Essays in Econometrics and Political Economy

Thesis by
Sergio Montero

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

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
(Defended May 12, 2016)

© 2016
Sergio Montero
All Rights Reserved

ACKNOWLEDGEMENTS

To the members of my dissertation committee, Federico Echenique, Alexander Hirsch, and especially Matthew Shum and Erik Snowberg, I am deeply indebted for their continued support, wisdom, and encouragement, without which I would not have made it this far.

To Jonathan Katz, Jean-Laurent Rosenthal, and especially Benjamin Gillen, I am extremely grateful for lasting contributions to my way of doing research and my career.

To my classmates, I am thankful for their friendship and countless valuable comments, discussions, and suggestions.

To Laurel Auchampaugh, we (HSS graduate students) owe so, so much...

To Laura Flower Kim, I am grateful for managing to turn navigating the bewildering U.S. immigration system into an (at times) enjoyable experience.

To my parents, Sergio and Ana María, I owe all my gratitude for their love and all the amazing opportunities they were able to create for me.

Most importantly, to my wife, Saralicia, I am forever thankful for her unwavering love, support, and for making me a better person. This accomplishment is as much hers as it is mine.

ABSTRACT

This dissertation comprises three essays in Econometrics and Political Economy offering both methodological and substantive contributions to the study of electoral coalitions (Chapter 2), the effectiveness of campaign expenditures (Chapter 3), and the general practice of experimentation (Chapter 4).

Chapter 2 presents an empirical investigation of coalition formation in elections. Despite its prevalence in most democracies, there is little evidence documenting the impact of electoral coalition formation on election outcomes. To address this imbalance, I develop and estimate a structural model of electoral competition that enables me to conduct counterfactual analyses of election outcomes under alternative coalitional scenarios. The results uncover substantial equilibrium savings in campaign expenditures from coalition formation, as well as significant electoral gains benefitting electorally weaker partners.

Chapter 3, co-authored with Benjamin J. Gillen, Hyungsik Roger Moon, and Matthew Shum, proposes a novel data-driven approach to the problem of variable selection in econometric models of discrete choice estimated using aggregate data. Our approach applies penalized estimation algorithms imported from the machine learning literature along with confidence intervals that are robust to variable selection. We illustrate our approach with an application that explores the effect of campaign expenditures on candidate vote shares in data from Mexican elections.

Chapter 4, co-authored with Abhijit Banerjee, Sylvain Chassang, and Erik Snowberg, provides a decision-theoretic framework in which to study the question of optimal experiment design. We model experimenters as ambiguity-averse decision makers who trade off their own subjective expected payoff against that of an adversarial audience. We establish that ambiguity aversion is required for randomized controlled trials to be optimal. We also use this framework to shed light on the important practical questions of rerandomization and resampling.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

This dissertation comprises three essays addressing questions from several areas of Econometrics and Political Economy. The essays offer both methodological and substantive contributions to the study of electoral coalitions (Chapter 2), the effectiveness of campaign expenditures (Chapter 3), and the general practice of experimentation (Chapter 4).

Chapter 2 presents an empirical investigation of coalition formation in elections. Despite its prevalence in most democracies, there is little evidence documenting the impact of electoral coalition formation on election outcomes. To address this imbalance, I study coalition formation in the context of legislative elections where coordination among coalition partners takes the form of joint nominations across distinct consituencies—e.g., electoral districts. Specifically, I develop and estimate a structural model of electoral competition in which: (i) parties can form coalitions to coordinate their candidate nominations, and (ii) parties invest in campaign activities in support of their candidates. The model is estimated using data from the 2012 Mexican Chamber of Deputies election, which offers district-level variation in coalition formation. I conduct counterfactual experiments to study election outcomes under alternative coalitional scenarios. The results uncover substantial equilibrium savings in campaign expenditures from coalition formation, as well as significant electoral gains benefitting electorally weaker partners.

Chapter 3, co-authored with Benjamin J. Gillen, Hyungsik Roger Moon, and Matthew Shum, proposes a novel approach to the problem of variable selection in econometric models of discrete choice estimated using aggregate data. Economists often study consumers' aggregate behavior across markets choosing from a menu of differentiated products. In this analysis, local demographic characteristics can serve as controls for market-specific heterogeneity in product preferences. Given rich demographic data, implementing these models requires specifying which variables to include in the analysis, an ad hoc process typically guided primarily by a researcher's intuition. We propose a data-driven approach to estimate these models applying penalized estimation algorithms imported from the machine learning literature along with confidence intervals that are robust to variable selection. Our application explores the effect of campaign expenditures on candidate vote shares in data from Mexican elections, a central question in Political Economy with an answer that is often sensitive to the choice of controls.

Chapter 4, co-authored with Abhijit Banerjee, Sylvain Chassang, and Erik Snowberg, provides a decision-theoretic framework in which to study the question of optimal experiment design. We model experimenters as ambiguity-averse decision makers who trade off their own subjective expected payoff against that of an adversarial audience. We establish that ambiguity aversion is required for randomized controlled trials to be optimal. Moreover, the model matches other stylized facts about experimental practice: randomization occurs when the sample is large enough, and when the weight on the experimenter's own subjective payoff is small. We use this framework to shed light on the important practical questions of rerandomization and resampling. Rerandomization creates a trade-off between subjective balance and robustness; however, the costs of rerandomization are very small. We propose a simple rule of thumb for using rerandomization and resampling in practice.

Chapter 2

GOING IT ALONE? AN EMPIRICAL STUDY OF COALITION FORMATION IN ELECTIONS

2.1 Introduction

Electoral coalitions are common in most democracies (Golder, 2006). In hopes of influencing election outcomes, like-minded political parties often coordinate their electoral strategies, typically by fielding joint candidates for office. This manipulation of the electoral supply—i.e., the alternatives available to voters—may significantly affect representation and post-election policy choices. Despite its prevalence, however, there is little evidence documenting the impact of coalition formation on election outcomes.[1]

To address this imbalance, this paper studies coalition formation in the context of legislative elections where coordination among coalition partners takes the form of joint candidate nominations across distinct constituencies—e.g., electoral districts. Most electoral coalitions throughout the world arise in this context (Ferrara and Herron, 2005; Golder, 2006). Specifically, I develop and estimate a structural model of electoral competition in which: (i) parties can make coalition formation commitments, which determine the menu of candidates competing in each constituency, and (ii) parties invest in campaign activities in support of their candidates. The model is used to simulate election outcomes under counterfactual coalitional scenarios. The goal is to quantitatively assess the tradeoffs involved in coalition formation as well as how it affects parties' campaign expenditures, voter behavior, and the post-election distribution of legislative power. To my knowledge, this paper is the first to address these questions empirically.

The model is estimated using data from the 2012 Mexican Chamber of Deputies

[1] The existing literature on electoral (also called pre-electoral) coalitions/alliances has focused on comparing electoral systems in terms of their conduciveness to coalition formation, or on the role of electoral coalitions in shaping government formation in parliamentary democracies (e.g., Ferrara and Herron, 2005; Golder, 2006; Carroll and Cox, 2007; Bandyopadhyay et al., 2011). With the exception of Kaminski (2001), there is no systematic evidence available, beyond informal media accounts, of the electoral significance of electoral coalitions.

election, which is appealing for two reasons. First, the Mexican Chamber of Deputies follows a mixed electoral system whereby three fifths of the seats in the chamber are contested in winner-takes-all district races, and the remaining seats are assigned to registered parties in accordance with a national proportional representation rule. While the specifics differ across countries, most legislative elections follow either a pure winner-takes-all system, a pure proportional representation system, or a mixed system like Mexico (Bormann and Golder, 2013). From an institutional design perspective, studying coalition formation in a mixed electoral system such as Mexico's can help shed light on the separate roles of the winner-takes-all and proportional-representation components of the election in shaping coalition formation incentives and its consequences.

Second, while elections in most democracies usually offer a single observation of coalition formation, parties in Mexico are allowed to form partial coalitions in national legislative elections: i.e., coalition partners may nominate joint candidates in only a fraction of the contested races, while running independently elsewhere.[2] In particular, in the 2012 Chamber of Deputies election, two parties, the Institutional Revolutionary Party (*Partido Revolucionario Institucional*, PRI) and the Ecologist Green Party of Mexico (*Partido Verde Ecologista de México*, PVEM), formed a partial coalition, nominating joint coalition candidates in only two thirds of the national electoral districts. As a result, the election offers a sample of district races, otherwise identical in terms of the underlying electoral environment, where outcomes with and without coalition candidates can be observed. The structural model leverages this variation and focuses on capturing the incentives driving PRI and PVEM's choice of coalition configuration. To quantify the tradeoffs entailed by this choice as well as its impact on the election, I use the model to conduct counterfactual experiments examining two extreme cases: I simulate the election outcomes that would have prevailed had PRI and PVEM either not formed a coalition or formed a total coalition instead (nominating joint candidates in all districts).

The estimation strategy proceeds in three stages, exploiting insights from the empirical industrial organization literature on entry and competition in markets with differentiated products. First, voters' preferences are estimated from district-level voting data following the aggregate discrete choice approach to

[2]This is not unique to Mexico: France (Blais and Indridason, 2007) and India (Bandyopadhyay et al., 2011), for instance, permit similar arrangements.

demand estimation (e.g., Berry, 1994; Berry et al., 1995). Second, the parameters of parties' payoffs driving campaign spending decisions are estimated by fitting predicted to observed campaign spending levels.[3] Lastly, the remaining parameters of parties' payoffs relevant for coalition formation decisions are partially identified from moment inequalities analogous to market entry conditions.[4] I follow the two-step procedure of Shi and Shum (2015) for inference in this setting where only a subset of the model's parameters is partially identified via moment inequalities.[5]

Reduced-form evidence and the results of the counterfactual experiments document substantial electoral gains from coalition formation. In terms of jointly held seats in the Chamber of Deputies, PRI and PVEM's partial coalition allowed them to close the gap to obtaining a legislative majority by almost half; and they would have been able to close it by 71% had they run together in all districts. These gains, however, accrue at the expense of the electorally stronger partner, PRI, due to institutional features of the election detailed in the following section. Relative to not forming a coalition, PRI lost 6% of its seats by running with PVEM as observed in the data, and would have lost an additional 3% by forming a total coalition. Thus, the results reveal that the partial coalition arrangement constituted a compromise in balancing net gains to the coalition with PRI's losses.

With regard to campaign expenditures, the counterfactual experiments uncover significant efficiency gains from coalition formation. The ratio between PRI and PVEM's joint spending and joint vote share provides a rough estimate of how much the two parties need to spend—in equilibrium—to produce 1 percentage point of joint vote share. On average across districts, this ratio drops from about 2,000 USD when they run independently to about 1,750 USD when they run together, which implies cost savings of 12.5% from joint nominations.[6] Moreover, average spending across parties also drops in response

[3] Having estimated voters' preferences, the model's spending predictions are obtained as parties' best responses to their rivals' observed spending in each district.

[4] These entry conditions require computation of the set of campaign spending equilibria. At the estimated parameter values obtained from the first two stages, the campaign spending game played by parties exhibits (strict) strategic complementarities, facilitating computation of all equilibria (Echenique, 2007).

[5] See, e.g., Chernozhukov et al. (2007), Beresteanu et al. (2011), and Pakes et al. (2015) for more on estimation and inference in partially identified models.

[6] All monetary quantities in this paper have been converted from Mexican *pesos* to U.S. dollars.

to joint PRI-PVEM nominations. This is consistent with the intuition that differentiation via campaign advertising becomes relatively more valuable in a more crowded—and hence less polarized—field, leading parties to invest more heavily (see, e.g., Ashworth and Bueno de Mesquita, 2009).

The paper proceeds as follows. Section 2.2 describes the institutional background and provides a preliminary analysis of the data to inform the structural model. Section 2.3 introduces the model and empirical strategy. Section 2.4 summarizes the estimation results, and Section 2.5, the counterfactual experiments. Section 2.6 discusses the main findings and concludes.

2.2 Mexican Elections: Background and Data

Mexico is a federal republic with 31 states and the capital, Mexico City. The executive branch of the federal government is headed by the president, and legislative power is vested in a bicameral Congress. Federal elections are held every 6 years to elect a new president and new members of both chambers of Congress. No incumbent can stand for consecutive re-election.[7] The lower chamber, the Chamber of Deputies, is further renewed following midterm federal elections in the third year of every presidential term.

For electoral purposes, Mexico is divided into 300 districts.[8] The Chamber of Deputies has 500 total members, 300 of whom directly represent a district after being elected by direct ballot under simple plurality voting. The remaining 200 seats in the chamber are assigned to the national political parties in accordance with a proportional representation (PR) rule: the votes cast across the 300 district races are pooled nationally, and each party is given a share of the 200 seats proportional to the share of votes received by the party's candidates in the district races.[9] Disproportionality restrictions preclude any party from obtaining more than 300 total seats or a share of seats that exceeds by more

[7] Presidential re-election is prohibited. Legislators can be re-elected in non-consecutive terms.

[8] The current district lines were drawn in 2005 by the national electoral authority with the objective of equalizing population while preserving state boundaries and ensuring each state a minimum of two districts.

[9] The division of seats follows the largest remainder method using Hare quotas (Bormann and Golder, 2013). Only parties that secure at least 2% of the national vote are eligible to hold seats in the legislature; otherwise, they lose their registration and their votes are annulled.

than 8 percentage points the party's national vote share, in which case the excess PR seats are divided proportionally among the remaining parties.[10]

In addition to the composition of the legislature, at stake in each Chamber of Deputies election is registered parties' funding for the subsequent three years. By law, Mexican parties are primarily funded from the federal budget. The baseline amount to be distributed yearly to the parties equals 65% of Mexico City's legal daily minimum wage multiplied by the number of registered voters in the country. In 2012, for example, this totaled about 250 million USD. For campaign purposes, an additional 50% of the year's baseline is provided to the parties in presidential election years, while 30% is provided in midterm election years. The final amount is distributed as follows: 30% is divided equally among all registered parties, and the remaining 70% is divided in proportion to their national vote shares in the most recent Chamber of Deputies election. To ensure the primacy of public funding, funds from outside sources such as member fees or private contributions are capped at 2% of the year's public funding. Thus, Mexican parties compete in this election to secure not only seats in the legislature but also their funding for day-to-day operations and campaign activities for the following three years.

Prior to each Chamber of Deputies election, parties are allowed to form coalitions, which enable them to coordinate their candidate nominations for the direct representation (DR) district races. Coalition partners may not, however, coordinate on the PR component of the election: national lists of up to 200 candidates for the PR seats in the chamber must be submitted independently by each party.

Coalition agreements are negotiated by the parties' national leadership and must be publicly registered before the national electoral authority, the National Electoral Institute (*Instituto Nacional Electoral*, INE). The agreements constitute binding commitments specifying, for each electoral district: (i) whether the coalition partners will nominate a joint candidate or independent candidates, and (ii) in the case of a joint nomination, from which party's ranks will the coalition candidate be drawn.[11] After the election, coalition agreements

[10]The adjustment is carried out only once: if a party exceeds the 8-percentage-points restriction after an initial round of adjustment, the process does not iterate.

[11]In 2012, prospective coalition partners formally had to choose from two available formats: they could either form a *partial* coalition, enabling them to nominate joint candidates in at most 200 districts, or they could from a *total* coalition, requiring them to nominate a

imply no formal obligations for coalition victors in the legislature, who retain their original party affiliation. Thus, by supporting a partner's candidate via a joint nomination, the remaining coalition partners forgo the corresponding district seat in the chamber.

A model of coalition formation in this environment must, therefore, capture this key feature of the decision problem faced by party leaders: by running independently in a district, coalition partners risk splitting the vote and losing the district to a rival party, but a joint nomination entails an agreement about which partners should stand down altogether. Moreover, while coalition partners may not coordinate on the PR component of the election, the decision of where to run together and independently may affect their national vote shares and, consequently, their PR performance and future funding.

When deciding whether to vote for a coalition candidate, Mexican voters in fact have some control over how their vote should be counted for PR (and funding) purposes. The ballots presented to the voters on election day feature one box per registered party containing the name of the party's candidate for that district.[12] If a candidate is nominated by a coalition, their name appears inside each of the coalition partners' boxes. To cast their vote in favor of a coalition candidate, voters can mark any subset of the coalition's boxes on the ballot. Regardless of the chosen subset, the vote is counted as a single vote in favor of the coalition candidate for the purpose of selecting that district's DR deputy. However, the vote is split equally among the chosen subset for PR purposes. For example, a citizen who wishes to vote for a candidate nominated by parties A, B, and C could mark all three boxes: while the candidate would receive 1 vote for the district seat, each party would get a third of her vote for PR purposes. The voter alternatively could mark A and B's boxes, in which case A and B would each get 50% of her vote but C would get zero. Or she could just mark party A's box giving A 100% of the vote. This feature of the Mexican Chamber of Deputies election contrasts with other PR systems where coalition partners are allowed to submit joint lists of PR candidates. In such

joint candidate in every district. The PRI-PVEM coalition that is the focus of this paper was not in practice bound by the size constraint on partial coalitions, and the counterfactual experiments of Section 2.5 consider only the no-coalition and total-coalition extremes, so this constraint is ignored in what follows.

[12] Independent candidacies or write-in campaigns are also allowed, but their vote shares are negligible. Moreover, voters supporting independent or write-in candidates forgo participation in the PR component of the election.

systems, votes in favor of a coalition are simply aggregated, and the number of PR seats each partner gets is determined by the composition of their joint list (i.e., the ranking of candidates), over which the partners bargain prior to the election. In Mexico, voters have more direct control over the PR component of the election.

Another key component of party leaders' decision problem concerns campaign expenditures. Due to the constraints on parties' outside funding, fundraising by candidates is effectively absent from the Chamber of Deputies election. As a result, the parties' national leadership directly fund their candidates' campaigns, making a centralized decision of how much to spend in each district. In the case of coalition candidates, partners may share campaign costs freely. Thus, pooling resources in favor of a joint candidate, instead of campaigning against each other, may also encourage coalition formation.

Events in an election year unfold as follows. First, as described, coalitions are publicly registered. Next, candidates are selected and nominated accordingly. Campaigns then take place within a fixed timeframe.[13] And, finally, ballots are cast.

2.2.1 The 2012 Election: Preliminary Data Analysis

In the 2012 Chamber of Deputies election, 2 parties, the National Action Party (*Partido Acción Nacional*, PAN) and the New Alliance Party (*Partido Nueva Alianza*, NA), participated independently; 3 parties, the Party of the Democratic Revolution (*Partido de la Revolución Democrática*, PRD), the Labor Party (*Partido del Trabajo*, PT), and the Citizens' Movement (*Movimiento Ciudadano*, MC), formed a total coalition called the Progressive Movement (*Movimiento Progresista*, MP); and PRI and PVEM formed a partial coalition called Commitment for Mexico (*Compromiso por México*, CM), joining forces in 199 of the 300 electoral districts. Of the 199 jointly contested districts, PRI and PVEM jointly nominated a PRI candidate in 156 districts and a PVEM candidate in the remaining 43 districts (see Figure 2.1).

As shown in Figure 2.2, which is based on a nationally representative poll of ideological identification conducted by a leading public opinion consultancy in 2012, the parties can be roughly placed on a one-dimensional ideology spec-

[13] Campaigns must end 3 days before election day, and they may last up to 90 days in presidential election years and up to 60 days in midterm election years.

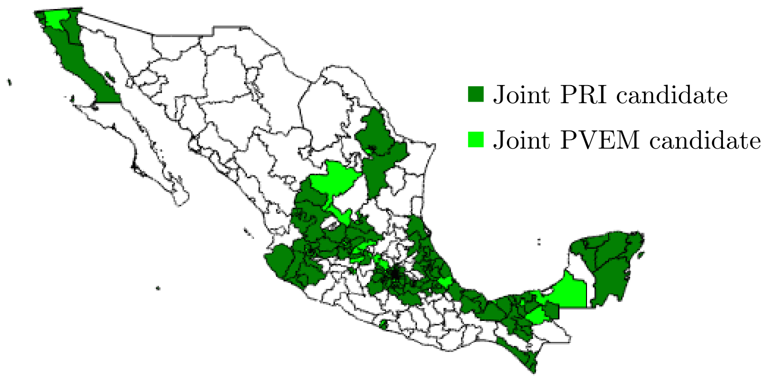
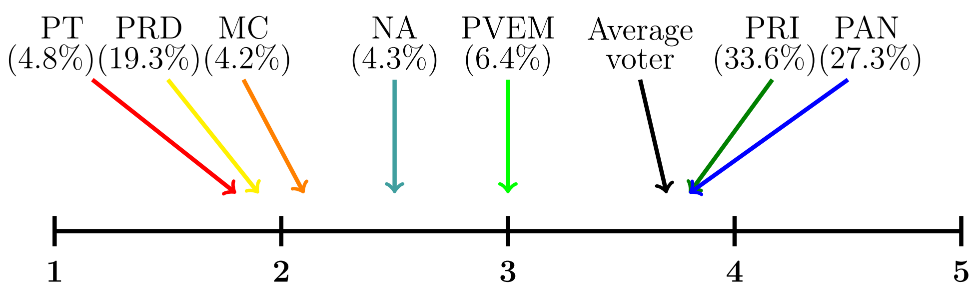


Figure 2.1: Districts with joint PRI-PVEM candidates

trum as follows; from left to right: the MP parties, NA, PVEM, PRI, and PAN. Figure 2.2 also presents the parties' national vote shares in the 2012 election to illustrate their relative strengths. PRI, PAN, and PRD are the main political forces, in that order; together they account for more than 80% of votes nationally. Of the smaller parties, the centrist PVEM is the strongest, with nearly a third of PRD's vote share. The shares in Figure 2.2, however, were shaped by the coalitions that formed prior to the election. The main objective of this paper is to quantify this effect: specifically, conditional on PAN and NA running independently and the MP parties forming a total coalition as observed, how would the election outcomes have changed had PRI and PVEM either not formed a coalition or formed a total coalition instead?



Source: Consulta Mitofsky (2012). One thousand registered voters were asked in December 2012 to place the parties and themselves on a five-point, left-right ideology scale. Arrows point to national averages. Parties' vote shares in parentheses.

Figure 2.2: Left-right ideological identification of Mexican parties and voters

District-level election outcomes are published by INE. The data include vote totals by distinct alternative available to voters in each district; that is, in districts with coalition candidates, vote totals for all subsets of the coalitions (as explained above) are available.

As a coalition, PRI and PVEM were very successful, winning 122 of the 199 districts they shared: 103 victories with a joint PRI candidate (out of 156 districts) and 16 victories with a joint PVEM candidate (out of 43). Independently, PRI obtained 52 additional victories, and PVEM obtained 3. The final composition of the Chamber of Deputies, including the PR seats, is presented in Table 2.1 (hereafter, I treat the total coalition MP as a single party). PRI's proportionally smaller share of the PR seats is a consequence of the restriction mentioned in Section 2.2 that a party's total share of seats cannot exceed by more than 8 percentage points its national vote share.[14] Without this constraint, PRI would have obtained 67 PR seats instead of 49.

Table 2.1: Chamber of Deputies composition after 2012 election

Party	Direct representation seats	Proportional representation seats	Total
PRI	158	49	207
PVEM	19	15	34
PAN	52	62	114
MP	71	64	135
NA		10	10
Total	300	200	500

Table 2.2 shows a breakdown of election outcomes by type of candidate ran by PRI and PVEM. Victory rates (percentage of districts won) and average vote shares are computed for each party. Table 2.2 suggests that, in terms of vote share, PVEM benefitted significantly from a joint nomination at the expense of PRI, with both parties doing better with a joint candidate drawn from their own ranks. In particular, PVEM benefitted from coalition supporters splitting their vote between the two parties—see Table 2.3—an important feature of the election captured by the model developed in Section 2.3. With respect to victory rates, joint PRI candidates were the most successful. The clear loser from a joint PRI-PVEM nomination appears to have been NA, while

[14] See Section 2.5 for details.

MP and PAN exhibit mixed effects. These crude comparisons, however, do not account for PRI and PVEM's strategic choice of where and how to run together, influenced by differences in the electorate across districts and in the parties' campaign strategies.

Table 2.2: Election outcomes by PRI-PVEM coalition configuration

Party	Districts with distinct PRI, PVEM candidates		Districts with joint PRI candidate		Districts with joint PVEM candidate	
	Victory rate (%)	Avg. vote share (%)	Victory rate (%)	Avg. vote share (%)	Victory rate (%)	Avg. vote share (%)
PRI	51.5	36.7	66.0	33.2	-	28.7
PVEM	3.0	4.9	-	7.0	37.2	7.7
PAN	22.8	27.6	10.9	26.4	27.9	28.4
MP	22.8	25.5	21.2	29.4	34.9	31.4
NA	0	5.3	0	3.9	0	3.8

Table 2.3: Votes in support of PRI-PVEM coalition candidates

Type of vote	Districts with joint PRI candidate	Districts with joint PVEM candidate
	Avg. vote share (%)	Avg. vote share (%)
PRI	30.0	25.7
PVEM	3.8	4.6
50-50 split	6.4	6.1

The first two rows represent voters who gave 100% of their vote to the corresponding party (see Section 2.2). Thus, adding half of the third row to the other two yields the parties' final vote shares as shown in Table 2.2.

To keep the structural model presented below as parsimonious as possible, I use only 4 broad demographics to describe the electorate: gender, age, education, and income.[15] The data are taken from the 2010 population census, which the

[15]Section 2.5.1 discusses a richer specification.

National Statistics and Geography Institute (*Instituto Nacional de Estadística y Geografía*, INEGI) makes available on a geo-electoral scale. For gender, as an indicator of the importance of women in the electorate, I use the percentage of households with a female head. Age is captured by the percentage of the voting age population aged 65 and older, and education is measured by average years of schooling (among population 15 and older). Income is not available in census data; as a proxy, I use the percentage of households that own an automobile. Table 2.4 provides a summary description of the electoral districts by type of candidate ran by PRI and PVEM, as in Table 2.2.

Table 2.4: District characteristics

	Districts with distinct PRI, PVEM candidates		Districts with joint PRI candidate		Districts with joint PVEM candidate	
Variable	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Female head of household (% of total)	23.8	3.1	24.7	4.3	26.8	5.1
Pop. over 64 (% of over 17)	10.6	2.6	9.5	2.8	10.1	2.5
Avg. years of schooling (pop. over 14)	7.8	1.3	8.3	1.5	9.1	1.6
Household owns a car (% of total)	45.3	17.7	41.0	14.5	47.9	14.0

Finally, Table 2.5 summarizes campaign spending in the district races—i.e., total expenditure in support of a candidate—by type of candidate ran by PRI and PVEM. The data can be requested directly from IFE.[16] While it would be preferable to obtain a detailed account of campaign activities (e.g., town hall meetings, media advertising, billboards, etc.), as well as information about the content of campaign advertising, the available data provide only a coarse description of monetary expenses. Consequently, I focus on total spending per candidate as a broad measure of the intensity of campaign efforts. A key

[16]Campaign spending data are self-reported by the parties to the electoral authority. These reports are subject to audits by IFE. However, audited data after 2006 are not yet available. For comparison, campaign expenditures were *over*reported by about 4% in 2006, while no discrepancies were found in 2003. I therefore ignore potential measurement error in the data and rely on the unaudited reports.

feature of the model presented in the following section is that parties' make strategic spending decisions on a district-by-district basis (as opposed to simply dividing up resources by state or regionally). To evaluate this assumption, Figure A.1 in Appendix A.3 maps each party's geographic distribution of campaign spending. As expected, there is substantial variation across neighboring districts, beyond anything that could be driven solely by differences in campaign costs considering that any relatively high-spending district for one party is a relatively low-spending district for another (and vice versa).[17]

Table 2.5: Campaign spending (in thousands of USD)

	Districts with distinct PRI, PVEM candidates		Districts with joint PRI candidate		Districts with joint PVEM candidate	
Party	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
PRI	54.9	11.0	80.6	27.3	94.3	40.9
PVEM	18.3	7.6				
PAN	38.0	10.4	41.4	12.7	44.6	14.2
MP	56.4	19.7	55.1	11.7	56.6	14.3
NA	19.7	8.5	16.7	4.4	19.1	8.5

2.3 Model and Empirical Strategy

Recall from Section 2.2 that the timing of events in the Mexican Chamber of Deputies election is as follows. First, parties make their coalition formation commitments. Conditional on these agreements, candidates for the district races are selected and registered, along with candidate lists for the PR assignment of seats. Campaigns then take place, and finally ballots are cast. The model I develop to examine the consequences of PRI and PVEM's partial coalition captures this timing in three stages: a coalition formation stage, a campaign stage, and a voting stage.

As mentioned previously, the analysis that follows focuses on the PRI-PVEM coalition while conditioning on all other parties running as observed in the data. The implicit assumption is that modifying PRI and PVEM's coalition

[17]In contrast, spending variation driven solely by cost differences would affect parties symmetrically; e.g., if campaigning is comparatively cheaper in a district, then all parties would be expected to spend relatively little there.

configuration wouldn't have affected the other parties' coalition formation decisions.[18] Accordingly, the coalition formation stage of the model is concerned only with PRI and PVEM's choice of where and how to run together. While it might be intriguing to consider, for example, a breakup of the MP coalition, or the formation of coalitions not in the data, this would require making strong assumptions regarding voting behavior. Only voting choices from three menus are observed: all three include an MP candidate, an NA candidate, and a PAN candidate, and they vary only with respect to PRI and PVEM's coalition configuration. Allowing for unrestricted coalition formation would involve predicting voting choices from menus of candidates not in the data. Rather than placing additional ad hoc structure on the model to accomplish this task, I leverage the information directly available in the data. The objective is thus to understand PRI and PVEM's strategic choice of coalition configuration and its effect on election outcomes.

I describe the model backwards from the voting stage. Before introducing the model, I develop some useful notation.

Notation. Districts are indexed by d, parties by p, and voters by i. The indicator $M_d \in \{\text{PRI}, \text{PVEM}, \text{IND}\}$ describes the menu of candidates available to voters in district d as a result of PRI and PVEM's coalition configuration: $M_d = \text{PRI}$ indicates that PRI and PVEM jointly nominate a PRI candidate in district d, $M_d = \text{PVEM}$ indicates that they jointly nominate a PVEM candidate, and $M_d = \text{IND}$ indicates that they nominate distinct candidates and thus run independently.

2.3.1 Model

Voting stage. There are two tactics available to parties by which they can hope to influence election outcomes. One is by manipulating the electoral supply via coalition formation. The other is through campaign advertising. To study their effectiveness, voters' preferences are modeled as menu-dependent and susceptible of persuasion.

Recall that, by casting their ballot, voters simultaneously select a candidate and a party list. If a candidate is nominated by a coalition, voters can de-

[18] Without this assumption, the results of Section 2.5 may nevertheless be interpreted as *partial* effects of the PRI-PVEM coalition on election outcomes.

cide how to split their vote among the nominating parties' lists (see Section 2.2). However, the selection of a candidate is the preeminent choice. I therefore model voting choices as a two-tier decision: voters first select a candidate and then, if necessary, how to split their vote. I describe the two tiers in turn.

When choosing a candidate, voters care about both the nominating party or coalition's policy platform and the candidate's quality (or valence). The policy platform summarizes the legislative objectives that the party or coalition hopes to achieve and that the candidate is expected to support if elected.[19] Quality, on the other hand, refers to characteristics of the candidate that all voters in the district may find appealing, such as charisma, intelligence, or competence (Groseclose, 2001); it may be interpreted as the candidate's ability to represent the district's interests in legislative bargaining. Lastly, voters may also care about the intensity of campaign efforts in support of a candidate; i.e., they can be persuaded by campaign advertising. Formally, voters' preferences take the following form: if the menu of candidates available to voters in district d is $M_d = m \in \{\text{PRI}, \text{PVEM}, \text{IND}\}$, voter i's utility from voting for candidate $j \in m$ is

$$\begin{aligned} u_{ijd}^m = \; & \underbrace{\alpha_1 c_{jd} + \alpha_2 c_{jd}^2}_{\text{effect of campaign spending}} + \underbrace{x_d' \beta_j^m}_{\text{policy platform}} \\ & + \underbrace{\xi_{jd}^m}_{\text{candidate quality}} + \underbrace{\epsilon_{ijd}^m}_{\text{random utility shock}} , \end{aligned} \tag{2.1}$$

where c_{jd} denotes campaign spending in support of candidate j, x_d is a vector of district demographics, ξ_{jd}^m measures candidate quality, and ϵ_{ijd}^m is a random utility shock that is independent of the other components of i's utility and captures individual heterogeneity.

