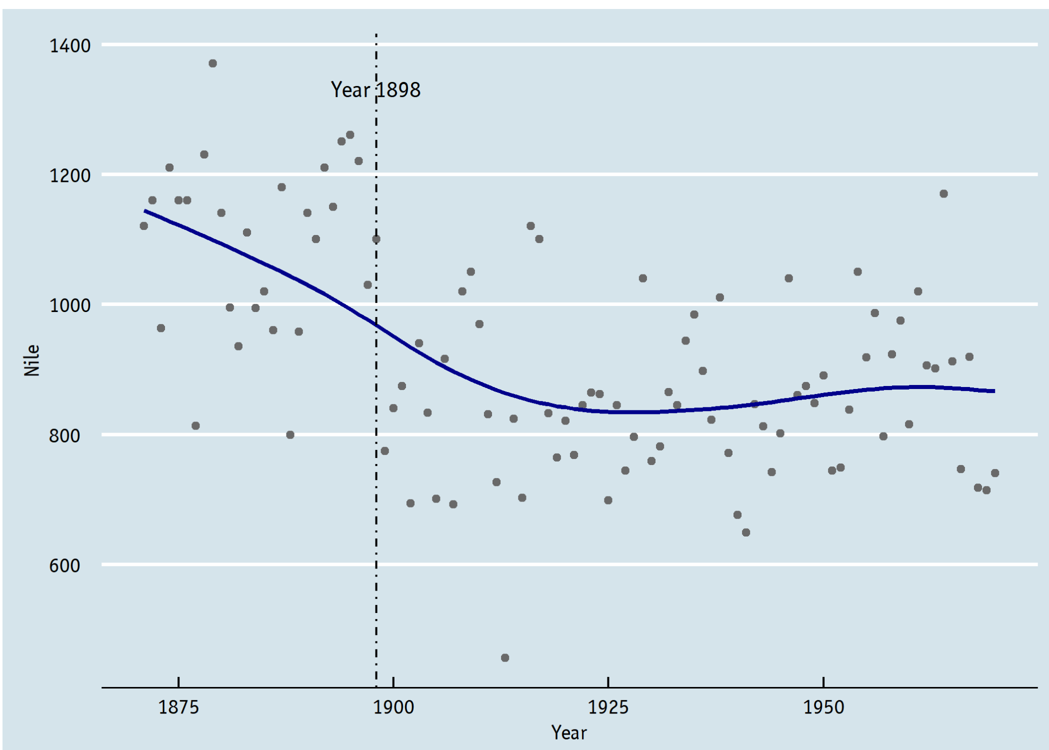


Figure C.2: Sequential Segmentation Spline Fit for the Nile Data

Having tested the Nile data with the spline method, we will now test the Bush poll data, the first example in Ratkovic and Eng (2010), with the classical methods, to see if they can also detect the two breaks at the 9/11 attacks and the start of the Iraq invasion, as does the spline method in Figure C.3.

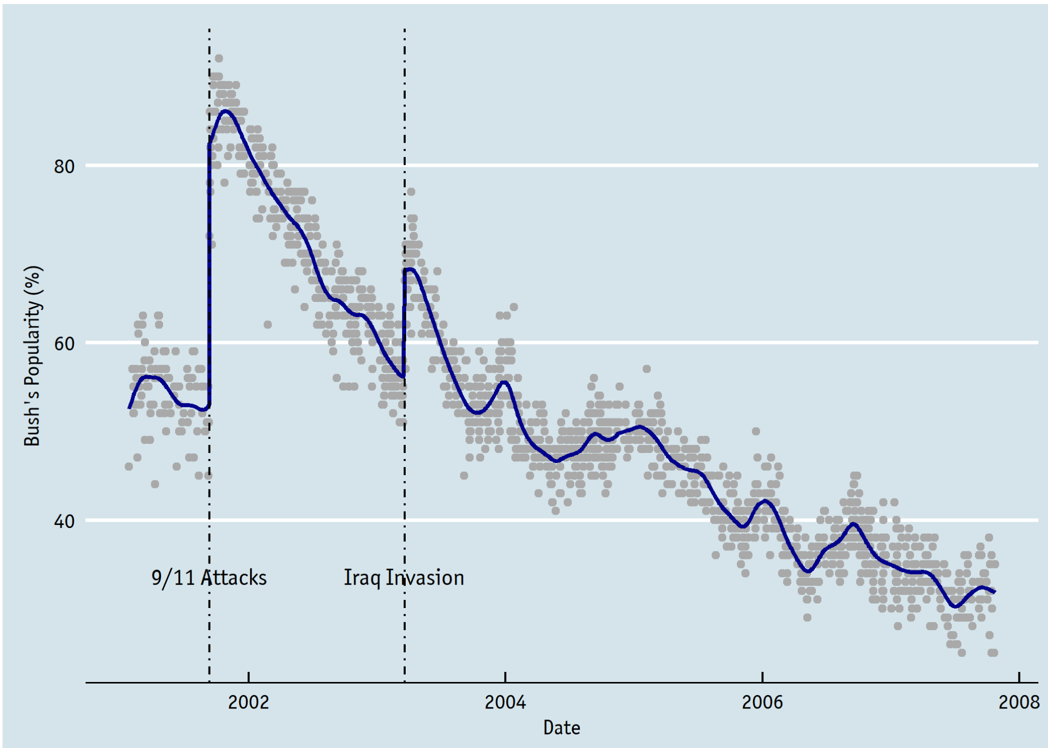


Figure C.3: Sequential Segmentation Spline Fit for the Bush Poll Data

Surprisingly, in Figure C.4, we see that the F statistics method fails to catch the 9/11, the more prominent of the jumps. While the method catches the Iraq invasion (more precisely, as March 17, 2003), it catches the following four more breaks: March 7, 2002, March 26, 2004, March 21, 2005, and June 11, 2006. The segmentations, however, do not seem to be an accurate summary of the bush poll's behavior, and misses the most obvious of the breaks.

We conclude that we cannot easily assess one method's "superiority" over another, as they are simply based on different assumptions about the underlying process of the data and have different strengths. For our purposes, the spline method is better than the classical structural break tests.