While parties are required by law to announce their policy platforms prior to the election, data on individual or district-level policy preferences are unavailable. Policy platforms are nevertheless allowed to influence local voting preferences by means of interactions between district demographics and menu-dependent party or coalition fixed effects. Thus, the term $x_d' \beta_j^m$ measures the relative appeal—with respect to other available candidates—of j's platform for the electorate of district d. In contrast, the coefficients α_1 and α_2 are fixed

[19]Data on candidates' individual policy positions are unavailable.

across candidates and menus. The underlying assumption is that all parties potentially have access to the same campaigning technology but spend varying amounts of effort—i.e., money—trying to persuade voters. The quadratic term $\alpha_2 c_{jd}^2$ is introduced to capture diminishing marginal returns to spending. Having common coefficients, however, does not imply a constant—across candidates, menus, or districts—marginal effect of campaign spending on vote shares (see (2.7) below). The effectiveness of a party's spending depends on all other components of voters' utilities.

In the style of probabilistic voting models with aggregate popularity shocks (see, for example, Persson and Tabellini, 2000, chap. 3), the random utility term is assumed to have the following structure:

$$\epsilon_{ijd}^m = \eta_j^m + e_{ijd}^m, \tag{2.2}$$

where η_j^m and e_{ijd}^m are independently distributed with a zero-mean, Type-I Extreme Value distribution. Thus,

$$\delta_{jd}^m = \alpha_1 c_{jd} + \alpha_2 c_{jd}^2 + x_d' \beta_j^m + \xi_{jd}^m + \eta_j^m \tag{2.3}$$

represents mean voter utility from voting for candidate j in district d. The term η_j^m is an aggregate—national—popularity shock. It can be viewed as a random component of voters' tastes for j's policy platform.

In every menu, voters additionally have available a compound outside option, $j = 0$, of either not voting, casting a null vote, or writing in the name of an unregistered candidate. As is standard in discrete-choice models, the mean utility of this outside option is normalized to zero: $\delta_{0d}^m = 0$. This normalization is without loss of generality for within-menu choices. Imposing a common normalization across menus provides a shared baseline against which to interpret the menu-dependent coefficients. How voters respond to changes in the electoral supply can thus be inferred directly from a comparison of coefficients across menus.

To complete the specification of the first tier of the voting stage, voters are assumed to behave expressively or sincerely, i.e., they choose the alternative they prefer the most without any strategic considerations. This defines a homogeneous logit model of demand in the spirit of Berry et al. (1995).[20] While

[20] In related work, Gillen et al. (2015) explore alternative formulations of vot-

accounting for strategic voting behavior is beyond the scope of this paper, the reward structure for parties following the Chamber of Deputies election (specifically, the proportional allocation of seats and future funding) arguably encourages sincere voting and warrants this assumption.[21] Indeed, Ferrara (2006) argues that this is to be expected in a single-ballot mixed system with national lists such as Mexico's for the election of deputies. Nevertheless, the menu-dependent structure of voters' preferences implicitly allows for potentially strategic responses to changes in the electoral supply.

Finally, the second tier of the voting stage takes a similar form: if the menu available to voters in district d contains a PRI-PVEM coalition candidate, i.e., $M_d = m \neq \text{IND}$, then voters can decide whether to split their vote 50-50 between the coalition partners or give 100% of their vote to one of them, where voter i's utility from choosing alternative j out of these three options is

$$u_{ijd}^{\text{ST},m} = x_d' \beta_j^{\text{ST},m} + \xi_{jd}^{\text{ST},m} + \epsilon_{ijd}^{\text{ST},m}. \tag{2.4}$$

Here, $\beta_j^{\text{ST},m}$ and $\xi_{jd}^{\text{ST},m}$ are the analogs of β_j^m and ξ_{jd}^m from the first tier, respectively, and $\epsilon_{ijd}^{\text{ST},m}$ has the same structure as in (2.2) above.[22] The only difference between the two tiers is that the second tier is unaffected by campaign spending. Campaigns are candidate-centric and, as such, are assumed to affect only the first-tier candidate choice.

Campaign stage. This stage follows the coalition formation stage and corresponding candidate nominations. The objective for parties is to decide how much to spend in support of their registered candidates. Given $M_d = m$, determined in the coalition formation stage, the candidate quality terms ξ_{jd}^m are commonly observed by all parties (but unobserved by the researcher). Hence, parties can tailor their spending to their candidates' relative strengths. The voters' random utility shocks, however, are unknown to the parties (and the researcher); only their distribution is known.

ing preferences—including the workhorse random coefficients logit model of demand estimation—using the same data from the 2012 Mexican Chamber of Deputies election. As discussed there, the results of Section 2.4 are fairly robust to different specifications and to variable selection of the demographic controls in x_d.

[21] See Kawai and Watanabe (2013) for a recent example of the challenges involved in identifying strategic voting.

[22] That is: $\epsilon_{ijd}^{\text{ST},m} = \eta_j^{\text{ST},m} + e_{ijd}^{\text{ST},m}$, independently distributed with a zero-mean, Type-I Extreme Value distribution.

As discussed in Section 2.2, parties care not only about winning district races but also about their final vote share, as it determines the number of PR seats they shall receive and also their future funding. Thus, when deciding how much to spend in each district, parties care about both their probability of winning the district seat and their expected vote share in the district. For analytical convenience, I assume that parties have a flexible national budget constraint. In particular, parties are assumed to make independent spending decisions across districts. The parties' payoff structure described below ensures that the spending levels predicted by the model conform to the levels observed in the data. But, rather than imposing a hard national budget constraint which would significantly complicate the analysis that follows, the model allows certain flexibility with respect to the parties' total spending under alternative scenarios. This assumption is not unreasonable, particularly for the 2012 election, which coincided with the senate and presidential elections. Indeed, parties are free to transfer resources between the elections. While the senate and presidential contests are outside the scope of this paper, any opportunity costs of such transfers are implicitly captured by the payoff structure described below.

The campaign stage therefore consists of parties playing an independent campaign spending game in each district (with complete information and simultaneous moves). The parties' payoffs are defined as follows. Given $M_d = m$, if party p enters a candidate in district d, its payoff is

$$\pi_{pd}^{m} = \theta_p^{\mathrm{PW}} \log\left(PW_{pd}^{m}\right) + \theta_p^{\mathrm{ES}} \log\left(ES_{pd}^{m}\right) - c_{pd}, \tag{2.5}$$

where PW_{pd}^{m} and ES_{pd}^{m} denote, respectively, party p's probability of winning and expected vote share in the district (derivations of which can be found in Appendix A.1), and c_{pd} is p's spending in support of its candidate. Thus, the coefficients θ_p^{PW} and θ_p^{ES} measure the monetary value of (the log of) PW_{pd}^{m} and ES_{pd}^{m}. This value represents—in relative terms to each party's available resources—not only the benefits derived from the election outcomes but also any opportunity costs of c_{pd} as discussed above.

For the PRI-PVEM coalition partners, if $p \in \{\mathrm{PRI}, \mathrm{PVEM}\}$ doesn't enter a candidate in district d, i.e., if $m \notin \{p, \mathrm{IND}\}$, then p's payoff is

$$\pi_{pd}^{m} = \theta_p^{\mathrm{NC}} + \theta_p^{\mathrm{ES}} \log\left(ES_{pd}^{m}\right) - c_{pd}. \tag{2.6}$$

In this case, the coefficient θ_p^{NC} measures the value of not fielding a candidate to support one's partner's candidate instead.[23]

When PRI and PVEM nominate a joint candidate, i.e., $M_d = m \neq \mathrm{IND}$, they must jointly decide how much to spend to support her. Only their joint spending $c_{\mathrm{PRI},d} + c_{\mathrm{PVEM},d}$ enters (2.1) and determines the candidate's probability of winning and their expected vote shares. Given the quasilinear structure of payoffs, I remain agnostic about how PRI and PVEM divide this amount between them and simply assume that it maximizes their joint surplus $\pi^m_{\mathrm{PRI},d} + \pi^m_{\mathrm{PVEM},d}$. In other words, joint spending is assumed to be Pareto optimal for the coalition. Hence, in districts where $M_d = m \neq \mathrm{IND}$, PRI and PVEM act as a single player in the spending game against other parties, who chooses $c_{\mathrm{PRI},d} + c_{\mathrm{PVEM},d}$ with joint payoff $\pi^m_{\mathrm{PRI},d} + \pi^m_{\mathrm{PVEM},d}$.

At the estimated parameter values of the voting stage and the parties' payoffs, and regardless of the menu of candidates, the resulting campaign spending game played in each district exhibits *strict* strategic complementarities (see Appendix A.1 for details). A formal definition of this class of games can be found in Echenique and Edlin (2004). It suffices here to point out three key properties of such games. First, existence of equilibrium is guaranteed (Vives, 1990). Second, mixed-strategy equilibria are not good predictions in these games, so their omission is justified (Echenique and Edlin, 2004). Third, the set of all pure-strategy equilibria can be feasibly computed (Echenique, 2007). This implies that full consideration of potential multiplicity of equilibria is feasible. At the estimated parameter values, however, the campaign spending games exhibit unique equilibria. Therefore, for ease of exposition, I proceed with the description of the model and the empirical strategy under the presumption that the spending game in each district has a unique equilibrium.[24]

Coalition formation stage. This stage completes the description of the model. As stated, the objective is to understand PRI and PVEM's choice of where and how to run together, conditional on all other parties running as observed in the data.

[23]This formulation allows partners to potentially prefer a joint nomination over having negligible chances of winning a district—by avoiding any fixed administrative or operational costs of candidate nominations.

[24]A note providing guidance on how to deal with multiplicity of equilibria is available on request from the author.

Recall that coalition formation decisions precede candidate nominations. In particular, I assume that, when PRI and PVEM choose their coalition configuration, they don't yet know the candidate qualities $(\xi_j^m)_{j,m}$, only their distribution.[25] This assumption is justified by the following observations. First, while party leaders may have some information regarding potential candidates, internal candidate selection processes are inherently random and thus difficult to anticipate precisely. Parties use a combination of procedures to select their candidates—most notably, primaries and appointments by local committees—which are beyond the direct control of the national leadership. Second, candidates for the district races are relatively inexperienced compared to candidates in the parties' PR lists. Career politicians with substantial influence within their party and national exposure generally don't seek district seats; they rather attempt to secure a favorable position on their party's PR list, which virtually guarantees them participation in the legislature. Indeed, this concern has fueled recent calls for reducing the number of PR legislators.[26]

Similarly to joint spending decisions, I therefore assume that M_d is chosen to maximize PRI and PVEM's ex-ante expected joint surplus—i.e., before candidate qualities are realized. Formally, $M_d \in \arg\max_m \mathbb{E}(\pi^m_{\text{PRI},d} + \pi^m_{\text{PVEM},d} | x_d)$. The expectation here is taken with respect to the campaign spending equilibrium and election outcomes induced by the candidate qualities.[27] This resembles nonnegative-profit entry conditions in models of market entry. A significant difference, however, is that the entry decision here conditions only on the observable district characteristics, x_d, not on the yet unknown product quality, ξ_j^m.

2.3.2 Empirical Strategy

Estimation of the model mirrors its three-stage structure. Step 1 deals with the voting stage and recovers the parameters of (2.1) and (2.4). Step 2 obtains the coefficients θ_p^{PW} and θ_p^{ES} of the parties' payoffs by matching the spending levels observed in the data with the model's predictions from the campaign stage.

[25] The ξ_j^m are independent, following a zero-mean, Normal distribution with standard deviation σ_j^m.

[26] One of the current president's campaign proposals in 2012 was reducing to 100 the number of PR seats in the Chamber of Deputies.

[27] Given the structure of the model, there is a unique optimal coalition configuration almost surely.

Finally, the entry conditions of the coalition formation stage are exploited in Step 3 to partially recover $\theta^{\mathrm{NC}}_{\mathrm{PRI}}$ and $\theta^{\mathrm{NC}}_{\mathrm{PVEM}}$.

Step 1. The voting stage is estimated following the discrete choice approach to demand estimation (Berry et al., 1995). Given that districts are large (>185,000 registered voters), by a law of large numbers approximation, candidate j's vote share, denoted S^m_{jd}, can be written in the familiar multinomial logit form:

$$S^m_{jd} = \frac{\exp(\delta^m_{jd})}{1 + \sum_{k\neq 0} \exp(\delta^m_{kd})}. \tag{2.7}$$

After taking logs and subtracting the (logged) share of the outside option, (2.7) yields the linear demand system:

$$\log(S^m_{jd}) - \log(S^m_{0d}) = \delta^m_{jd} = \alpha_1 c_{jd} + \alpha_2 c^2_{jd} + x'_d \beta^m_j + \xi^m_{jd} + \eta^m_j. \tag{2.8}$$

The second-tier coefficients of (2.4) are recovered analogously: letting $S^{\mathrm{ST},m}_{pd}$ and $S^{\mathrm{ST},m}_{0d}$ denote, respectively, the shares of PRI-PVEM coalition supporters who give their vote to $p \in \{\mathrm{PRI}, \mathrm{PVEM}\}$ or who split their vote 50-50, it follows that

$$\log(S^{\mathrm{ST},m}_{pd}) - \log(S^{\mathrm{ST},m}_{0d}) = \delta^{\mathrm{ST},m}_{jd} = x'_d \beta^{\mathrm{ST},m}_j + \xi^{\mathrm{ST},m}_{jd} + \eta^{\mathrm{ST},m}_j. \tag{2.9}$$

Identification of the voting stage parameters is obtained as follows. First, recall from the coalition formation stage that PRI and PVEM's choice of M_d conditions only on the observable district characteristics, x_d; all other components of the right-hand sides of (2.8) and (2.9) are unknown to PRI and PVEM at the time of their decision. This selection on observables implies that realized vote shares are independent of M_d conditional on x_d, ensuring that voters' preferences relative to menu m can be directly estimated from the subsample of districts where $M_d = m$.

Second, since parties tailor their spending to their candidates' qualities, c_{jd} is correlated with ξ^m_{jd} and so is endogenous. Instrumental variables are therefore necessary to identify the effect of campaign spending on candidates' vote shares. I use average spending by rival parties in neighboring districts (with the same menu of candidates) to instrument for the endogenous c_{jd}. Parties best respond to their rivals' spending, and campaigning costs are likely to be similar in neighboring districts (e.g., wages and transportation costs). By averaging rivals' spending in nearby districts, the presence of local cost shifters provides exogenous variation in spending levels with which to identify α_1 and α_2.[28]

Estimation of (2.8)-(2.9) and inference proceed using standard methods for linear random-effects (due to the aggregate popularity shocks) panel data models. The residuals of (2.8) and (2.9)—demeaned for each (j, m) pair to difference out the random effects η_j^m and $\eta_j^{\mathrm{ST},m}$—deliver consistent estimates of ξ_{jd}^m and $\xi_{jd}^{\mathrm{ST},m}$ for the district races as observed in the data, which are required for Step 2. Moreover, the standard deviations of these residuals yield estimates of their population counterparts—recall that $\xi_{jd}^m \overset{\text{i.i.d.}}{\sim} N(0, (\sigma_j^m)^2)$, and similarly $\xi_{jd}^{\mathrm{ST},m} \overset{\text{i.i.d.}}{\sim} N(0, (\sigma_j^{\mathrm{ST},m})^2)$—which are necessary to simulate counterfactuals.

Step 2. The parameters θ_p^{PW} and θ_p^{ES} of the parties' payoff functions are estimated by fitting predicted to observed campaign spending levels. For each party $p \notin \{\mathrm{PRI}, \mathrm{PVEM}\}$, let $\hat{c}_p = (\hat{c}_{pd})_{d\in\{1,\ldots,300\}}$ denote the party's spending levels as observed in the data, and let $\tilde{c}_p = (\tilde{c}_{pd})_{d\in\{1,\ldots,300\}}$ denote their predicted counterparts. These predictions are computed as follows. Given the estimates of the voting stage and candidate qualities obtained in Step 1, and for each possible value of $\theta_p = (\theta_p^{\mathrm{PW}}, \theta_p^{\mathrm{ES}}) \in \mathbb{R}_+^2$, I simulate p's best responses to its rivals' observed spending in each district, collected in $\tilde{c}_p$. I omit the dependence of these predictions on the estimates from Step 1 and simply write $\tilde{c}_p = \tilde{c}_p(\theta_p)$. Then θ_p is estimated by minimizing the distance between $\hat{c}_p$ and $\tilde{c}_p(\theta_p)$, i.e., by minimizing the norm:

$$Q_p(\theta_p) = \big(\hat{c}_p - \tilde{c}_p(\theta_p)\big)' W_p \big(\hat{c}_p - \tilde{c}_p(\theta_p)\big),$$

where W_p is a positive definite, diagonal weighting matrix. I initially estimate θ_p using the identity as weighting matrix. I then re-weight each district d by the reciprocal of the variance of the estimation error for the subsample of districts with the same menu of candidates as d.

For PRI and PVEM, $\theta_{\mathrm{PRI}} = (\theta_{\mathrm{PRI}}^{\mathrm{PW}}, \theta_{\mathrm{PRI}}^{\mathrm{ES}})$ and $\theta_{\mathrm{PVEM}} = (\theta_{\mathrm{PVEM}}^{\mathrm{PW}}, \theta_{\mathrm{PVEM}}^{\mathrm{ES}})$ are estimated similarly. Let $\hat{c}$ be a stacking of PRI and PVEM's observed joint spending levels along with their observed individual spending levels. That is, $\hat{c}$ contains 199 observations corresponding to the districts where PRI and PVEM ran together, plus 2×101 observations corresponding to the 101 districts where

[28]The results are robust to alternative choices of instruments, including lagged spending from the 2009 election. See Gillen et al. (2015) for details.

they ran independently. Let $\tilde{c}$ contain their predicted counterparts. Then θ_{PRI} and θ_{PVEM} are estimated by minimizing

$$Q_{\text{PRI-PVEM}}(\theta_{\text{PRI}}, \theta_{\text{PVEM}}) = \big(\hat{c} - \tilde{c}(\theta_{\text{PRI}}, \theta_{\text{PVEM}})\big)' W_{\text{PRI-PVEM}} \big(\hat{c} - \tilde{c}(\theta_{\text{PRI}}, \theta_{\text{PVEM}})\big),$$

as before.

Standard errors for these estimates are obtained by bootstrapping.

Step 3. Finally, the parameters $\theta^{\text{NC}}_{\text{PRI}}$ and $\theta^{\text{NC}}_{\text{PVEM}}$ are partially identified from the moment inequalities implied by the optimality of M_d for the PRI-PVEM coalition in each district. Recall from Section 2.3.1 that $M_d \in \arg\max_m \mathbb{E}(\pi^m_{\text{PRI},d} + \pi^m_{\text{PVEM},d} | x_d)$. This implies that $\mathbb{E}(\pi^{M_d}_{\text{PRI},d} + \pi^{M_d}_{\text{PVEM},d} | x_d) \geq \mathbb{E}(\pi^m_{\text{PRI},d} + \pi^m_{\text{PVEM},d} | x_d)$ for all $m \in \{\text{PRI}, \text{PVEM}, \text{IND}\}$, which in turn implies the unconditional moment inequality

$$\mathbb{E}\big(\pi^{M_d}_{\text{PRI},d} + \pi^{M_d}_{\text{PVEM},d} - (\pi^m_{\text{PRI},d} + \pi^m_{\text{PVEM},d})\big) \geq 0 \tag{2.10}$$

for each m. Computation of (2.10) is via simulation, and it involves the estimates from Steps 1 and 2.

Shi and Shum (2015) propose a simple inference procedure for models with such a structure, i.e., models where a subset of parameters is point identified and estimated in a preliminary stage—in this case, Steps 1 and 2—and the remaining parameters are related to the point-identified parameters through inequality/equality restrictions—in this case, the inequalities in (2.10). To implement their procedure, which requires both equalities and inequalities, I introduce slackness parameters as suggested by Shi and Shum: for each m, (2.10) becomes an equality restriction,

$$\mathbb{E}\big(\pi^{M_d}_{\text{PRI},d} + \pi^{M_d}_{\text{PVEM},d} - (\pi^m_{\text{PRI},d} + \pi^m_{\text{PVEM},d})\big) + \gamma_m = 0,$$

and the slackness parameter satisfies $\gamma_m \geq 0$. A criterion function is constructed as follows. With a slight abuse of notation, let β be a vector collecting the output of Steps 1 and 2, and let $\theta = (\theta^{\text{NC}}_{\text{PRI}}, \theta^{\text{NC}}_{\text{PRI}}, \gamma_{\text{PRI}}, \gamma_{\text{PVEM}}, \gamma_{\text{IND}})$. Then, following Shi and Shum's notation, define $g^e(\theta, \beta) = (g^e_m(\theta, \beta))_{m \in \{\text{PRI,PVEM,IND}\}}$ by

$$g^e_m(\theta, \beta) = \mathbb{E}\big(\pi^{M_d}_{\text{PRI},d} + \pi^{M_d}_{\text{PVEM},d} - (\pi^m_{\text{PRI},d} + \pi^m_{\text{PVEM},d})\big) + \gamma_m,$$

and let $g^{ie}(\theta) = (g^{ie}_m(\theta))_{m \in \{\text{PRI,PVEM,IND}\}} = (\gamma_m)_{m \in \{\text{PRI,PVEM,IND}\}}$. Thus, g^e summarizes the equality restrictions involving all the parameters of the model, and g^{ie} summarizes the inequality restrictions involving only θ. Letting β_0 denote the true value of β, the identified set of θ is

$$\Theta_0 = \{\theta : g^e(\theta, \beta_0) = 0 \text{ and } g^{ie}(\theta) \geq 0\}.$$

The criterion function is defined by

$$Q(\theta, \beta; W) = g^e(\theta, \beta)' W g^e(\theta, \beta),$$

where W is a positive definite matrix. It follows that $\Theta_0 = \arg\min_\theta Q(\theta, \beta_0; W)$ subject to $g^{ie}(\theta) \geq 0$. Shi and Shum show that the following is a confidence set of level $\alpha \in (0, 1)$ for θ:

$$CS = \{\theta : g^{ie}(\theta) \geq 0 \text{ and } Q(\theta, \hat{\beta}, \hat{W}) \leq \chi^2_{(3)}(\alpha)/N\},$$

where $\chi^2_{(3)}(\alpha)$ is the α-th quantile of the χ^2 distribution with 3 degrees of freedom (the number of restrictions in g^e), $\hat{\beta}$ is the estimate of β_0 from Steps 1 and 2, N is the number of observations used to estimate $\hat{\beta}$, and

$$\hat{W} = \left[G(\theta, \hat{\beta}) \hat{V}_\beta G(\theta, \hat{\beta})'\right]^{-1}$$

with $G(\theta, \hat{\beta}) = \partial g^e(\theta, \hat{\beta}) / \partial \beta'$ and $\hat{V}_\beta$ a consistent estimate of the asymptotic variance of $\hat{\beta}$.

As $g^e(\theta, \beta)$ and $g^{ie}(\theta)$ are in fact linear in θ (recall (2.6)), $Q(\theta, \hat{\beta}; \hat{W})$ has a unique minimizer subject to $g^{ie}(\theta) \geq 0$, which provides a useful point estimate for the counterfactual experiments of Section 2.5. Moreover, CS is convex, so upper and lower bounds of marginal confidence intervals for $\theta^{\mathrm{NC}}_{\mathrm{PRI}}$ and $\theta^{\mathrm{NC}}_{\mathrm{PRI}}$ can be computed by optimizing $f_p(\theta) = \theta^{\mathrm{NC}}_p$ subject to $\theta \in CS$.[29]

2.4 Estimation Results

This section summarizes the main estimation results. The discussion follows the structure of the model, beginning with the voting stage. A goodness of fit evaluation of the model is also provided.

[29]As discussed by Shi and Shum, the slackness parameters γ_m are nuisance parameters which may lead to conservative confidence sets for the parameters of interest. This does not seem to be a problem in this application, however, as the confidence intervals reported in Section 2.4 are fairly tight.

Estimates of voters' preferences. Tables A.1-A.5 in Appendix A.3 present estimates of the coefficients β_j^m capturing voters' menu-dependent preferences for candidate j's policy platform across the three menus $M_d = m \in \{\text{PRI}, \text{PVEM}, \text{IND}\}$. The estimates are overall consistent with the interpretation of district demographics introduced in Section 2.2.1: male-dominated districts with an older electorate are more likely to prefer candidates nominated by the right-wing parties PAN and PRI, and they are less likely to prefer a left-wing candidate from MP (although the effects for the latter are generally imprecise). Higher income is also associated with a preference for PAN candidates and a disliking of MP candidates. The intermediate parties in the ideology spectrum, NA and PVEM, exhibit mixed effects.

For a closer look at the menu-dependence of preferences, I discuss each party's coefficients in turn.

Preference for an MP candidate. The MP coefficients are the most stable in magnitude across menus, indicating that MP supporters were the least affected by the PRI-PVEM coalition. This is not surprising given that MP is the most ideologically distant from the coalition partners. The only significant cross-menu differences are with respect to gender and education. While female-dominated districts strongly support MP candidates when a PVEM candidate is in the race, i.e., when M_d = IND or M_d = PVEM, this support fades when M_d = PRI.[30] This suggests that the presence of more ideologically-close competitors potentially splitting the vote—i.e., both NA and PVEM—drives female-dominated districts to rally behind MP.[31] With respect to education, support for MP candidates considerably increases in better-educated districts when PRI and PVEM nominate a joint coalition candidate.[32] This perhaps reveals a desire to counterbalance the strength of the PRI-PVEM coalition. Consistently across menus, higher income depresses support for MP candidates.

Preference for an NA candidate. When PRI and PVEM run independently,

[30]Notice that, while the gender coefficients are negative for all other parties, the estimates for MP are positive and significant when M_d = IND or M_d = PVEM. For the cross-menu differences in the MP coefficients, the p-value is 0.037 for M_d = IND versus M_d = PRI, and 0.066 for M_d = PVEM versus M_d = PRI.

[31]Abortion, for example, is only broadly allowed in Mexico City, an MP stronghold.

[32]The MP coefficients are positive, significant, and largest in magnitude among all parties when M_d = PRI or M_d = PVEM. The p-values for the differences are 0.002 and 0.052 for M_d = IND versus M_d = PRI and M_d = PVEM, respectively.

NA has substantial support in better-educated districts.[33] This support shifts to MP, however, when PRI and PVEM nominate a joint candidate.[34] As discussed, this might be due to a desire to counterbalance the PRI-PVEM coalition by deserting the weaker NA for the stronger ideological neighbor MP. In contrast, while higher-income districts dislike NA candidates when $M_d = \text{IND}$, this effect disappears when PRI and PVEM run together.

Preference for a PAN candidate. Districts with an older electorate strongly support PAN candidates, and their support intensifies in response to a joint nomination from PRI and PVEM—particularly when they nominate a joint, ideologically closer, PRI candidate.[35] Similarly, support for PAN candidates in higher-income districts rises when PRI and PVEM nominate a joint candidate.[36] This suggests that conservative districts rally behind PAN to counterbalance the PRI-PVEM coalition.

Preference for a PRI candidate. PRI candidates have weak but consistent support in better-educated districts. Coalition PRI candidates gain considerable support relative to independent PRI candidates in older, male-dominated districts, indicating that the PRI-PVEM coalition primarily competes with PAN to attract conservative voters.[37]

Preference for a PVEM candidate. The most striking cross-menu differences relate to PVEM candidates. While independent PVEM candidates are strongly disliked in districts with an older electorate, coalition PVEM candidates obtain considerable support.[38] Higher-income districts also increase their support for coalition PVEM candidates substantially relative to independent PVEM candidates. Again, this suggests that the PRI-PVEM coalition primarily competes with PAN for supporters.

Regarding the second-tier choice for PRI-PVEM coalition supporters of how to split their vote between the two parties, Table A.6 shows estimates of the coefficients describing the choice of giving PRI 100% of the vote, and Table A.7

[33] The coefficient is positive, significant, and largest in magnitude among all parties.

[34] The p-values for the differences are 0.006 and 0.002 for $M_d = \text{IND}$ versus $M_d = \text{PRI}$ and $M_d = \text{PVEM}$, respectively.

[35] The age coefficients for PAN are positive and largest in magnitude among all parties. The p-value for the difference between $M_d = \text{IND}$ and $M_d = \text{PRI}$ is 0.023.

[36] The p-values are 0.078 and 0.003 for $M_d = \text{IND}$ versus $M_d = \text{PRI}$ and $M_d = \text{PVEM}$, respectively.

[37] The p-value for the gender difference is 0.021, and for the age difference is 0.067.

[38] The age coefficients for PVEM are the largest and second-largest in magnitude when $M_d = \text{IND}$ and $M_d = \text{PVEM}$, respectively. The p-value for the difference is 0.014.

shows the analogous estimates for PVEM. The outside option here is splitting the vote 50-50 between the parties. Female-dominated districts are generally more likely to split their vote between the parties, while districts with an older electorate are more likely to give 100% of their vote to PRI.

Finally, Table 2.6 reports estimates of α_1 and α_2, the parameters describing the persuasive effect of campaign spending on voters' preferences. The first column presents ordinary least squares (OLS) estimates, while the second column controls for the endogeneity of spending via two-stage least squares (2SLS) as explained in Section 2.3.2. Both sets of estimates indicate that campaign spending has a significant and initially positive persuasive effect on voters' preferences, which is mitigated by diminishing marginal returns. In fact, the OLS and 2SLS estimates agree with respect to the point at which the effect turns negative: spending more than 197,000 USD is detrimental to a candidate, revealing potential voter fatigue from excessive advertising. OLS considerably underestimates the overall persuasiveness of campaign spending. The 2SLS estimates imply that, for a candidate with an average vote share ($\sim$23%) and average spending ($\sim$45,000 USD), a 1% increase in campaign spending raises her vote share by about 0.95%, almost a one-to-one relationship. In contrast, the same calculation using the OLS estimates yields an increase of only 0.22%.

Estimates of parties' payoffs. Table 2.7 shows estimates of the coefficients θ_p^{PW} and θ_p^{ES} of parties' payoffs (measured in tens of thousands of USD). With the sole exception of NA, the results suggest that parties care only about their expected vote share when deciding how much to spend in a district. This is not surprising considering that their funding for the three following years and the number of PR seats they receive are both tied to their final vote share in the election. NA appears to have placed substantial weight on its probability of winning, though it was ultimately unsuccessful in the district races.

Table 2.8 reports 95% confidence intervals for the partially identified parameters $\theta_{\text{PRI}}^{\text{NC}}$ and $\theta_{\text{PVEM}}^{\text{NC}}$ of PRI and PVEM's payoffs when they don't enter a candidate in a district. Point estimates, which are necessary for the counterfactual experiments of Section 2.5, can be obtained as $\theta_{\text{PRI}}^{\text{NC}} = -1.555$ and $\theta_{\text{PVEM}}^{\text{NC}} = -0.443$. These values can be interpreted as direct compensation the parties demand in exchange for supporting their partner's candidate, revealing

Table 2.6: Structural estimates of persuasive effect of campaign spending (in tens of thousands of USD)

	(OLS)	(2SLS)
Coefficient	Estimate (St. Error)	Estimate (St. Error)
α_1	0.118	0.511
	(0.019)	(0.174)
α_2	-0.006	-0.026
	(0.001)	(0.010)
F-statistic (first stage)		34.570

Ordinary and two-stage least squares estimates of random effects model (2.8) with robust standard errors clustered by candidate's party affiliation and PRI-PVEM's coalition configuration.

their relative bargaining power in the choice of coalition candidates.

Goodness of fit. To evaluate the performance of the model, Table 2.9 provides a comparison of the model's main predictions with their counterparts in the data. The predictions are computed from an ex-ante perspective—i.e., before candidate qualities are known—as follows. Conditional on PRI and PVEM's observed coalition configuration, one thousand elections are simulated by drawing candidate qualities for each district, calculating the campaign spending equilibria played by the parties, and computing the resulting election outcomes. From these simulations, 95% confidence intervals are constructed for each party's final vote share and seat count, as shown in Table 2.9.

Despite its parsimonious structure, the model overall fits the data well; it only slightly overestimates PRI and PVEM's performance at the expense of NA's.