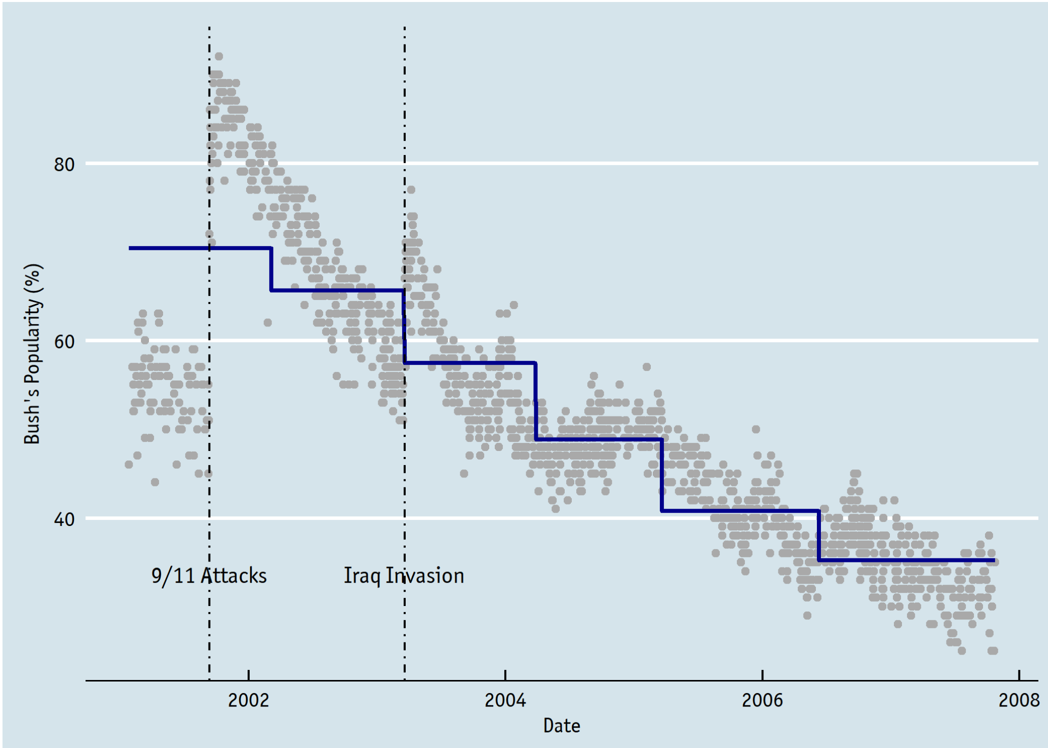


Figure C.4: Optimal Fitted Model for the Bush Poll Data, F Statistics Method

C.2 Primary Forecasts for All Candidates from FiveThirtyEight

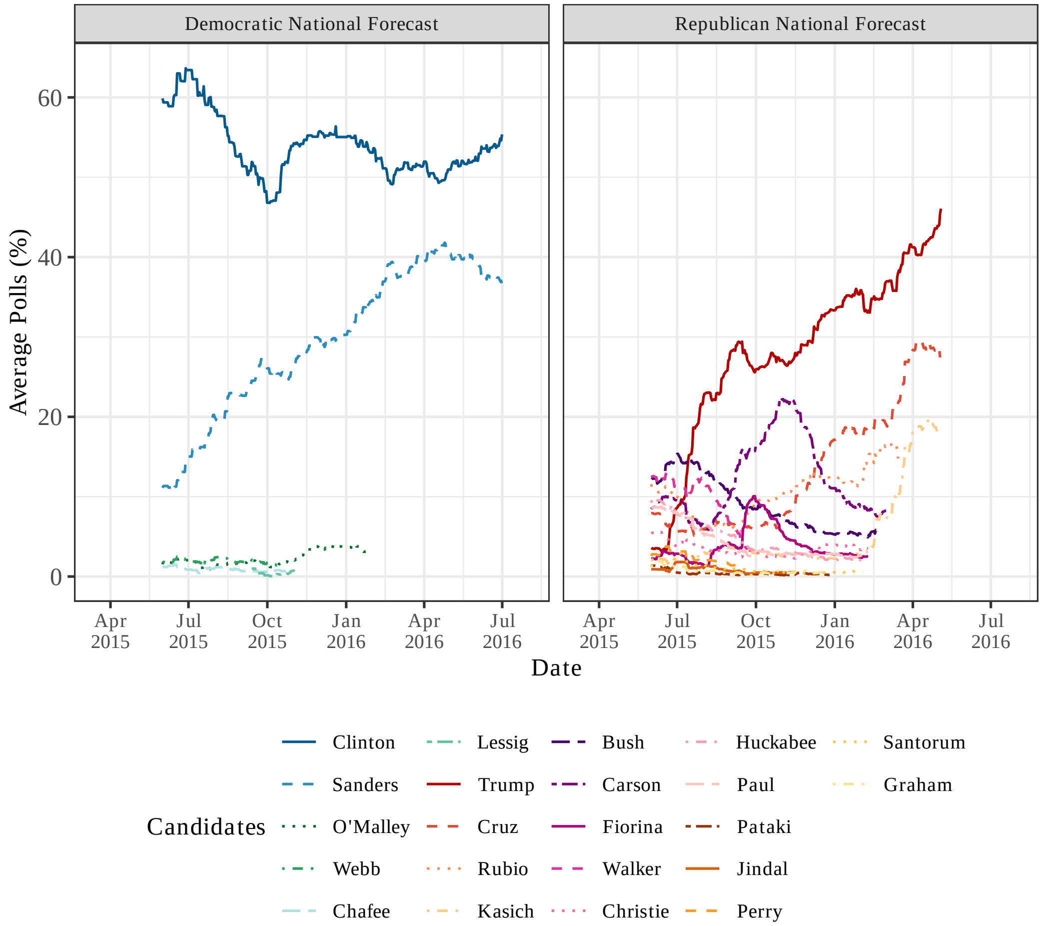


Figure C.5: National Primary Polls by Party, FiveThirtyEight, 2016 (All Candidates)