2.5 Counterfactual Experiments

The primary objective of this paper is to quantify the extent to which PRI and PVEM's coalition affected election outcomes and the composition of the

Table 2.7: Structural estimates of parties' payoffs

	θ_p^{PW}	θ_p^{ES}
Party	Estimate (St. Error)	Estimate (St. Error)
PRI	0.077 (0.354)	8.027 (3.028)
PVEM	0.101 (0.438)	3.203 (1.172)
PAN	0.142 (0.533)	5.554 (1.700)
MP	0.007 (0.065)	7.460 (2.556)
NA	1.759 (0.834)	0.921 (0.423)

First column corresponds to probability of winning. Second column corresponds to expected vote share. Bootstrapped standard errors in parentheses.

Table 2.8: Structural estimates of $\theta_{\mathrm{PRI}}^{\mathrm{NC}}$ and $\theta_{\mathrm{PVEM}}^{\mathrm{NC}}$.

	θ_p^{NC}
Party	Confidence interval (95%)
PRI	$[-1.961, -0.210]$
PVEM	$[-1.419, -0.270]$

Chamber of Deputies in 2012. To this end, I conduct two counterfactual experiments. First, I study what would have happened had PRI and PVEM not formed a coalition. That is, I simulate election outcomes (as described in Section 2.4) imposing M_d = IND in all districts where PRI and PVEM nominated a joint coalition candidate. Second, at the other extreme, I examine

Table 2.9: Goodness of fit: observed versus predicted seats and vote shares

	Vote share (%)		Seats	
Party	Observed	Predicted (95% conf. interval)	Observed	Predicted (95% conf. interval)
PRI	33.6	$[33.8, 35.5]$	207	$[209, 217]$
PVEM	6.4	$[6.2, 6.9]$	34	$[36, 48]$
PAN	27.3	$[25.5, 27.3]$	114	$[99, 119]$
MP	28.3	$[27.3, 29.3]$	135	$[115, 138]$
NA	4.3	$[3.8, 4.2]$	10	$[8, 10]$

the effects of constraining PRI and PVEM to form a *total* coalition. For this experiment, in all districts where PRI and PVEM ran independently, I force PRI and PVEM to run together by restricting the choices available to them in the coalition formation stage of the model to $M_d \in \{\text{PRI}, \text{PVEM}\}$. Thus, PRI and PVEM are constrained to run together in all districts, but they optimally select the party affiliation of their coalition candidates.

Table 2.10 presents the results of these experiments. For comparison, the first column reproduces the outcomes observed in the data. The second column reports predicted counterfactual vote shares and seats for each party under the no PRI-PVEM coalition treatment, and the third column reports their counterparts under the total PRI-PVEM coalition treatment. Individually, PRI and PVEM faced opposing consequences of their coalition: while PVEM benefitted greatly, both in terms of seats and vote share, these benefits accrued at the expense of PRI. Relative to not forming a coalition, by running with PRI as observed in the data, PVEM managed to secure almost thrice as many seats—13 versus 34—and to increase its vote share by about 42%—from 4.5% to 6.4%. Forming a total coalition would have given PVEM 10 additional seats and raised its vote share to 6.9%. On the other hand, by running as observed, PRI lost 6% of its seats—221 versus 207—and 7% of its vote share—36.3% versus 33.6%. By running together with PVEM in all districts, PRI would have additionally lost 5 seats and 1.1 percentage points in vote share. Overall, however, the PRI-PVEM coalition obtained net gains in terms of jointly held seats in the chamber. By running as observed, PRI and PVEM closed the gap to obtaining a legislative majority (i.e., 251 seats) by almost half—from 17 seats to 10; and they would have closed it by 71% had they run together in all districts—from 17 to 5. Thus, the results reveal that the observed coalition configuration constituted a compromise in balancing net gains to the coalition with PRI's individual losses.

Table 2.10: Counterfactual outcomes under no coalition or total coalition

	Vote share (%)		
Party	Observed	No coalition	Total coalition
PRI	33.6	$(+2.7 =)$ 36.3	$(-1.1 =)$ 32.5
PVEM	6.4	$(-1.9 =)$ 4.5	$(+0.5 =)$ 6.9
PAN	27.3	$(-0.1 =)$ 27.2	$(+0.7 =)$ 28.0
MP	28.3	$(-1.3 =)$ 27.0	$(+0.1 =)$ 28.4
NA	4.3	$(+0.7 =)$ 5.0	$(-0.2 =)$ 4.1
	Seats		
Party	Observed	No coalition	Total coalition
PRI	207	$(+14 =)$ 221	$(-5 =)$ 202
PVEM	34	$(-21 =)$ 13	$(+10 =)$ 44
PAN	114	$(+8 =)$ 122	$(+3 =)$ 117
MP	135	$(-2 =)$ 133	$(-7 =)$ 128
NA	10	$(+1 =)$ 11	$(-1 =)$ 9

Differences in parentheses are with respect to first column. Second and third columns correspond to counterfactual outcomes had PRI and PVEM run independently or together in all districts, respectively.

The source of PRI's losses is voting behavior. Table 2.10 shows that PRI and PVEM command roughly 40% of the national vote share, regardless of how they run. But PRI's individual vote share drops substantially when PRI and PVEM join forces, as a consequence of the way in which PRI-PVEM coalition supporters split their vote between the two parties. When PRI and

PVEM jointly nominate a PRI candidate, on average, 74.5% of coalition supporters give 100% of their vote to PRI, 9.6% give their vote to PVEM, and 15.8% split their vote 50-50. These percentages change to 71.4%, 12.9%, and 15.7%, respectively, when PRI and PVEM jointly nominate a PVEM candidate. As a result, PRI's expected vote share with a joint PRI candidate is about 33%, and with a joint PVEM candidate, about 31.7%. Compared to its vote share of 36.3% when it runs independently, PRI experiences considerable losses from joint nominations. Lost vote share then translates into seat losses for PRI via the restriction on the PR assignment of seats discussed in Section 2.2. The exact form of the restriction is: if a party's vote share is S_p, it cannot hold more than $\lfloor 500(S_p + 0.08) \rfloor$ total seats. Notice that, across the three columns of Table 2.10, PRI is bound by this restriction.[39] Thus, PRI is severely capacity constrained in the Chamber of Deputies election.

For a closer look at the district races, Table 2.11 breaks down the seat counts in Table 2.10 by type of seat—i.e., direct representation (DR) seats and PR seats. The DR seat counts reveal that there are few competitive districts in Mexico. Relative to not forming a coalition, there are only 9 districts that PRI and PVEM can steal from their competition by joining forces; and this number is independent of whether they run as observed in the data or together in all districts. Consistent with the discussion of Section 2.4, Table 2.11 shows that PAN was the most affected by the PRI-PVEM coalition: of the 9 additional victories that PRI and PVEM obtained by running as observed in the data, 8 were from PAN-leaning districts. Interestingly, forcing PRI and PVEM to run together in all districts would cause the coalition to target MP-leaning districts via joint PVEM candidate nominations. Against a total PRI-PVEM coalition, PAN would only lose 4 districts, but MP would lose 5 instead of just 1.

In addition to its effects on seats and vote share, I consider how the PRI-PVEM coalition affected campaign spending in the district races. While the model refrains from specifying PRI and PVEM's individual shares of spending in support of coalition candidates, it is possible to examine whether there were any aggregate financial gains for the parties from coalition formation. First, in terms of total surplus for the coalition partners, which takes into account

[39] That is, $207 = \lfloor 500(0.335953 + 0.08) \rfloor$, $221 = \lfloor 500(0.363 + 0.08) \rfloor$, and $202 = \lfloor 500(0.325 + 0.08) \rfloor$.

Table 2.11: Counterfactual outcomes by type of seat

	Observed		No coalition		Total coalition	
Party	DR seats	PR seats	DR seats	PR seats	DR seats	PR seats
PRI	158	49	165	56	148	54
PVEM	19	15	3	10	29	15
PAN	52	62	60	61	56	61
MP	71	64	72	61	67	61
NA		10		11		9

their preferences over election outcomes as expressed by their payoffs in (2.5) and (2.6), the model estimates the value of running as observed in the data, relative to running independently, at about 1.8 million USD.[40] This amount equals approximately 8% of PRI and PVEM's observed total spending in the election, and it can be interpreted as a willingness-to-pay measure of the value of the coalition for the two parties. In contrast, the value of forming a total coalition, relative to running independently, is only half: about 900,000 USD. Though substantially smaller, this value reveals that, if PRI and PVEM had been constrained to choose only between not forming a coalition or forming a total one, they would have nonetheless formed a coalition, and the election outcomes would have been those in the third column of Table 2.10.

As a more direct measure of financial gains from coalition formation, the ratio between PRI and PVEM's joint spending and joint vote share provides a rough estimate of how much the two parties need to spend—in equilibrium—to produce 1 percentage point of joint vote share. On average across districts, this ratio is about 2,036 USD when PRI and PVEM run independently, 1,812 USD when they nominate a joint PRI candidate, and 1,701 USD when they nominate a joint PVEM candidate, which implies cost savings of 11% to 16% from joint nominations. Thus, by not having to campaign against each other, joint nominations allow PRI and PVEM to internalize externalities, substantially increasing the effectiveness of their spending.

[40] Total surplus is calculated as the sum of $\mathbb{E}(\pi^m_{\mathrm{PRI},d} + \pi^m_{\mathrm{PVEM},d} | X_d)$ over all districts, with the appropriate value of m for each scenario.

Finally, Table 2.12 shows how average spending across parties would have changed under the two counterfactual scenarios. It is interesting to note that, with the sole exception of MP, spending is increasing in the number of competing candidates. This is consistent with the intuition that differentiation via campaign advertising becomes relatively more valuable in a more crowded—and hence less polarized—field, leading parties to invest more heavily (see, for example, Ashworth and Bueno de Mesquita, 2009; Iaryczower and Mattozzi, 2013). Indeed, total spending in the election is highest when PRI and PVEM run independently in all districts and lowest when they form a total coalition.

Table 2.12: Counterfactual spending (in thousands of USD)

	Average spending per district		
Party	Observed	No coalition	Total coalition
PRI+PVEM	80.1	88.6	78.3
PAN	40.7	40.9	40.4
MP	55.8	55.1	56.1
NA	18.1	19.7	17.6

2.5.1 Robustness to Richer Specification

As explained in Section 2.3.2, one of the key identifying assumptions underpinning the counterfactuals presented above is that PRI and PVEM's optimal choice of where and how to run together conditions only on the observable district characteristics. If PRI and PVEM had additional information on which to base their decision, the results could be severely biased. One way in which this concern can be addressed is via a richer specification of the model. Specifically, expanding x_d with other district characteristics related to voting behavior can help mitigate any potential omitted-variables bias.

Accordingly, to evaluate the robustness of the results, I estimate a richer version of the model with 14 demographics instead of 4, including regional fixed effects to capture any spatial features of voting preferences. For a complete list of the variables used, see Appendix A.2. As shown in Table A.8, which reproduces Table 2.10 using the predictions of the richer model, the main results of the counterfactual experiments remain virtually unchanged.[41]

2.6 Discussion

The results of the previous section provide a rich picture of the effects of electoral coalition formation on voter behavior, political campaigns, and election outcomes. The counterfactual experiments document the willingness of an electorally strong but capacity-constrained party to sacrifice its individual position—in terms of both seats and vote share—in order to substantially build up a weaker partner. The results also reveal considerable financial gains from coalition formation: by supporting common candidates instead of campaigning against each other, coalition partners can increase the effectiveness of their campaign expenditures.

While post-election legislative bargaining is not explicitly considered in this paper, the results are suggestive of the importance of electoral coalition formation as a preliminary stage of the legislative bargaining process. Parties can use electoral coalitions to pre-select and foster legislative bargaining partners. Indeed, a cursory look at legislative voting data for the Mexican Chamber of Deputies following the 2012 election reveals an extremely high degree of (but not complete) coherence between PRI and PVEM legislators. However, electoral coalitions are not mergers, and post-election disagreements among electoral coalition partners are not uncommon. Further research is needed to fully understand the role of electoral coalitions in shaping both electoral and legislative output.

The potential for financial incentives in coalition formation had been previously unrecognized. In settings where parties and candidates are not publicly funded, these incentives may even be stronger, as coalition partners can share the burdens of fundraising. Moreover, potential donors may be more willing to back coalition candidates with broader support, further prompting parties to make joint nominations. Understanding the role of fundraising in coalition formation is an interesting open question for future research.

Lastly, the results indicate that coalition formation can lead to an overall reduction in total campaign expenditures. The net welfare impact of this ef-

[41] Detailed results from the richer model are available on request from the author.

fect hinges on whether campaign advertising provides valuable information to voters. Martin (2014) finds, using data from U.S. Senate and gubernatorial elections, that the informational content of campaign advertising is limited: political campaigns have a primarily persuasive—rather than informative—effect on voter behavior. If the social opportunity cost of resources devoted to political campaigns is believed to be high, then the results suggest that electoral coalition formation can deliver a welfare-improving reduction of campaign expenditures. As noted by Iaryczower and Mattozzi (2013), however, this conclusion may be sensitive to the institutional environment.

Chapter 3

BLP-LASSO FOR AGGREGATE DISCRETE-CHOICE MODELS APPLIED TO ELECTIONS WITH RICH DEMOGRAPHIC COVARIATES

3.1 Introduction

When analyzing aggregated data about consumers' choices in different regional markets, researchers must account for the demographic characteristics of local markets that might drive observable variability in consumers' preferences and firms' pricing policies. The abundance of such variables, whether from census data, localized search trends, or local media viewership surveys, immediately confronts researchers defining which model to adopt for this analysis with difficult questions. Which variables should be included in the model? Which controls can be excluded from the analysis without introducing omitted variable bias? How sensitive are the estimated effects of a firm's pricing policy on their market share to these specification decisions?

In the current paper, we hope to help answer these questions by providing data-driven algorithms for addressing model selection in analyzing consumer demand data. Our main contribution here is to apply recent econometric results from the variable selection literature to a popular nonlinear aggregate demand model. Specifically, our technique generalizes procedures from Belloni et al. (2012) and Belloni et al. (2013a) for selecting variables to a nonlinear Berry et al. (1995) model of consumer demand with random coefficients. Our asymptotic results extend Gillen et al. (2014)'s results by adopting techniques from Fan and Liao (2014)'s analysis of penalized GMM estimators, particularly the conditions for an oracle property that ensures all necessary variables are included in the model.

The specific problem of interest addresses high-dimensional demographic data for local markets that may help characterize local preferences. To address this problem, we adopt techniques proposed by the literature on machine learning to identify the demographic characteristics that exert the most important influence on observed market shares. As we discuss in section 3.2.1, these innovative algorithms present powerful devices for variable selection that require some care in their implementation. When properly deployed through multiple iterations of variable selection with appropriate penalization, these algorithms identify all the variables necessary for valid inference in the model.

We conduct an empirical investigation of campaign expenditures' influence on election outcomes, utilizing structural models inspired by the industrial organization and marketing literatures. In elections, the "consumers" are the voters, the "market" is the voting district, and the set of available "products" is the set of political parties running in the district. We use our technique to analyze the impact of campaign expenditures on candidate vote shares in Mexican elections. With access to the full census records for Mexico, we have rich demographic data for each voting district. We also have some variability in the "market structure" or the set of political parties competing in each district, since Mexican elections allow parties to form partial coalitions. In districts where parties coordinate, the number of competing candidates is smaller than in those districts where they compete. Our analysis yields the robust finding that campaign expenditures significantly influence voter preferences.

We present our inferential technique in a series of four progressively more complex models. We introduce the simple discrete choice approach to testing the influence of campaign expenditures on voting in section 3.4. This simple setting assumes voters vote for their most preferred candidate, with voter preferences represented by a linear utility function. This simplicity admits a standard generalized linear model, which we refer to as Model LM-F (Linear Model with Fixed Controls), for candidate vote shares, providing a first look at the influence of campaign expenditures on vote shares in Mexican elections with a pre-selected model. The results show that campaign expenditures make a positive and significant contribution to a candidate's vote share, though this contribution is mitigated by diminishing marginal returns.

We then introduce data-driven variable selection for the demographic controls we include in the model. In so doing, section 3.5 applies the techniques proposed in Belloni et al. (2012) and Belloni et al. (2013b) to develop our second empirical specification, Model LM-S (Linear Model with Selected Controls). In this discussion, we explicitly present the sparsity assumptions required for consistent and valid inference and describe the exact algorithm for performing that analysis. Implementing Model LM-S for the Mexican voting data, we illustrate how the algorithm can provide an agnostic characterization of the robustness of our results to model specification. In particular, the variable selection algorithm is governed by only two "tuning" parameters, both of which are constrained by theory to lie in a reasonably small band of candidate values. Varying these tuning parameters allows us to verify the robustness of our empirical findings from Model LM-F to different specifications.

Next, Model RC-F (Random Coefficients with Fixed Controls) allows heterogeneity in voters' responses to campaign expenditures in section 3.6. This specification corresponds to the "BLP" random coefficients logit model (Berry et al. (1995)), which is a workhorse model in empirical industrial organization. We fit Model RC-F to Mexican data using the same pre-specified set of demographic variables for controls used in the linear Model LM-F. Though we find very little evidence of heterogeneous impressionability, the impact of campaign expenditures on vote shares from the linear specification remains robust.

Our main innovation lies in the final specification, Model RC-S (Random Coefficients with Selected Controls). This model extends the results from Gillen et al. (2014) to a setting where the number of potential parameters grows exponentially with the sample size. Estimation and inference here poses some conceptual and computational challenges. The analytical development incorporates asymptotic results from Fan and Liao (2014), which builds on Caner and Zhang (2013) and Belloni et al. (2012) by establishing an oracle property for penalized GMM estimators in the ultra-high dimensional setting. Computationally, we start by fitting Model RC-F with the variables selected by Model LM-S to recover latent mean utilities and optimal instruments from the nonlinear BLP voting model. We then select a series of additional variables for robust inference, specifically including controls for observable heterogeneity in the optimal instruments. Finally, we verify the first order conditions of the penalized GMM-estimator to ensure we find a local optimum to which we can apply Fan and Liao (2014)'s oracle property.

3.2 Related Literature

The current paper sits at the intersection of political science, economics, and statistics. Our application addresses a well-worn question on how expenditures by a political campaign influence the outcome of an election. The inferential model we use to investigate this question is grounded in structural econometric methods used for consumer demand estimation by researchers in industrial organization and marketing. Finally, the statistical techniques we apply utilize recent innovations in machine learning developing automated techniques for variable selection.

3.2.1 Model Selection and Inference

Data-driven approaches to variable selection represents one of the most active areas of statistical research today. Tibshirani (1996)'s Lasso estimator ushered in a new approach to estimation in high-dimensional settings by incorporating convex penalties to least-squares objective functions. The penalized estimation technique has been further developed by Fan and Li (2001)'s SCAD penalty, Zou and Hastie (2005)'s elastic net and Huang et al. (2008)'s Bridge estimator, Bickel et al. (2009)'s infeasible lasso, and Zhang (2010)'s minimax concave penalty. This literature has also inspired several closely related estimators, including Candes and Tao (2007)'s Dantzig selector and Gautier and Tsybakov (2011)'s feasible Dantzig selector as well as Belloni et al. (2011)'s Square-Root Lasso. Each of these estimators incorporate some form of $L1$-regularization to the objective function's maximization problem, selecting variables for the model by imposing a large number of zero coefficients on the solution.

For an estimator that imposes a large number of zero coefficients in the solution to be consistent, it must be the case that a large number of zero coefficients are present in true model for the data generating process. This restriction on the true parameters of the model takes the form of a sparsity assumption. In its early formulations, the sparsity restriction was stated as an upper bound on the $L0$ or $L1$ norm of the true coefficients.[1] If an estimator classifies zero and non-zero coefficients with perfect accuracy as the sample grows, the estimator satifies an oracle property. In order to establish an oracle property, the sparsity restrictions need to be coupled with a minimum absolute value for non-zero coefficients to ensure they are selected by the penalized estimator. Intuitively, the variability of a single residual (which could be explained by an erroneously included explanatory variable) needs to be dominated by the penalty, which

[1] Generalized notions of sparsity appear in Zhang and Huang (2008) and Horowitz and Huang (2010), which allow for local perturbations in which the zero-coefficients are very small. A similar approach appears in Belloni et al. (2012) and Belloni et al. (2013a) characterizing inference under an approximate sparsity condition that constrains the error in a sparse representation of the true data generating process.

in turn needs to be dominated by the effect of a non-zero regressor (to justify the penalty associated with the coefficient's non-zero value).

Performing inference after model selection, even with an estimator that satisfies the oracle property, has presented a non-trivial challenge to interpreting the results of estimators that incorporate these techniques. Leeb and Pötscher (2005, 2006, 2008) present early critiques of the sampling properties for naïvely-constructed test statistics after model selection, illustrating the failure of asymptotic normality to hold uniformly and the fragility of the bootstrap for computing standard errors in the selected model. Lockhart et al. (2014), published with a series of comments, propose significance tests for lasso estimators that perform well on "large" coefficients but are less effective for potentially "small" coefficients for which the significance tests are not pivotal due to the randomness of the null hypothesis. In a series of papers, Belloni et al. (2013a) and Belloni et al. (2012) propose techniques for inference on treatment effects in linear, instrumental variables, and logistic regression problems. These techniques incorporate multiple stages of variable selection with data-driven penalties that ensure the relevant controls are included in the econometric model before performing inference in an unpenalized post-selection model. By focusing on inference for a predefined, fixed-dimensional, subset of coefficients, the selected models represent a desparsified data generating process, with inference results from van de Geer et al. (2014) providing uniformly valid confidence intervals.

Extending these techniques from least squares regression models to more general settings presents additional challenges. Fan and Li (2001), Zou and Li (2008), Bradic et al. (2011), and Fan and Lv (2011) propose methods for analyzing models defined by quasi-likelihood. Our application focuses on GMM estimators, whose properties in high-dimensions are considered by Caner (2009), Caner and Zhang (2013), Liao (2013), Cheng and Liao (2015), and Fan and Liao (2014). Several of these papers address the issue of moment selection, as in Andrews (1999) and Andrews and Lu (2001). As our application considers an environment with a fixed set of instruments, our analysis does not require moment selection but makes heavy use of the oracle properties established by Fan and Liao (2014).

Our model builds directly on Gillen et al. (2014)'s analysis of demand models with complex products. The Gillen et al. (2014) application considers aggregate demand models where the dimension of the vector of product characteristics is large, on the same order of magnitude as the number of observations. Our current application utilizes variable selection to mitigate an incidental parameter problem in characterizing voter preferences. This utilization is similar in motivation to Harding and Lamarche (2015), who use a penalized quantile regression to allow for heterogeneity in individual nutritional preferences when analyzing a household grocery consumption data. In addition to applying Gillen et al. (2014)'s approach to an interactive fixed-effects model, our analysis allows for material weaker conditions on the data generating process than in Gillen et al. (2014). Notably, by applying Fan and Liao (2014)'s oracle properties, we can allow for the number of parameters to grow exponentially with the number of observations, rather than linearly.

3.2.2 Structural Models of Campaign Spending and Voting

Empirical analysis of voting data presents a particularly challenging exercise for political scientists due to the large number of factors driving voter behavior, endogeneity induced by party competition and candidate selection, and behavioral phenomena driving individual voter decisions. Including early work from Rothschild (1978) and Jacobson (1978), a number of political scientists have explored the effect of campaign spending on aggregate vote shares, often coming to different conclusions on its importance in influencing vote share by informing, motivating, and persuading voters. These inconclusive results arise in part due to challenges in identifying valid and relevant instruments (Jacobson, 1985; Green and Krasno, 1988; Gerber, 1998). Gordon et al. (2012) discuss several challenges to this research agenda, highlighting the value of incorporating historically underutilized empirical methods from marketing researchers.

A nascent literature in political science adopts structural approaches to inference for analyzing political data. Discrete choice approaches to analyzing voting data date back to Poole and Rosenthal (1985) and King (1997). Among the early adopters of this approach are Che et al. (2007), who utilize a nested logit model that takes advantage of individual voter data to identify the impact of advertisement exposure on their behavior. The problem we consider is closest to Rekkas (2007), Milligan and Rekkas (2008), and Gordon and Hartmann (2013), who apply a Berry et al. (1995) model to infer the impact of campaign expenditures on aggregate voting data. The analysis presented in Gordon and Hartmann (2013) provides an excellent motivation for our proposed inference technique. Though they find a robust evidence that campaign spending on advertisement positively contributes to a candidate's vote share, the magnitude of this contribution varies by a factor of 3 depending on the specification of controls adopted. Their extremely large sample allows them to adopt very rich models of fixed effects and, in the most flexible models, the significance of the contribution of campaign expenditures to vote shares drops to 10%. A natural concern is that this loss of significance is in part due to an excessively conservative model of control variables. Our data-driven approach to selecting these control variables provides an agnostic approach to addressing some of the inherent ambiguity in determining which of these estimates is "most correct."

A number of other researchers have also adopted a structural approach to analyze voting behavior. Degan and Merlo (2011) present a structural model for analyzing multiple concurrent elections in US Congressional and Presidential elections, with an extension by Levonyan (2013) analyzing the influence of the Presidential election on the outcome of California's Proposition 8. Kawai and Watanabe (2013) adopt a structural approach in investigating strategic voting behavior using Japanese general election data. Kawai (2014) provides a dynamic extension of Erikson and Palfrey (2000)'s model of fund-raising and campaigning to analyze elections for US House of Representatives while adopting a control function approach to mitigate unobserved heterogeneity in voter behavior. Our application is clearly most closely related to Montero (2015)'s structural analysis of the incentives for coalition formation in Mexican elections.

Beyond structural approaches for analyzing equilibrium outcomes, a massive body of empirical research investigates the influence of campaign expenditures on vote shares using natural and field experiments. These investigations are particularly valuable in their ability to differentiate how different styles of campaign advertising influences voter behavior. Gerber (2011) surveys much of this literature. Though our inference technique is derived in the context of a structural model of voting, the approach to selecting demographic control variables could be readily adopted to these environments.

3.3 Voting in Mexico

We utilize the data analyzed by Montero (2015)'s investigation of the coalition formation incentives of political parties. To control for observable heterogeneity in voter preferences by district, we have access to rich demographic data—over 200 variables—from the 2010 population census.

3.4 Model LM-F: Homogeneous Voters with Fixed Controls

We begin by introducing the structural model for voting in a setting free of complication by variable selection and nonlinear effects. This approach allows us to describe the economic environment for voting decisions and preferences before focusing on the methodological issues introduced by model specification tests. We then estimate a pre-selected model for control variables and instruments for the voting data in Mexico.

3.4.1 The Structural Model and Estimation Strategy

In district t, we observe vote shares based on individual voters (indexed by i) who choose from among the candidates competing in the district (indexed by $j = 1, \ldots, J$). We represent the option to not vote or to write in a non-party candidate as an "outside good" indexed by $j = 0$. To characterize preferences for a representative voter in the district, we observe a vector of K_0 demographic characteristics for the district, denoted x_{0t} and K_1 characteristics describing the candidate for party j in that district, denoted x_{1jt}. The endogenous treatment variable of interest, campaign spending in the district by a candidate, is represented by p_{jt}.[2] Finally, we allow exogenous unmodelled variation in voters' preferences through a product-market specific latent shock, ξ_{jt}.

We begin by introducing preferences to our model in a restrictively homogeneous setting without random effects in preference characteristics, though we will relax these assumptions later. In district t, suppose consumer preferences are homogeneous up to an idiosyncratic, individual specific shock, denoted

[2] For expositional purposes, we treat p_{jt} as a scalar, though it could be interpreted as a fixed-dimensional vector of treatment variables. Our empirical specification will allow for campaign expenditures to exert both a linear and quadratic influence on voter latent utilities.

ϵ_{ijt}. This simplification allows us to represent consumer i's latent utility from voting for candidate j:

$$u_{ijt} = x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt} + \epsilon_{ijt}. \tag{3.1}$$

Note that, though the candidate-specific characteristics influence voter preferences in a common way across parties, district-specific demographics act as interactive fixed-effects, impacting voter preferences differently for different parties. The latter form of heterogeneity allows national party platform positions to influence local voting preferences depending on the district's demographic composition.

Assuming the individual ϵ_{ijt} shocks are independently distributed with a Type-I Extreme Value distribution and normalizing the utility of not voting to 0, the probability of a randomly-selected district t voter choosing candidate j is given by the usual logit form:

$$\Pr\{y_{ijt} = j\} = \frac{\exp\{x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}\}}{1 + \sum_{r=1}^{J} \exp\{x'_{0t}\beta_{0r} + x'_{1rt}\beta_1 + p_{rt}\beta_{ip} + \xi_{rt}\}\kappa_{rt}}. \tag{3.2}$$

The modification $\kappa_{rt} \equiv \mathbf{1}\{\text{Party } r \text{ runs in District } t\}$ reflects the impact coalition formation has on the menu of parties available to voters in each district. This formulation assumes that voters cast their ballots "sincerely" in favor of their most preferred candidate, without any strategic considerations. While accounting for strategic voting is beyond the scope of this paper, the proportional nature of the post-election allocation of seats and future funding among parties described in Section 3 provides support for the sincere voting assumption.[3]

Let candidate j's vote share in district t be denoted by s_{jt}. The current setting, in which the expected vote share simply equals the choice probabilities, admits a linear "demand" system. Denoting the share of voters abstaining or writing-in candidates by s_{0t}, the logged vote shares (given a large number of voters) take the form:

$$\mathcal{S}_{jt} \equiv \log s_{jt} - \log s_{0t} = x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}. \tag{3.3}$$

Among other sources, endogeneity arises from parties' consideration of unobserved local shocks to voter preferences when determining expenditures. This

[3] See Kawai and Watanabe (2013) for an example of the challenges involved in identifying strategic voting.

reaction induces correlation between the unobserved shock ξ_{jt} and spending levels p_{jt}. However, these expenditures also respond to L exogenous instruments z_{jt}, allowing us to identify the causal relationship between campaign expenditures and their impact on vote shares. Imposing an (admittedly restrictive) linear structural relationship on expenditures and these features, suppose:

$$p_{jt} = x_{0t}'\pi_{0j} + x_{1jt}'\pi_1 + z_{jt}'\pi_z + \nu_{jt},\ \mathbb{E}[\nu_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0. \tag{3.4}$$

Finally, suppose that district-specific shocks to preferences take a linear form:

$$\xi_{jt} = \rho\nu_{jt} + \eta_{jt},\ \mathbb{E}[\eta_{jt}|\nu_{jt}] = 0. \tag{3.5}$$

Assumption 3.1. *Linear Logit Truthful Voting Structural Model:*

1. *In each of T districts, a large and representative sample of voters truthfully vote for their most-preferred candidate under equation (3.1)'s utility specification.*

2. *Logged vote shares are linear in K_0 district characteristics x_{0t}, K_1 candidate characteristics x_{1jt}, and campaign spending p_{jt} according to equation (3.3).*

3. *Campaign spending is linear in the district and candidate characteristics x_{0t} and x_{1jt} as well as L exogenous instruments z_{jt}, as in equation (3.4).*

4. *Residual vote share correlates endogenously with campaign spending, as in equation 3.5, but is exogenous with respect to instruments: $\mathbb{E}[\eta_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0$.*

We consolidate the above statements about the data generating process for observed vote shares in Assumption 3.1. This setting presents a simple model for evaluating the influence of campaign spending on voter preferences, in which instrumental variables via two-stage least squares estimation can be used for accurate inference. With standard regularity conditions, consistent inference on β_p can proceed using a standard IV regression of the model:

$$\mathcal{S}_{jt} = x_{0t}'\beta_{0j} + x_{1jt}'\beta_1 + p_{jt}\beta_p + \xi_{jt},\ \mathbb{E}[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0. \tag{3.6}$$

Fitting the first stage regression from equation 3.4 gives fitted values for p_{jt},

$$\hat{p}_{jt} = x'_{0t}\hat{\pi}_{0j} + x'_{1jt}\hat{\pi}_1 + z'_{jt}\hat{\pi}_z. \tag{3.7}$$

The second stage then fits the regression using these fitted values:

$$\mathcal{S}_{jt} = x'_{0t}\hat{\beta}_{0j} + x'_{1jt}\hat{\beta}_1 + \hat{p}_{jt}\hat{\beta}_p + \xi_{jt},\ \mathbb{E}[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}, \hat{p}_{jt}] = 0. \tag{3.8}$$

Under standard conditions, these estimates will be asymptotically normal with the usual variance-covariance matrix, admitting standard hypothesis tests for inference.

3.4.2 A Preselected Model for Returns to Campaign Spending

As a first-pass in our empirical analysis, we present the results for the linear model using a preselected set of control variables. The selected control variables, summarized in table 3.1, relate to regional dummies, economic characteristics, education levels, and household structures. We allow for both a linear and quadratic impact of campaign spending on candidate vote share, with the latter reflecting diminishing marginal returns to campaign expenditure. As instruments for campaign spending, we include lagged campaign expenditures, campaign expenditures by competitors in nearby districts, and, as cost shifters, the population density of the district and the percent of the population with internet access.

Table 3.1: Pre-Selected Controls for Fixed Model of Voting

Regional Dummies	Demographics	Economic Status
Region 1	% of Pop Age 18-24	Unemployment
Region 2	% of Pop Age 65+	% of Households w/Car
Region 3	% of Pop that's Married	% of Households w/Refrigerators
Region 4	Average Years of Education	% of Households w/o Basic Utils
Region 5	% of Pop with Elementary Ed	% of Households w/Female Head

This table presents demographic control variables taken from the census measured at the district-level that are included in a pre-specified model of voter preferences. Each of these controls is associated with a party-specific fixed effect, x_{0t} in the utility model.

The main results in table 3.2's Panel A indicate a positive first-order return to campaign spending, with the linear contribution of campaign expenditure to vote share indicating a 0.63% expected increase from raising campaign spending by $10,000 USD. The coefficient for squared campaign spending indicates the second-order effect diminishes this contribution by 0.03%. Both of these effects are highly statistically significant.

Investigating the interactive fixed-effects for demographic characteristics shows that many of these fixed-effects are statistically insignificant. Panel B in table 3.2 reports the p-Value for the t-Statistics associated with each of these individual fixed effects. The majority (77%) of these coefficients do not have a statistically significant effect on expected vote shares. This result motivates our application of variable selection techniques in the current problem. Can we reduce the number of parameters we need to estimate? Will doing so provide more robust results?

3.5 Model LM-S: Variable Selection and Inference

Without pre-specifying which demographic controls to include in the analysis, our application includes more potential parameters in the model than we have available observations. These demographic characteristics may be either irrelevant or redundant in describing voting preferences, with many controls reporting similar information with slightly different measures. As such, it's both reasonable and necessary to ignore those demographic variables that have little explanatory power. This variable selection exercise, however, has the potential to distort inference on the effect of campaign spending on vote shares and so our goal is to conduct this exercise while maintaining consistent and valid inference.

As discussed in section 3.2.1, a considerable literature in econometrics and statistics explores the effect of model selection on inference in cases where the number of variables exceeds the number of observations. Performing inference in this environment depends critically on restricting the data generating process to satisfy some form of "sparsity." Even though there may be a large number of possible parameters, only a relatively small number of those parameters are truly non-zero in a sparse model. Consequently, estimation requires selecting which variables are actually relevant to the estimation problem and excluding the irrelevant variables. We now formally introduce the notion of sparsity we assume and review existing results for consistent inference in the

Table 3.2: Estimated Returns to Campaign Spending in Pre-Selected Model (Model LM-F)

Panel A: Main Results

	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.627	0.213	2.95	0%**
Expenditures^2	-0.033	0.011	3.03	0%**

Panel B: Significance of Demographic Controls (p-Values)

Party	Region 1	Region 2	Region 3	Region 4	Region 5
MP	45%	0%**	63%	15%	80%
NA	2%*	12%	29%	51%	92%
PAN	44%	0%**	4%*	1%*	26%
PRI	88%	48%	86%	37%	73%
PVEM	8%	79%	1%**	87%	8%
CM	11%	98%	5%	18%	14%

			% of Households with		
Party	Unempl	Car	Refrigerator	Utilities	Female Head
MP	19%	0%**	9%	17%	1%*
NA	26%	41%	0%**	0%**	47%
PAN	39%	13%	1%*	41%	88%
PRI	36%	1%**	23%	9%	3%*
PVEM	52%	8%	59%	50%	35%
CM	11%	81%	63%	64%	0%**

Party	Pop 18-24	Pop 65+	Married	Element Ed +	Avg Years School
MP	66%	73%	61%	3%*	1%**
NA	10%	5%*	3%*	9%	8%
PAN	91%	1%**	31%	4%*	29%
PRI	38%	86%	32%	6%	77%
PVEM	2%*	35%	68%	92%	4%*
CM	17%	48%	1%**	37%	49%

Panel A reports the return to campaign expenditures and squared campaign expenditures using the linear logit voting model estimated from equation (3.8) with interactive fixed effects between the political party and demographic controls listed in Table 3.1. Panel B reports the significance of each of the interactive fixed-effects by party, with * and ** indicating significance at the 5% and 1% levels, respectively.

simple environment described above.

3.5.1 Sparsity Assumptions and Regularity Conditions

Sampling approximations for high-dimensional inference require an asymptotic framework that allows the number of parameters (here, $JK_0 + K_1$) to grow with the number of observations. We treat each district as the unit of observation, due to obvious correlation in the vote shares across parties within a given district. For convenience, we'll assume that number of candidate-specific characteristics, K_1, and the number of excluded instruments, L, remain fixed but allow the number of coefficients associated with district demographic characteristics, K_0, to become large as $T \to \infty$. For measurement purposes, introducing sample-size dependent variable specifications requires indexing the data generating process, which we'll denote $\mathcal{P}_T$, by the sample size. Within this sequence of data generating processes, we allow $K_T \equiv JK_{0T} + K_1$ to be much larger than sample size T, requiring only $\log(K_T) = o\left(T^{1/3}\right)$.[4]

Assumption 3.2. *Exact Sparsity in Preferences and Spending:*

Let $\theta = [\beta'_{01}, \ldots, \beta'_{0J}, \beta'_1, \pi'_{01}, \ldots, \pi'_{0J}, \pi'_1, \pi'_z]'$. Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^{\infty}$ has K_T possible parameters. Allowing for the possibility that $K_T > T$, only $k_T < T$ of these parameters are non-zero. We also allow for the possibility that both $K_T \underset{T\to\infty}{\to} \infty$ and $k_T \underset{T\to\infty}{\to} \infty$, but we fix the number of excluded instruments in z at $L \geq 2$.

1. *The parameter space isn't too large, with $\log(K_T) = o\left(T^{-1/3}\right)$.*

2. *The model is sparse, so the number of non-zero variables $k_T^2 \log^2(K_T \vee T) = o\left(T^{-1}\right)$*

3. *The Gram matrix satisfies a sparse eigenvalue condition that ensures finite-sample identification of the sparse model.*

4. *Non-zero coefficients are bounded away from zero.*

5. *The distribution of controls, instruments, and vote shares have exponential tails.*

A more detailed exposition of Assumption 3.2 appears in Appendix B.2.2, along with additional commentary. This assumption consolidates the restrictions in Belloni et al. (2013a)'s (ASTE) condition with Belloni et al. (2012)'s (AS) condition in an exactly sparse specification given a fixed number of excluded instruments. The first two conditions ensure that the true model has

[4]Note that our limits are taken with respect to a large number of markets ($T \to \infty$) for a fixed number of products competing in each market. Other analyses of demand data consider the problem where $J \to \infty$; we maintain J as fixed. We could allow J to grow with the main restriction that $\log(JK_0 + K_1) = o\left((JT)^{1/3}\right)$, which is already satisfied by Assumption 3.2.1.

sufficient structure and enough zeros to be estimable with available data. The third assumption ensures that, even though the complete empirical Gram matrix ("$X'X$") will be non-invertible (because $K_T > T$), the sub-matrices formed from the variables associated with non-zero coefficients are almost surely well-behaved. The fourth assumption ensures that non-zero coefficients satisfy the conditions for an oracle inequality to be selected by a Lasso estimator. The last assumption allows the application of large deviation theory to bound the probability of mis-classifying zero-coefficients with a fixed penalty weighting.

Additionally, we require regularity conditions on variability in the data generating process summarized in Assumption 3.3 for well-behaved asymptotic properties of the post-selection estimator. For completeness, the technical details and additional discussion appear in Appendix B.2. These conditions present a special case of the regularity conditions embedded in Belloni et al. (2013a) and Belloni et al. (2012)'s condition RF. In their detailed remarks, BCH and BCCH present a number of plausible sufficient conditions that illustrate these assumptions are not overly restrictive. These restrictions are sufficient to apply the asymptotic results for post-selection estimators established in Belloni et al. (2012).

Assumption 3.3. *High-Dimensional Linear Logit Regularity Conditions:*

1. *Sufficient moments for unmodeled variability in the data admit a LLN and CLT.*
2. *Variability in observables and their impact on unobservables is bounded.*
3. *Regularity conditions for asymptotic theory with i.n.i.d. sampling.*
4. *Regularity conditions for optimal instruments from first-stage regression.*

3.5.2 Selecting Control Variables for Inference

Our inference strategy proceeds in two stages of penalized estimation for selecting control variables followed by a two-state-least-squares estimator using the selected controls in an unpenalized model. The two stages of selection reflect our need to model conditional expectations for the expected impact of control variables on both (i) the campaign spending treatment variable and (ii) the vote share outcome. We perform each of the variable selection exercises using a lasso regression, which minimizes the sum of squared residuals

Algorithm 1 Post-Selection Estimation and Inference on Treatment Effects after Double-Selection of Controls in High-Dimensional Logit Voting Model

I. **Select controls for expected vote share.** Let $x^I \equiv \left\{x|\tilde{\beta}_I(x) \neq 0\right\}$, where:

$$\tilde{\beta}_I = \underset{\beta \in \mathbb{R}^{K_T+1}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p\right)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_\beta \beta\|_1.$$

II. **Select controls for campaign spending.** Let $x^{II} \equiv \{x|\tilde{\omega}(x) \neq 0\}$, where:

$$\tilde{\omega} = \underset{\omega \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1\right)^2 + \frac{\lambda_\omega}{T} \|\hat{\Upsilon}_\omega \omega\|_1.$$

III. **Post-Selection Estimation and Inference.** Let $\tilde{x} = x^I \bigcup x^{II}$ and compute the unpenalized IV regression:

$$\mathcal{S}_{jt} = \tilde{x}'_{0t}\beta_{0j} + \tilde{x}'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}, \ \mathbb{E}[\xi_{jt}|p_{jt}, \tilde{x}_{0t}, \tilde{x}_{1jt}, z_{jt}] = 0.$$

Details: $\lambda_{(\beta)} = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2K_T + 2))$ and $\lambda_{(\omega)} = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2K_T))$, with $c = 1.1$ and $\gamma = \frac{0.05}{\log(K_T \vee T)}$, satisfying the restrictions $c > 1$ and $\gamma = o(\log(K_T \vee T))$. $\hat{\Upsilon}_\beta$ is a diagonal matrix whose ideal (k,k) entry is $\sqrt{\bar{E}\left[x^2_{k,jt}\epsilon^2_{jt}\right]}$, with x_k representing the k^{th} regressor and residuals $\epsilon_{jt} = \mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p$. $\hat{\Upsilon}_\omega$ is defined analogously for the regression in step (II). Since the residuals are unobserved, these penalty loadings are feasibly calculated using the iterative algorithm presented in Appendix B.1.

subject to an L_1 penalty on the regression coefficients. For ease of reference, we summarize our inference approach in Algorithm 1.

Our variable selection exercise identifies the controls that are important for predicting vote shares and campaign spending. For both of these stages, we apply the iterated Lasso estimator proposed in Belloni et al. (2012) for heteroskedastic, non-Gaussian models. In short, candidate and demographic characteristics that drive variation in p but have little explanatory power for vote shares tend not to be selected by the first lasso (Step I), inviting an omitted variable bias. The second stage of selection (Step II) mitigates this potential distortion by explicitly modeling the campaign spending process.

The Lasso estimator's loss function represents a convex optimization problem with a penalty that enforces many of the estimated coefficients to be exactly zero. As such, it presents a computationally tractable and convenient model selection device. We note here that we allow for the demographic variables selected as relevant interactive fixed-effects to differ across parties. For instance, Unemployment may be a relevant control for the PRI party but not for the other parties. This yields a very flexible control strategy without introducing six new parameters for each demographic variable under consideration.

3.5.3 Post-Selection Estimation and Inference

The previous subsections focus on penalization strategies for estimating sparse models, particularly as a device for selecting which variables to include and which to exclude from the model. We now take these selected variables and consolidate them into a post-selection model for unpenalized estimation via a control function approach. We collect the controls selected by either of the variable screening devices into $\tilde{x} = x^I \cup x^{II}$, the number of which we denote by $\tilde{k}_T$. Note that we allow each of the parties to have fixed-effects for different demographic variables and, as such, no longer distinguish between demographic and candidate-specific control variables.

After selection, we can use the usual 2SLS approach to estimate the effect of campaign spending on vote share. Estimating the unpenalized, post-selection, first-stage regression:

$$p_{jt} = \tilde{x}'_{jt}\pi_x + z'_{jt}\pi_z + \tilde{\nu}_{jt}, \text{ gives fitted values } \tilde{p}_{jt} = \tilde{x}'_{jt}\tilde{\pi}_x + z'_{jt}\tilde{\pi}_z. \tag{3.9}$$

This first-stage regression gives Belloni et al. (2012)'s optimal instruments in the presence of a high-dimensional vector of controls with a fixed number of instruments. To characterize the residual variation in $\tilde{p}$ after controlling for the selected demographic variables, which will characterize the denominator for our standard errors, compute the regression:

$$\tilde{p}_{jt} = \tilde{x}'_{jt}\psi_x + \breve{p}_{jt}, \text{ which gives fitted residuals } \hat{\breve{p}}_{jt} = \tilde{p}_{jt} - \tilde{x}'_{jt}\tilde{\psi}_x. \tag{3.10}$$

Finally, we use the first-stage estimate of exogenous variation in campaign spending and the selected control variables in our second-stage regression.

$$\mathcal{S}_{jt} = \tilde{p}_{jt}\beta_p + \tilde{x}'_{jt}\beta_x + \xi_{jt}, \ \mathbb{E}[\xi_{jt}|\tilde{x}_{jt}, \tilde{p}_{jt}] = 0. \tag{3.11}$$

The second-stage results provide a consistent estimator for the treatment effect of campaign spending on candidate vote share. Our regularity conditions imposed thus far are sufficient to yield asymptotic normality of the estimate for β_p, allowing the application of standard hypothesis tests. This result follows immediately from Theorem 3 in Belloni et al. (2012).

Theorem 3.1 (Inference on Returns to Campaign Spending under Sparsity)**.** *Suppose Assumptions 1-3 hold. The estimated treatment effect of campaign spending on candidate vote share, $\hat{\beta}_p$, from fitting equation (3.11) is asymptotically normal:*

$$V_p^{\infty-1/2}\sqrt{JT}\left(\hat{\beta}_p - \beta_p\right) \rightarrow_d N\left(0,1\right), \ \textit{with}\ V_p^{\infty} = \bar{E}\left[\breve{p}_{jt}^2\right]^{-2}\bar{E}\left[\breve{p}_{jt}^2\xi_{jt}^2\right].$$

Here $\breve{p}_{jt}$ is the residual exogenous variation in the optimal instrument after regressing $\tilde{p}_{jt}$ on $\tilde{x}_{jt}$. Letting $\hat{\xi}_{jt} = \mathcal{S}_{jt} - \tilde{p}_{jt}\hat{\beta}_p - \tilde{x}'_{jt}\hat{\beta}_x$ represent the residuals from the regression (3.11), define:

$$\hat{\Omega} \equiv \frac{1}{JT}\sum_{j,t=1}^{J,T}\hat{\breve{p}}_{jt}^2 \rightarrow_p \bar{E}\left[\breve{p}_{jt}^2\right]^{-2},$$
$$\hat{\Sigma} \equiv \frac{1}{JT}\sum_{j,t=1}^{J,T}\sum_{k,s=1}^{J,T}\rho_{\{jt,ks\}}\hat{\breve{p}}_{jt}\hat{\breve{p}}_{ks}\hat{\xi}_{jt}\hat{\xi}_{ks} \rightarrow_p \bar{E}\left[\breve{p}_{jt}^2\xi_{jt}^2\right],$$

where $\rho_{\{jt,ks\}}$ represents the regularization coefficient for a HAC variance estimator. Consequently, V_p^{∞} can be consistently estimated using $\hat{V}_p^{\infty} \equiv \hat{\Omega}^{-1}\hat{\Sigma}\hat{\Omega}^{-1}$, and replacing $V_p^{\infty-1/2}$ with $\hat{V}_p^{\infty-1/2}$ preserves the t-statistic's asymptotic standard normal distribution.

3.5.4 Returns to Campaign Expenditures after Control Selection

We now investigate the robustness of our findings on campaign spending from section 3.4.2 to the fixed specification. Instead of restricting our attention to a fixed set of control variables, we allow for fixed effects driven by any of the 210 demographic variables captured in the Mexican census either linearly, quadratically, or in logs. With six parties' fixed effects, the most unstructured representation of this model allows for over 3,780 parameters when we have only 300 markets and 1,301 total observed market shares.

Table 3.3's Panel A presents the headline results for the flagship specification with the tuning parameters $c = 1.1$ and $\gamma = 0.05/\log(T \vee K_T)$. This double-lasso model, which includes 20 interactive fixed effects, is much more parsimonious than the pre-selected model, which allowed for 90 such free parameters. With the reduced number of included controls, the magnitude of the effect of \$10,000 in campaign expenditures on vote share declines from 0.63% to 0.42%. However, this reduction in magnitude brings with it much lower standard errors, resulting in a much more significant t-Statistic. The second-order effect is also somewhat reduced, from -0.033% to -0.019%, which remains significant at reasonable confidence levels though slightly less so than in the pre-selected model.

Panel B in Table 3.3 reflects the sensitivity of these estimates to the tuning parameters c and γ, illustrating the robustness of these results. Under reasonable specifications of the penalty, the estimated first-order effect ranges from 0.378% to 0.426% with all t-Statistics greater than 3. The second-order effect maintains its significance in these specifications as well, with an effect size ranging from -0.014% to -0.027%. At extreme levels of penalization, the magnitude of the contribution diminishes severely and loses statistical significance in a model that includes only three interactive fixed-effects. These findings are consistent with the guidance by Belloni et al. (2012) that the value for c should be set close to unity and that the selection mechanism may encounter problems when c becomes too large.

Finally, Panel C reports the number of selected control variables under the flagship variable-selection specification. These results illustrate how large penalty terms effectively reduce the model to one in which no controls are included to mitigate observed heterogeneity in voting preferences and campaign expenditures. Interestingly, the number of included controls roughly matches the number of significant controls from the pre-specified model. However, we note that these controls are all different.

3.6 Model RC-F: Heterogeneous Impressionability with Fixed Controls

The linear vote share model imposes a strong homogeneity restriction on voter preferences within any given district with potentially material empirical implications. In the literature on demand estimation, one approach to allowing for heterogeneity in preferences allows for random-coefficients in the individual's utility model. We incorporate this heterogenity here by allowing voters to have heterogeneous impressionability, introducing a random-coefficient to the individual influence of campaign spending. Among other sources, these ran-

Table 3.3: Campaign Expenditure and Vote Shares with Control Selection (Model LM-S)

Panel A: Main Results

	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.423	0.098	4.32	0%
Expenditures^2	-0.019	0.006	2.97	0%

Panel B: Robustness to Tuning Parameters

	Expenditure - Coefficient (Std Err)					Expenditure^2 - Coefficient (Std Err)			
	$\gamma\log(K_T \vee T)$					$\gamma\log(K_T \vee T)$			
c	0.01	0.05	0.1	0.2	c	0.01	0.05	0.1	0.2
1	0.423 (0.098)	0.421 (0.098)	0.421 (0.098)	0.491 (0.149)	1	-0.019 (0.006)	-0.019 (0.006)	-0.019 (0.006)	-0.027 (0.009)
1.1	0.426 (0.097)	0.423 (0.098)	0.423 (0.098)	0.423 (0.098)	1.1	-0.019 (0.006)	-0.019 (0.006)	-0.019 (0.006)	-0.019 (0.006)
1.25	0.381 (0.098)	0.407 (0.098)	0.398 (0.096)	0.426 (0.097)	1.25	-0.014 (0.007)	-0.016 (0.007)	-0.017 (0.006)	-0.019 (0.006)
1.5	0.378 (0.101)	0.379 (0.098)	0.381 (0.099)	0.381 (0.099)	1.5	-0.014 (0.007)	-0.014 (0.007)	-0.014 (0.007)	-0.014 (0.007)
1.75	0.412 (0.125)	0.387 (0.102)	0.397 (0.102)	0.378 (0.101)	1.75	-0.015† (0.008)	-0.015 (0.007)	-0.015 (0.007)	-0.014 (0.007)
2	0.102† (0.130)	0.182† (0.139)	0.225† (0.119)	0.412 (0.125)	2	0.004† (0.009)	0.000† (0.010)	-0.003† (0.008)	-0.015† (0.008)

† Statistically insignificant at the 5% Level

Panel C: Number of Selected Interactive Fixed Effects

	c					
$\gamma\log(K_T \vee T)$	1.00	1.10	1.25	1.50	1.75	2.00
0.01	20	19	15	11	8	3
0.05	21	20	17	12	11	5
0.1	21	20	18	14	10	6
0.2	24	20	19	14	11	8

This table reports the contribution of campaign expenditures to a candidate's vote share after data-driven selection of demographic controls for interactive fixed-effects using Algorithm 1. Panel A reports the first- and second-order contribution for a benchmark specification with tuning parameters $c = 1.1$ and $\gamma\log(K_T \vee T) = 0.1$. Panel B reports the robustness of this result with respect to these tuning parameters. Panel C indicates the number of interactive fixed-effects included under each of the model specifications.

dom coefficients could reflect heterogeneous levels of attention paid by different voters in the population.

3.6.1 GMM Estimation with the BLP Model

We begin by focusing on the structural model for preferences without addressing the variable selection step necessary to address the high dimensionality of the problem. To model heterogeneity in preferences, let an individual voter's preference for candidate j be defined as:

$$u_{ijt} = x_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_{0p} + p_{jt}b_{ip} + \xi_{jt} + \epsilon_{ijt}, b_{ip} \sim N\left(0, v_p^2\right). \qquad (3.12)$$

Conditional on b_{ip}, when ϵ_{ijt} has the usual Type-I extreme value distribution, voter i's decision will be governed by the logit choice probabilities:

$$Pr\left\{y_{ijt} = j|b_i\right\} = \frac{\exp\left\{x_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + p_{jt}b_i + \xi_{jt}\right\}}{1 + \sum_{r=1}^{J} \exp\left\{x_{0t}\beta_{0r} + x'_{1rt}\beta_1 + p_{rt}\beta_p + p_{rt}b_i + \xi_{rt}\right\}}. \qquad (3.13)$$

Since these individual shocks to voter sensitivity aren't observed, we need to integrate equation 3.13 to compute the expected vote share for a candidate in the district. Letting Φ represent the standard normal distribution's cumulative density, gives:

$$s_{jt} = \int Pr\left\{y_{ijt} = j|b_i\right\} d\Phi\left(b_i/v_p\right). \qquad (3.14)$$

The nonlinear form of the expected vote share in equation 3.14 complicates inference because we can no longer transform the vote shares into a generalized linear model. As such, we follow Berry et al. (1995)'s (BLP) approach to identify the model by exploiting the exogeneity of the party-district specific shocks to expected utility. Given a candidate specification for parameter values θ, Berry et al. (1995) show that a contraction mapping recovers these shocks, which we denote $\xi_{jt}\left(\theta, X, p, s\right)$. Under the true values for θ, the instruments are orthogonal to these shocks, i.e., $\mathbb{E}[\xi_{jt}\left(\theta, X, p, s\right)|z_{jt}] = 0$, so that θ is estimated using a GMM objective function with weighting matrix W:

$$Q\left(\theta, x, z, p, s\right) = \frac{1}{JT}\xi_{jt}\left(\theta, X, p, s\right)' zWz'\xi_{jt}\left(\theta, X, p, s\right). \qquad (3.15)$$

In the standard setting with a fixed number of controls and instruments, for any positive-definite W, minimizing equation 3.15 provides an asymptotically normal estimator for the parameters in θ. To address numerical issues in the evaluation of this estimator, Dube et al. (2012) present an MPEC algorithm, which we also use here.

One last sensitivity associated with the GMM objective function above relates to the instruments themselves. Berry et al. (1999) present an early discussion on the importance of using Chamberlain (1987) optimal instruments in evaluating (3.15). Gandhi and Houde (2015) illustrate how to utilize vote shares themselves as valuable instruments. Reynaert and Verboven (2014) illustrate how sensitive the estimator is to implementation with optimal instruments, particularly with respect to estimating the variance parameters v_p. Our implementation adopts this latter approach, since the Reynaert and Verboven (2014) instruments are easily recovered from the gradient of the constraints in the MPEC algorithm.

Assumption 3.4. *BLP Truthful Voting Structural Model:*

1. *In each of T districts, a large number of voters truthfully vote for the candidate they most prefer given the utility specification for u_{ijt} in equation 3.12.*

2. *Expected vote shares are non-linear in K_0 district characteristics x_{0t}, K_1 candidate characteristics x_{1jt}, and campaign spending p_{jt} as in equation 3.14.*

3. *Campaign spending is linear in the district and candidate characteristics x_{0t} and x_{1jt} as well as L exogenous instruments z_{jt}, as in equation 3.4.*

4. *Residual shocks to expected voter preferences in a district are endogenously correlated with unmodeled variation in campaign spending, as in equation 3.5, but have a zero conditional expectation given district and candidate district characteristics and observable instruments: $\mathbb{E}[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0$.*

Assumption 3.5 contains regularity conditions, which are fairly standard for GMM estimation with i.n.i.d. data (additional technical details for these conditions are presented in Appendix B.2). By textbook analysis, assumptions 3.4 and 3.5 are sufficient to establish the usual consistency and asymptotic normality results for the value of θ that minimizes equation (3.15).

Assumption 3.5. *Regularity Conditions for GMM Estimator:*

1. *Compactness of parameter space: The true parameter values $\theta_0 \in \Theta_{K_T}$, where $\Theta_K \subset \mathbb{R}^{K_T+2}$ is compact, with a compact limit set $\Theta_\infty \equiv \lim_{T\to\infty} \Theta_K$.*

2. *Continuity and differentiability of sample-analog and population moment conditions in parameter space.*

3. *Letting* $g_{jt}(\theta) \equiv \left[x'_{0t}, x'_{1jt}, z'_{jt}\right]' \xi_{jt}(\theta)$*, a uniform law of large numbers ensures the sample-analog,* $\frac{1}{JT}\sum_{j,t=1}^{JT} g_{jt}(\theta)$*, converges to the population moment condition.*

4. *A uniform law of large numbers applies to Hessian of sample analog to the population moment condition,* $\hat{G}_T(\theta) \equiv \frac{1}{JT}\sum_{j,t=1}^{J,T} \frac{\partial g_{jt}(\theta)}{\partial \theta'}$.

5. *The weighting matrix,* W_T*, is positive definite and converges to* W*, a symmetric, positive definite, and finite matrix.*

6. *The expected outer product of the score,* $\Omega \equiv \lim_{T\to\infty} (JT)^{-1} \sum_{j,t=1}^{J,T} \mathbb{E}\left[g_{jt}(\theta)\, g_{jt}(\theta)'\right]$*, is a positive definite, finite matrix.*

7. *The matrix* $\Sigma \equiv G(\theta_0)' \Omega^{-1} G(\theta_0)$ *is almost surely positive definite and finite.*

3.6.2 Campaign Spending with Heterogeneous Impressionability

We again begin our empirical analysis of heterogeneous impressionability in voting for Mexico using the pre-specified set of controls considered in table 3.1. These results allow us to differentiate the influence of heterogeneity in the model from the effects of control selection.

Table 3.4's Panel A reports the expected coefficients and standard deviation of coefficients associated with campaign expenditures' influence on voters' latent utility. The results indicate that heterogeneous impressionability is not a prominent feature of preferences, as revealed through the low variance of the coefficients themselves, which are not statistically distinguishable from zero.

Partly as a consequence of this limited heterogeneity, the expected coefficients are rather similar to those reported in Table 3.3. Estimating standard errors with the sandwich covariance matrix we find weaker, but still significant, evidence for the significance of campaign expenditure on candidate vote share. The linear term's p-value rises to 4% and the negative quadratic effect loses statistical significance with a p-value of 17%.

As in our analysis of the linear model, 3.4's Panel B reports the significance of the demographic controls included in the model with heterogeneous impressionability. With the higher variance of the estimates under the random coefficients specification, we find a slightly smaller share of the interactive fixed effects achieve the threshold for a statistically significant influence on voter preferences.

Table 3.4: Campaign Expenditure and Vote Shares with Heterogeneous Impressionability (Model RC-F)

Panel A: Main Results

Expected Coefficients	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.77	0.38	2.04	4%
Expenditures^2	-0.04	0.03	-1.39	17%
Variance of Coefficients	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.11	0.40	0.27	79%
Expenditures^2	0.00	0.04	0.00	100%

Panel B: Significance of Demographic Controls (p-Values)

Party	Region 1	Region 2	Region 3	Region 4	Region 5
MP	15%	0%*	84%	35%	85%
NA	2%*	11%	52%	41%	53%
PAN	0%*	0%*	4%*	10%	47%
PRI	62%	46%	89%	28%	58%
PVEM	11%	43%	1%*	95%	10%
CXM	32%	98%	10%	29%	27%
			% of Households with		
Party	Unempl	Car	Refrigerator	Utilities	Female Head
MP	29%	0%*	14%	18%	5%
NA	24%	51%	0%*	0%*	71%
PAN	50%	30%	0%*	44%	96%
PRI	40%	2%*	22%	10%	5%*
PVEM	35%	16%	35%	58%	44%
CXM	28%	90%	64%	86%	4%*
Party	Pop 18-24	Pop 65+	Married	Element Ed +	Avg Years School
MP	81%	92%	50%	13%	2%*
NA	6%	2%*	2%*	12%	7%
PAN	95%	2%*	23%	6%	46%
PRI	36%	99%	35%	13%	77%
PVEM	1%*	30%	65%	70%	4%*
CXM	38%	52%	10%	58%	50%

Panel A reports the return to campaign expenditures and squared campaign expenditures using the nonlinear BLP voting model with heterogeneous impressionability estimated from equation (3.15) with interactive fixed effects between the political party and demographic controls listed in Table 3.1. Panel B reports the significance of each of the interactive fixed-effects by party, with * and ** indicating significance at the 5% and 1% levels, respectively.

3.7 Model RC-S: Variable Selection in the BLP Voting Model

We now address the implications of the high-dimensional setting on the BLP model, particularly when there are more control variables than observations. Inference in this setting is non-trivial, since the model is unidentified in finite samples. That is, there exists a multiplicity of values for the parameters θ for which the residual shock to preferences, ξ_{jt}, can be equal to zero for all observations. As in the linear setting, a sparsity condition on the coefficients suggests incorporating a penalty to the criterion function that admits consistent inference. Our approach builds on the proposed technique from Gillen et al. (2014), which presents a model for inference in demand models after selection from a high-dimensional set of product characteristics when the number of control variables is of the same order of magnitude as the number of markets.

Our analysis here extends the Gillen et al. (2014) approach to a "non-polynomial" setting that allows the number of possible control variables to grow exponentially with the number of markets. Implementing selection in this "ultra-high" dimensional setting must confront two significant complications. The first is computational, as optimizing nonlinear objective functions with a non-polynomial number of parameters is simply infeasible in most circumstances. The second is analytical, in that we must apply the oracle properties established in Fan and Liao (2014)'s analysis of penalized estimation in high-dimensional GMM problems. We begin this section by discussing the conditions for valid inference under a penalized GMM objective function before turning to the computational issues and empirical results.

3.7.1 Sparsity Assumptions and Regularity Conditions

We begin by laying out the sparsity conditions required for establishing the oracle properties from Fan and Liao (2014)'s analysis of penalized estimation in high-dimensional GMM problems. These properties allow us to generalize the results from Gillen et al. (2014) to apply to a higher-dimensional setting. Gillen et al. (2014) relied on the asymptotic theory of Caner and Zhang (2013), who establish oracle properties for Zou and Hastie (2005)'s Elastic Net in an environment where the number of parameters grows more slowly than the number of observations, so that $K_T/T \to 0$. In contrast, the oracle properties in Fan and Liao (2014) apply to the setting where the dimensionality grows non-polynomially with the sample size, requiring only that $\log(K_T)/T^{1/3} \to 0$. These requirements are summarized in Assumption 3.6, which slightly tightens the restrictions of the linear model in Assumption 3.2, with the technical details relegated to Appendix B.2.