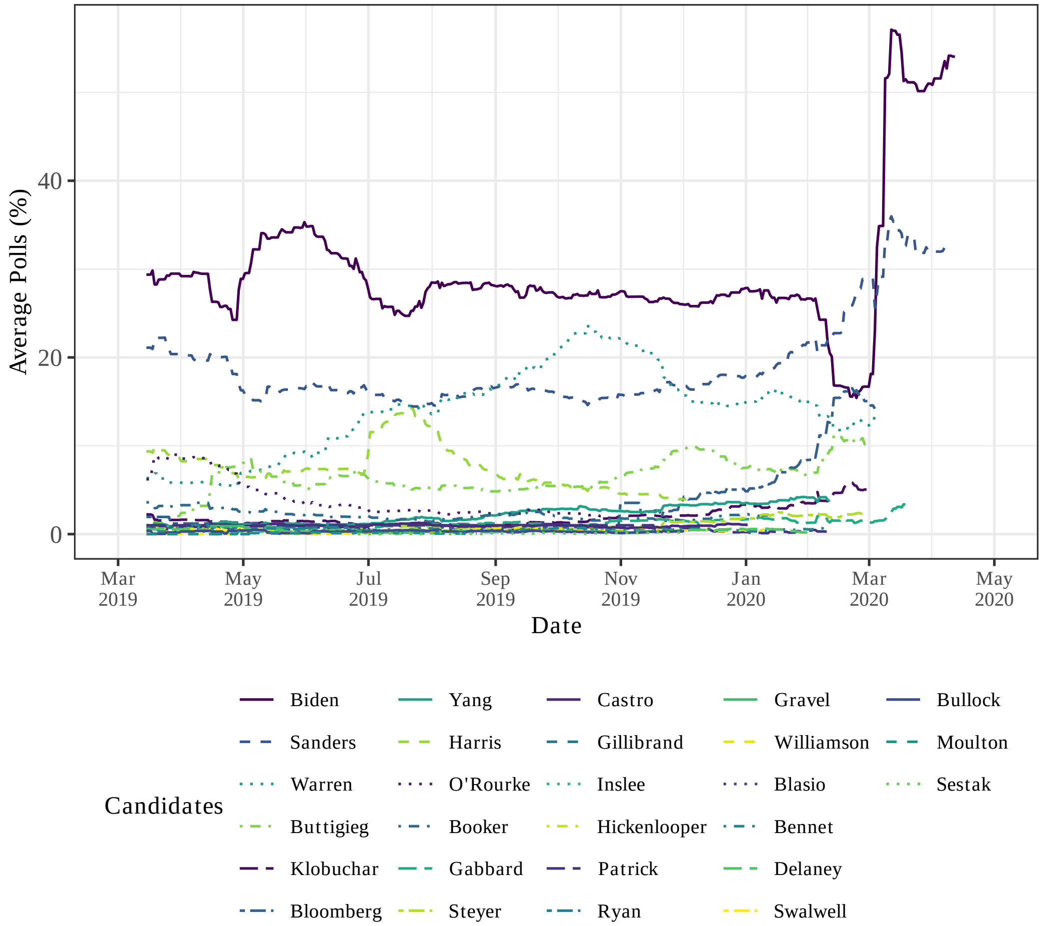


Figure C.6: National Primary Polls by Party, FiveThirtyEight, 2020 (All Candidates)