Assumption 3.6. *Sparsity Assumptions for High-Dimensional BLP Model:*

Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^{\infty}$, has $K_T > T$ possible parameters, $k_T < T$ of which are non-zero, where both $K_T \underset{T\to\infty}{\to} \infty$ and $k_T \underset{T\to\infty}{\to} \infty$. Further, the number of excluded instruments in z is fixed at $L \geq 2$.

1. *The parameter space isn't too large, with $\log(K_T) = o\left(T^{-1/3}\right)$.*
2. *The model is sparse, with the number of non-zero variables, $k_T^3 \log k_T = o\left(T^{-1}\right)$*
3. *The Gram matrix for controls with non-zero influence on vote shares is almost surely postive definite with finite eigenvalues.*
4. *The Hessian of the objective function with respect to non-zero variables is almost surely positive definite.*
5. *Non-zero coefficients are bounded away from zero.*
6. *The marginal distributions for controls, instruments, and residual vote shares have exponentially decaying tails.*

These assumptions are sufficient to establish an oracle property for the lasso-penalized GMM estimator, ensuring that the penalized estimator accurately identifies all non-zero coefficients and effectively thresholds all irrelevant coefficients at zero. Coupled with the continuity and uniform laws of large numbers of the GMM objective function from assumption 3.5, the results from Fan and Liao (2014) establish sufficient conditions on the penalty term for the penalized GMM estimator to achieve the near oracle convergence rate.

The first stage of inference requires solving a penalized GMM objective function with a lasso penalty. Similar to the linear demand model, we apply a data-dependent penalization that is robust to heteroskedasticity in sampling across markets:

$$\tilde{\theta} = \underset{\theta \in \mathbb{R}^{K_T+2}}{\arg\min}\, Q\left(\theta, x, z, p, s\right) + \frac{\lambda_\theta}{T} \|\hat{\Upsilon}_\theta \theta\|_1. \tag{3.16}$$

Our penalty loading, $\lambda_\theta = 2c\sqrt{T}\Phi^{-1}\left(1-\gamma/(2\left(K_T+4\right)\right)$ satisfies the restrictions in Fan and Liao (2014) for a lasso penalty loading to be $k_T\sqrt{k_T}/\sqrt{T} \prec \lambda_\theta/T \prec 1/\sqrt{k_T}$ when $c = 0.05$ and $\gamma = 0.1/\log\left(K_T \vee T\right)$. As in the linear model, Υ_θ is a diagonal matrix, with the ideal weights for $\beta_{0j,k}$ equal to $\sqrt{\bar{E}\left[x_{0t,k}^2\xi_{jt}^2\right]}$, for $\beta_{1,k}$ equal to $\sqrt{\bar{E}\left[x_{1jt,k}^2\xi_{jt}^2\right]}$, and for β_p equal to $\sqrt{\bar{E}\left[p_{jt}^2\xi_{jt}^2\right]}$. For the heterogeneity coefficient, v_p, the ideal value in the Υ_θ matrix is $\sqrt{\bar{E}\left[\frac{\partial\xi_{jt}(\theta,x,z,p,s)}{\partial v_p}^2\xi_{jt}^2\right]}$. Since ξ_{jt} is unobserved, Appendix B.1 reports the feasible iterated algorithm used to calculate Υ_θ.

3.7.2 Implementing Variable Selection via Penalized GMM

We now describe the approach we use for selecting variables using a penalized GMM estimator, since it is computationally infeasible to directly optimize the GMM objective function in extremely high-dimensional problems. We begin by fitting the nonlinear model with heterogeneous impressionability to the model selected by Algorithm 1. Within this fitted model, we can approximate the optimal instruments for the nonlinear features of the model, selecting the relevant demographic controls for observed heterogeneity in these features. We augment the selected demographic controls from Algorithm 1 with the controls for the latent utilities and optimal instruments for nonlinear features of the model. With this robustly augmented set of control variables, we compute unpenalized GMM estimates for the selected variables. We then verify that the first order conditions for the selected model are satisfied in the larger model with all included controls. The steps presented in Algorithm 2 summarize our approach to this implementation.

Our implementation picks up from where Algorithm 1 leaves off, defining the post-selection model including demographic controls $\tilde{x}$. With this model, we solve the GMM objective function without any penalization:

$$\tilde{\theta} = \arg\min Q\left(\theta, \tilde{x}, z, p, s\right). \tag{3.17}$$

Given the solution $\tilde{\theta}$, we can recover the latent mean utilities:

$$\tilde{\delta}_{jt} = \tilde{x}_{jt}'\tilde{\beta}_j + x_{1jt}'\tilde{\beta}_1 + p_{jt}\tilde{\beta}_p + \tilde{\xi}_{jt}.$$

These provide the outcome variable for which we need to select the relevant demographic controls using another application of the lasso. In parallel to the linear model's treatment, let $x^{III} \equiv \left\{x|\tilde{\phi}\left(x\right) \neq 0\right\}$, where:

$$\tilde{\phi} = \underset{\phi \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\tilde{\delta}_{jt} - x_{0t}'\phi_{0j} - x_{1jt}'\phi_1\right)^2 + \frac{\lambda_\phi}{T} \|\hat{\Upsilon}_\phi \phi\|_1. \tag{3.18}$$

The penalization term λ_ϕ has the same expected form as previous applications. The Υ_ϕ matrix requires a slight adjustment to account for estimation error in the $\tilde{\delta}_{jt}$'s. Defining $\epsilon_{\delta,jt} \equiv \delta_{jt} - \tilde{\delta}_{jt}$ and $\epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x_{0t}'\phi_{0j} - x_{1jt}'\phi_1$, the ideal weight for $\zeta_{0j,k}$ is equal to $\sqrt{\bar{E}\left[x_{0t,k}^2\left(\epsilon_{\delta,jt} + \epsilon_{\phi,jt}\right)^2\right]}$ and $\sqrt{\bar{E}\left[x_{1jt,k}^2\left(\epsilon_{\delta,jt} + \epsilon_{\phi,jt}\right)^2\right]}$ for $\beta_{1,k}$. The additional residuals can be characterized by using the asymptotic covariance matrix for $\tilde{\theta}$, which can be consistently estimated using the sandwich covariance matrix from the penalized GMM estimator.

We note that, by applying Algorithm 1, we have already selected the demographic controls necessary to explain observable variation in campaign expenditure. Now we select the demographic controls that explain variation across districts in the heterogeneity of impressionability. To do this, we need the optimal instruments for the heterogeneity parameters to identify the relevant controls for their first-order impact on model fit. Using the fitted model from equation (3.17), we compute the derivative of the objective function with respect to v_p:

$$\tilde{z}_{v,jt} = \frac{\partial}{\partial v_p} \xi_{jt}\left(\theta, \tilde{x}, z, p, s\right)|_{\theta = \tilde{\theta}}.$$

The formula for $\tilde{z}_{v,jt}$ from Berry et al. (1999) is presented in Nevo (2000)'s appendix and easily recovered from the Jacobian of the constraint for the MPEC objective function. We can then select demographic control variables that explain heterogeneity in this optimal instrument using a last application of the lasso estimator. Let $x^{IV} \equiv \left\{x|\tilde{\zeta}\left(x\right) \neq 0\right\}$, where:

$$\tilde{\zeta} = \underset{\zeta \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\tilde{z}_{v,jt} - x_{0t}'\zeta_{0j} - x_{1jt}'\zeta_1\right)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1. \tag{3.19}$$

Since $\tilde{z}_{v,jt}$ represents a generated regressor, we may wish to incorporate the variance induced by estimation error in its definition when determining the adapted penalty factor for equation (3.19), as in Υ_ϕ. However, by defining $\tilde{z}_{v,jt}$ as our identifying instrument, we only need to select demographic controls for variation in the *generated* $\tilde{z}_{v,jt}$ without regard to the population $z_{v,jt}$, which plays no direct role in our estimation. Consequently, when computing the values for Υ_ζ, we ignore the generated-regressors problem.

3.7.3 Post-Selection Inference via Unpenalized GMM

Combining the selected controls from Algorithm 1, $\tilde{x}$ with x^{III} and x^{IV}, define $\tilde{x}^* = \tilde{x} \bigcup x^{III} \bigcup x^{IV}$. We then compute the unpenalized, post-selection estimator

$$\tilde{\theta}^* = \arg\min Q\left(\theta, \tilde{x}^*, z, p, s\right). \tag{3.20}$$

To maximize the efficiency of our estimates, we first compute the optimal instruments for the Berry et al. (1995) model as discussed in Berry et al. (1999) and Reynaert and Verboven (2014). For the demographic controls and candidate characteristics, the selected variables themselves present the optimal instruments. We computed the optimal instruments for heterogeneity, $\tilde{z}_v$, in the variable selection stage. Finally, the optimal instruments for campaign expenditures can be easily estimated by an unpenalized first-stage regression which contains the selected controls and excluded instruments as regressors:

$$\tilde{z}_{p,jt} = \tilde{x}^{*\prime}_{jt}\hat{\pi}_x + z^{\prime}_{jt}\hat{\pi}_z.$$

Denoting the optimal instruments by $\tilde{z}$ and the selected control variables by $\tilde{x}$, we then compute the post-selection estimator for the voting model with heterogeneous impressionability as the solution to:

$$\theta = \underset{\theta \in \mathbb{R}^{k^*_T+2}}{\arg\min}\, Q\left(\theta, \tilde{x}, \tilde{z}, p, s\right). \tag{3.21}$$

The last step then verifies that this solution also satisfies the first-order conditions for the penalized objective function (3.16) to ensure we have not erroneously excluded any variables. We perform this test sequentially, evaluating the first-order conditions with respect to each excluded variable and verifying that they are dominated by the magnitude of the penalty term. As when calculating the optimal instruments for the variance parameters in the model, these gradients can be recovered from the Jacobian of the constraint in the MPEC objective function:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q\left(\tilde{\theta}^*, \tilde{x}^k, z, p, s\right) < \lambda_\theta \upsilon_k, k = 1, \ldots, K_0, j = 1, \ldots, J.$$

Any variables whose first-order conditions dominate the penalty should be included within the selected model. This requirement leads to an iterative process that, in our experience, converges within two iterations.

Given assumptions 3.5 and 3.6, the post-selection estimator for the treatment effects β_p and their heterogeneity v_p will be consistent and asymptotically normal. It is straightforward to show that the oracle property established in Fan and Liao (2014) satisfies the High-Level Model Selection condition in Belloni et al. (2012), giving the asymptotic result:

Theorem 3.2 (Inference on Returns to Campaign Spending under Sparsity). *Suppose Assumptions 4-6 hold. The estimated effect of campaign spending on mean latent voter utilities is asymptotically normal. That is:*

$$V_p^{\infty -1/2}\sqrt{JT}\left(\hat{\beta}_p - \beta_p\right) \to_d N\left(0,1\right).$$

The asymptotic variance V_p^{∞} is the element in the sandwich covariance matrix for $\tilde{\theta}^*$ corresponding to β_p. Specifically, defining the limit of the i.n.i.d. expectations:

$$G_0 \equiv \bar{E}\left[\frac{\partial g_{jt}\left(\theta\right)}{\partial \theta'}\right]|_{\theta=\theta_0},\ \Omega_0 \equiv \bar{E}\left[g_{jt}\left(\theta_0\right)g_{jt}\left(\theta_0\right)'\right],\ \textit{and}\ \Sigma_0 \equiv G\left(\theta_0\right)'\Omega^{-1}G\left(\theta_0\right),$$

then $V_\theta^{\infty} \equiv \left(G_0'G_0\right)^{-1}\Sigma_0\left(G_0'G_0\right)^{-1}$, which can be estimated with:

$$\hat{G}_T \equiv \frac{1}{JT}\sum_{j,t=1}^{J,T}\frac{\partial g_{jt}\left(\theta\right)}{\partial \theta'}|_{\theta=\tilde{\theta}^*} \to_p G_0,\ \textit{and}\ \hat{\Omega} \equiv \frac{1}{JT}\sum_{j,t=1}^{J,T} g_{jt}\left(\tilde{\theta}^*\right)g_{jt}\left(\tilde{\theta}^*\right)' \to_p \Omega_0.$$

Consequently, $\hat{\Sigma}_T \equiv \hat{G}_T'\hat{\Omega}\hat{G}_T \to_p \Sigma_0$ and

$$\hat{V}_\theta^{\infty} \equiv \left(\hat{G}_T'\hat{G}_T\right)^{-1}\hat{\Sigma}_T\left(\hat{G}_T'\hat{G}_T\right)^{-1} \to V_\theta^{\infty},$$

so replacing $V_p^{\infty-\frac{1}{2}}$ with $\hat{V}_p^{\infty-\frac{1}{2}}$ preserves the t-statistic's asymptotic normal distribution.

Further, if $v_p > \delta > 0$, the estimated variance of campaign spending's impact on voter utlities is also asymptotically normal with asymptotic variance of the estimate given by the sandwich covariance matrix.

3.7.4 Heterogeneous Impressionability after Control Selection

Even with only validating local optimality conditions for the selected model, practical implementation of Algorithm 2 is quite computationally intensive. In analyzing the Mexican voting data, fitting a single penalty specification requires approximately 80 core-hours of computation. Due to the intensive computational resources required for estimation, we focus our empirical analysis on the benchmark penalty specification where $c = 1.1$ and $\gamma \log\left(K_T \vee T\right) = 0.10$.

Panel A of Table 3.5 reports the impact of campaign expenditures on mean utilities. The estimated first-order mean effect of 0.465 is a bit lower than the pre-selected model's result of 0.767. This result matches the impact of variable selection on the estimated effect in the linear model, which dropped to 0.423 from 0.627 after variable selection. The standard error of the model with variable selection is lower than the pre-specified model's, yielding very similar t-Statistics for both the first- and second-order effects. As in the pre-selected model, the variance coefficients reflecting heterogeneity in preferences are indistinguishable from zero, indicating that there may not be much heterogeneity in voter impressionability.

Table 3.5: Campaign Expenditures and Candidate Vote Share with Heterogeneous Impressionability and Selected Controls (Model RC-S)

Panel A: Main Results

Expected Coefficients	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.465	0.198	2.35	2%
Expenditures^2	-0.020	0.015	-1.33	18%
Variance of Coefficients	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.05	0.67	0.07	94%
Expenditures^2	0.00	0.06	0.01	99%

Panel B: Selected Demographic Controls

MP	NA	PAN
Party FE (+)**	Party FE (–)	Party FE (–)**
Female Popn >60 (–)**	% Female HoH (–)	
	% Pop 18-24 (–)	
	% Pop Married (+)	
PVEM	% Pop w/Elementary Ed (–)*	**CXM**
Party FE (–)**	Female Popn >60 (–)	Party FE (–)**
% HH w/o Utils (–)	Total Popn >65 (–)	Region 4 Dummy (–)
% HH w/o Fridge (–)	Male Pop >15 w/Ed (–)*	% Female HoH (–)**
Total Popn >65 (–)	Popn in Private Dwell (+)**	Female Pop w/o Ed (+)**
Popn in Private Dwell (+)**	Avg Num People/Dwell (–)	Pop w/Social Medicine (+)
Avg Num People/Dwell (–)**		Travel Time to Polls (+)
	PRI	
	Party FE (–)**	

Panel B of Table 3.5 reports the actual demographic controls affecting voters' preferences for each party's candidates. For the two largest parties, PAN and PRI, only the party fixed effects were selected to control for voter preferences. For the coalition parties, more controls were incorporated to reflect heterogeneity in preferences. Interestingly, these demographic controls were most important for characterizing voter preferences for parties that sit in the middle of the policy spectrum. Panel B also indicates the significance and sign of the controls' impact on voter preferences. However, we caution against drawing too many conclusions from their selection. These controls were selected to identify the impact of campaign expenditures on voter preferences, not as independent causal elements driving voter preferences. As such, their selection may only be as proxies representing the effect of other variables on preferences and spending.

3.8 Generalizations and Extensions

Our results imposed an exact sparsity condition, requiring that the negligible coefficients in the model are exactly equal to zero. This restriction is quite a bit stronger than the approximate sparsity condition presented in Belloni et al. (2012) and Belloni et al. (2013a). The sampling properties of the penalized GMM estimator presented in Fan and Liao (2014) are robust to local perturbations, suggesting that exact sparsity could be weakened to the generalized forms of sparsity considered by Zhang and Huang (2008) and Horowitz and Huang (2010). Some additional regularity conditions may be required on the nonlinear features recovering latent mean utilities from observed market shares, but such an extension should be viable.

Though our analysis focused on GMM as an approach for estimating the voting model, alternative estimation strategies could also be considered after performing a selection step. Empirical likelihood approaches following Kitamura (2001) and Donald et al. (2003) have been adapted to demand estimation by Moon et al. (2014) and Conlon (2013). The Fan and Liao (2014) asymptotic analysis also applies to empirical likelihood estimators that could be mapped onto these techniques.

Though we allow for correlation among vote shares across parties within a district, we note that our analysis leans heavily on an independence assumption for sampling across districts. Limited spatial correlation could be accounted for by computing robust standard errors in estimating the covariance matrix of residuals. As long as strong-mixing and ergodicity conditions are met, this sort of dependence should not preclude effective variable selection.

There is some tension between our assumption of a linear campaign financing rule in light of a structural model of competition between parties. Indeed, Montero (2015) solves the equilibrium campaign financing rule in the Mexican election environment and shows it to be highly nonlinear. One way to address this issue characterizes the linearized campaign finance rule as an approximation to the structural finance rule, bounding the approximation error relative to instrumental variability, and showing that the approximation error doesn't affect variable selection and inference. Another strategy might adopt a control function approach to estimation, perhaps following Kawai (2014)'s strategy of incorporating techniques from production function estimation.

3.9 Conclusion

We present several results in high-dimensional inference and apply these techniques in an empirical analysis of voting behavior in Mexican elections. Our analysis applies high-dimensional inference techniques for estimating aggregate demand models with a very large number of demographic covariates. Though our statistical analysis is largely informed by previously established properties of these techniques, the extensions to the specific application are not trivial.

Our results show, robustly, that campaign expenditures have a significant and positive impact on voters' latent utilities for a candidate, with indications that the impact of these expenditures diminishes with the amount of campaign spending. Strikingly, we find little evidence of heterogeneity in voters' response to campaign expenditure, perhaps because limited variability in the slate of candidates provides little opportunity for this heterogeneity to impact vote shares.

Algorithm 2 Post-Selection Estimation and Inference with Double-Selection from High-Dimensional Controls in a Voting Model with Heterogeneous Impressionability

I. **Apply Algorithm 1 to select $\tilde{x} = x^I \bigcup x^{II}$ as the controls for a homogeneous model.**

II. **Compute GMM estimates for heterogeneous model using selected controls.** Let $\tilde{\delta}_{jt} \equiv \tilde{x}'_{jt}\tilde{\beta}_j + x'_{1jt}\tilde{\beta}_1 + p_{jt}\tilde{\beta}_p + \xi_{jt}\left(\tilde{\theta}, x, z, p, s\right)$, where $\tilde{\theta} \equiv \left[\tilde{\beta}'_1, \ldots, \tilde{\beta}'_J, \tilde{\beta}_1, \tilde{v}_p\right]'$:

$$\tilde{\theta} = \arg\min Q\left(\theta, \tilde{x}, z, p, s\right).$$

III. **Estimate optimal instruments for heterogeneity in impressionability.** Compute the derivative of the moment condition with respect to the variability parameter v_p:

$$\tilde{z}_{v,jt} = \frac{\partial}{\partial v_p}\xi_{jt}\left(\theta, \tilde{x}, z, p, s\right)|_{\theta=\tilde{\theta}}.$$

IV. **Select controls for mean utilities.** Let $x^{III} \equiv \left\{x|\tilde{\phi}\left(x\right) \neq 0\right\}$, where:

$$\tilde{\phi} = \underset{\phi\in\mathbb{R}^{K_T}}{\arg\min}\, \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1\right)^2 + \frac{\lambda_\phi}{T}\|\hat{\Upsilon}_\phi\phi\|_1.$$

V. **Select controls for optimal nonlinear instruments.** Let $x^{IV} \equiv \left\{x|\tilde{\zeta}\left(x\right) \neq 0\right\}$, where:

$$\tilde{\zeta} = \underset{\zeta\in\mathbb{R}^{K_T}}{\arg\min}\, \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1\right)^2 + \frac{\lambda_\zeta}{T}\|\hat{\Upsilon}_\zeta\zeta\|_1.$$

VI. **Post-selection estimation and inference.** Let $\tilde{x}^* = \tilde{x} \bigcup x^{III} \bigcup x^{IV}$ and compute the unpenalized GMM estimate:

$$\tilde{\theta}^* = \arg\min Q\left(\theta, \tilde{x}^*, z, p, s\right).$$

VII. **Verify First Order Conditions in Unselected Model.** For each excluded demographic control x_{0k}, define $\tilde{x}^k = \tilde{x}^* \bigcup x_{0k}$. Verify the first order improvement in the objective function from including this variable x_{0k} for any party is dominated by the penalty:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q\left(\tilde{\theta}^*, \tilde{x}^k, z, p, s\right) < \lambda_\theta \Upsilon_{\theta,(k,k)}, k = 1, \ldots, K_0, j = 1, \ldots, J.$$

VIII. **Add improperly excluded variables to the model and iterate.** Define the set of controls that fail to satisfy first order conditions in step (VII) as $x^V = \left\{x_k : q_k > \lambda_\theta \Upsilon_{\theta,(k,k)}\right\}$. Redefine $\tilde{x} = \tilde{x}^* \bigcup x^V$ and return to Step (II) until there are no changes in the set of included variables.

Details: $\lambda_\phi = \lambda_\zeta = 2c\sqrt{T}\Phi^{-1}\left(1 - \gamma/(2K_T)\right)$, and $\lambda_\theta = 2c\sqrt{T}\Phi^{-1}\left(1 - \gamma/(2K_T + 8)\right)$ with $c = 1.1$ and $\gamma = \frac{0.05}{\log(K_T \vee T)}$. The details for calculating the diagonal factor loading matrices $\hat{\Upsilon}_{(\cdot)}$, whose ideal entries reflect the square root of the expected product of the squared residual and control variable, are discussed in the text. The iterative algorithms by which we feasibly calculate these values are detailed in Appendix B.1.

Chapter 4

A THEORY OF EXPERIMENTERS

4.1 Introduction

As the use of experiments spreads in academia, business, and public policy, there has been a growing need for tools to increase the reliability of experimental findings. The experimental community has responded through the introduction of registries, and spirited discussions of practices such as pre-analysis plans, rerandomization, and statistical techniques (Bruhn and McKenzie, 2009; Deaton, 2010; Duflo et al., 2008; Humphreys et al., 2013; Imbens, 2010; Olken, 2015). Quite surprisingly, however, there is no comprehensive decision theoretic framework for optimal experiment design to guide these efforts. This paper seeks to provide such a framework.

A framework for optimal experiment design requires a model of experimenters and their objectives. While models of information acquisition feature prominently in modern microeconomic theory (Rothschild, 1974; Grossman and Stiglitz, 1980; Aghion et al., 1991; Bergemann and Välimäki, 1996; Persico, 2000; Bergemann and Välimäki, 2002, 2006), they fail to predict a key feature of the way scientists learn: by running randomized controlled trials (RCTs; see Kasy, 2013). The reason for this is that much of applied microeconomic theory models decision makers using subjective expected utility (Savage, 1954). Mixed strategies are never strictly optimal for such a decision maker. Since RCTs are mixed strategies over experimental assignments, they can never be strictly optimal.

As experimenters do, quite often, randomize their experimental allocations, it is important to understand why in order to establish experimental practices that are optimal *from their point of view.* We propose to replace subjective expected utility with ambiguity averse preferences, specifically minmax expected utility as axiomatized by Gilboa and Schmeidler (1989). In our model, an ambiguity averse decision-maker must make a binary policy choice $a \in \{0, 1\}$ affecting a population of individuals with characteristics $x \in X \subset \mathbb{R}^m$ and conditionally independent outcomes. To improve the quality of her decision-making, the decision-maker runs an experiment that assigns a given number N of participants to either treatment and control. Each experimental participant obtains an outcome $y \in \{0, 1\}$ observed by the decision-maker. The decision-maker's final policy choice depends on her beliefs and the experimental assignment and outcomes.

Under an innocuous assumption, we show that the decision maker can be thought of as maximizing the weighted average of a Bayesian subjective expected utility term and an adversarial maxmin-expected-utility term. We interpret this decomposition as a decision maker who has a subjective view of the world but must also satisfy an adversarial audience with veto power. The relative weights that the decision maker places on the subjective and minmax terms permit informative comparative static exercises.

The paper reports two main sets of results. Our first set shows describes when RCTs are optimal for an ambiguity averse decision maker. If the decision maker places non-zero weight on satisfying her adversarial audience, then, for sufficiently large sample sizes, it is always strictly optimal for the decision maker to use an RCT. Such trials permit robust, prior-free inference, and achieve assignment losses of order $\frac{1}{\sqrt{N}}$. Additionally, for any sample size, deterministic experiments are generically strictly optimal when the decision maker puts sufficiently high weight on the expected utility term.

These results are descriptively accurate of experimental practice: Randomized experiments are optimal when the decision maker puts high value on convincing an adversarial audience (science, pharmaceutical companies), or when the decision maker can afford large samples (experiments in online marketing). Whenever data points are expensive and the decision maker puts little weight on satisfying an adversarial audience, optimal experiments are deterministic, and finely optimize the subjective decision-making value of each acquired data point (firms testing new products in select markets, politicians testing platforms in specific states, etc...). Both in practice and in theory, RCTs are not always optimal, and our model fits the observed heterogeneity well.

Our second set of results exploits our framework to shed light on the important practical question of rerandomization (Morgan and Rubin, 2012). Rerandomization consists of drawing multiple treatment assignments, and choosing one that maximizes the balance between groups on some covariates. For example, a medical researcher may want to ensure that treatment and control groups are similar in terms of gender, age, race, and baseline health variables such as blood pressure and weight. Despite the practical ease of using rerandomization to ensure balance, there is the concern that it may affect the reliability of findings (Bruhn and McKenzie, 2009).

We show that the trade-offs at the heart of rerandomization are well captured in our framework. Successive rerandomizations improve balance, as captured by the subjective expected utility component of preferences. However, rerandomization reduces robustness, as captured by the adversarial component of preferences. In the extreme case where the allocation is rerandomized until perfect balance is achieved, the allocation is effectively deterministic and the adversarial term remains bounded away from first best.

While there exists a tradeoff between balance and robustness, we show that the costs of rerandomization are small. In order to significantly affect the robustness of decision making, the number of rerandomizations needs to be on the order of exponential in the sample size. Thus, at the conclusion of this set of results, we suggest a rule of thumb for rerandomization that will markedly improve balance, while keeping the robustness costs small.

In addition to the practical considerations just described, our work suggests that models of information acquisition need to take non-Bayesian perspectives seriously. As our work emphasizes, this allows for satisfactory microfoundations for some elements of experimental practice, and makes useful positive predictions. However, the usefulness of this perspective does not stop there: valid normative conclusions require accurate assessments of the underlying preferences and motives of experimenters and decision makers.

4.2 A Framework for Optimal Experiment Design

Decisions and payoffs. A decision maker chooses whether or not to implement a policy that provides a treatment $\tau \in \{0, 1\}$ to a unit mass of individuals indexed by $i \in [0, 1]$. Potential outcomes for subject i with treatment status $\tau_i \in \{0, 1\}$ are random variables $Y_i^\tau \in \{0, 1\}$; $Y = 1$ is referred to as a success. Each individual i is associated with covariates $x_i \in X \subset \mathbb{R}^m$, where X is finite. Covariates $x \in X$ are observable and affect the distribution of outcomes Y. The distribution $q \in \Delta(X)$ of covariates in the population is known and has full support, and outcomes Y_i are i.i.d. conditional on covariates. The probability of success given covariate x is denoted by $p_x^\tau \equiv \text{prob}(Y_i^\tau = 1 | x_i = x)$.

The state of the world is described by the the finite-dimensional vector p of success probabilities conditional on covariates, $p = (p_x^0, p_x^1)_{x \in X} \in [0,1]^{2X} \equiv P$. Note that state space P is compact, convex, and finite-dimensional. Given a state p and a policy decision $a \in \{0,1\}$, the decision maker's payoff $u(p,a)$ is

$$u(p,a) \equiv \mathbb{E}_p Y^a = \sum_{x \in X} q(x) p_x^a.$$

Experiments and strategies. To maximize her odds of making the correct policy choice, the decision maker can run an experiment on N participants. For simplicity, we assume that N is even, and exogenously given. Formally, an experiment is a tuple $e = (x_i, \tau_i)_{i \in \{1,\dots,N\}} \in (X \times \{0,1\})^N \equiv E$. Experiment e generates outcome data $y = (y_i)_{i \in \{1,\dots,N\}} \in \{0,1\}^N \equiv \mathcal{Y}$, with y_is independent realizations of $Y_i^{\tau_i}$ given (x_i, τ_i).

The decision maker's strategy consists of both an experimental design $\mathcal{E} \in \Delta(E)$—a mixed strategy over experimental assignments—and an allocation rule $\alpha : E \times \mathcal{Y} \to \Delta(\{0,1\})$ which maps experimental data (e,y) to policy decisions $a \in \{0,1\}$. We denote by $\mathcal{A}$ the set of such mappings. A stratified RCT, assigning a share $\pi \in (0,1)$ of participants to treatment $\tau = 1$, corresponds to a strategy $(\mathcal{E}_1, \alpha_1)$ such that:

- $\mathcal{E}_1$ samples N subjects with covariates independently drawn according to q;
- $\mathcal{E}_1$ assigns treatment $\tau_i = \mathbf{1}_{i \le \pi N}$;
- $\alpha_1(e,y) \equiv \mathbf{1}_{\overline{y}^1 \ge \overline{y}^0}$, where $\overline{y}^\tau \equiv \sum_{i=1}^N y_i \mathbf{1}_{\tau_i = \tau} \big/ \sum_{i=1}^N \mathbf{1}_{\tau_i = \tau}$.

Preferences. The decision maker is ambiguity averse with standard maxmin preferences (Gilboa and Schmeidler, 1989). She chooses a strategy $(\mathcal{E}, \alpha)$ that solves

$$\max_{\substack{\alpha \in \mathcal{A} \\ \mathcal{E} \in \Delta(E)}} U(\mathcal{E}, \alpha), \quad \text{where} \quad U(\mathcal{E}, \alpha) \equiv \min_{h \in \widehat{H}} \mathbb{E}_{h,\mathcal{E}}[u(p, \alpha(e,y))] \tag{4.1}$$

and $\widehat{H}$ is a convex set of priors $h \in \Delta(P)$ over states $p \in P$. This can be thought of as a zero-sum game in which nature picks distribution $h \in \widehat{H}$ after the decision maker picks a strategy $(\mathcal{E}, \alpha)$. Randomizations in mixed strategies are independent of moves by nature.

We use the usual statistical distance $d(h, h') \equiv \sup_{\substack{A \subset P \\ A \text{ meas.}}} |h(A) - h'(A)|$ on distributions whenever making genericity statements. Almost-surely statements are made with respect to the Lebesgue measure.

Equivalent experiments. As we are assuming that subjects are exchangeable conditional on covariates, experiments that differ only by a permutation of subjects with identical covariates are equivalent from a decision-making perspective. It is useful to formalize this point in the context of maxmin preferences.

Definition 4.1 (equivalent experiments)**.** *Two experiments* $e = (x_i, \tau_i)_{i \in \{1,\dots,N\}}$ *and* $e' = (x'_i, \tau'_i)_{i \in \{1,\dots,N\}}$ *are* equivalent*, denoted by* $e \sim e'$*, if there exists a permutation* $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$ *of the subjects' labels such that* $(x_i, \tau_i) = (x'_{\sigma(i)}, \tau'_{\sigma(i)})$ *for all* i*. The equivalence class of an experiment* e *is denoted by* $[e]$*.*[1] *We denote by* $[E]$ *the partition of possible experiments in equivalence classes. We say that two experimental designs* $\mathcal{E}$ *and* $\mathcal{E}'$ *are equivalent, denoted by* $\mathcal{E} \sim \mathcal{E}'$*, if they induce the same distribution over* $[E]$

Lemma 4.1. *Whenever* $\mathcal{E} \sim \mathcal{E}'$*,* $\max_{\alpha \in \mathcal{A}} U(\mathcal{E}, \alpha) = \max_{\alpha \in \mathcal{A}} U(\mathcal{E}', \alpha)$*.*

Thus, equivalent experiments guarantee the decision maker the same utility.