C.3 Additional Figures by State

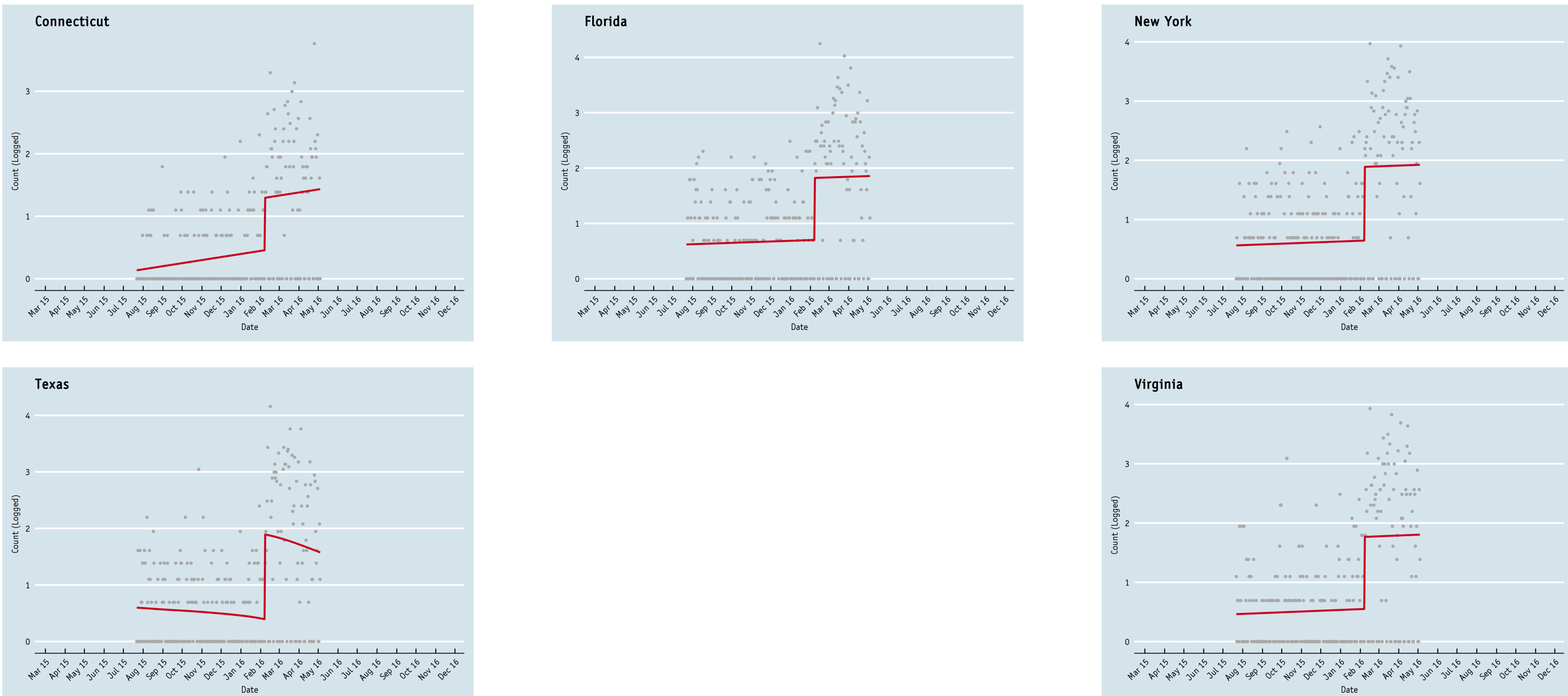


Figure C.7: States with a Jump at the New Hampshire Primary, Kasich

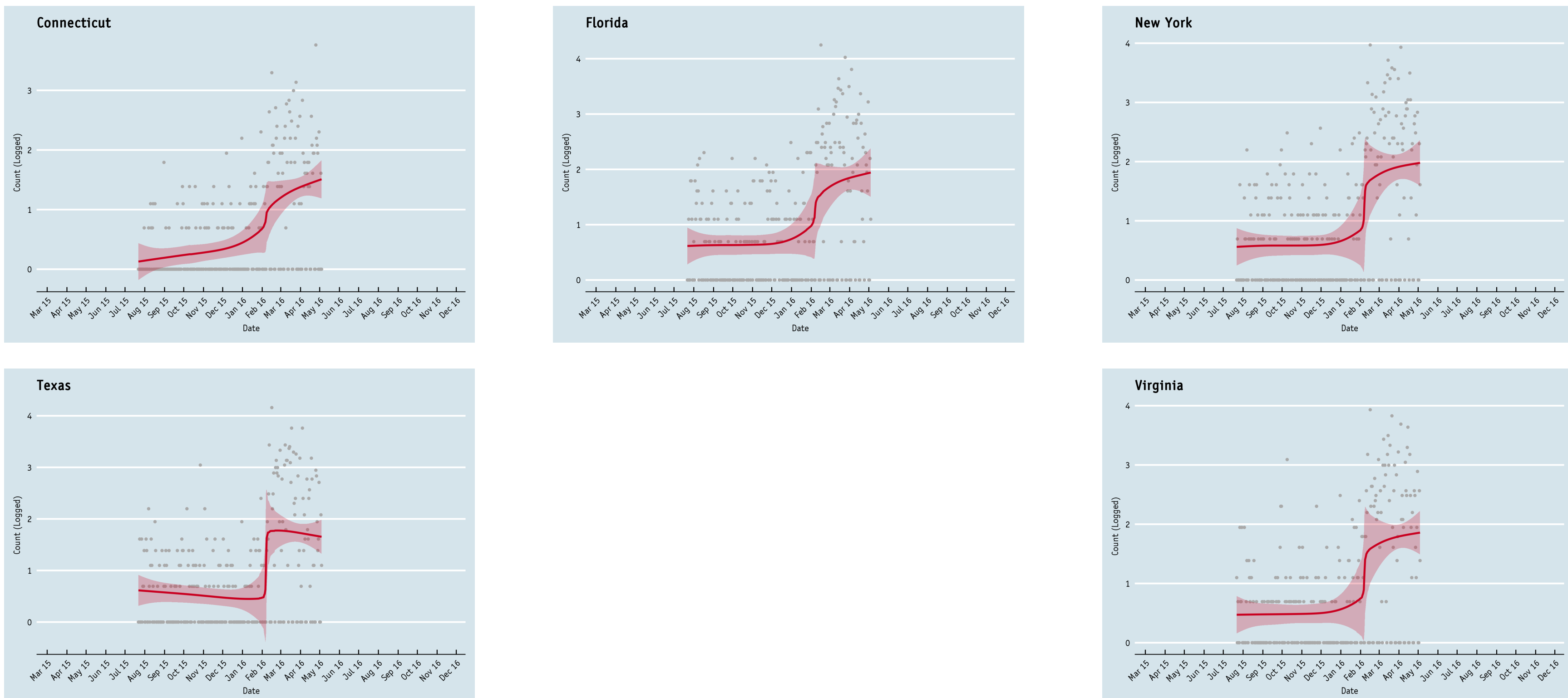


Figure C.8: Figure C.7, Bootstrapped

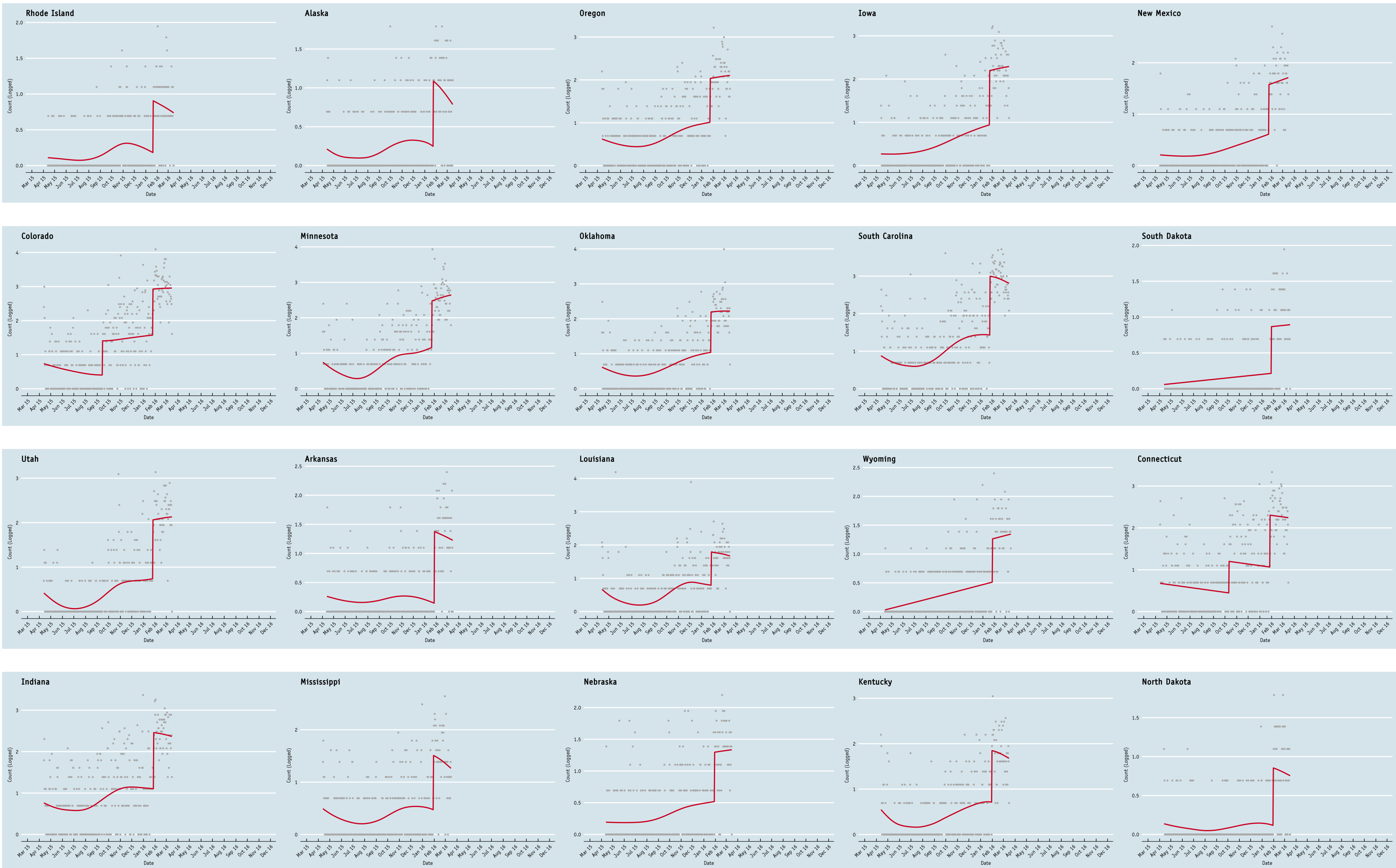