4.2.1 Key Assumptions

We place two additional assumptions on the model of Section 4.2. The first is innocuous, and allows the decision marker's objective to be expressed as a weighted average of a Bayesian subjective expected utility term and a maxmin expected utility term.

The second is more substantial: it ensures that the set of possible priors entertained by the decision-maker is rich enough that for any given experimental assignment, there exists a prior under which this assignment does not permit efficient decision-making. Note that the order of quantifiers is important: given a realized experimental assignment, we can find such a prior h. Indeed, we show in Proposition 4.3 that randomized experiments yield approximately efficient decisions for all priors.

[1] It is convenient to include distributions $\mathcal{E}$ with support in $[e]$ in the equivalence class of e.

Decomposition of Maxmin Preferences

The following assumption leads to a useful decomposition of the decision maker's preferences.

Assumption 4.1 (absolute continuity)**.** *There exist* h_0 *and* $\lambda \in (0, 1)$ *such that for every prior* $h \in \widehat{H}$ *and almost every state* $p \in P$,

$$h(p) \geq \lambda h_0(p). \tag{4.2}$$

Absolute continuity requirement (4.2) implies that every prior $\widehat{h} \in \widehat{H}$ can be written as $\widehat{h} = \lambda h_0 + (1-\lambda)h$, where $h \in H \equiv \frac{1}{1-\lambda}(\widehat{H} - \lambda h_0)$. Condition (4.2) moreover implies that elements $h \in H$ are themselves probability distributions over states $p \in P$, and the set H is also compact and convex.

Altogether, this implies that the decision maker's objective (4.1) can be rewritten as

$$U(\mathcal{E}, \alpha) \equiv \lambda \mathbb{E}_{h_0,\mathcal{E}}[u(p, \alpha(e, y))] + (1-\lambda) \min_{h \in H} \mathbb{E}_{h,\mathcal{E}}[u(p, \alpha(e, y))]. \tag{4.3}$$

Keeping h_0 and H fixed, parameter λ provides a convenient and continuous measure of the decision maker's degree of ambiguity aversion. With $\lambda = 1$ this nests standard subjective expected utility maximization; thus, we sometimes refer to such a decision maker as Bayesian.

This yields a useful interpretation: The decision maker wants to make a decision that is successful under her own subjective prior h_0, but also satisfies an audience of players with priors $h \in H$. Weights λ and $1-\lambda$ represent the respective weights that the decision maker places on her own subjective utility and that of her audience. In a traditional model of ambiguity aversion, the possible priors are all internal to the decision maker; here we add an additional interpretation, that they come from an adversarial audience.[2]

Limited Extrapolation

Throughout we assume that $N \leq |X|$ so that, even though there are finitely many covariate profiles $x \in X$, assigning each of them to treatment and control is not feasible. This condition is assumed to hold even as we take N to be large.

[2]Note that if the audience has veto power and enjoys some outside option, then the weight ratio $\frac{1-\lambda}{\lambda}$ would be interpreted as the appropriate Lagrange multiplier on satisfying the audience's individual rationality constraint.

This allows us to impose the following *limited extrapolation* condition on X, N, and H. Denote by $p^a \equiv \sum_{x \in X} q(x) p_x^a$ the expected probability of success given policy $a \in \{0,1\}$. Given an experiment $e = (\tau_i, x_i)_{i \in \{1,\cdots,N\}}$, denote by $\overline{p}_e \equiv (p_{x_i}^{\tau_i})_{i \in \{1,\cdots,N\}}$ the subset of success rates for experimental subjects. Vector $\overline{p}_e$ is thus an upper bound to the information generated by experiment e.

Assumption 4.2 (limited extrapolation)**.** *There exists $\epsilon > 0$ such that, for all $e \in E$, there exists a prior $h \in \arg\min_{h \in H} \mathbb{E}_h(\max_{a \in \{0,1\}} p^a)$ such that, for almost every $\overline{p}_e$,*

$$\min \left\{ \mathbb{E}_h \left[\max_{a \in \{0,1\}} p^a - p^0 \middle| \overline{p}_e \right], \mathbb{E}_h \left[\max_{a \in \{0,1\}} p^a - p^1 \middle| \overline{p}_e \right] \right\} > \epsilon.$$

Limited extrapolation implies that for any realized experimental assignment there exists a prior $h \in H$ under which that assignment does not allow for first-best decision-making. That is, conditional on the data generated by any experiment, their exists a prior such that the residual uncertainty about which policy maximizes population-level outcomes remains bounded away from 0.

4.3 Optimal Design and Randomization

We now use our framework to characterize optimal experimental design.

4.3.1 Bayesian Experimentation

When $\lambda = 1$, the decision maker is a standard subjective expected utility maximizer. It is immediate, and well understood, that in this case deterministic experiments are weakly optimal. We show that for generically every prior (that is, for an open and dense set of priors), deterministic experiments are in fact strictly optimal when λ is close to one.

Proposition 4.2 (near-Bayesians do not randomize)**.** *For generically every prior h_0, there exist $\underline{\lambda} \in (0,1)$ and a unique equivalence class of experiments $[e^*]$ such that for all $\lambda > \underline{\lambda}$, a (potentially mixed) experiment $\mathcal{E} \in \Delta(E)$ solves* (4.3) *if and only if* $\operatorname{supp} \mathcal{E} \subset [e^*]$.

In recent work, Kasy (2013) uses a specialization of Proposition 4.2 to the case where $\lambda = 1$ to conclude that RCTs are suboptimal. We believe that rather than invalidating the use of RCTs, Proposition 4.2 highlights the limits of subjective expected utility maximization as a suitable positive model of experimenters. We argue instead that the adversarial framework of (4.3) is much more successful in explaining the range of information acquisition strategies observed in practice.

4.3.2 Adversarial Experimentation

We now assume that the decision maker puts a fixed positive weight on satisfying her audience, and study comparative statics as the sample size becomes large.

Proposition 4.3. *Take weight $\lambda \in (0,1)$ as given. There exists $\underline{N}$ such that for all $N \geq \underline{N}$, any optimal experiment is randomized. More precisely, the following hold:*

(i) For any N, any optimal experiment $\mathcal{E}^*$ satisfies

$$\max_{\alpha} \min_{h \in H} \mathbb{E}_{h,\mathcal{E}^*} \left[u(p, \alpha(e,y)) \right] \geq \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \sqrt{\frac{1}{N}}.$$

(ii) For any N, all deterministic experiments $e \in E$ are bounded away from first-best:

$$\forall e \in E, \quad \max_{\alpha \in \mathcal{A}} \min_{h \in H} \mathbb{E}_{h,e} \left[u(p, \alpha(e,y)) \right] < \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \epsilon.$$

The first part of the proposition shows that the optimal experiment produces an erroneous policy decision less than $\sqrt{1/N}$ of the time. The second part shows that a deterministic experiment produces the incorrect policy ϵ of the time. Thus, as N grows, the optimal experiment cannot be deterministic: it must be randomized. Intuitively, the decision maker is playing a zero-sum game against nature (with probably $1 - \lambda$). After the decision maker picks an experiment, nature picks the prior which maximizes the chance of picking the wrong policy, given that experimental design. If there is any clear pattern in the decision maker's assignment of treatment, nature will exploit these. Randomization eliminates patterns for nature to exploit.

Figure 4.1 maps out implications of Propositions 4.2 and 4.3 for practical experiment design. Proposition 4.2 shows that when sample points are scarce, or when the decision maker does not put much weight on satisfying anyone

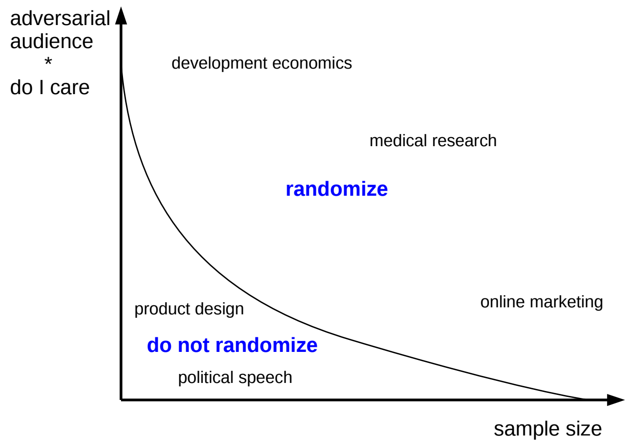


Figure 4.1: Maximizing subjective decision-making value vs. adversarial, prior-free inference.

else (λ close to 1), optimal experimentation will be Bayesian. That is, the experimenter will focus on assigning treatment and control observations to the subjects from whom she expects to learn the most. This is the case, for example, when a firm is implementing a costly new process in a handful of production sites: The firm will focus on a few teams where it can learn the most. Similarly, a politician trying out platforms will do so at a few carefully chosen venues in front of carefully chosen audiences.

However, when the decision maker must satisfy an adversarial audience, or has a sufficiently large sample, she will randomize as in Proposition 4.3. The former is the case in scientific research. The latter is the case for firms dealing online with many end users: Although the firm only needs to convince itself of the effectiveness of a particular ad or UI design, observations are plentiful so randomization is cheap, and used.

Proposition 4.3, in particular, has constructive applications to experimental practice. First, it implies that a decision maker who randomizes even without understanding all its ramifications—why she is randomizing, what audience the experiment is meant to satisfy—will nevertheless produce an almost-optimal experiment for large values of N. Even if someone (or her own doubts) produces a particularly challenging prior, the decision rule is still likely to be close to optimal. Further, this proposition highlights the importance of actually randomizing. An experiment that adopts a protocol that is only "nearly" random, such as assignment based on time of day of an experimental session (see Green and Tusicisny, 2012, for a critique), or the first letter of an experimental subject's name (as was the case in the deworming study of Miguel and Kremer, 2004; see Deaton, 2010 for a critique), will tend to find a skeptical prior in its audience. Randomization provides a defense against the most skeptical priors, but near-randomization offers no such protection.

4.4 Rerandomization and Resampling

Proposition 4.3 established that randomization is essential to guarantee successful prior-free inference. Although the outcome of that randomization is not relevant to the inferential process, in practice, experimenters often care about whether a sample is *balanced*—that is, whether there are differences between the samples on some set of covariates. This may be because unbalanced samples can lead to re-evaluation and criticism of a study's findings Banerjee et al. (forthcoming); Gerber and Green (2000); Imai (2005). A common tool to achieve balance is the practice of rerandomization—redrawing an assignment until an acceptable balance is achieved. However, concerns about the effects of rerandomization on the robustness of inference from RCTs lead many scholars to hide the fact that they have rerandomized (Bruhn and McKenzie, 2009). While scholars are often advised to use stratification or matching to achieve balance instead, this is often quite difficult to implement when stratifying on multiple continuous covariates; thus, researchers turn to rerandomization. Our framework lets us formalize the robustness loss to rerandomization, and leads to a useful rule of thumb for experimental practice.

Definitions. For all results in this section, we assume that the experimenter seeks to find a realized experimental design, out of some set of experimental designs of cardinality K, that maximizes some arbitrary balance function, $B(e)$.[3] A focal balance function is

$$B(e) = \left\| \sum_{i|\tau_i=1} x^i - \sum_{i|\tau_i=0} x^i \right\| \tag{4.4}$$

[3]Rerandomization rules may also use a stopping time to endogenously pick the number of randomizations (Morgan and Rubin, 2012). Provided the stopping time has an upper bound K, all our results apply for this bound.

for some appropriate norm $||\cdot||$ in $\mathbb{R}^m$.[4] However, $B(e)$ may be any criterion that reflects preferences over realized assignments. These preferences may come from the experimenter, or from research partners, funders, etc. We provide examples in Section 4.5.

In addition, we assume that the policy function α is set to the near-optimum for a randomized experiment used in Proposition 4.3. That is, $\alpha(e,y) \equiv \arg\max_{a\in\{0,1\}} \overline{y}^a - \overline{y}^{1-a}$.

4.4.1 Rerandomization

Given K, rerandomized experiment $\mathcal{E}_K$ proceeds as follows:

1. For a fixed sample of xs drawn according to population distribution $q \in \Delta(X)$, independently draw a set of K assignments $\{e_1, \cdots, e_K\}$ with each (possibly stratified) $e_k = (x_i, \tau_{i,k})$ such that a fraction $\pi \in (0,1)$ of participants receives treatment $\tau = 1$;
2. Select an assignment $e_K^* \in \arg\max_{e\in\{e_1,\cdots,e_K\}} B(e)$ that maximizes balance function $B(e)$, breaking ties randomly;
3. Run the experiment e_K^*.

It is immediate that $B(e_K^*)$ first-order stochastically dominates $B(e_{K-1}^*)$. This is the value of rerandomization. The question, therefore, is whether rerandomization can adversely affect robustness. We show that it can:

Proposition 4.4. *Consider rerandomized experiment $\mathcal{E}_K$. There exists $\underline{\rho} > 0$ such that for generically every h_0, and for every N, if $K \geq (2|X|)^N$, then*

$$\max_{\alpha} \min_{h\in H} \mathbb{E}_{h,\mathcal{E}_K} \left[u(p, \alpha(e,y))\right] < \min_{h\in H} \mathbb{E}_h \left(\max_{a\in\{0,1\}} u(p,a) \right) - \underline{\rho}\epsilon.$$

Intuitively, when K is sufficiently large, the allocation is essentially deterministic, which by Proposition 4.3 precludes first-best robustness. However, the number of rerandomizations K needed for Proposition 4.4 to apply is extremely large: it is exponential in the sample size. This suggests that the cost of rerandomization may be quite small for reasonable numbers of rerandomizations. Indeed, this is the case:

[4] This nests the Mahalanobis distance commonly used in multivariate matching (Rubin, 1980; Cochrane and Rubin, 1973; Rubin, 1979).

Proposition 4.5. *Given $K \geq 2$, consider a rerandomized experiment $\mathcal{E}_K$ assigning treatment to a proportion $\pi \in (0,1)$ of participants. Then,*

$$\min_{h \in H} \mathbb{E}_{h,\mathcal{E}_K} \left[u(p, \alpha(e,y)) \right] \geq \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \sqrt{\frac{\log(K)}{\kappa N}},$$

where $\kappa = \min\{\pi, 1-\pi\}$.

Comparing this with Proposition 4.3, frequentist decision making is optimal up to a loss bounded by $\sqrt{\frac{\max\{1, \log(K)/\kappa\}}{N}}$. The additional loss from rerandomization, $\sqrt{\log(K)}$, is rleatively small: between 1.5 and 3 for sample sizes between 10 and 10,000. However, the additional loss to robustness due to increasing K falls off relatively slowly. On the other hand, balance is established very quickly: K rerandomizations guarantee that the final sample will be within the group of 5% most balanced samples with probability $1 - 0.95^K$, and the improvement in balance falls off relatively quickly. Observing that $1 - 0.95^{100} > 0.99$, we suggest the following simple rule of thumb for rerandomization:

Rule of Thumb. Set $K = \min\{N, 100\}$.

4.4.2 Resampling

Although rarely used in practice, we examine resampling as it allows us to make a subtle point: more randomization is not always better. Resampling reformulates the first step in the process outlined in the previous subsection:

1.’ Independently draw a set of K assignments $\{e_1, \cdots, e_K\}$ with each $e_k = (x_{i,k}, \tau_{i,k})$ with xs (possibly stratified) drawn according to population distribution $q \in \Delta(X)$, and a (possibly stratified) fraction $\pi \in (0,1)$ of participants receives treatment $\tau = 1$.

The relative merits of different sampling schemes may depend on the balance function used to choose from the set of possible experiments. To examine a worst-case scenario, we define a “demon” function that always selects the worst possible experiment from the set of available experiments, breaking ties randomly:

$$\mathcal{D}_K \left(p, (e_k, y_k)_{k \in \{1,\ldots,K\}} \right) \in \underset{\{e_1, \cdots, e_K\}}{\arg\min}\, u(p, \alpha(e_k, y_k)).$$

In practice, such a function will not exist. However, it is useful for establishing the following result:

Corollary 4.6. *Suppose the experimental implementation is chosen according to $\mathcal{D}_K$. Then, the expected loss from resampling is greater than from rerandomization, and this differential grows with K.*

The intuition behind this result is simple, but subtle. Under rerandomization, the sample is fixed, and therefore there may simply not exist an allocation of subjects to treatment and control that causes the decision maker to make a mistaken policy decision. However, with resampling, a new sample is drawn each time, increasing the chances that there is a sample-allocation pair that will cause the decision maker to make a mistake. The result points out that, while randomization generally protects against poor interpretation of the data, it will not always have that effect.

4.5 Simulations

We illustrate the usefulness of our rerandomization results using a simple simulation with a single covariate $x \in X = \{1, 2, \ldots, 10{,}000\}$. Even covariates are twice as likely as odd covariates, and the treatment effect is small and negative for even covariates, and large and positive for odd covariates. Specifically, for $n \in \{1, 2, \ldots, 5{,}000\}$,

$$q(2n-1) = \frac{q(2n)}{2} = \frac{2}{3|X|}, \quad p^1_{2n-1} = 4p^0_{2n-1} = \frac{4}{5}, \quad \text{and} \quad p^1_{2n} = \frac{p^0_{2n}}{2} = \frac{1}{4}.$$

Thus, on aggregate, $u(p, 1) = \frac{13}{30} > \frac{2}{5} = u(p, 0)$, so treatment is beneficial, and $\alpha = 1$ is the “correct” decision. This setup is meant to make attempts to balance the sample likely to cause inferential mistakes—balancing will tend to pair odd observations with the more numerous even observations, which are not an appropriate comparison group.[5]

In all of the simulations that follow, we use stratified random sampling with $\pi = \frac{1}{2}$—that is, exactly half of the sample is assigned to treatment and half to control.

[5]Indeed, using pairwise matching to assign treatment and control status increases inferential errors, but does so equally for randomization, rerandomization, and resampling.

Figure 4.2 examines the error rates (that is, percent of the time that $\alpha = 0$ is chosen based on the outcome of the experiment), and balance, according to (4.4) using the l_1 norm, of three different ways of selecting a sample: randomization, rerandomization, and resampling. As can be seen in the first panel, all three give roughly the same error rate. This is because the chosen balance function, $B(e)$, in these simulations is very unlikely to select a more biased sample allocation. While in any specific application the interaction of the model parameters and the balance function may produce different results, it appears quite difficult to find a balance function that might actually be used and is particularly pernicious.

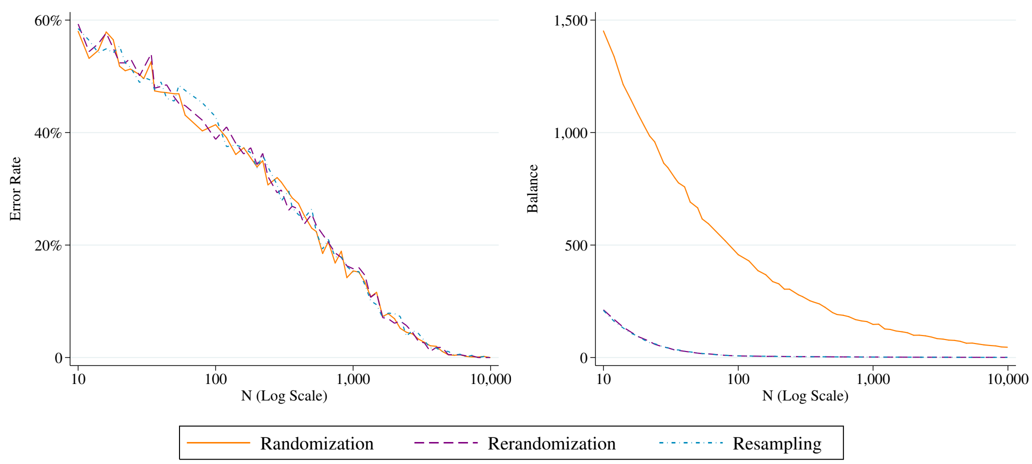


Figure 4.2: In simulations, resampling and rerandomization substantially increase balance with no cost to robustness.

On the other hand, rerandomization and resampling both substantially improve the balance of the samples. This is particularly true for small and moderate sample sizes, up to the order of 1,000, although even with 10,000 sample points there is an improvement in balance, even though we only re-draw the allocation (and sample) 100 times. However, resampling adds no additional balance to simple rerandomization.

This difficulty in finding a balance function that results in a substantial loss of robustness motivates the use of the demon function $\mathcal{D}_K$, defined in the last subsection, that chooses a treatment assignment that produces an erroneous decision whenever one is available. As expected, this substantially increases the error rates of both resampling and rerandomization, as shown in Figure 4.3. However, in keeping with Corollary 4.6, it can be seen that the error rate is substantially lower under rerandomization until sample sizes are of order 10,000. Indeed, for small to moderate sample sizes, the demon function is always able to find a sample and allocation that induces a mistake under resampling, but this is not the case with rerandomization.

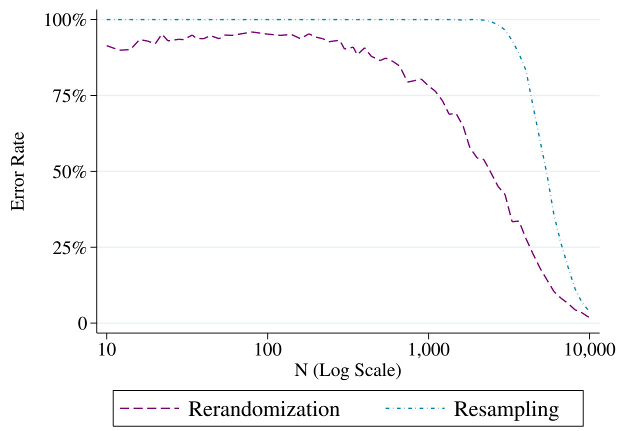


Figure 4.3: Rerandomization is less likely to produce errors than resampling.

4.6 Discussion

We have shown that a maxmin framework for experimental design produces reliable positive results, and can be used to inform practice. Elsewhere, we informally apply this framework to discuss other aspects of experimental design, including registration, pre-analysis plans, and external validity (Banerjee et al., forthcoming).

Before closing, we make a final comment that can tie together the findings from both sets of our results. In Section 4.3 we discuss the issues with quasi-random assignment, and discuss as an example the seminal work on field trials in development by Miguel and Kremer (2004), and the criticism of that work by Deaton (2010). At issue was the authors use of treatment assignment to every third village, alphabetically.[6] This was done to satisfy implementing partners, who were uncomfortable with randomization. Our proposal for rerandomization could provide a middle ground: specifically, Miguel and Kremer could have drawn up $K = N = 75$ lists of possible allocations, and allowed the

[6]All villages eventually got the treatment, but in three waves. The first, fourth, seventh, etc. villages were assigned to Wave I, and the second, fifth, eighth, etc. to Wave II, and the third, sixth, ninth, etc to Wave III.

implementing partner to pick which one they preferred. This would come at little cost of robustness to the trial, and may have satisfied the implementing partners need for a feeling of control. However, to be sure of this, one would need a theory of implementing partners, which is beyond the scope of this paper.

Appendix A

APPENDIX TO CHAPTER 2

A.1 Campaign Stage Details

Closed-form expressions for the candidates' probability of winning can be obtained as follows. Recall from (2.7) that candidate j's vote share can be written as

$$S_{jd}^{m} = \frac{\exp(\bar{\delta}_{jd}^{m} + \eta_j^m)}{1 + \sum_{k \neq 0} \exp(\bar{\delta}_{kd}^{m} + \eta_k^m)}, \tag{A.1}$$

where $\bar{\delta}_{kd}^{m} = \alpha_1 c_{kd} + \alpha_2 c_{kd}^2 + x_d' \beta_k^m + \xi_{kd}^m$. Since $(\eta_k^m)_{k \neq 0}$ are i.i.d. with a Type-I Extreme Value distribution, ties occur with probability zero, and j's probability of winning also takes a multinomial logit form:

$$\begin{aligned} PW_{jd}^{m} &= \Pr\left(\cap_{k \notin \{j,0\}} \{S_{jd}^m > S_{kd}^m\}\right) \\ &= \Pr\left(\cap_{k \notin \{j,0\}} \{\bar{\delta}_{jd}^m + \eta_j^m > \bar{\delta}_{kd}^m + \eta_k^m\}\right) \\ &= \frac{\exp(\bar{\delta}_{jd}^m)}{\sum_{k \neq 0} \exp(\bar{\delta}_{kd}^m)}. \end{aligned} \tag{A.2}$$

Candidate j's expected vote share is obtained by integrating (A.1) with respect to the distribution of $(\eta_k^m)_{k \neq 0}$, which can be easily simulated.

Computation of PRI and PVEM's individual vote shares when they nominate a joint coalition candidate involves the second tier of the voting stage. As in the first tier, let $\delta_{jd}^{\mathrm{ST},m} = \bar{\delta}_{jd}^{\mathrm{ST},m} + \eta_j^{\mathrm{ST},m} = x_d' \beta_j^{\mathrm{ST},m} + \xi_{jd}^{\mathrm{ST},m} + \eta_j^{\mathrm{ST},m}$ represent mean voter utility from selecting alternative j in the second tier of the voting stage. Normalizing to zero the mean utility of splitting the vote 50-50 between PRI and PVEM, and denoting by $j = p \in \{\mathrm{PRI}, \mathrm{PVEM}\}$ the option of giving party p 100% of the vote, p's vote share is given by (again using a law of large numbers approximation)

$$S_{pd}^{m} = S_{\text{PRI-PVEM},d}^{m} \left(\frac{0.5 + \exp(\bar{\delta}_{pd}^{\mathrm{ST},m} + \eta_p^{\mathrm{ST},m})}{1 + \sum_{j \in \{\mathrm{PRI},\mathrm{PVEM}\}} \exp(\bar{\delta}_{jd}^{\mathrm{ST},m} + \eta_j^{\mathrm{ST},m})} \right), \tag{A.3}$$

where $S_{\text{PRI-PVEM},d}^{m}$ is the coalition candidate's total share of votes in accordance with (A.1). Integration of (A.3) with respect to the distribution of $(\eta_k^m)_{k \neq 0}$ and $(\eta_j^{\mathrm{ST},m})_{j \in \{\mathrm{PRI},\mathrm{PVEM}\}}$ yields p's expected vote share.

Games with strategic complementarities. Formally, the campaign spending game played in each district is described as follows. As discussed in Section 2.3.1, the set of players is composed of all 5 parties when $M_d = \text{IND}$, and PRI and PVEM act as a single player when $M_d \neq \text{IND}$. The strategy space available to each player is $\mathbb{R}_+$, the set of nonnegative expenditure levels, and the players' payoffs are defined in (2.5) and (2.6).

While I refer the reader to Echenique and Edlin (2004) for a formal definition of games with strict strategic complementarities (GSSC), I discuss here properties of the parties' payoff functions, satisfied at the estimated parameter values, which imply that the spending games belong to this class. First, since $\alpha_1 > 0 > \alpha_2$ (see Table 2.6), the effect of campaign spending on $\bar{\delta}^m_{jd}$ is maximized at $\bar{c} = -\alpha_1/(2\alpha_2)$. Given that candidate j's vote share and probability of winning are strictly increasing in $\bar{\delta}^m_{jd}$, they are also maximized at $\bar{c}$. It then follows that spending more than $\bar{c}$ is a strictly dominated strategy for all players in the spending games. That is, regardless of their rivals' spending, each player's payoff is higher at $\bar{c}$ than at any level exceeding $\bar{c}$. Thus, the players' effective strategy space is $[0, \bar{c}]$, a compact interval, which satisfies condition 1 of the definition of GSSC in Echenique and Edlin (2004). Second, it can be verified that, at the estimated parameter values, the parties' payoff functions are twice differentiable with positive cross partial derivatives, which implies the remaining conditions of the definition of GSSC.

As mentioned in Section 2.3.1, GSSC have three useful properties. First, existence of equilibrium is guaranteed (Vives, 1990). Second, mixed-strategy equilibria are unstable, so their omission is justified (Echenique and Edlin, 2004). Lastly, Echenique (2007) provides a simple and fast algorithm for computing the set of all pure-strategy equilibria. This set has an additional key property; it has a largest and a smallest equilibrium, providing a simple test of uniqueness: if the largest and smallest equilibria coincide, the resulting strategy profile is the unique equilibrium of the game. These extremal equilibria can be easily computed by iterating best responses; see Echenique (2007) for details. As previewed in Section 2.3.1, the largest and smallest equilibria of the campaign spending games analyzed in this paper always coincide.

A.2 Richer Specification

For the results of Section 2.5.1, I expand x_d by incorporating the following. First, the electoral authority groups districts into 5 electoral regions (see Figure A.2). Accordingly, I use regional fixed effects to capture any spatial features of voting preferences. In addition, I consider the unemployment rate, marriage (as a percentage of population 15 and older), and I enrich the description of age, education, and income by including: the percentage of the voting-age population aged 24 and under, the percentage of the voting-age population without post-elementary education, the percentage of households with refrigerators, and the percentage of households without basic utilities (power and plumbing).

A.3 Figures and Tables

Table A.1: Structural parameters β_j^m of candidate choice j = MP

	m = IND	m = PRI	m = PVEM
Coefficient	Estimate (St. Error)	Estimate (St. Error)	Estimate (St. Error)
Intercept	-3.297 (1.503)	-3.800 (1.428)	-4.719 (1.562)
Female head of household	0.032 (0.016)	-0.010 (0.013)	0.037 (0.020)
Age (over 64)	-0.017 (0.020)	-0.014 (0.015)	-0.030 (0.027)
Schooling	0.099 (0.050)	0.292 (0.038)	0.258 (0.064)
Income (owns a car)	-0.028 (0.004)	-0.026 (0.003)	-0.024 (0.005)

Two-stage least squares estimates of random effects model (2.8) with robust standard errors clustered by candidate's party affiliation and PRI-PVEM's coalition configuration. Demographics as in Table 2.4.

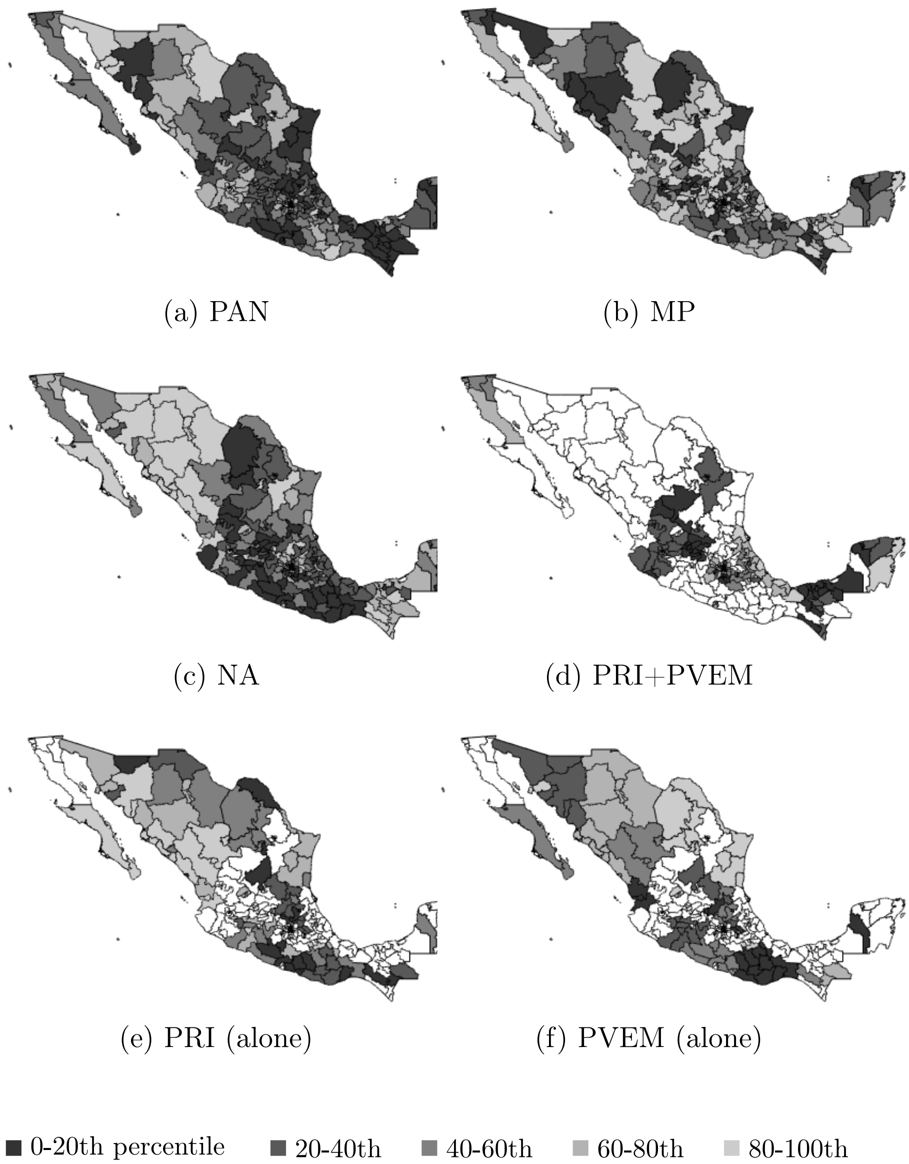


Figure A.1: Geographic distribution of campaign spending by party

Table A.2: Structural parameters β_j^m of candidate choice $j = \text{NA}$

Coefficient	$m = \text{IND}$ Estimate (St. Error)	$m = \text{PRI}$ Estimate (St. Error)	$m = \text{PVEM}$ Estimate (St. Error)
Intercept	-3.839	-3.251	-3.681
	(1.462)	(1.310)	(1.349)
Female head of household	-0.065	-0.041	-0.012
	(0.022)	(0.012)	(0.017)
Age (over 64)	-0.003	0.014	0.016
	(0.029)	(0.014)	(0.024)
Schooling	0.274	0.056	-0.015
	(0.071)	(0.038)	(0.062)
Income (owns a car)	-0.012	0.001	0.004
	(0.005)	(0.003)	(0.005)

Two-stage least squares estimates of random effects model (2.8) with robust standard errors clustered by candidate's party affiliation and PRI-PVEM's coalition configuration. Demographics as in Table 2.4.

Table A.3: Structural parameters β_j^m of candidate choice $j = \text{PAN}$

Coefficient	$m = \text{IND}$ Estimate (St. Error)	$m = \text{PRI}$ Estimate (St. Error)	$m = \text{PVEM}$ Estimate (St. Error)
Intercept	-2.666	-2.216	-1.419
	(1.389)	(1.336)	(1.355)
Female head of household	-0.085	-0.008	-0.040
	(0.020)	(0.015)	(0.020)
Age (over 64)	0.024	0.085	0.077
	(0.020)	(0.017)	(0.029)
Schooling	0.234	-0.147	-0.197
	(0.052)	(0.056)	(0.074)
Income (owns a car)	-0.002	0.006	0.018
	(0.004)	(0.004)	(0.005)

Two-stage least squares estimates of random effects model (2.8) with robust standard errors clustered by candidate's party affiliation and PRI-PVEM's coalition configuration. Demographics as in Table 2.4.

Table A.4: Structural parameters β_j^m of candidate choice j = PRI

	m = IND	m = PRI	m = PVEM
Coefficient	Estimate (St. Error)	Estimate (St. Error)	Estimate (St. Error)
Intercept	-2.931	-2.670	
	(1.467)	(1.689)	
Female head of household	-0.020	-0.055	
	(0.012)	(0.011)	
Age (over 64)	0.024	0.057	
	(0.015)	(0.010)	
Schooling	0.103	0.135	
	(0.037)	(0.048)	
Income (owns a car)	-0.009	-0.007	
	(0.003)	(0.003)	

Two-stage least squares estimates of random effects model (2.8) with robust standard errors clustered by candidate's party affiliation and PRI-PVEM's coalition configuration. Demographics as in Table 2.4.

Table A.5: Structural parameters β_j^m of candidate choice j = PVEM

	m = IND	m = PRI	m = PVEM
Coefficient	Estimate (St. Error)	Estimate (St. Error)	Estimate (St. Error)
Intercept	-1.942		-2.586
	(1.450)		(1.734)
Female head of household	-0.036		-0.002
	(0.023)		(0.019)
Age (over 64)	-0.041		0.038
	(0.027)		(0.018)
Schooling	0.062		0.005
	(0.067)		(0.069)
Income (owns a car)	-0.025		-0.009
	(0.005)		(0.006)

Two-stage least squares estimates of random effects model (2.8) with robust standard errors clustered by candidate's party affiliation and PRI-PVEM's coalition configuration. Demographics as in Table 2.4.

Table A.6: Structural parameters $\beta_j^{\mathrm{ST},m}$ of party choice $j=$ PRI conditional on voting for PRI-PVEM coalition candidate

	$m=$ PRI	$m=$ PVEM
Coefficient	Estimate (St. Error)	Estimate (St. Error)
Intercept	2.091	2.111
	(1.293)	(1.311)
Female head of household	-0.047	-0.054
	(0.009)	(0.012)
Age (over 64)	0.059	0.086
	(0.010)	(0.017)
Schooling	-0.007	-0.038
	(0.026)	(0.040)
Income (owns a car)	0.003	0.006
	(0.002)	(0.003)

Generalized least squares estimates of random effects model (2.9) with robust standard errors clustered by party and coalition candidate's party affiliation. Outside option is 50-50 vote split between PRI and PVEM. Demographics as in Table 2.4.

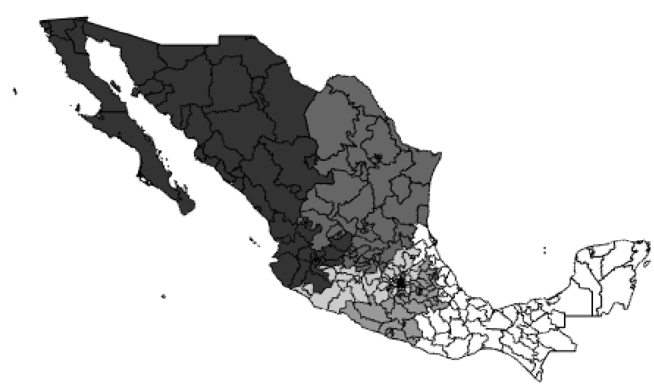


Figure A.2: Mexican electoral regions (color-coded) and districts (delimited)

Table A.7: Structural parameters $\beta_j^{\mathrm{ST},m}$ of party choice j = PVEM conditional on voting for PRI-PVEM coalition candidate

Coefficient	m = PRI Estimate (St. Error)	m = PVEM Estimate (St. Error)
Intercept	-0.204 (1.326)	0.503 (1.433)
Female head of household	0.008 (0.018)	-0.058 (0.029)
Age (over 64)	0.001 (0.021)	0.036 (0.041)
Schooling	-0.091 (0.054)	0.078 (0.096)
Income (owns a car)	0.003 (0.004)	-0.009 (0.008)

Generalized least squares estimates of random effects model (2.9) with robust standard errors clustered by party and coalition candidate's party affiliation. Outside option is 50-50 vote split between PRI and PVEM. Demographics as in Table 2.4.

Table A.8: Counterfactual outcomes using richer model (see Appendix A.2)

	Vote share (%)		
Party	Observed	No coalition	Total coalition
PRI	33.6	$(+3.1 =)$ 36.7	$(-0.7 =)$ 32.9
PVEM	6.4	$(-2.1 =)$ 4.3	$(+0.5 =)$ 6.9
PAN	27.3	$(-0.3 =)$ 27.0	$(+0.6 =)$ 27.9
MP	28.3	$(-1.4 =)$ 26.9	$(+0.1 =)$ 28.4
NA	4.3	$(+0.7 =)$ 5.0	$(-0.4 =)$ 3.9
	Seats		
Party	Observed	No coalition	Total coalition
PRI	207	$(+16 =)$ 223	$(-3 =)$ 204
PVEM	34	$(-20 =)$ 14	$(+11 =)$ 45
PAN	114	$(+5 =)$ 119	$(+1 =)$ 115
MP	135	$(-3 =)$ 132	$(-8 =)$ 127
NA	10	$(+2 =)$ 12	$(-1 =)$ 9

Differences in parentheses are with respect to first column. Second and third columns correspond to counterfactual outcomes had PRI and PVEM run independently or together in all districts, respectively.

Appendix B

APPENDIX TO CHAPTER 3

B.1 Iterative Computation of Penalty Loadings

The penalized estimators apply data-dependent factor loadings for each of the coefficients included in the model. The data-dependent factor loadings scale the penalty for each coefficient according to the variability of the associated coefficient and the model residual. These loadings appear in Υ_β and Υ_ω in Algorithm 1, Υ_θ in equation (3.16), Υ_ϕ in equation (3.18), and Υ_ζ in equation (3.19). Here we review the application of Belloni et al. (2013a)'s iterative approach to computing these penalty loadings.

B.1.1 Iterative Computation for Linear Models

Recalling the formula for step I of Algorithm 1:

$$\min_{\beta \in \mathbb{R}^{K_T+1}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p\right)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_\beta \beta\|_1.$$

As discussed in the details, $\lambda_\beta = 2c\sqrt{JT}\Phi^{-1}\left(1 - \gamma/2(K_T+1)\right)$, with Belloni et al. (2013a)'s recommended values being $c = 1.1$ and $\gamma = 0.05/\log(K_T + 1 \vee T)$. The k^{th} diagonal entry in $\hat{\Upsilon}_\beta$ scales the penalty according to the variability in the k^{th} regressor, which we'll denote $x_{k,jt}$, and the residual $\epsilon_{jt} \equiv \mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p$. The infeasible ideal sets $\hat{\Upsilon}_{\beta,\{k,k\}} = \sqrt{\mathbb{E}\left[x^2_{k,jt}\epsilon^2_{jt}\right]}$. The iterative Algorithm B.1 initializes Υ_β with the expected squared value of each regressor, fits the lasso regression, recovers the residuals, and uses these residuals to compute the sample analog to the ideal value. This algorithm extends immediately to Υ_ω. Defining the residual $\varepsilon_{jt} \equiv p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1$, the infeasible ideal penalty values for this problem are $\hat{\Upsilon}_{\omega,\{k,k\}} = \sqrt{\mathbb{E}\left[x^2_{k,jt}\varepsilon^2_{jt}\right]}$. For completeness, the calculation is detailed in Algorithm B.2.

B.1.2 Iterative Computation for Nonlinear Models

The selection in nonlinear models requires accounting for the additional estimation error introduced by selection on a generated regressor. Consequently,

Algorithm B.1 Iterative Algorithm for Υ_β

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. For $\mathcal{I} = 1, ..., \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\hat{\beta} = \underset{\beta \in \mathbb{R}^{K_T+1}}{\arg\min} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p\right)^2 + \frac{\lambda_\beta}{T}\|\Upsilon^{\mathcal{I}-1}\beta\|_1$.

b) Compute the Residuals: $\hat{\epsilon}_{jt} \equiv \mathcal{S}_{jt} - x'_{0t}\hat{\beta}_{0j} - x'_{1jt}\hat{\beta}_1 - p_{jt}\hat{\beta}_p$.

c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\hat{\epsilon}^2_{jt}}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\beta = \Upsilon^{\mathcal{I}}$.

Algorithm B.2 Iterative Algorithm for Υ_ω

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. For $\mathcal{I} = 1, ..., \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\hat{\omega} = \min_{\omega \in \mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1\right)^2 + \frac{\lambda_\omega}{T}\|\hat{\Upsilon}_\omega\omega\|_1$.

b) Compute the Residuals: $\hat{\varepsilon}_{jt} \equiv p_{jt} - x'_{0t}\hat{\omega}_{0j} - x'_{1jt}\hat{\omega}_1$.

c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\hat{\varepsilon}^2_{jt}}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\omega = \Upsilon^{\mathcal{I}}$.

the residual with which to scale the regressor's variability must be augmented by the variance of the generated selection target. Recall the selection problem in equation (3.18):

$$\tilde{\phi} = \underset{\phi \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1\right)^2 + \frac{\lambda_\phi}{T}\|\hat{\Upsilon}_\phi\phi\|_1.$$

The Υ_ϕ matrix requires a slight adjustment to account for estimation error in the $\tilde{\delta}_{jt}$'s. Defining

$$\epsilon_{\delta,jt} \equiv \delta_{jt} - \tilde{\delta}_{jt} = \tilde{\delta}_{jt} = \tilde{x}'_{jt}\left(\tilde{\beta}_j - \beta_j\right) + x'_{1jt}\left(\tilde{\beta}_1 - \beta_1\right) + p_{jt}\left(\tilde{\beta}_p - \beta_p\right) + \tilde{\xi}_{jt} - \xi_{jt}$$

and $\epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1$, the ideal weight for $\phi_{0j,k}$ is equal to $\sqrt{\bar{E}\left[x^2_{0t,k}\left(\epsilon_{\delta,jt}+\epsilon_{\phi,jt}\right)^2\right]}$ and $\sqrt{\bar{E}\left[x^2_{1jt,k}\left(\epsilon_{\delta,jt}+\epsilon_{\phi,jt}\right)^2\right]}$ for $\phi_{1,k}$.

We can define $\bar{x}_{jt} = \left[\tilde{x}'_{jt}x'_{1jt}, p_{jt}\right]'$ and Σ_j as the rows and columns of the variance-covariance matrix for $\tilde{\beta}$ computed using the sandwich covariance matrix from the solution to (3.17):

$$\hat{\epsilon}^2_{\delta,jt} = E\left[\epsilon^2_{\delta,jt}\right] = \bar{x}'_{jt}\Sigma_j\bar{x}_{jt} + \sigma^2_\xi.$$

For feasible implementation, we again initialize the Υ_ϕ matrix with the diagonal variances of the regressors. We then recursively solve (3.18) to recover the residuals $\epsilon_{\phi,jt}$ and update the Υ_ϕ accordingly.

Algorithm B.3 Iterative Algorithm for Υ_ϕ

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. Compute $\hat{\epsilon}^2_{\delta,jt} = \bar{x}'_{jt}\Sigma_j\bar{x}_{jt} + \sigma^2_\xi$ from the solution to the feasible GMM problem (3.17).

III. For $\mathcal{I} = 1, ..., \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

   a) Solve $\tilde{\phi} = \arg\min_{\phi\in\mathbb{R}^{K_T}} \dfrac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1\right)^2 + \dfrac{\lambda_\phi}{T}\|\Upsilon^{\mathcal{I}-1}\phi\|_1$.

   b) Compute the Residuals: $\hat{\epsilon}_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t}\hat{\phi}_{0j} - x'_{1jt}\hat{\phi}_1$.

   c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\left(\hat{\epsilon}^2_{\phi,jt} + \hat{\epsilon}^2_{\delta,jt}\right)}$, $k = 1, \ldots, K_T$.

IV. Set $\hat{\Upsilon}_\phi = \Upsilon^{\mathcal{I}}$.

The approach above doesn't apply as readily to the solution for (3.19), as we cannot easily characterize the variance of the optimum instruments for the nonlinear features of the model. However, we don't need to account for the population variance of the asymptotic optimal instruments in our selection of controls. Importantly, the estimated optimal instruments provide the only source of exogenous variation used to identify the heterogeneity in voter impressionability. Consequently, performing selection on the utilized instruments as if they represented the population optimal instruments suffices to control for observable heterogeneity. Recalling the penalization problem:

$$\tilde{\zeta} = \underset{\zeta \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1\right)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1$$

and defining the residual $\varepsilon_\zeta = \tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1$, the ideal $(k,k)^{th}$ entry in $\Upsilon_\zeta = \mathbb{E}\left[x^2_{k,jt}\varepsilon^2_\zeta\right]$. We can then apply the approach from Algorithms B.1 and B.2.

Algorithm B.4 Iterative Algorithm for Υ_ζ

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. For $\mathcal{I} = 1, ..., \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\hat{\zeta} = \underset{\zeta \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1\right)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1$.

b) Compute the Residuals: $\hat{\varepsilon}_{\zeta,jt} = \tilde{z}_{v,jt} - x'_{0t}\hat{\zeta}_{0j} - x'_{1jt}\hat{\zeta}_1$.

c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\hat{\varepsilon}^2_{\zeta,jt}}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\zeta = \Upsilon^{\mathcal{I}}$.

B.1.3 GMM Penalty for Verifying First Order Conditions

While we do not directly evaluate the objective function in the global parameter space for equation (3.16), we do need to verify the first-order conditions for the local solution based on the selected model in the last step of Algorithm 2:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q\left(\tilde{\theta}^*, \tilde{x}^k, z, p, s\right) < \lambda_\theta \upsilon_k, k = 1, \ldots, K_0, j = 1, \ldots, J.$$

As discussed in the text, the infeasible ideal value of $\upsilon_k = \sqrt{\bar{E}\left[x^2_{0t,k}\xi^2_{jt}\right]}$. Here, we are already working from a (putative) local optimum, so we can take the estimated values $\xi\left(\tilde{\theta}^*, \tilde{x}, z, p, s\right)$ to estimate the empirical analog to the expectation:

$$\hat{\upsilon}_k = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{J,T} x^2_{0t,k}\tilde{\xi}^2_{jt}}.$$

This calculation has the added benefit of being computable variable-by-variable to mitigate memory and computational limitations.

B.2 Detailed Statements of Model Assumptions

B.2.1 Notation

- $\bar{E}_T$ represents the average expectation of a series, for example, $\bar{E}_T\left[x_{0t}x'_{0t}\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[x_{0t}x'_{0t}]$.
- $\mathcal{E}_T$ represents the empirical average of a series, for example, $\mathcal{E}_T\left[x_{0t}x'_{0t}\right] = \frac{1}{T}\sum_{t=1}^{T}x_{0t}x'_{0t}$.
- $\phi_{min}\left(c\right)\{\Sigma\}$ and $\phi_{max}\left(c\right)\{\Sigma\}$ represent the c^{th} smallest and largest eigenvalues of the matrix Σ.

B.2.2 Exact Sparsity in Preferences and Spending

Assumption B.2. *Exact Sparsity in Preferences and Spending:*

Let $\theta = [\beta'_{01}, \ldots, \beta'_{0J}, \beta'_1, \pi'_{01}, \ldots, \pi'_{0J}, \pi'_1, \pi'_z]'$. *Each data generating process in the sequence* $\{\mathcal{P}_T\}_{T=1}^{\infty}$, *has* $K_T > T$ *possible parameters,* $1 \leq k_T < T$ *of which are non-zero, where both* $K_T \underset{T\to\infty}{\to} \infty$ *and* $k_T \underset{T\to\infty}{\to} \infty$. *Further, the number of excluded instruments in* z *is fixed at* $L \geq 2$. *Finally, there exists a sequence* $\{\delta_T, \Delta_T\} \to_{T\to\infty} 0$.

1. *The parameter space isn't too large:* $\log K_T \leq \left(T\delta_T\right)^{1/3}$.
2. *The model is sufficiently sparse:* $k_T^2 \log^2\left(K_T \vee T\right)/T \leq \delta_T$.
   a) *The number of variables explicitly included,* $K_1 + L$ *in the model is fixed.*
   b) *In equation (3.1), the true coefficients have* $\sum_{j=1}^{J}\|\beta_{0j}\|_0 \leq k_T$.
   c) *In the campaign spending equation (3.4), the true coefficients have* $\sum_{j=1}^{J}\|\pi_{0j}\|_0 \leq k_T$.
3. *Sparse eigenvalues for Gram matrix: There exists a sequence* $\ell_T \to 0$, κ', *and* κ'' *such that, with probability* $1 - \Delta_T$*:*

   $$0 \leq \kappa' \leq \phi_{min}\left(\ell_n k_T\right)\left\{\bar{E}_T\left[x_{0t}x'_{0t}\right]\right\} \leq \phi_{max}\left(\ell_n k_T\right)\left\{\bar{E}_T\left[x_{0t}x'_{0t}\right]\right\} \leq \kappa'' < \infty.$$

4. *Detectable Non-zero Coefficients:* $\min\left\{|\theta_k|\,|\theta_k \neq 0\right\} > \delta_T$.
5. *Exponential tails: There exists* $b > 0$ *and* $r > 0$ *such that, for any* $\tau > 0$,

a) $P(|\xi_{jt}| > \tau) \leq \exp\left(-(\tau/b)^r\right);$

b) $\forall k, P(|x_{0t,k}| > \tau) \leq \exp\left(-(\tau/b)^r\right).$

As discussed in the main text, assumptions 2.1-2.4 consolidate the restrictions in Belloni et al. (2012) and Belloni et al. (2013a) for the voting application with a fixed number of excluded instruments. The restriction on exponential tails isn't strictly necessary for application in the linear model, as we adopt the Belloni et al. (2012) penalization strategy that applies moderate deviation theory for self-normalized sums to bound deviations in the maximal element of the score vector. However, we maintain the restriction, as it allows us to use Fan and Liao (2014)'s results, which require results from large deviation theory, in the nonlinear GMM setting.

B.2.3 Regularity Conditions for High-Dimensional Linear Model

Assumption B.3. *Linear Logit DGP Regularity Conditions:*

Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^{\infty}$, *has* $K_T > T$ *possible parameters,* $1 \leq k_T < T$ *of which are non-zero, where both* $K_T \underset{T\to\infty}{\to} \infty$ *and* $k_T \underset{T\to\infty}{\to} \infty$, *but the number of excluded instruments in* z *is fixed at* $L \geq 2$. *Finally, there exists a sequence* $\{\delta_T, \Delta_T\} \to_{T\to\infty} 0$ *and fixed constants* $0 < c < C < \infty$.

1. *Sufficient moments for unmodeled variability in the data admit a LLN and CLT:*
   - *a)* $\bar{E}\left[|\xi_{jt}|^q + |\nu_{jt}|^q\right] \leq C,$
   - *b)* $c \leq \bar{E}\left[\xi_{jt}^2 | x_{jt}, \nu_{jt}\right] \leq C,$ *a.s., and,*
   - *c)* $c \leq \bar{E}\left[\nu_{jt}^2 | x_{jt}, z_{jt}\right] \leq C,$ *a.s.*
2. *Variability in observables and their impact on unobservables is bounded:*
   - *a) Demographics Controls:* $\max_{t\leq T} \|x_{0t}^2\|_\infty k_T T^{\left(-\frac{1}{2}+\frac{2}{q}\right)} \leq \delta_T$ *w.p.* $1-\Delta_T$.
   - *b) Candidate Characteristics:* $\bar{E}\left[|x_{1jt,k}|^q\right] \leq C,$ *and* $|\beta_{1,k}| < C, k = 1, \ldots, K_1$.
   - *c) Campaign Expenditure and Impact:* $\bar{E}\left[|p_{jt}|^q\right] \leq C,$ *and* $|\beta_p| < C$.
3. *Additional regularity restrictions for asymptotic theory with i.n.i.d. sampling:*

a) $\max_{j\leq J}\bar{E}[|\xi_{jt}|^q]+\bar{E}[|\mathcal{S}_{jt}|^q]+\max_{k\leq K_T}\{\bar{E}\left[x^2_{jt,k}\mathcal{S}^2_{jt}\right]+\bar{E}\left[|x^3_{jt,k}\xi^3_{jt}|\right]+1/\bar{E}\left[x^2_{jt,k}\right]\}\leq C$.

b) $\max_{k\leq K_T, j\leq J}\{|(\mathcal{E}_T-\bar{E})\left[x^2_{jt,k}\xi^2_{jt}\right]|+|(\mathcal{E}_T-\bar{E})\left[x^2_{jt,k}\mathcal{S}^2_{jt}\right]|\}+\max_{t\leq T}\|x_{0t}\|^2_\infty\frac{k_T\log(T\vee K_T)}{T}\leq \delta_T$.

4. *Regularity restrictions for first-stage regression: Let* $\tilde{p}_{jt}\equiv p_{jt}-\bar{E}\left[p_{jt}\right]$,

a) $\max_{j\leq J}\bar{E}\left[\tilde{p}^2_{jt}\right]+\max_{l\leq L}\left\{\bar{E}\left[z^2_{jt,l}\tilde{p}^2_{jt}\right]+1/\bar{E}\left[z^2_{jt,l}\nu^2_{jt}\right]\right\}$
$+\max_{k\leq K_0}\left\{\bar{E}\left[x^2_{0t,k}\tilde{p}^2_{jt}\right]+1/\bar{E}\left[x^2_{0t,k}\nu^2_{jt}\right]\right\}$
$+\max_{k\leq K_1}\left\{\bar{E}\left[x^2_{1jt,k}\tilde{p}^2_{jt}\right]+1/\bar{E}\left[x^2_{1jt,k}\nu^2_{jt}\right]\right\}\lesssim 1.$

b) $\max_{j\leq J}\left\{\max_{k\leq K_0}\bar{E}\left[|x^3_{0t,k}\nu^3_{jt}|\right]+\max_{k\leq K_1}\bar{E}\left[|x^3_{1jt,k}\nu^3_{jt}|\right]+\max_{l\leq L}\bar{E}\left[|z^3_{jt,l}\nu^3_{jt}|\right]\right\}\lesssim \kappa_T$ *with* $\kappa^2_T\log^3\left(K_T\vee T\right)=o\left(T\right)$.

c) $\max_{t\leq T, j\leq J}\left(\max_{l\leq L}z^2_{jt,l}+\max_{k\leq K_0}x^2_{0t,k}+\max_{k\leq K_1}x^2_{1jt,k}\right)\left[s\log\left(K_T\vee T\right)\right]/T\to_p 0.$

d) $\max_{j\leq J}\max_{l\leq L}\{|(\mathcal{E}-\bar{E})\left[z^2_{jt,l}\nu^2_{jt}\right]|+|(\mathcal{E}-\bar{E})\left[z^2_{jt,l}\tilde{p}^2_{jt}\right]|\}$
$+\max_{k\leq K_0}\{|(\mathcal{E}-\bar{E})\left[x^2_{0t,k}\nu^2_{jt}\right]|+|(\mathcal{E}-\bar{E})\left[x^2_{0t,k}\tilde{p}^2_{jt}\right]|\}+$
$\max_{k\leq K_1}\{|(\mathcal{E}-\bar{E})\left[x^2_{1jt,k}\nu^2_{jt}\right]|+|(\mathcal{E}-\bar{E})\left[x^2_{1jt,k}\tilde{p}^2_{jt}\right]|\}\to_p 0.$

The regularity conditions here require somewhat cumbersome notation to specify explicitly. The conditions presented in assumption 3.3 are sufficient to ensure a law of large numbers and central limit theorem apply to the post-selection estimator with heteroskedastic and non-Gaussian residuals. With the exponential tails assumption in B.2.5, the assumptions restricting the sup-norm of regressors (B.3.2(a) and 3(b)) can be weakened. In our discussion of the GMM estimator, many of these restrictions are subsumed by simply assuming a uniform law of large numbers applies to the score of the objective function. Assumption B.3.3 provides sufficient conditions for such a ULLN to apply in the linear environment. The regularity conditions for the first-stage regression in Assumption B.3.4 come from Belloni et al. (2012) and ensure the existence of optimal instruments for the endogenous campaign spending and the ability to consistently estimate these instruments via the first-stage regression.

B.2.4 Regularity Conditions for GMM Estimator

Assumption B.5. *Regularity Conditions for GMM Estimator:*

Let $\theta = [\beta_{01}', \ldots, \beta_{0J}', \beta_1', \pi_{01}', \ldots, \pi_{0J}', \pi_1', \pi_z', \upsilon_b]'$*. For all* T*, each data generating process in the sequence* $\{\mathcal{P}_T\}_{T=1}^{\infty}$*, satisfies the following restrictions:*

1. *Compactness of Parameter Set: The true parameter values* $\theta_0 \in \Theta_{K_T}$*, where* $\Theta_K \subset \mathbb{R}^{K_T+2}$ *is compact, with a compact limit set* $\Theta_\infty \equiv \lim_{T\to\infty} \Theta_K$*.*

2. *Continuity of Moment Conditions:*

   a) *The unconditional moment condition* $E\left[z_{jt,l}\xi_{jt}(\theta)\right]$ *is continuously differentiable in* $\theta, \forall j \leq J, t \leq T,$ *and,* $l \leq L$*.*

   b) *For the full-sample moment condition* $m_{lT}(\theta) \equiv \bar{E}\left[\frac{1}{J}\sum_{j=1}^{J} E\left[z_{jt,l}\xi_{jt}(\theta)\right]\right]$*,*

      i. $m_{lT}(\theta) \underset{T\to\infty}{\to} m_l(\theta)$ *uniformly for* $\theta \in \Theta_{K_T}$*, for all* K_T*,*

      ii. $m_{lT}(\theta)$ *is continuously differentiable and its limit* $m_l(\theta)$ *is continuous in* θ*, and,*

      iii. $m_{lT}(\theta_0) = 0$ *and* $m_{lT}(\theta) \neq 0, \forall \theta \neq \theta_0$*.*

3. *Uniform LLN for Sample Analog: Let* $g_{jt}(\theta) \equiv \left[x_{0t}', x_{1jt}', z_{jt}'\right]' \xi_{jt}(\theta)$*, the following uniform law of large numbers applies:*

$$\sup_{k\leq K_T} \sup_{\theta\in\Theta_k} \left\| \frac{1}{JT} \sum_{j,t=1}^{J,T} \left(g_{jt}(\theta) - \bar{E}\left[g_{jt}(\theta)\right]\right) \right\| \underset{T\to\infty}{\longrightarrow}_p 0.$$

4. *Define the* $L_T \times K_T$ *matrix* $\hat{G}_T(\theta) \equiv \frac{1}{JT}\sum_{j,t=1}^{J,T} \frac{\partial g_{jt}(\theta)}{\partial \theta'}$*:*

   a) *A uniform law of large numbers holds in a neighborhood of* θ_0 *for all* K_T*:*

$$\|\hat{G}_T(\theta) - G(\theta)\|_2^2 \underset{T\to\infty}{\to_p} 0.$$

   b) *The limiting matrix* $G(\theta)$ *is continuous in* θ *and* $G(\theta_0)$ *has full column rank* K_T*.*

5. W_T *is a positive definite matrix with* $\|W_T - W\|_2^2 \to_p 0$*, with* W *a symmetric, positive definite, and finite matrix.*

6. *The expected outer product of the score,* $\Omega \equiv \lim_{T\to\infty} (JT)^{-1} \sum_{j,t=1}^{J,T} E\left[g_{jt}(\theta) g_{jt}(\theta)'\right]$ *is a positive definite, finite matrix.*

7. *The minimal and maximal eigenvalues of* $\Sigma \equiv G(\theta_0)' \Omega^{-1} G(\theta_0)$*, denoted* $\underline{e}$ *and* $\bar{e}$*, are finite and bounded between finite constants* $0 < c \leq \underline{e} \leq \bar{e} \leq C < \infty$*.*

The GMM regularity conditions are very similar to those in Gillen et al. (2014), reflecting fairly standard restrictions on GMM estimators. The discussion in Caner and Zhang (2013) and Fan and Liao (2014) provide more primitive conditions for these results, though their restrictions are stated for i.i.d. sampling environments, requiring some additional notation to extend to the i.n.i.d. setting here. These assumptions are fairly standard in the literature on GMM estimation, with references to Newey (1990), Newey (1993), Caner (2009), and Newey and Windmeijer (2009).

B.2.5 Sparsity Assumptions for High-Dimensional BLP Model

Assumption B.6. *Sparsity Assumptions for High-Dimensional BLP Model:*

Let $\theta = [\beta_{01}', \ldots, \beta_{0J}', \beta_1', \pi_{01}', \ldots, \pi_{0J}', \pi_1', \pi_z' \upsilon_p]'$. *Each data generating process in the sequence* $\{\mathcal{P}_T\}_{T=1}^{\infty}$, *has* $K_T > T$ *possible parameters,* $1 \leq k_T < T$ *of which are non-zero, where both* $K_T \underset{T\to\infty}{\to} \infty$ *and* $k_T \underset{T\to\infty}{\to} \infty$. *Further, the number of excluded instruments in* z *is fixed at* $L \geq 2$. *Finally, there exists a sequence* $\{\delta_T, \Delta_T\} \to_{T\to\infty} 0$ *with* $\sqrt{\frac{k_T \log k_T}{T}} > \delta_T > \frac{k_T}{T}$.

1. *The parameter space isn't too large, with* $\log(K_T) = o\left(T^{-1/3}\right)$.

2. *The model is sparse, with the number of non-zero variables,* $k_T^3 \log k_T = o\left(T^{-1}\right)$.

3. *The Gram matrix for controls satisfies restricted eigenvalues of Assumption 2.3.*

4. *The Hessian of the objective function with respect to non-zero variables is almost surely positive definite. There exists a sequence* $\ell_T \to 0$, κ', *and* κ'' *such that, with probability* $1 - \Delta_T$*:*

$$0 \leq \kappa' \leq \phi_{min}(\ell_n k_T)\{\Omega\} \leq \phi_{max}(\ell_n k_T)\{\Omega\} \leq \kappa'' < \infty.$$

5. *Non-zero coefficients are bounded away from zero:* $\min\{|\theta_k| \;\; s.t. \; \theta_k \neq 0\} > 2\delta_T$.