Figure C.9: States with a Jump at New Hampshire Primary, Rubio

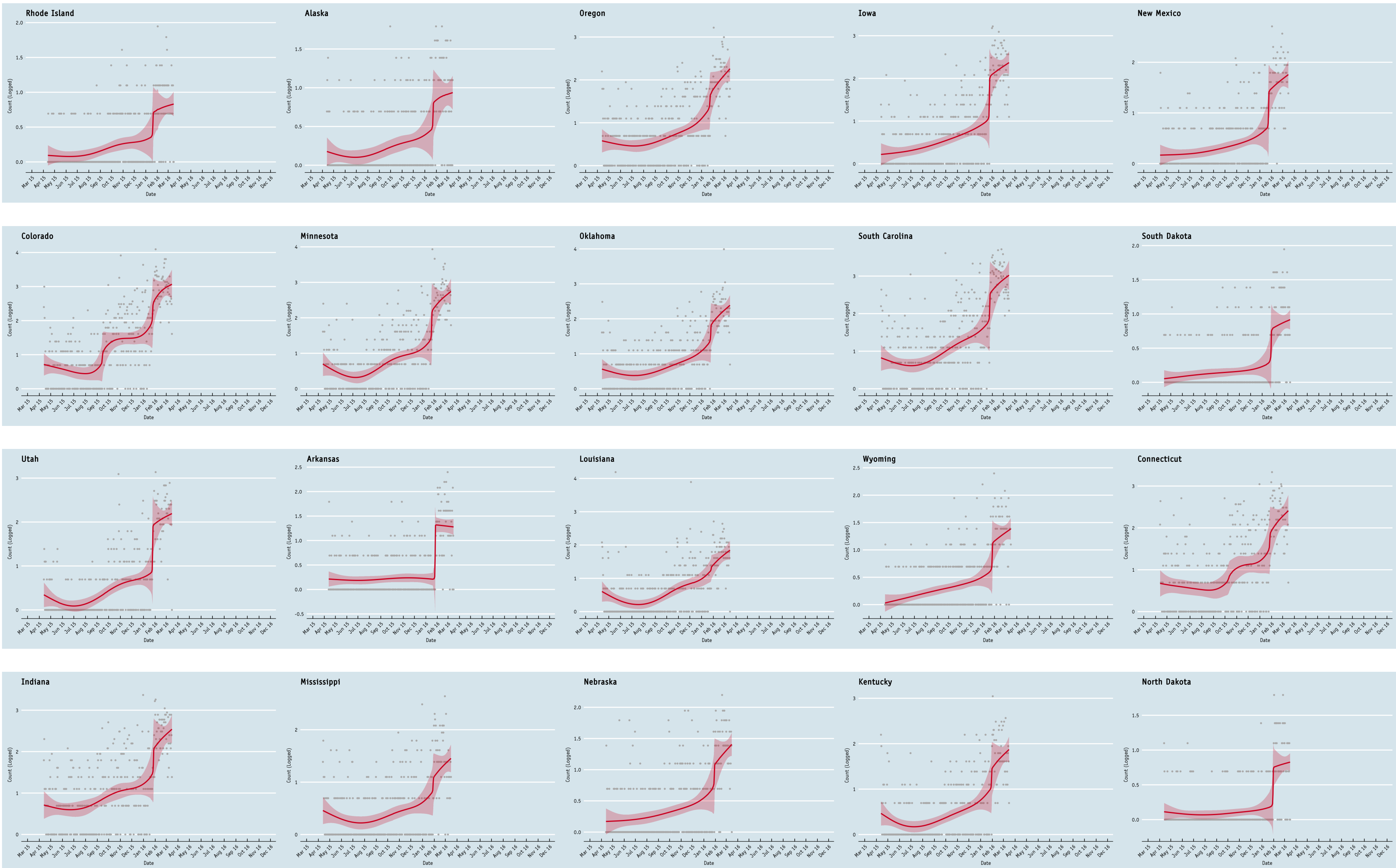
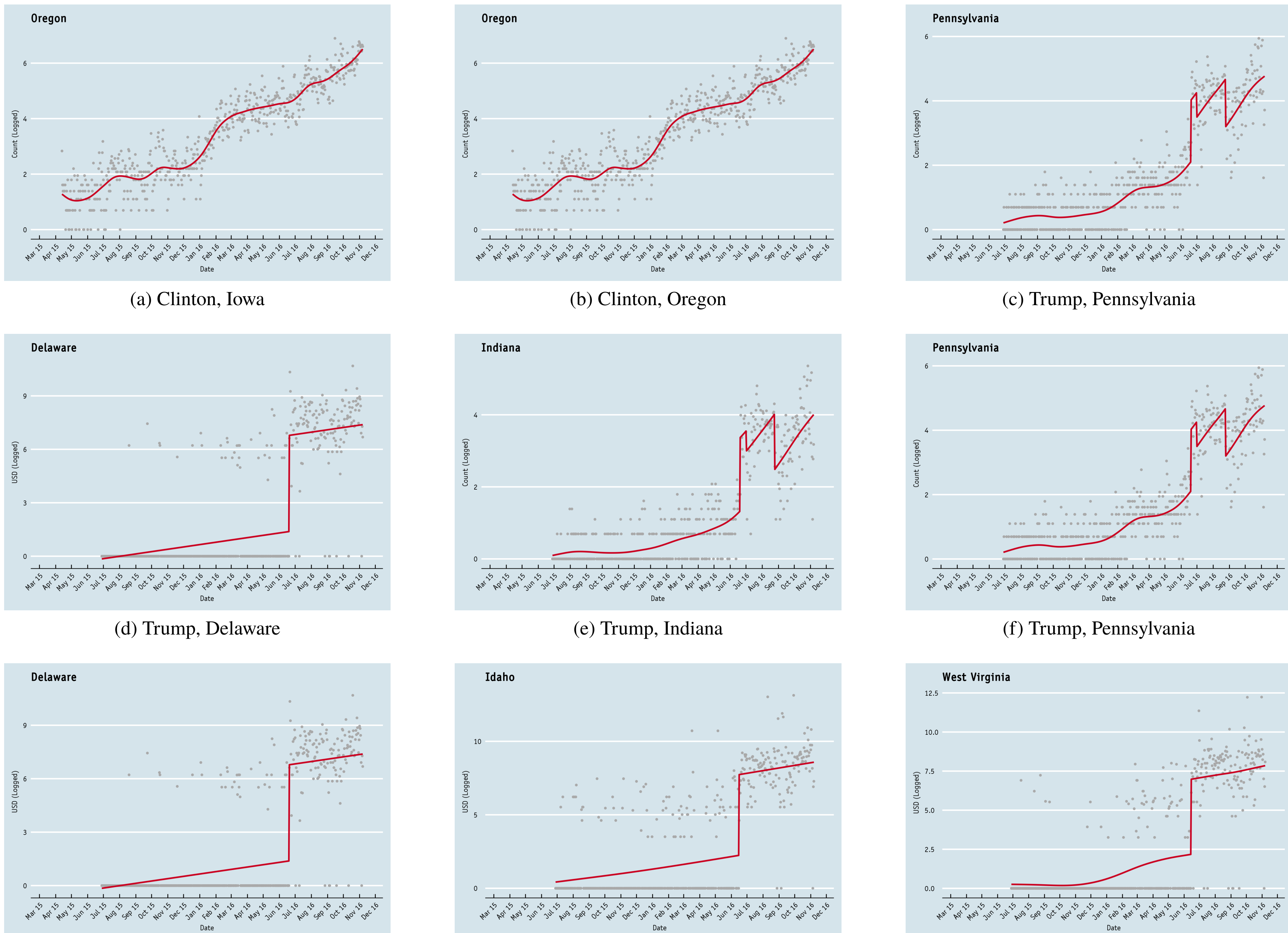