6. *The marginal distributions for controls, instruments, and residual vote shares have exponentially decaying tails.*

The additional sparsity restrictions for the high-dimensional BLP model are not that different from those in the linear model. The sparsity restriction is a bit tighter, accounting for the need to select variables on estimated optimal instruments. The restricted eigenvalue assumption needs to be extended to the outer product of the gradients, and the exponential tail restriction is extended to the optimal instruments.

Appendix C

APPENDIX TO CHAPTER 4

C.1 Proofs

Proof of Lemma 4.1. The decision-maker's indirect utility from running experiment $\mathcal{E}$, $V(\mathcal{E}) \equiv \max_{\alpha\in\mathcal{A}} U(\mathcal{E},\alpha) = \max_{\alpha\in\mathcal{A}} \min_{h\in\widehat{H}} \mathbb{E}_{h,\mathcal{E}}[u(p,\alpha(e,y))]$, can be viewed as the value of a zero-sum game where player 1, the decision-maker, chooses an allocation rule $\alpha \in \mathcal{A}$, and player 2, nature, chooses a prior $h \in \widehat{H}$. By standard arguments, this game has a Nash equilibrium, implying that $V(\mathcal{E}) = \min_{h\in\widehat{H}} \max_{\alpha\in\mathcal{A}} \mathbb{E}_{h,\mathcal{E}}[u(p,\alpha(e,y))]$.

Given h, the decision-maker's payoff from running experiment $\mathcal{E}$ can be written as

$$\begin{aligned}
\max_{\alpha\in\mathcal{A}} \mathbb{E}_{h,\mathcal{E}}\left[u(p,\alpha(e,y))\right] &= \max_{\alpha\in\mathcal{A}} \sum_{e\in E} \mathcal{E}(e)\mathbb{E}_{p\sim h}\left[\sum_{y\in\mathcal{Y}} \operatorname{prob}(y|p,e)u(p,\alpha(e,y))\right] \\
&= \sum_{e\in E} \mathcal{E}(e) \sum_{y\in\mathcal{Y}} \max_{a\in\{0,1\}} \mathbb{E}_{p\sim h}\left[\operatorname{prob}(y|p,e)u(p,a)\right] \\
&= \sum_{e\in E} \mathcal{E}(e)v(h,e),
\end{aligned}$$

where $v(h,e) \equiv \sum_{y\in\mathcal{Y}} \max_{a\in\{0,1\}} \mathbb{E}_{p\sim h}\left[\operatorname{prob}(y|p,e)u(p,a)\right]$. Since $v(h,e) = v(h,e') \equiv v(h,[e])$ for all $e' \in [e]$, it follows that $V(\mathcal{E}) = \min_{h\in\widehat{H}} \sum_{[e]\in[E]} \mathcal{E}([e])v(h,[e])$. Thus, if $\mathcal{E}$ and $\mathcal{E}'$ induce the same distribution over $[E]$, $V(\mathcal{E}) = V(\mathcal{E}')$. ■

Proof of Proposition 4.2. We first show the result in the specific case where $\lambda = 1$ — the more general result then obtains by continuity.

As in the proof of Lemma 4.1, the decision-maker's payoff from running experiment $\mathcal{E}$ can be written as

$$\max_{\alpha\in\mathcal{A}} \mathbb{E}_{h_0,\mathcal{E}}\left[u(p,\alpha(e,y))\right] = \sum_{[e]\in[E]} \mathcal{E}([e])v(h_0,[e]) \leq \max_{[e]\in[E]} v(h_0,[e]).$$

Therefore, $\mathcal{E}$ solves (4.3) if and only if $\operatorname{supp}\mathcal{E} \subset \arg\max_{[e]\in[E]} v(h_0,[e])$.

Let us now show that $\arg\max_{[e]\in[E]} v(h_0,[e])$ is generically a singleton. We first show that the set of priors h_0 such that there is a uniquely optimal equivalence class of experiments is open. Suppose that $[e_0]$ is uniquely optimal under h_0. Since E is finite, there exists $\eta > 0$ such that $v(h_0, [e]) < v(h_0, [e_0]) - \eta$ for all $[e] \neq [e_0]$. Since $v(h, e)$ is continuous in h, this implies that there exists a neighborhood H_0 of h_0 such that, for all $h \in H_0$, $v(h, [e]) < v(h, [e_0]) - \eta/2$. Hence, $[e_0]$ is the uniquely optimal design for all priors $h \in H_0$.

We now prove that the set of priors h_0 such that there is a uniquely optimal equivalence class of experiments is dense. The proof is by induction on the number of equivalence classes $[e_0] \in \arg\max_{[e]\in[E]} v(h_0, [e])$. We show that if there exist n such equivalence classes, then in any neighborhood of h_0 there exists a prior h such that there are at most $n-1$ equivalence classes in $\arg\max_{[e]\in[E]} v(h, [e])$.

Indeed, assume that $[e_0] \neq [e_1]$ both belong to $\arg\max_{[e]\in[E]} v(h_0, [e])$. For $\theta > 0$, consider the polynomial $M_\theta(p)$ in $p \in P$ defined by

$$M_\theta(p) = v\left((1-\theta)h_0 + \theta p, [e_0]\right) - v\left((1-\theta)h_0 + \theta p, [e_1]\right),$$

where $(1-\theta)h_0 + \theta p$ denotes the mixture probability measure that places mass $1-\theta$ on h, and mass θ on the Dirac mass at p. Since $[E]$ is finite, for all $\theta > 0$ small enough, it must be that

$$\arg\max_{[e]\in[E]} v((1-\theta)h_0 + \theta p, [e]) \subset \arg\max_{[e]\in[E]} v(h_0, [e]).$$

Consider such a $\theta > 0$. The fact that $[e_0] \neq [e_1]$ implies that $M_\theta(p)$ is not identically equal to 0. Hence, there exists p such that $v\left((1-\theta)h_0 + \theta p, [e_0]\right) \neq v\left((1-\theta)h_0 + \theta p, [e_1]\right)$. This implies that the inductive step holds at prior $\tilde{h} = (1-\theta)h_0 + \theta p$. Using the fact that $[E]$ is finite and $v(h, [e])$ is continuous in h, this implies that the inductive step holds at a prior that admits a density against the Lebesgue measure. Thus, when $\lambda = 1$, deterministic experiments are generically strictly optimal.

Now, given any λ, h, and $[e]$, since the decision maker's utility only takes values in $[0, 1]$, letting $\alpha_0 \in \arg\max_{\alpha\in\mathcal{A}} \mathbb{E}_{h_0,e}[u(p, \alpha(e, y))]$ we have

$$\begin{aligned} v(\lambda h_0 + (1-\lambda)h, [e]) &\leq \lambda v(h_0, [e]) + (1-\lambda)v(h, [e]) \leq v(h_0, [e]) + (1-\lambda) \quad \text{and} \\ v(\lambda h_0 + (1-\lambda)h, [e]) &\geq \lambda v(h_0, [e]) + (1-\lambda)\mathbb{E}_{h,e}[u(p, \alpha_0(e, y))] \geq v(h_0, [e]) - (1-\lambda). \end{aligned}$$

As there are finitely many experiments, if $[e_0]$ is the unique maximizer of $v(h_0, [e])$, there exists $\eta > 0$ such that, for all $[e] \neq [e_0]$, $v(h_0, [e_0]) > v(h_0, [e]) +$

η. Together, this implies that there exists $\underline{\lambda} \in (0,1)$ such that, for all $\lambda > \underline{\lambda}$, objective (4.3) is maximized at $\mathcal{E}$ if and only if $\operatorname{supp}\mathcal{E} \subset [e_0]$. ■

Proof of Proposition 4.3. To establish point (i) we use the strategy $(\mathcal{E}_1, \alpha_1)$ such that

- $\mathcal{E}_1$ consists of sampling N subjects with covariates independently drawn according to q and assigning treatments $\tau_i = \mathbf{1}_{i \leq N/2}$;
- $\alpha_1(e, y) \equiv \mathbf{1}_{\overline{y}^1 > \overline{y}^0}$, where $\overline{y}^\tau$ is the sample average of outcomes among subjects with treatment status τ.

Losses $L(p)$ from first best, given state of the world p, are defined as

$$L(p) \equiv \max_{a \in \{0,1\}} p^a - \mathbb{E}_{p,\mathcal{E}_1}\left[p^{\mathbf{1}_{\overline{y}^1 - \overline{y}^0 > 0}}\right].$$

By symmetry, it suffices to treat the case where $p_1 - p_0 > 0$. In this case, we have $L(p) = (p_1 - p_0)\mathrm{prob}_{p,\mathcal{E}_1}(\overline{y}^1 - \overline{y}^0 \leq 0)$. The probability of choosing the suboptimal policy can be bounded using McDiarmid's inequality.[1] By applying McDiarmid's inequality to $f(y) \equiv \frac{2}{N}\sum_{i=1}^{N/2} y^0_{i+N/2} - y^1_i$, we obtain

$$\begin{aligned} \mathrm{prob}_{p,\mathcal{E}_1}(\overline{y}^1 - \overline{y}^0 \leq 0) &= \mathrm{prob}_{p,\mathcal{E}_1}\left(\overline{y}^0 - \overline{y}^1 - (p^0 - p^1) \geq (p^1 - p^0)\right) \\ &\leq \exp\left(-(N/2)(p^1 - p^0)^2\right). \end{aligned}$$

For any $a > 0$, $x \mapsto x\exp(-ax^2)$ is log-concave and maximized at $x = (2a)^{-1/2}$. This implies that

$$\max_{a \in \{0,1\}} p^a - \mathbb{E}_{p,\mathcal{E}_1}\left[p^{\mathbf{1}_{\overline{y}^1 - \overline{y}^0 > 0}}\right] \leq \frac{1}{\sqrt{N}}. \tag{C.1}$$

An analogous argument delivers (C.1) also for the case where $p^1 - p^0 \leq 0$. Hence, given any $h \in H$,

$$\mathbb{E}_h\left(\max_{a \in \{0,1\}} u(p,a)\right) - \mathbb{E}_{h,\mathcal{E}_1}\left[u(p, \alpha_1(e,y))\right] \leq \frac{1}{\sqrt{N}}.$$

To establish point (ii), fix a deterministic experiment $e \in E$. From Assumption 4.2, there exists $h \in H$ such that for almost every $\overline{p}_e$,

[1] McDiarmid's (1989) inequality can be stated as follows. Let $X_1, \ldots, X_n$ be independent random variables, with X_k taking values in a set A_k for each k. Suppose that the (measurable) function $f : \times_k A_k \to \mathbb{R}$ satisfies $|f(x) - f(x')| \leq c_k$ whenever x and x' differ only in the kth coordinate. Then, for any $t > 0$, $\mathrm{prob}\left(f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \geq t\right) \leq \exp\left(-2t^2/\sum_k c_k^2\right)$.

$$\min\left\{\mathbb{E}_h\left[\max_{a\in\{0,1\}} p^a - p^0|\overline{p}_e\right], \mathbb{E}_h\left[\max_{a\in\{0,1\}} p^a - p^1|\overline{p}_e\right]\right\} > \epsilon. \text{ Hence,}$$

$$\begin{aligned}\max_{\alpha}\mathbb{E}_{h,e}\left[u(p,\alpha(e,y))\right] &\leq \mathbb{E}_{h,e}\left[\max_{a\in\{0,1\}}\mathbb{E}_{h,e}\left[u(p,a)|\overline{p}_e\right]\right]\\ &\leq \mathbb{E}_{h,e}\left[\max_{a\in\{0,1\}} u(p,a)\right] - \epsilon.\end{aligned}$$

■

Proof of Proposition 4.4. Fix any $e^\dagger \in \arg\max_{e\in\operatorname{supp}\mathcal{E}_K} B(e)$. The kth rerandomized trial, $k \in \{1,\ldots,K\}$, selects each experiment in its support with probability at least $r \equiv (\underline{q}\underline{\pi})^N$, where $\underline{q} \equiv \min_{x\in X} q(x) \leq 1/|X|$ and $\underline{\pi} \equiv \min\{\pi, 1-\pi\} \leq 1/2$. Therefore, the odds of rerandomization picking experiment $e^\dagger$ are at least $\rho \equiv 1-(1-r)^K$. For $K \geq (2|X|)^N$,

$$\rho = 1 - \exp(K\log(1-r)) \sim 1 - \exp(-Kr) \geq 1 - 1/\exp > 0.$$

Hence, there exists $\underline{\rho} > 0$ such that, for all N, rerandomized experiment $\mathcal{E}_K$ selects deterministic experiment $e^\dagger$ with probability at least $\underline{\rho}$.

The proof of Proposition 4.3 implies that there exists $h^\dagger \in H$ such that

$$\begin{aligned}\forall e\in E,\ \max_{\alpha\in\mathcal{A}}\mathbb{E}_{h^\dagger,e}[u(p,\alpha(e,y))] &\leq \min_{h\in H}\mathbb{E}_h\left(\max_{a\in\{0,1\}} u(p,a)\right),\\ \text{and } \max_{\alpha\in\mathcal{A}}\mathbb{E}_{h^\dagger,e^\dagger}[u(p,\alpha(e^\dagger,y))] &\leq \min_{h\in H}\mathbb{E}_h\left(\max_{a\in\{0,1\}} u(p,a)\right) - \epsilon.\\ \text{Hence, } \max_{\alpha\in\mathcal{A}}\min_{h\in H}\mathbb{E}_{h,\mathcal{E}_K}[u(p,\alpha(e,y))] &\leq \min_{h\in H}\mathbb{E}_h\left(\max_{a\in\{0,1\}} u(p,a)\right) - \underline{\rho}\epsilon.\end{aligned}$$

■

Proof of Proposition 4.5. Denote by $(\overline{y}_{0,k}, \overline{y}_{1,k})$ the sample average of outcomes by treatment group for experiment e_k, and by $(\overline{y}_0^*, \overline{y}_1^*)$ the sample average of outcomes by treatment group for the experimental design e_K^* selected by rerandomized experiment $\mathcal{E}_K$.

Losses $L(p)$ from first best given state of the world p are defined as $L(p) \equiv \max_{a\in\{0,1\}} p^a - \mathbb{E}_{p,\mathcal{E}_K}\left[p^{\mathbf{1}_{\overline{y}_1^*-\overline{y}_0^*>0}}\right]$. By symmetry, it suffices to treat the case where $p^1 - p^0 > 0$. In this case, we have

$$\begin{aligned}
L(p) &= (p^1 - p^0)\mathrm{prob}_{p,\mathcal{E}_K}(\overline{y}_1^* - \overline{y}_0^* \le 0) \\
&\le (p^1 - p^0)\mathrm{prob}_{p,\mathcal{E}_K}\left(\min_{k\in\{1,\dots,K\}} \overline{y}_{1,k} - \overline{y}_{0,k} \le 0\right) \\
&\le (p^1 - p^0)\min\left\{1, \sum_{k=1}^{K} \mathrm{prob}_{p,\mathcal{E}_K}(\overline{y}_{1,k} - \overline{y}_{0,k} \le 0)\right\} \\
&\le (p^1 - p^0)\min\left\{1, K\exp\left(-2\pi(1-\pi)(p^1 - p^0)^2 N\right)\right\} \\
&\le (p^1 - p^0)\min\left\{1, K\exp\left(-\kappa(p^1 - p^0)^2 N\right)\right\},
\end{aligned}$$

where the second-to-last step used McDiarmid's inequality (McDiarmid, 1989), already invoked in the proof of Proposition 4.3, applied to $f(y) \equiv \overline{y}_{0,k} - \overline{y}_{1,k}$, and the last step follows from $2\pi(1-\pi) \ge \kappa \equiv \min\{\pi, 1-\pi\}$.

We have that $K\exp(-\kappa(p^1 - p^0)^2 N) \le 1 \iff p^1 - p^0 \ge \sqrt{\frac{\log(K)}{\kappa N}}$, which implies that

$$L(p) \le \begin{cases} p^1 - p^0 & \text{if} \quad p^1 - p^0 < \sqrt{\frac{\log(K)}{\kappa N}}, \\ K(p^1 - p^0)\exp(-\kappa(p^1 - p^0)^2 N) & \text{if} \quad p^1 - p^0 \ge \sqrt{\frac{\log(K)}{\kappa N}}. \end{cases} \tag{C.2}$$

The mapping $x \mapsto x\exp(-\kappa N x^2)$ is log-concave and maximized at $x = \sqrt{\frac{1}{2\kappa N}}$. Since $K \ge 2$, we have that $\sqrt{\frac{\log(K)}{\kappa N}} \ge \sqrt{\frac{1}{2\kappa N}}$, which implies that both terms on the right-hand side of (C.2) are maximized at $p^1 - p^0 = \sqrt{\frac{\log(K)}{\kappa N}}$. This implies that indeed $L(p) \le \sqrt{\frac{\log(K)}{\kappa N}}$. Identical reasoning applies in the case where $p^1 - p^0 < 0$. ■

Proof of Corollary 4.6. By symmetry, it suffices to treat the case where $p^1 - p^0 > 0$. Since $\mathcal{D}_K$ picks an assignment that leads to incorrect inference whenever possible, we have

$$\begin{aligned}
L(p) &= (p^1 - p^0)\mathrm{prob}_{p,\mathcal{D}_K}\left(\min_{k\in\{1,\dots,K\}} \overline{y}_{1,k} - \overline{y}_{0,k} \le 0\right) \\
&= (p^1 - p^0)\left[1 - \mathrm{prob}_{p,\mathcal{D}_K}\left(\cap_{k\in\{1,\dots,K\}}\{\overline{y}_{1,k} - \overline{y}_{0,k} > 0\}\right)\right].
\end{aligned}$$

Under rerandomization, $\overline{y}_{1,k} - \overline{y}_{0,k}$ are i.i.d. conditional on the initial draw of subjects from the population, that is, conditional on $(x_i, y_i^0, y_i^1)_{i\le N}$, thus

$$\begin{aligned}
&\mathrm{prob}_{p,\mathcal{D}_K}\left(\cap_{k\in\{1,\dots,K\}}\{\overline{y}_{1,k} - \overline{y}_{0,k} > 0\}\right) \\
&\qquad = \mathbb{E}_{p,\mathcal{D}_K}\left[\mathrm{prob}_{\mathcal{D}_K}\left(\overline{y}_{1,1} - \overline{y}_{0,1} > 0 \,|\, (x_i, y_i^0, y_i^1)_{i\le N}\right)^K\right].
\end{aligned}$$

Under resampling, $\overline{y}_{1,k} - \overline{y}_{0,k}$ are unconditionally i.i.d., so

$$\begin{aligned}
&\text{prob}_{p,\mathcal{D}_K}\big(\cap_{k\in\{1,\ldots,K\}} \{\overline{y}_{1,k} - \overline{y}_{0,k} > 0\}\big) \\
&\qquad = \mathbb{E}_{p,\mathcal{D}_K}\left[\text{prob}_{\mathcal{D}_K}\left(\overline{y}_{1,1} - \overline{y}_{0,1} > 0 \,|\, (x_{i,1}, y^0_{i,1}, y^1_{i,1})_{i\leq N}\right)\right]^K .
\end{aligned}$$

The corollary follows by Jensen's inequality. ■

BIBLIOGRAPHY

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation," *The Review of Economic Studies*, 58, 621–654.

Andrews, D. W. (1999): "Consistent moment selection procedures for generalized method of moments estimation," *Econometrica*, 67, 543–563.

Andrews, D. W. and B. Lu (2001): "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models," *Journal of Econometrics*, 101, 123–164.

Ashworth, S. and E. Bueno de Mesquita (2009): "Elections with platform and valence competition," *Games and Economic Behavior*, 67, 191–216.

Bandyopadhyay, S., K. Chatterjee, and T. Sjöström (2011): "Pre-Electoral Coalitions and Post-Election Bargaining," *Quarterly Journal of Political Science*, 6, 1–53.

Banerjee, A., S. Chassang, and E. Snowberg (forthcoming): "Decision Theoretic Approaches to Experiment Design and External Validity," in *Handbook of Field Experiments*, ed. by E. Duflo and A. Banerjee, Elsevier.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80, 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2013a): "Inference on treatment effects after selection amongst high-dimensional controls," *Review of Economic Studies*.

Belloni, A., V. Chernozhukov, and L. Wang (2011): "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, 98, 791–806.

Belloni, A., V. Chernozhukov, and Y. Wei (2013b): "Honest confidence regions for logistic regression with a large number of controls," *arXiv preprint arXiv:1304.3969*.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

Bergemann, D. and J. Välimäki (1996): "Learning and Strategic Pricing," *Econometrica*, 64, 1125–1149.

——— (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70, 1007–1033.

——— (2006): "Bandit Problems," Cowles Foundation discussion paper.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

——— (1999): "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *American Economic Review*, 89, 400–430.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009): "Simultaneous analysis of Lasso and Dantzig selector," *The Annals of Statistics*, 1705–1732.

Blais, A. and I. H. Indridason (2007): "Making Candidates Count: The Logic of Electoral Alliances in Two-Round Legislative Elections," *Journal of Politics*, 69, 193–205.

Bormann, N.-C. and M. Golder (2013): "Democratic Electoral Systems around the world, 1946-2011," *Electoral Studies*, 32, 360–369.

Bradic, J., J. Fan, and W. Wang (2011): "Penalized composite quasi-likelihood for ultrahigh dimensional variable selection," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 325–349.

Bruhn, M. and D. McKenzie (2009): "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 1, 200–232.

Candes, E. and T. Tao (2007): "The Dantzig selector: statistical estimation when p is much larger than n," *The Annals of Statistics*, 2313–2351.

Caner, M. (2009): "Lasso-type GMM estimator," *Econometric Theory*, 25, 270–290.

Caner, M. and H. H. Zhang (2013): "Adaptive elastic net for generalized methods of moments," *Journal of Business & Economic Statistics*.

Carroll, R. and G. W. Cox (2007): "The Logic of Gamson's Law: Pre-election Coalitions and Portfolio Allocations," *American Journal of Political Science*, 51, 300–313.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Che, H., G. Iyer, and R. Shanmugam (2007): "Negative advertising and voter choice," Tech. rep., working paper, University of Southern California.

Cheng, X. and Z. Liao (2015): "Select the valid and relevant moments: An information-based LASSO for GMM with many moments," *Journal of Econometrics*, 186, 443–464.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

Cochrane, W. G. and D. B. Rubin (1973): "Controlling Bias in Observational Studies," *Sankhyā: The Indian Journal of Statistics, Series A*, 35, 417–446.

Conlon, C. T. (2013): "The Empirical Likelihood MPEC Approach to Demand Estimation," *Available at SSRN*.

Consulta Mitofsky (2012): "Geometría electoral en México (Electoral geometry in Mexico)," Public Opinion Poll, `http://www.consulta.mx/web/images/MexicoOpina/2013/NA_GEOMETRIA_ELECTORAL.pdf` (in Spanish).

Deaton, A. (2010): "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 48, 424–455.

Degan, A. and A. Merlo (2011): "A structural model of turnout and voting in multiple elections," *Journal of the European Economic Association*, 9, 209–245.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117, 55–93.

Dube, J.-P., J. Fox, and C.-L. Su (2012): "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, forthcoming.

Duflo, E., R. Glennerster, and M. Kremer (2008): "Using Randomization in Development Economics Research: A Tool Kit," in *Handbook of Development Economics, Vol. 4*, ed. by T. P. Schultz and J. Strauss, Amsterdam: Elsevier, 3895–3962.

Echenique, F. (2007): "Finding All Equilibria in Games of Strategic Complements," *Journal of Economic Theory*, 135, 514–532.

Echenique, F. and A. Edlin (2004): "Mixed equilibria are unstable in games of strategic complements," *Journal of Economic Theory*, 118, 61–79.

Erikson, R. S. and T. R. Palfrey (2000): "Equilibria in Campaign Spending Games: Theory and Data," *American Political Science Review*, 94, 595–609.

Fan, J. and R. Li (2001): "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, 96, 1348–1360.

Fan, J. and Y. Liao (2014): "Endogeneity in high dimensions," *Annals of statistics*, 42, 872.

Fan, J. and J. Lv (2011): "Nonconcave penalized likelihood with NP-dimensionality," *Information Theory, IEEE Transactions on*, 57, 5467–5484.

Ferrara, F. (2006): "Two in one: Party competition in the Italian single ballot mixed system," *Electoral Studies*, 25, 329–350.

Ferrara, F. and E. S. Herron (2005): "Going It Alone? Strategic Entry under Mixed Electoral Rules," *American Journal of Political Science*, 49, 16–31.

Gandhi, A. and J. F. Houde (2015): "Measuring Substitution Patterns in Differentiated Products Industries: The Missing Instruments," *Mimeo.*

Gautier, E. and A. Tsybakov (2011): "High-dimensional instrumental variables regression and confidence sets," *arXiv preprint arXiv:1105.2454*.

Gerber, A. S. (1998): "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables," *American Political Science Review*, 92, 401–411.

——— (2011): "Field Experiments in Political Science," in *Cambridge Handbook of Experimental Political Science*, ed. by J. N. Druckman, D. P. Green, J. H. Kuklinski, and A. Lupia, Cambridge University Press.

Gerber, A. S. and D. P. Green (2000): "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment," *American Political Science Review*, 94, 653–663.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Gillen, B., S. Montero, H. R. Moon, and M. Shum (2015): "BLP-Lasso for Aggregate Discrete-Choice Models Applied to Elections with Rich Demographic Covariates," Working Paper. California Institute of Technology.

Gillen, B. J., H. R. Moon, and M. Shum (2014): "Demand Estimation with High-Dimensional Product Characteristics," *Bayesian Model Comparison*, 301–24.

Golder, S. N. (2006): *The Logic of Pre-Electoral Coalition Formation*, Columbus: Ohio State University Press.

Gordon, B. R. and W. R. Hartmann (2013): "Advertising effects in presidential elections," *Marketing Science*, 32, 19–35.

Gordon, B. R., M. J. Lovett, R. Shachar, K. Arceneaux, S. Moorthy, M. Peress, A. Rao, S. Sen, D. Soberman, and O. Urminsky (2012): "Marketing and politics: Models, behavior, and policy implications," *Marketing Letters*, 23, 391–403.

Green, D. P. and J. S. Krasno (1988): "Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections," *American Journal of Political Science*, 32, 884–907.

Green, D. P. and A. Tusicisny (2012): "Statistical Analysis of Results from Laboratory Studies in Experimental Economics: A Critique of Current Practice," Working Paper. Columbia University.

Groseclose, T. (2001): "A Model of Candidate Location When One Candidate Has a Valence Advantage," *American Journal of Political Science*, 45, 862–886.

Grossman, S. J. and J. E. Stiglitz (1980): "On the Impossibility of Informationally Efficient Markets," *The American Economic Review*, 70, 393–408.

Harding, M. and C. Lamarche (2015): "Penalized Quantile Regression with Semiparametric Correlated Effects: An Application with Heterogeneous Preferences," *Mimeo.*

Horowitz, J. L. and J. Huang (2010): "The adaptive LASSO under a generalized sparsity condition," *Manuscript, Northwestern University.*

Huang, J., J. L. Horowitz, and S. Ma (2008): "Asymptotic properties of bridge estimators in sparse high-dimensional regression models," *The Annals of Statistics*, 36, 587–613.

Humphreys, M., R. S. de la Sierra, and P. Van der Windt (2013): "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration," *Political Analysis*, 21, 1–20.

Iaryczower, M. and A. Mattozzi (2013): "On the Nature of Competition in Alternative Electoral Systems," *Journal of Politics*, 75, 743–756.

Imai, K. (2005): "Do Get-out-the-vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments," *American Political Science Review*, 99, 283–300.

Imbens, G. W. (2010): "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423.

Jacobson, G. C. (1978): "The Effects of Campaign Spending in Congressional Elections," *American Political Science Review*, 72, 469–491.

——— (1985): "Money and votes reconsidered: Congressional elections, 1972–1982," *Public choice*, 47, 7–62.

Kaminski, M. M. (2001): "Coalitional Stability of Multi-Party Systems: Evidence from Poland," *American Journal of Political Science*, 45, 294–312.

Kasy, M. (2013): "Why Experimenters Should Not Randomize, and What They Should Do Instead," Working Paper. Harvard University.

Kawai, K. (2014): "Campaign Finance in U.S. House Elections," Working Paper. NYU Stern.

Kawai, K. and Y. Watanabe (2013): "Inferring Strategic Voting," *American Economic Review*, 103, 624–662.

King, G. (1997): "A solution to the ecological inference problem," .

Kitamura, Y. (2001): "Asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica*, 69, 1661–1672.

Leeb, H. and B. M. Pötscher (2005): "Model selection and inference: Facts and fiction," *Econometric Theory*, 21, 21–59.

——— (2006): "Can one estimate the conditional distribution of post-model-selection estimators?" *The Annals of Statistics*, 2554–2591.

——— (2008): "Sparse estimators and the oracle property, or the return of Hodges' estimator," *Journal of Econometrics*, 142, 201–211.

Levonyan, V. (2013): "What Led to the Ban on Same-Sex Marriage in California?: Structural Estimation of Voting Data on Proposition 8," *Mimeo.*

Liao, Z. (2013): "Adaptive GMM shrinkage estimation with consistent moment selection," *Econometric Theory*, 29, 857–904.

Lockhart, R., J. Taylor, R. J. Tibshirani, and R. Tibshirani (2014): "A significance test for the lasso," *Annals of statistics*, 42, 413.

Martin, G. J. (2014): "The Informational Content of Campaign Advertising," Working Paper. Emory University.

McDiarmid, C. (1989): "On the Method of Bounded Differences," *Surveys in Combinatorics*, 141, 148–188.

Miguel, E. and M. Kremer (2004): "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72, 159–217.

Milligan, K. and M. Rekkas (2008): "Campaign spending limits, incumbent spending, and election outcomes," *Canadian Journal of Economics/Revue canadienne d'économique*, 41, 1351–1374.

Montero, S. (2015): "Going It Alone? An Empirical Study of Coalition Formation in Elections," *Mimeo*.

Moon, H. R., M. Shum, and M. Weidner (2014): "Estimation of random coefficients logit demand models with interactive fixed effects," Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.

Morgan, K. L. and D. B. Rubin (2012): "Rerandomization to Improve Covariate Balance in Experiments," *The Annals of Statistics*, 40, 1263–1282.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9, 513–548.

Newey, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica: Journal of the Econometric Society*, 809–837.

——— (1993): "16 Efficient estimation of models with conditional moment restrictions," *Handbook of statistics*, 11, 419–454.

Newey, W. K. and F. Windmeijer (2009): "Generalized method of moments with many weak moment conditions," *Econometrica*, 77, 687–719.

Olken, B. A. (2015): "Promises and Perils of Pre-Analysis Plans," *The Journal of Economic Perspectives*, 29, 61–80.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334.

Persico, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68, 135–148.

Persson, T. and G. Tabellini (2000): *Political Economics: Explaining Economic Policy*, MIT Press.

Poole, K. T. and H. Rosenthal (1985): "A spatial model for legislative roll call analysis," *American Journal of Political Science*, 357–384.

Rekkas, M. (2007): "The Impact of Campaign Spending on Votes in Multiparty Elections," *Review of Economics and Statistics*, 89, 573–585.

Reynaert, M. and F. Verboven (2014): "Improving the performance of random coefficients demand models: the role of optimal instruments," *Journal of Econometrics*, 179, 83–98.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185–202.

Rothschild, M. L. (1978): "Political Advertising: A Neglected Policy Issue in Marketing," *Journal of Marketing Research*, 15, 58–71.

Rubin, D. B. (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 74, 318–328.

——— (1980): "Bias Reduction using Mahalanobis-metric Matching," *Biometrics*, 36, 293–298.

Savage, L. J. (1954): *The Foundations of Statistics*, Courier Corporation.

Shi, X. and M. Shum (2015): "Simple Two-Stage Inference for a Class of Partially Identified Models," *Econometric Theory*, 31, 493–520.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014): "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models," *The Annals of Statistics*, 42, 1166–1202.

Vives, X. (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305–321.

Zhang, C.-H. (2010): "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, 894–942.

ZHANG, C.-H. AND J. HUANG (2008): "The sparsity and bias of the lasso selection in high-dimensional linear regression," *The Annals of Statistics*, 1567–1594.

ZOU, H. AND T. HASTIE (2005): "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320.

ZOU, H. AND R. LI (2008): "One-step sparse estimates in nonconcave penalized likelihood models," *Annals of statistics*, 36, 1509.