Figure C.10: Figure C.9, Bootstrapped



(a) Clinton, Iowa

(b) Clinton, Oregon

(c) Trump, Pennsylvania

(d) Trump, Delaware

(e) Trump, Indiana

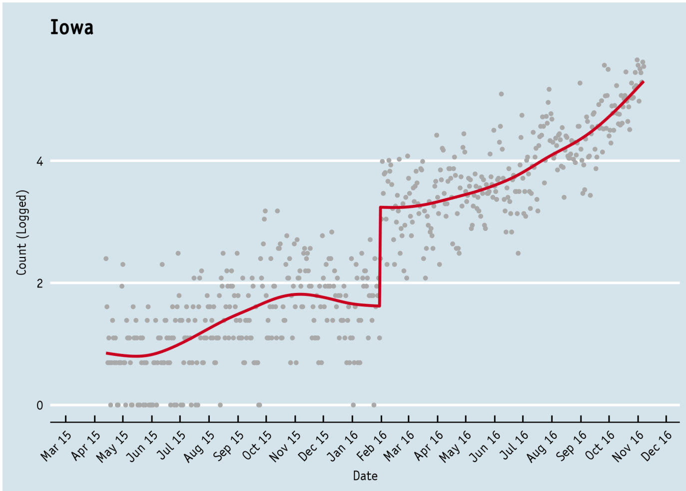
(f) Trump, Pennsylvania

(g) Trump, Delaware

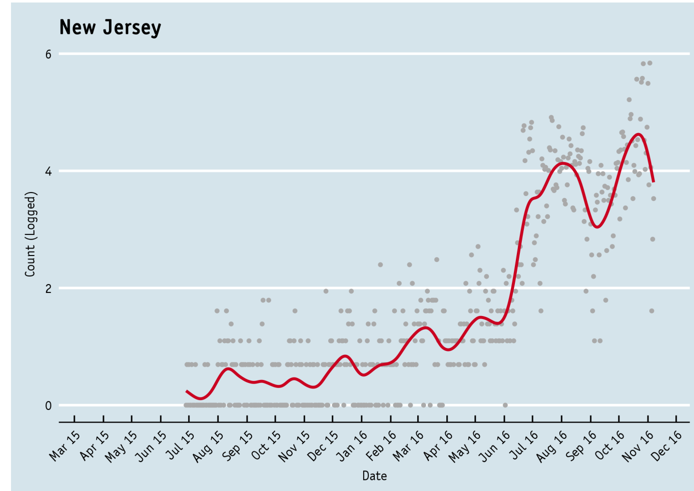
(h) Trump, Idaho

(i) Trump, West Virginia

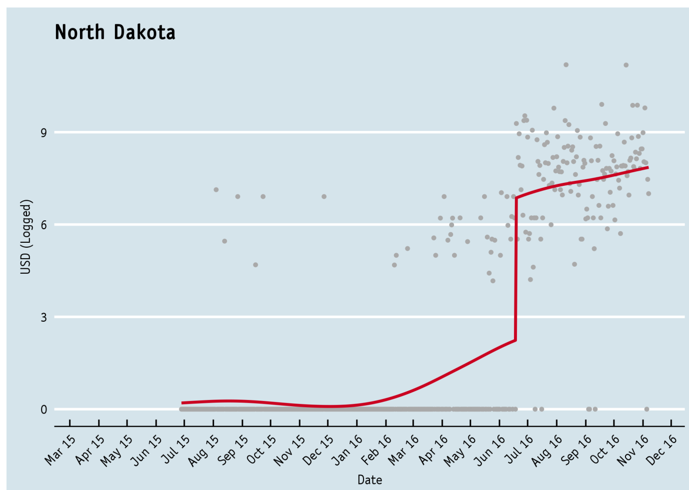
Figure C.11: Figure 3.8 When Jumps Not Modeled



(a) Clinton, Iowa (Feb 1, 2016)



(b) Trump, New Jersey (Jun 7, 2016)



(c) Trump, North Dakota (Apr 1, 2016)

Figure C.12: Figure 3.9 When Jumps Not Modeled